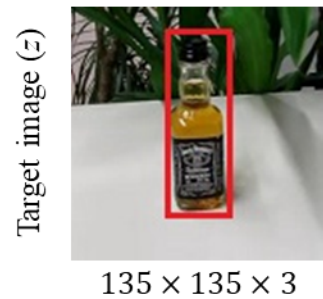
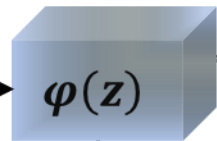


Stacked channel-spatial attention block (4)

Feature extractor (z)

Channel attention

Spatial attention

Skip connection

 $7 \times 7 \times 512$ 

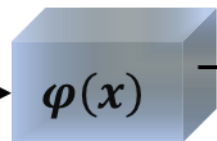
Elementwise multiplication



Elementwise summation



Cross-correlation

Feature extractor (x) $23 \times 23 \times 512$

$17 \times 17 \times 1$
Response map

