



$3 \times H \times W$

CNN backbone
up to stage $L - 1$
(total stride 16)

$256 \times H / 16 \times W / 16$

Stage L

$512 \times H / 32 \times W / 32$

scoring

$21 \times H / 32 \times W / 32$

Bilinear
Upsampling $\times 2$

$21 \times H / 16 \times W / 16$

scoring

+

Bilinear
Upsampling $\times 16$

$21 \times H / 16 \times W / 16$



$21 \times H \times W$