

Part	Input \rightarrow Output Shape	Layer Information
Transform Layer	$(\frac{h}{4}, \frac{w}{4}, 256 + c_i) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
Residual Blocks	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 256)$	CONV-(N256, K3, S1, P1), IN, ReLU
Mask Prediction Layer	$(\frac{h}{4}, \frac{w}{4}, 256) \rightarrow (\frac{h}{4}, \frac{w}{4}, 1)$	CONV-(N1, K7, S1, P3), TanH