

Module	Details
Embedding ($\hat{f} \rightarrow \bar{f}$)	$\begin{bmatrix} \text{conv}1 \times 1 \\ \text{BN} \\ \text{ReLU} \\ \text{Dropout} \end{bmatrix} \times 3$
Embedding ($\hat{\mathcal{P}} \rightarrow \bar{\mathcal{P}}$)	$\begin{bmatrix} \text{conv}1 \times 1 \\ \text{ReLU} \end{bmatrix} \times 3$
Spatial attention mask ($\bar{\mathcal{P}} \rightarrow \mathcal{P}^n$)	$\begin{bmatrix} \text{conv}1 \times 1, 256 \\ \text{ReLU} \\ \text{conv}1 \times 1, 128 \\ \text{ReLU} \\ \text{conv}1 \times 1, 1 \end{bmatrix}$

Table 1: Building blocks of our SAPNet. We use $\text{conv}1 \times 1$ to enable efficient computations.