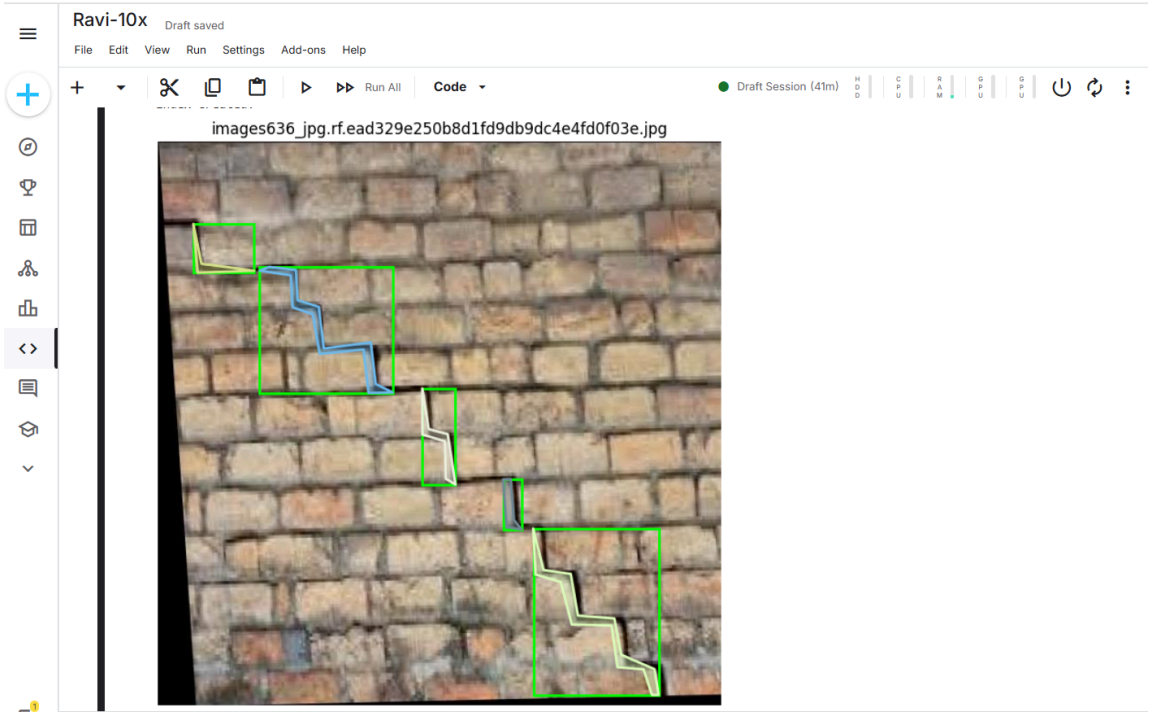
Project: Steps to train a text-guided segmentation model on custom data

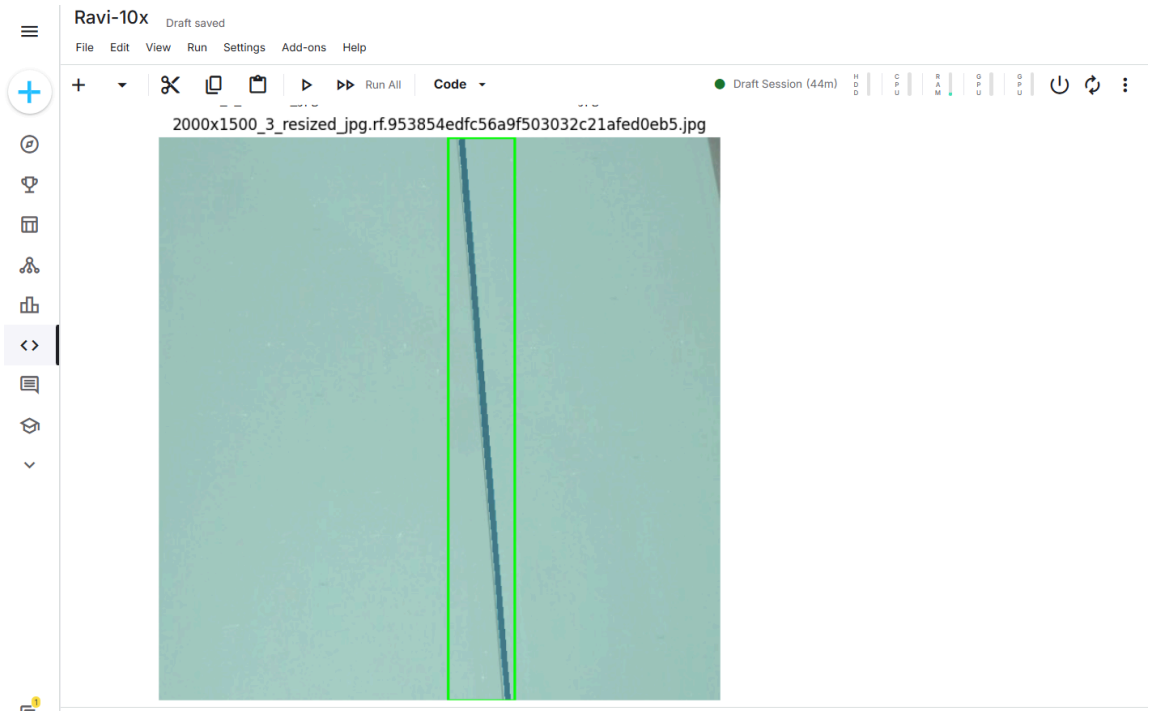
1. Visualize data from roboflow:
   Crack - seg masks + bbox masks
   Taping area - bbox masks

   Have annotation information in coco format



2. Using SAM model (Segment Anything Model) to generate seg masks for taping area given bbox coordinates
   https://docs.ultralytics.com/models/sam/#key-features-of-the-segment-anything-model-sam

   Saved masks as coco format
   (Preprocessing Notebook in the repo)

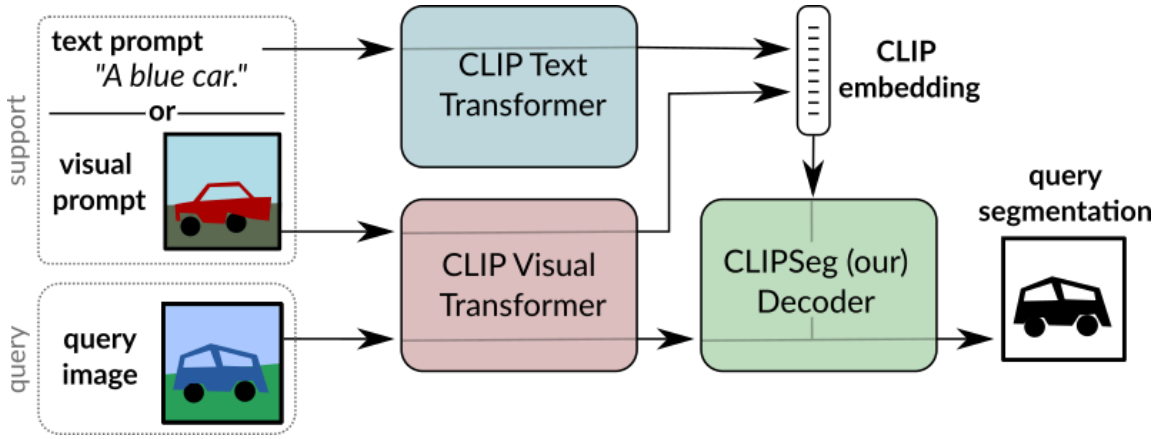


3. Data preprocessing and DataLoader
   Data Augmentation done in Roboflow, while extracting dataset
   Using Train/Val split as provided in the dataset

4. Using CLIPSeg model:
   https://huggingface.co/docs/transformers/en/model_doc/clipseg

   In warmup epochs, training only the decoder (SEG head) and, clip.text_projection, clip.visual_projection to align text embedding and vision embedding to decoder space

   Then, training all params
   Model Size: Total params: 150.75M



5. Training Strategy and loss func:
   Started with BCE (binary cross-entropy) and dice loss
   Added Focal loss later

   Changed image input dimension from 352 to 576

6. Model Training:
   V0: Warmup epochs - 3
   Full training epochs - 15

   Loss: Dice + BCE
   Image size - 352

   Runtime per epoch - 14 min
   Avg inference time/image - 0.6 second

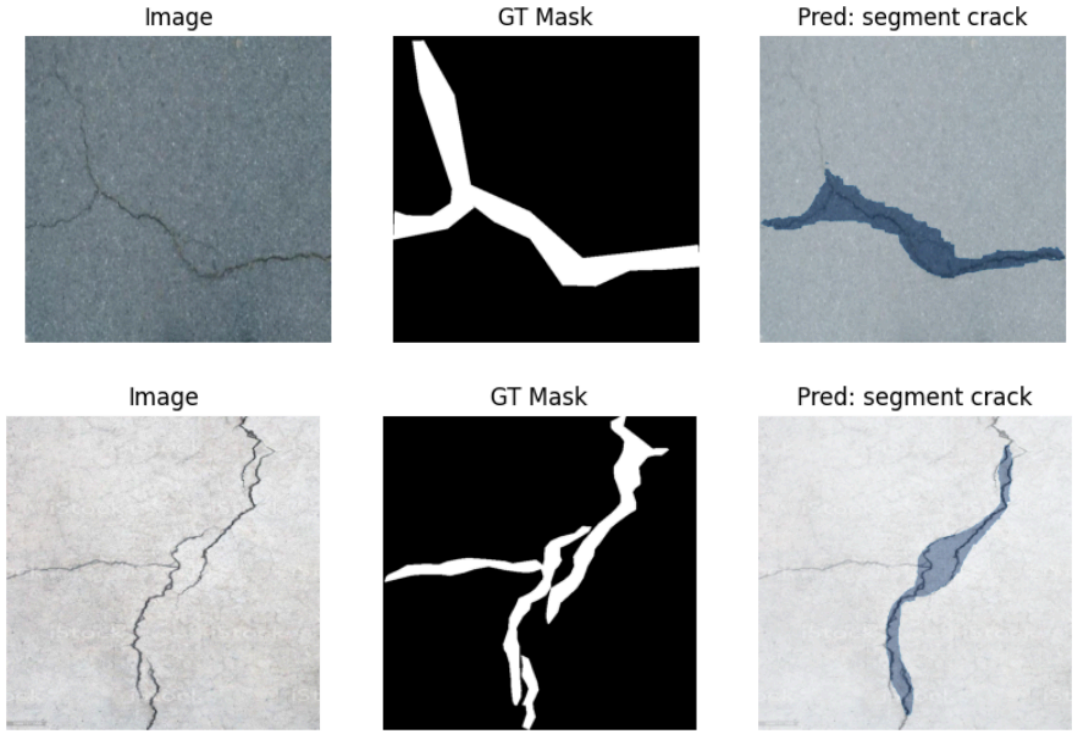
   V1: Warmup epochs - 5
   Full training epochs - 15

   Loss: Dice +Focal BCE
   Image size - 576
   LR scheduler - CosineAnnealingLR

   Runtime per epoch - 20 min
   Avg inference time/image - 1 second

```
Epoch 1/40:   0%|          | 0/1858 [00:00<?, ?it/s]huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling par
allelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
/tmp/ipykernel_114/4208557573.py:42: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with torch.cuda.amp.autocast(enabled=scaler is not None):
Epoch 1/40:  54%|█████     | 1000/1858 [09:53<08:28,  1.69it/s]
  Batch 1000/1858
Epoch 1/40: 100%|██████████| 1858/1858 [18:22<00:00,  1.68it/s]
Epoch 1/40 | Train loss: 0.4540
Validation Epoch 1:   0%|          | 0/68 [00:00<?, ?it/s]huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabli
ng parallelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
Validation Epoch 1: 100%|██████████| 68/68 [01:09<00:00,  1.02s/it]
Unused or unrecognized kwargs: truncation, padding.
/tmp/ipykernel_114/1531635077.py:21: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detac
h().requires_grad_(True), rather than torch.tensor(sourceTensor).
  mask_t = torch.tensor(mask).unsqueeze(0).to(device)
Val Loss: 0.5281 | Val mIoU: 0.3858 | Val Dice: 0.5265
```

Some results while training



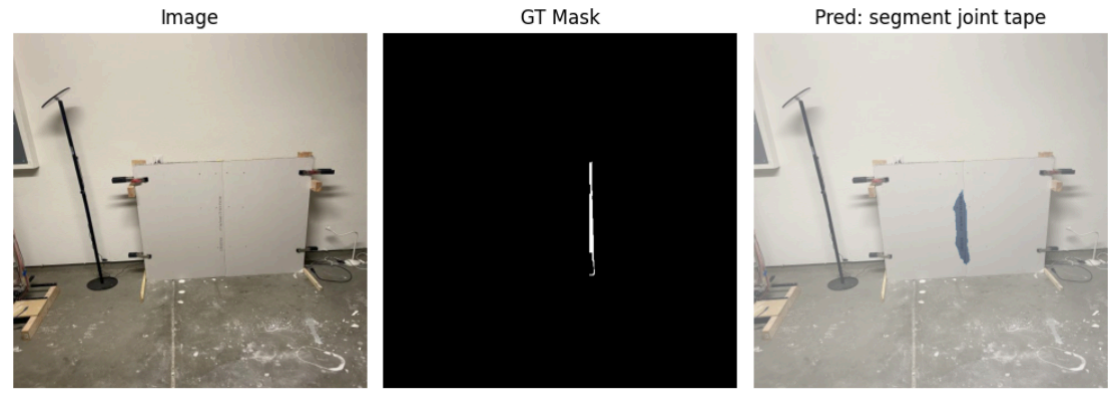


Image
GT Mask
Pred: segment joint tape

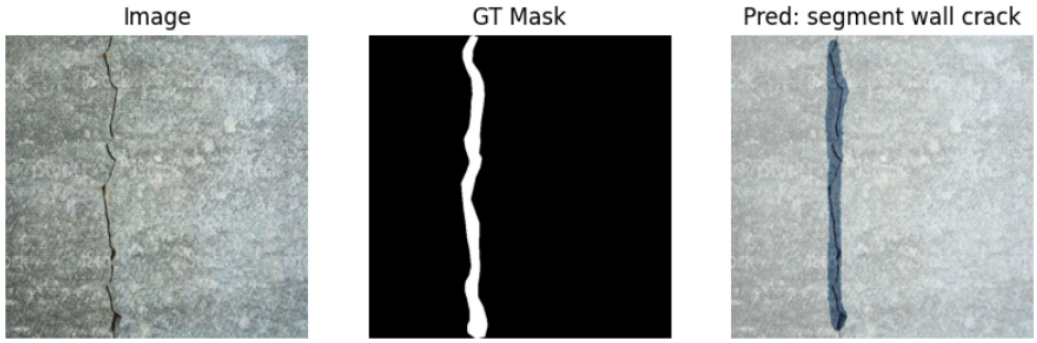


Image
GT Mask
Pred: segment wall crack

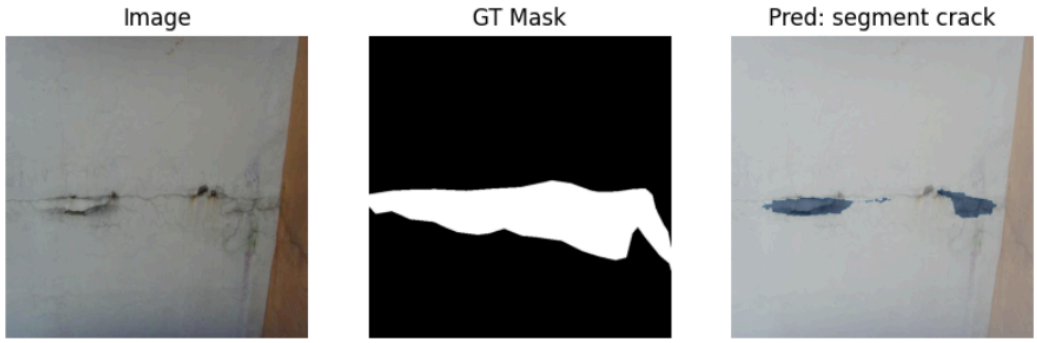
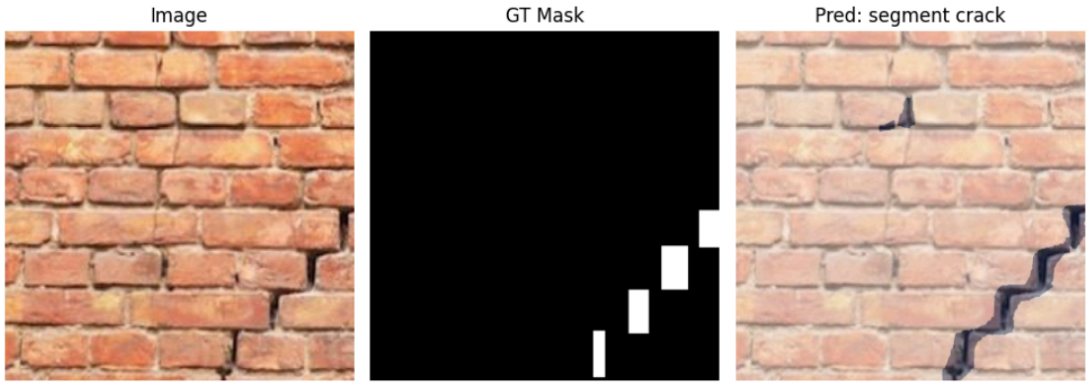
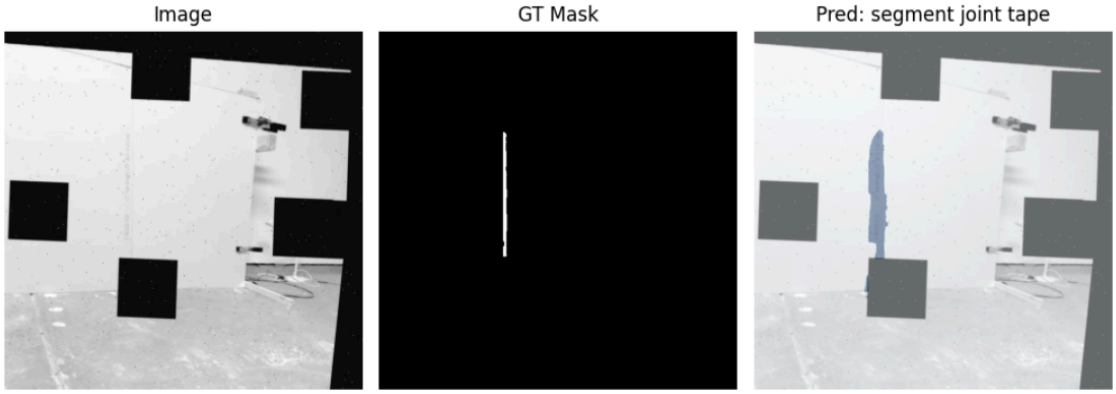


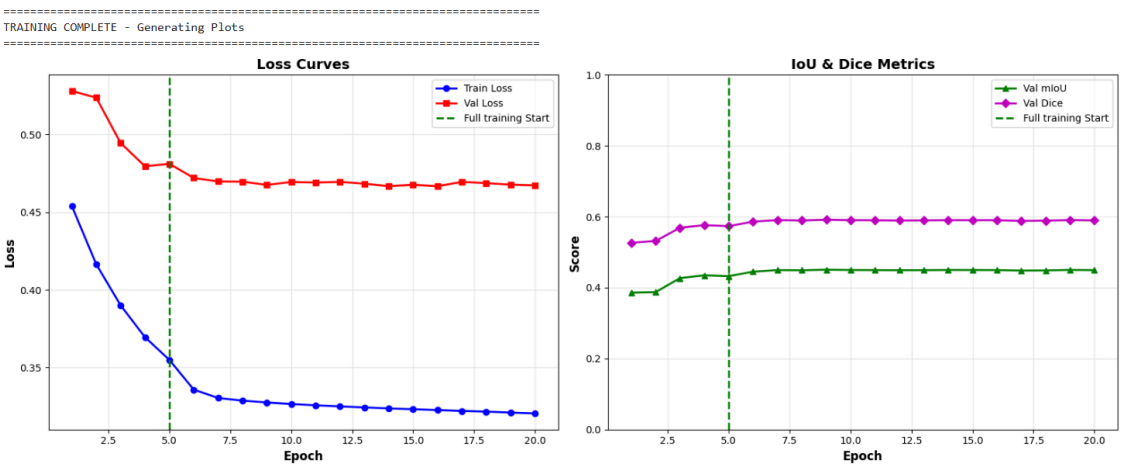
Image
GT Mask
Pred: segment crack





Failure Notes:
Size mismatch, python libraries not compatible, etc

7. Loss plots and metrics on validation set:




```
BEST VALIDATION METRICS
Best mIoU:       0.4507 at epoch 9
Best Dice:       0.5916 at epoch 9
Lowest Val Loss: 0.4668 at epoch 16
```

8. Uploaded models to HuggingFace:

V0 - https://huggingface.co/RaviKush/clipseg_finetuned_dice_bce

V1: https://huggingface.co/RaviKush/clipseg_focal_loss_v1

- Ravi Kumar Kushawaha