$\mathcal{I}$ $\mathcal{I} \odot \text{mask}(l_{1:t})$ $g_{\text{AVP}}(\theta | \mathbf{y}_{1:t}, l_{1:t})$

$l_{1:t}$ CNN

$\text{mask}(l_{1:t})$ $\text{embedding}(\mathcal{I}, l_{1:t})$