In the original paper - <https://github.com/LeapLabTHU/Dynamic_Perceiver>

The architecture is implemented with ResNetY, MobileNet and ResNet

For this project, the backbone model is going to be YOLOV8 which supports both detection and segmentation pipeline.