Problem 1.

$$L = -\log\left(\exp(y_{gt}) / \Sigma_{i=1}^{n} \exp(y_i) \right)$$

$$= -\log(\exp(y_{gt})) + \log(\Sigma_{i=1}^{n} \exp(y_i))$$

if $i = gt$ then

$$L = -\log(\exp(y_i)) + \log(\Sigma_{i=1}^{n} \exp(y_i))$$

$$\frac{\partial L}{\partial y_i} = -1 + \frac{\exp(y_i)}{\Sigma_{i=1}^{n} \exp(y_i)} = -1 + p_i$$

if $i \neq gt$ then

$$L = -\log(\exp(y_{gt})) + \log(\Sigma_{i=1}^{n} \exp(y_i))$$

$$\frac{\partial L}{\partial y_i} = \frac{\exp(y_i)}{\Sigma_{i=1}^{n} \exp(y_i)} = p_i$$

so

$$\frac{\partial L}{\partial y_i} = \begin{cases} -1 + p_i, & i = gt \\ p_i, & i \neq gt \end{cases}$$

Problem 2.

$$z^{(t)} = Wx^{(t)} + Vh^{(t-1)} + b$$
$$h^{(t)} = \tanh(z^{(t)})$$

$$\nabla_b L = \sum_t \left(dh^{(t)} / db^{(t)} \right)^T \nabla_{h^{(t)}} L$$

$$= \sum_t \mathrm{diag}(1-(h(t))^2)\, \nabla_{h^{(t)}} L$$

$$\nabla_{x^{(t)}} L = \frac{dL}{dh^{(t)}} \frac{dh^{(t)}}{dx^{(t)}} = \sum_t \left(dh^{(t)} / dx^{(t)} \right)^T \nabla_{h^{(t)}} L$$

$$= W^T \nabla_{h^{(t)}} L \sum_t \mathrm{diag}(1-(h^{(t)})^2)$$

$$\nabla_{h^{(t-1)}} L = \frac{dL}{dh^{(t)}} \cdot \frac{dh^{(t)}}{dh^{(t-1)}} = \sum_t \left(dh^{(t)} / dh^{(t-1)} \right)^T \nabla_{h^{(t)}} L$$
$$= V^T \nabla_{h^{(t)}} L \sum_t \mathrm{diag}(1-(h(t))^2)$$

$$dL/dW = \frac{dL}{dh(t)} \frac{dh(t)}{dW} = \sum_t \sum_i \left(\frac{dL}{dh_i^{(t)}} \right) \nabla_{W^t} h_i^{(t)}$$

$$= \sum_t \mathrm{diag}(1-(h^{(t)})^2)\, \nabla_{h^{(t)}} L \; x^{(t)T}$$

$$dL/dV = \frac{dL}{dh(t)} \frac{dh(t)}{dV} = \sum_t \mathrm{diag}(1-(h^{(t)})^2)\, \nabla_{h^{(t)}} L \; h^{(t-1)T}$$

Problem 3.

$$y(k,i,j) = \sum_{t=0}^{T-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(t, i \times s + m, j \times s + n)\, W_k(t,m,n) + b_k$$

$$\frac{dL}{db_k} = \frac{dy(k,i,j)}{db_k} \cdot \frac{dL}{dy(k,i,j)}$$

$$= \cdot 1 \cdot \frac{dL}{dy(k,i,j)} \qquad = \cdot \frac{dL}{dy(k,i,j)}$$

$$\frac{dL}{dW(t,m,n)} = \sum_i \sum_j \sum_k \frac{dy(k,i,j)}{dW(t,m,n)} \times \frac{dL}{dy(k,i,j)}$$

$$= \sum_i \sum_j X(t, i \times s + m, j \times s + n) \cdot \frac{dL}{dy}$$

$$\frac{dL}{dx(0,p,q)} = \sum_i \sum_j \sum_k \frac{dy}{dx} \cdot \frac{dL}{dy}$$

$$0 = t \qquad p = i \times s + m \implies m = p - i \times s$$

$$q = j \times s + n \implies n = q - j \times s$$

$$\frac{dL}{dw} = \sum_i \sum_j w(0, p - i \times s, q - j \times s) \cdot \frac{dL}{dy}$$