Hard Gate: $G_x^f = \mathbf{1}_{\{q_x^f > 0\}}$ **or Soft Gate:** $G_x^f = 1/(1 + \exp(-\beta \cdot q_x^f))$

copy gates

$x \rightarrow$ | $q_x^f \rightarrow$ (ReLU) $\rightarrow \max(0, q_x^f) \rightarrow \hat{y}_{\Theta^f}(x)$

$x \rightarrow$ | $q_x^v \rightarrow (\cdot) \rightarrow q_x^v \cdot G_x^f \rightarrow \hat{y}_{\Theta^{DGN}}(x)$

Gating/Feature Network : Θ^f

Weight/Value Network : Θ^v