

Input image (RGB)

$$[H \times W \times 3]$$

Layer 1 feature maps

$$[H/4 \times W/4 \times C_1]$$

Layer 2 feature maps

$$[H/8 \times W/8 \times C_2]$$

