

## Appendix

### A Detailed Architecture

#### A.1 Image Encoder

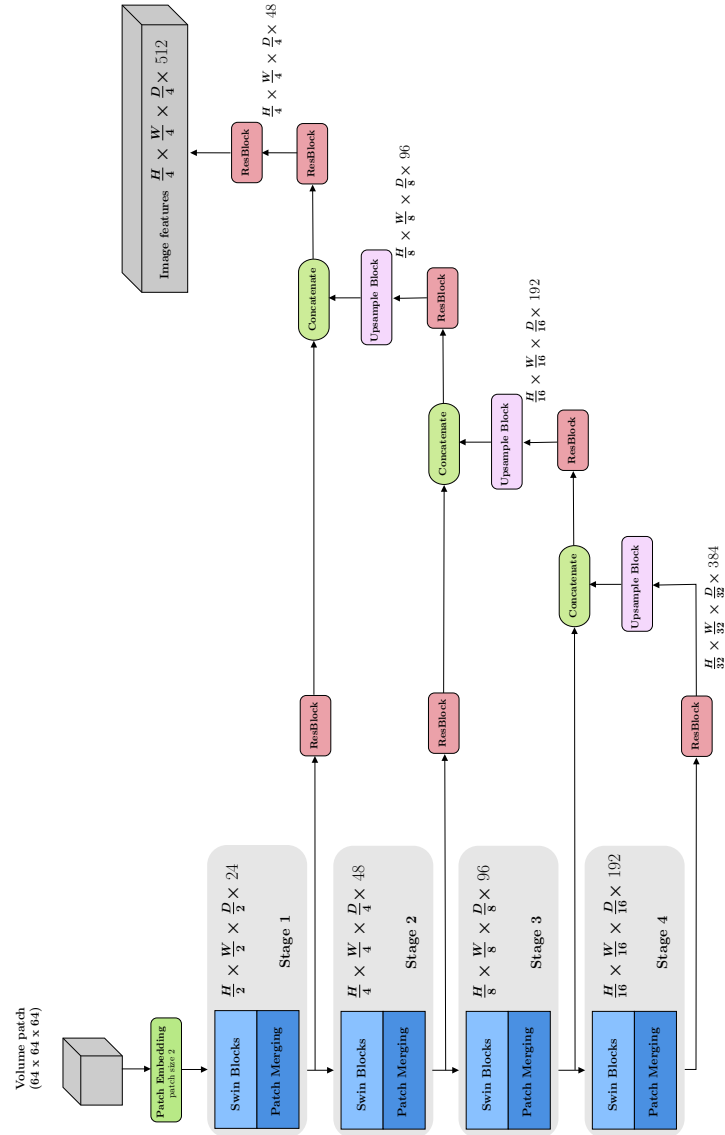


Fig. 1: Detailed architecture of the image encoder.

## A.2 DETR Decoder

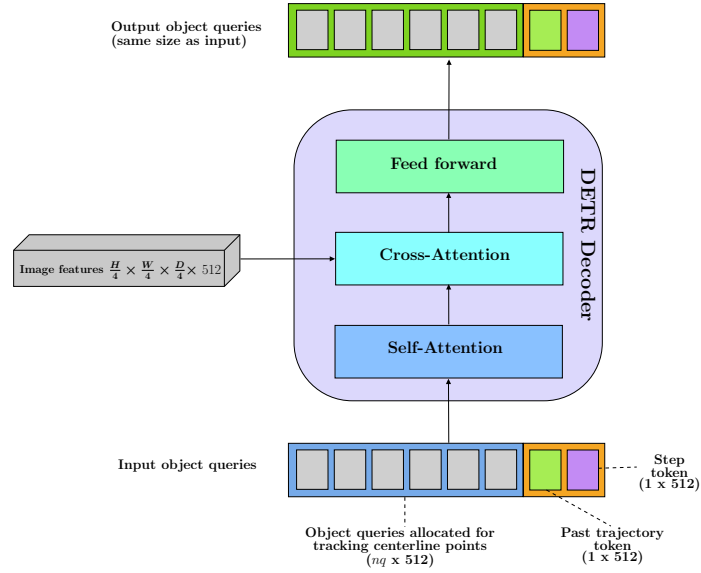


Fig. 2: Detailed architecture of the DETR decoder. Here  $nq$  is the number of object queries which is set dynamically depending on the number of branches being tracked.

## A.3 Prediction Heads

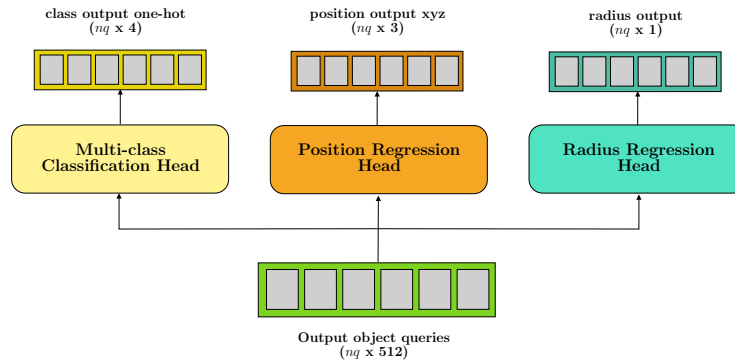


Fig. 3: Detailed architecture of the prediction heads.