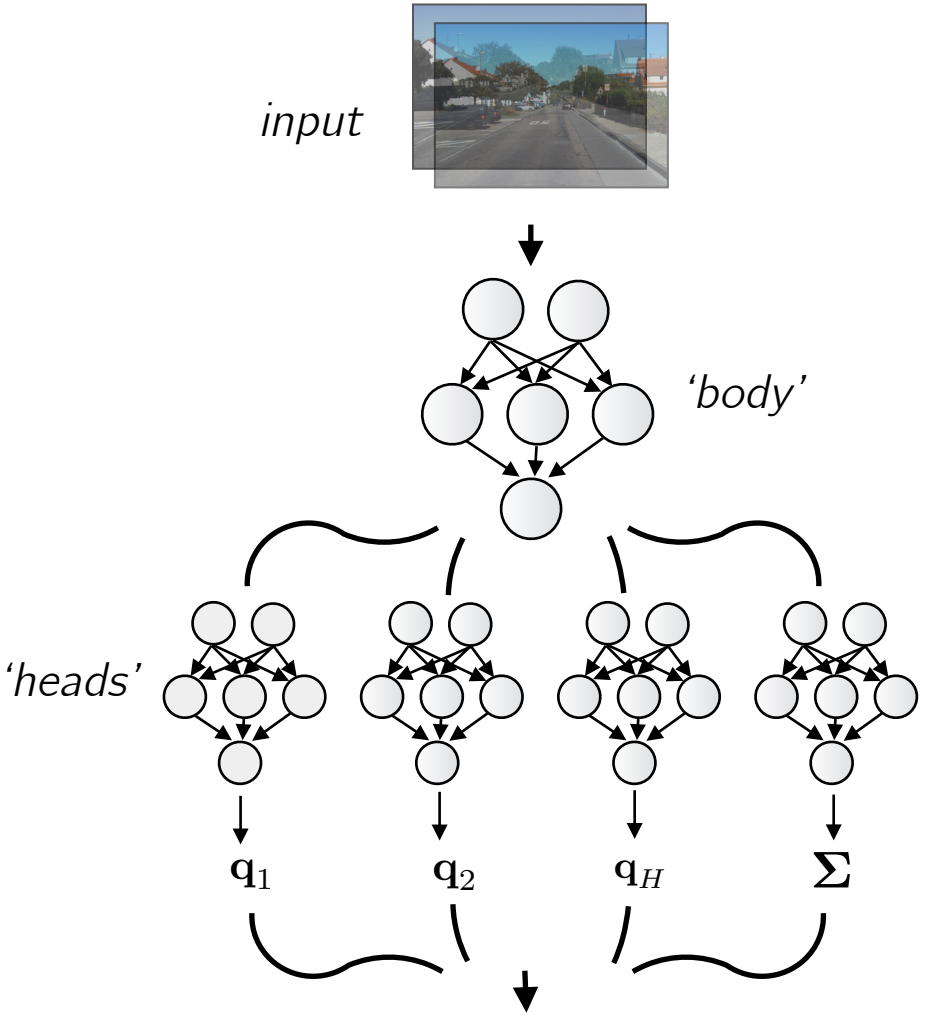

1. Evaluate loss for H heads: $\mathcal{L} = \sum_{h=1}^{H} \mathcal{L}_{\text{NLL}}(\mathbf{q}_h, \mathbf{q}_t, \boldsymbol{\Sigma})$

$$\mathcal{L}_{\text{NLL}}(\mathbf{q}, \mathbf{q}_t, \boldsymbol{\Sigma}_a) = \frac{1}{2}\boldsymbol{\phi}^{\mathsf{T}}\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\phi} + \frac{1}{2}\log\det(\boldsymbol{\Sigma}_a),$$

$$\boldsymbol{\phi} = \text{Log}\left(\mathbf{q} \otimes \mathbf{q}_t{}^{-1}\right)$$