

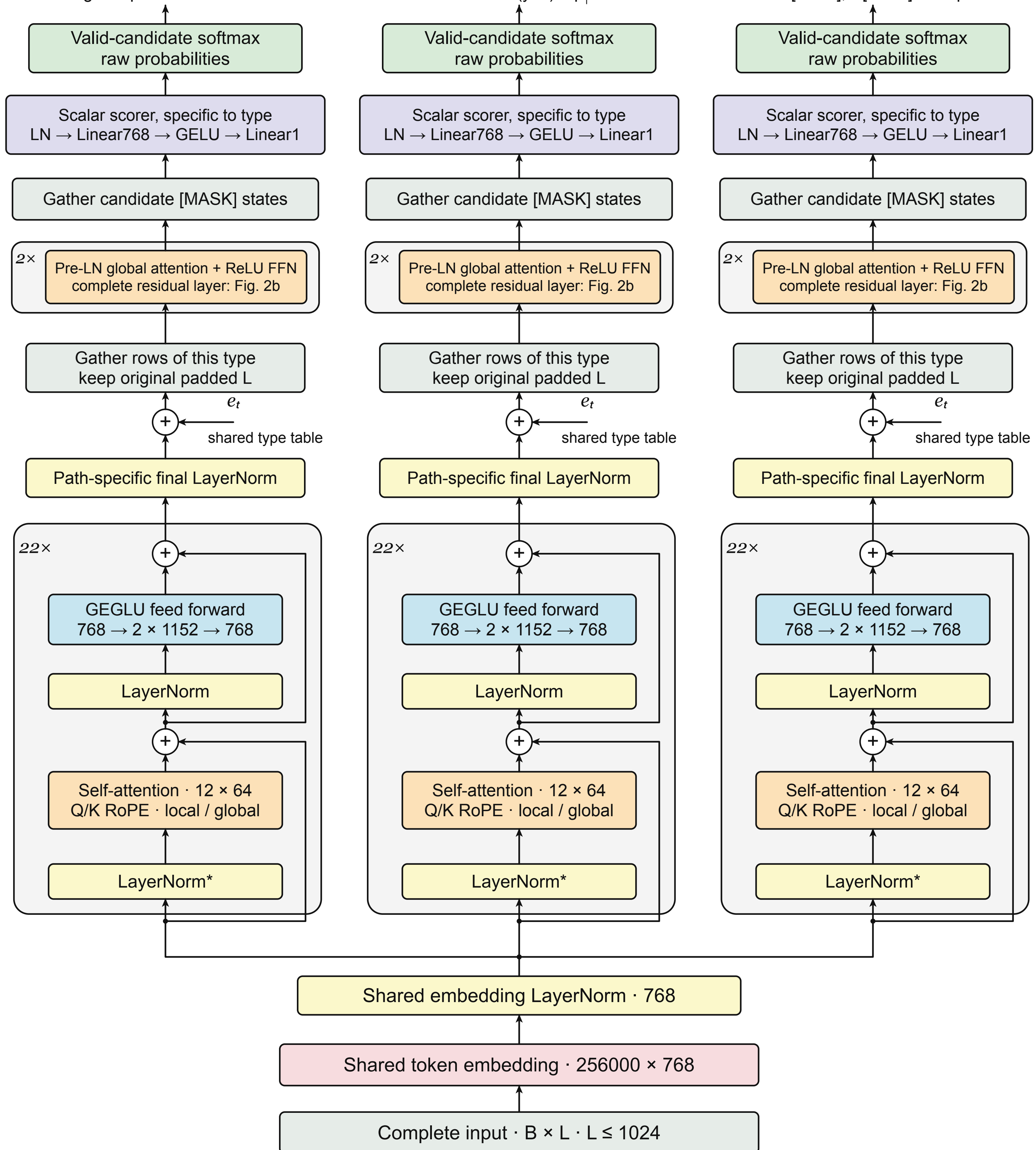
Choice

 $\text{argmax } p \rightarrow \text{candidate ID}$

Noul

 $K = 2 \cdot P(\text{yes}) = p_1$

Score

 $E[\text{index}], E[\text{value}] \text{ from } p$ Each needed encoder keeps full $B \times L$; only interaction heads gather type rows.571,909,635 parameters $\cdot$ dynamic candidates $\cdot$ complete 1K input

* Layer 1 attention normalization is Identity in every encoder path.