Detachable Encoder Transformer for SCR detection

Ashuta Bhattarai, Chandra Kambhamettu, Jing Jin
March 15, 2023

Goals

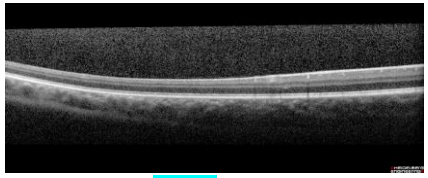
1. Improve SCR detection accuracy in complex cases
2. Look at multiple adjacent b-scans at once to confirm SCR
3. Train the deep learning model to learn where to look in the image while comparing the adjacent b-scans

Two step method

- Step 1: Pre-training
- Step 2: Detection

Pre-training

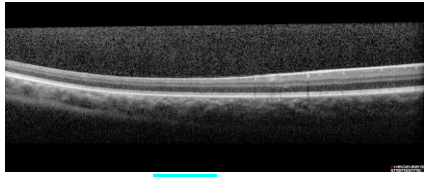
- Primary goal: Train the deep learning model to learn where to look in the image while comparing the adjacent or successive b-scans
- Hypothesis: A deep learning model (M) can be trained such that, given two input b-scans (i1 and i2), M identifies with reasonable accuracy whether i1 and i2 are successive b-scans from the same OCT or they are unrelated b-scans from different patients.
- This process is self-supervised because we don't need to provide annotations. The annotations can be generated from the dataset itself.
- Why this works:
    - Adjacent b-scans from the same OCT tests are structurally similar to each other but b-scans from different people and different OCT tests are structurally different.



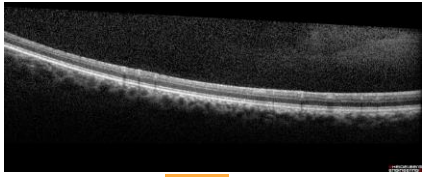
121R1006

121R1008

VS.



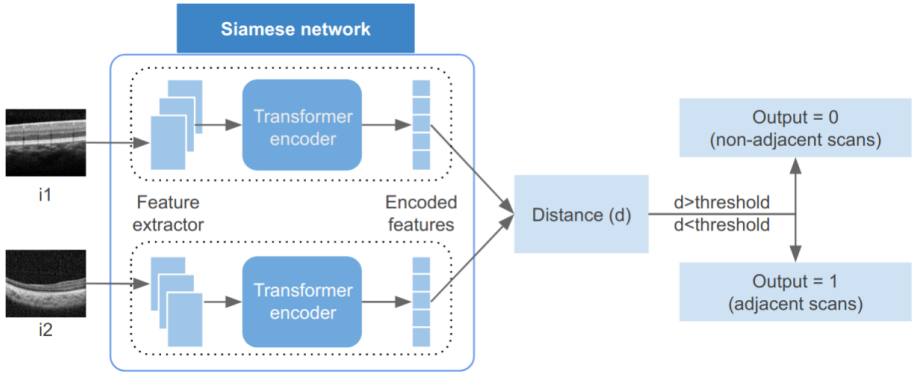
121R1006



106R1006

Pre-training

- Procedure:
  - Train a siamese (twin) network.
    - Contains transformer encoder: The encoder generates embeddings vectors that represents the most salient features of the image.
  - The input images go through a siamese network that outputs a learned feature vector (v1 and v2) for each input.
  - Calculate distance between v1 and v2
  - Threshold the distance to get binary outputs ‘0’ and ‘1’. ‘0’ indicates that the input images were from different people and ‘1’ indicates that the input images were adjacent or successive to each other.



Pre-training

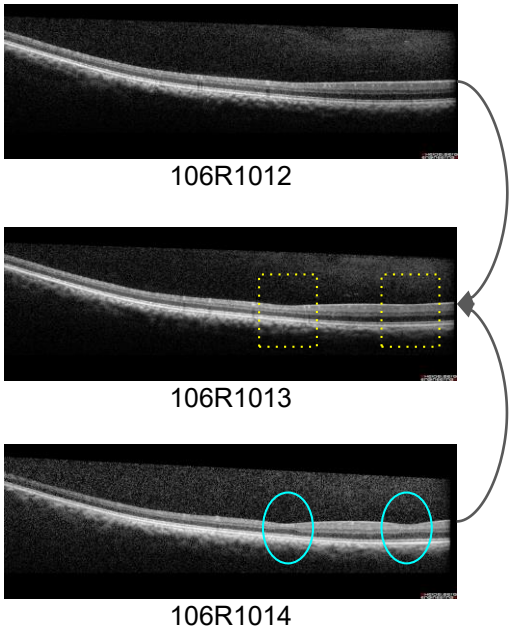
- Benefits:
  - Encoder learns where to look while comparing two b-scans
  - Encoder learns what features in the images make two b-scans similar
  - These qualities will be useful in the detection phase when this encoder will be used to identify SCR based on the features in adjacent b-scans.

- Test results:
  - Tested on 40,012 positive and 52,046 negative examples
  - Accuracy = 73.5%
  - Precision = 75.4%
  - Recall = 65.01%

		Ground truth	
		True	**False**
Prediction	**Positive**	28,231	9,203
	Negative	15,193	39,431

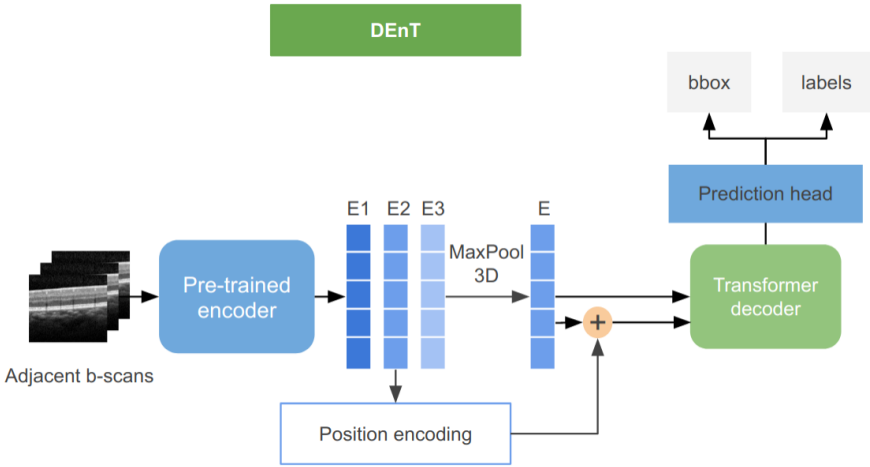
SCR Detection using Detachable Encoder Transformer (DEnT) network

- Primary goals:
    - Improve SCR detection accuracy in complex cases
    - Look at multiple adjacent b-scans at once to confirm SCR
- Hypothesis: A deep learning model (M) can be trained such that, given three adjacent b-scans (i1, i2, i3), M can detect SCR in i2 by encapsulating features from all three images: i1, i2 and i3.
- Why this works:
    - The experts confirm SCR by looking at specific locations in adjacent b-scans.



SCR detection

- Procedure:
  - Train a transformer-based object detection network. The network contains encoder and decoder blocks. The encoder block in the transformer can be replaced by the pre-trained encoder from step 1.
  - The pre-trained encoder generates three feature vectors (E1, E2, E3) for three input images i1, i2, i3.
  - An ensemble feature vector (E) is generated by combining E1, E2 and E3
  - The decoder queries E for significant SCR-like features in the positions or coordinates indicated by E2.
  - The decoder outputs bounding boxes and class predictions



SCR detection

- Benefits:
  - Using pre-trained encoder reduces training time since the encoder part does not need to be trained again
  - Features of the adjacent input images is utilized to predict outputs

- Test results:
  - Tested on 10,052 adjacent b-scans from 40 OCT volumes of 24 patients.
  - Precision = 81.0%
  - Recall = 74.0%
  - Comparison with existing state of the art networks showed superior performance.

	Precision	Recall
YOLO v5	0.76	0.68
Faster RCNN	0.78	0.64
DETR	0.79	0.74
DEnT	0.81	0.74