post process

input image

CNN backbone (resnet101)

feature vector

W

H

3

feature map

H_1

W_1

2048
2048

gem pooling

norm