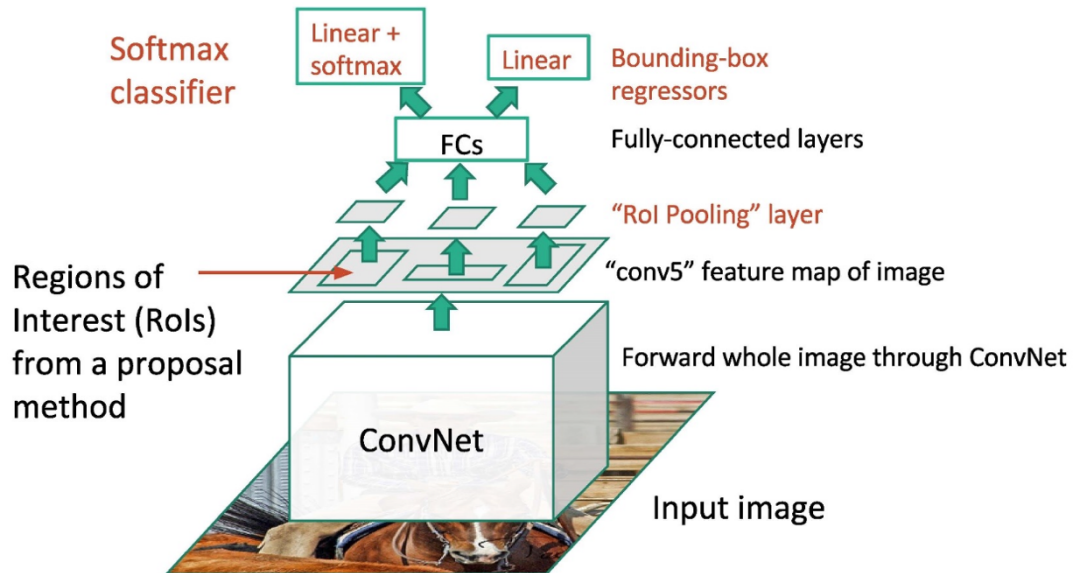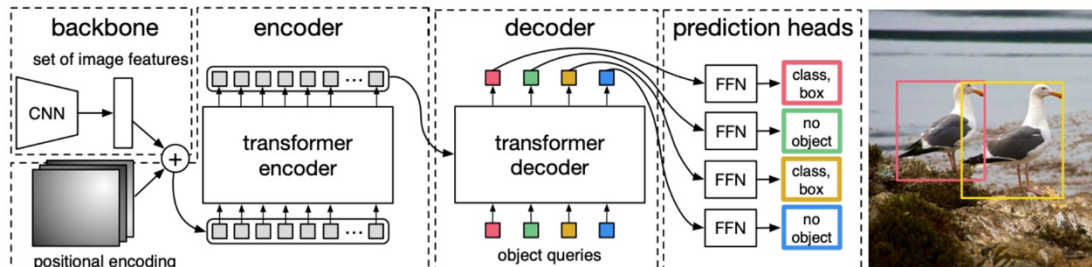1. (5%) Draw the architectures for both CNN-based and Transformer-based methods
   1. CNN-based model: FRCNN

   2. Transformer-based model: DETR

2. (10%) Report and compare the performance of two methods on validation set

Models/Acc	mAP@[50:5:95]	mAP@50	mAP@75
FRCNN	0.4183	0.7187	0.4345
DETR	0.2662	0.5377	0.2108

3. (10%) Report the implementation details of both methods
   1. CNN-based:

      • Data augumentation: horizaonal flip rate: 0.5, vertical flip rate: 0.5, color jitter rate: 0.2
      • Loss function: SGD

- Learning rate: 1e-3, Weight decay: 1e-4,
- Epochs: 50, early stop: 25 epochs, train / valid batch size: 1
- Cross validation: 0.2

2. Transformer-based
- Loss function: AdamW
- Learning rate: 1e-4, Backbone learning rate: 1e-4, Weight decay: 1e-5
- Total steps: 10000, train / valid batch size: 2
- accumulate grad batches: 4
- Cross validation: 0.2

4. (5%) Visualization: draw the bounding boxes of **two** methods on this **test** image.

- Left: CNN-bsased
- Right: Transformer-based

Reference

1. [FRCNN model](#)
2. [FRCNN totorial](#)
3. [data argument](#)

4. [DETR model](#)
5. [DETR totorial](#)