Gradients
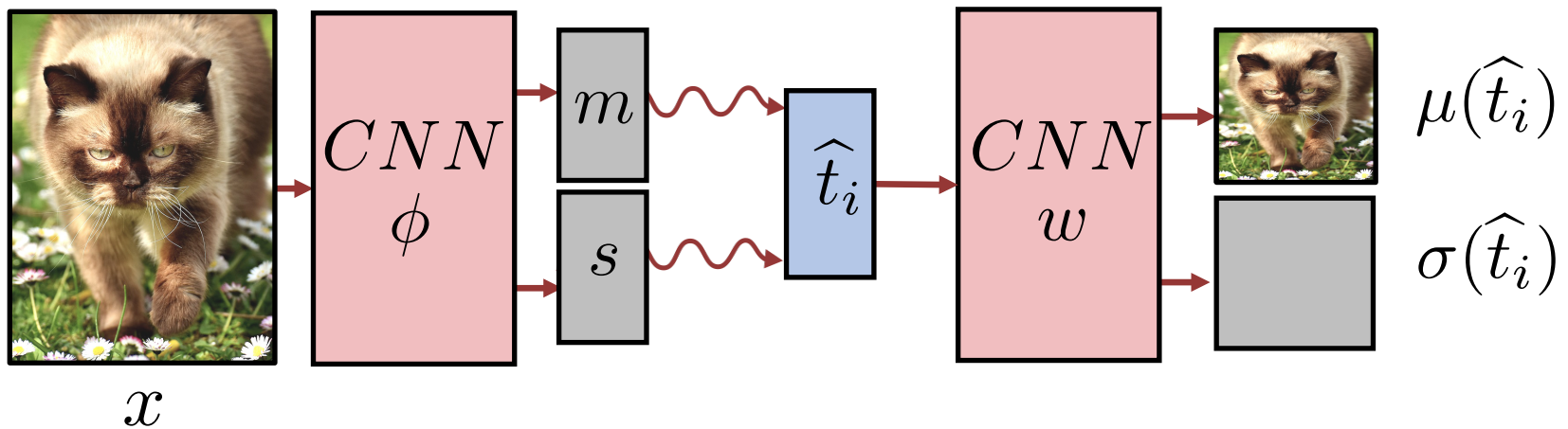
$$\nabla_w f(w, \phi) \approx \sum_{i=1}^{N} \nabla_w \log p(x_i \mid \widehat{t}_i, w)$$

$$\approx \frac{N}{n} \sum_{s=1}^{n} \nabla_w \log p(x_{i_s} \mid \widehat{t}_{i_s}, w)$$

$$\widehat{t}_i \sim \mathcal{N}(m(x_i, \phi), \mathrm{diag}(s^2(x_i, \phi)))$$