The method is based on traditional ResNet [1]. Specifically, the data are resized to 512*384. The augmentation includes rotation and flip. There are no other special tricks.

[1] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.