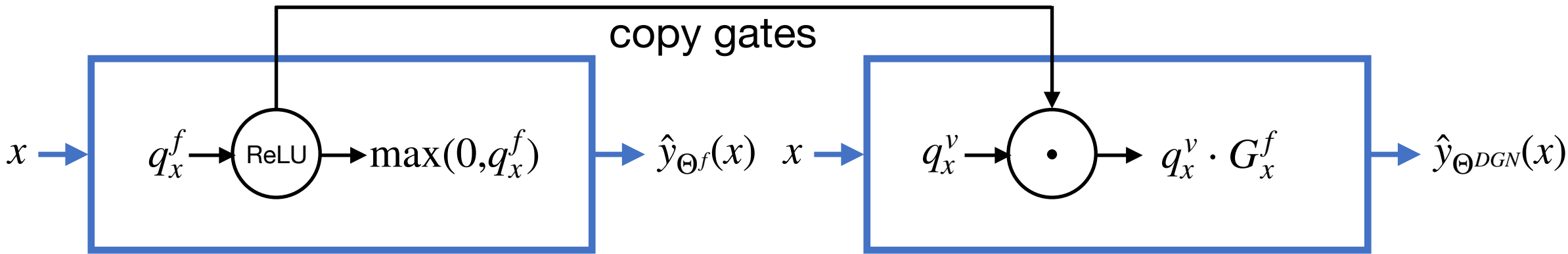


Hard: $G_x^f = \mathbf{1}_{\{q_x^f > 0\}}$ **or Soft:** $G_x^f = 1/(1 + \exp(-\beta \cdot q_x^f))$



Gating/Feature Network

Weight/Value Network