4.(a) $$\frac{\partial y_L}{\partial y_{L-1}} = \frac{\partial}{\partial y_{L-1}}\left(A_{w_L} y_{L-1} + b_L \underline{1}_{n_L}\right)$$

$$= A_{w_L} \checkmark$$

$$\left(\frac{\partial y_\ell}{\partial y_{\ell-1}}\right)_{ij} = \frac{\partial}{\partial (y_{\ell-1})_j}\left(\sigma\left(\sum_k (A_{w_\ell})_{ik}(y_{\ell-1})_k + b_\ell\right)\right)$$

$$= (A_{w_\ell})_{ij} \cdot \sigma_i'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)$$

$$\Rightarrow \frac{\partial y_\ell}{\partial y_{\ell-1}} = \mathrm{diag}\left(\sigma'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)\right) A_{w_\ell} \quad \blacksquare \checkmark$$

(see HW4.6(a) equivalent workings)

$$\left(\frac{\partial y_L}{\partial \underline{w}_\ell}\right)_{1,i} = \sum_{k=1}^{n_\ell}\left(\frac{\partial y_L}{\partial y_\ell}\right)_{1,k}\left(\frac{\partial y_\ell}{\partial \underline{w}_\ell}\right)_{k,i} \qquad i = 1, \dots, f_\ell$$

$$= \sum_{k=1}^{n_\ell}\left(\frac{\partial y_L}{\partial y_\ell}\right)_{1,k} \sigma_k'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)(y_{\ell-1})_{k+i-1}$$

$$= \sum_{k=1}^{n_{\ell-1}-f_\ell+1}\left(\frac{\partial y_L}{\partial y_\ell}\right)_{1,k} \sigma_k'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)(y_{\ell-1})_{k+i-1}$$

$$\dots ?$$ (see revised below)

$$= \sum_{m=1}^{n_{\ell-1}} (C_{v_\ell^T})_{im}(y_{\ell-1})_m = (C_{v_\ell^T} y_{\ell-1})_i$$

$$= (C_{v_\ell^T} y_{\ell-1})^T_{1,i}$$

$\frac{\partial y_L}{\partial y_\ell} \in \mathbb{R}^{1\times n_\ell}$

$\mathrm{diag}(\sigma'(\dots)) \in \mathbb{R}^{n_\ell \times n_\ell}$

$y_{\ell-1} \in \mathbb{R}^{n_{\ell-1}\times 1}$

$v_\ell \in \mathbb{R}^{1\times n_\ell}$

$v_\ell^T \in \mathbb{R}^{n_\ell \times 1}$

$C_{v_\ell^T} \in \mathbb{R}^{f_\ell \times n_{\ell-1}}$

$\frac{\partial y_L}{\partial \underline{w}_\ell} \in \mathbb{R}^{1\times f_\ell} \ni (C_{v_\ell^T} y_{\ell-1})^T$

$C_{v_\ell^T} y_{\ell-1} \in \mathbb{R}^{f_\ell \times 1}$

$$\left(\frac{\partial y_\ell}{\partial \underline{w}_\ell}\right)_{ij} = \frac{\partial (y_\ell)_i}{\partial (\underline{w}_\ell)_j} = \frac{\partial}{\partial (\underline{w}_\ell)_j}\left(\sigma\left(\sum_{k=1}^{n_{\ell-1}} (A_{w_\ell})_{ik}(y_{\ell-1})_k + b_\ell\right)\right)$$

$$= \delta_{k,i+j-1}(y_{\ell-1})_k \, \sigma_i'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)$$

$$= (y_{\ell-1})_{i+j-1} \, \sigma_i'\left(A_{w_\ell} y_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)$$

$$\frac{\partial (A_{w_\ell})_{ik}}{\partial (\underline{w}_\ell)_j} = \delta_{k,i+j-1}$$

$(w_\ell)_j$ located in positions in A_{w_ℓ}: $(1,j)$, $(2,j+1)$, $\vdots$, $(i, j+(i-1))$

$$A_{w_\ell} = \begin{pmatrix} (w_\ell)_1 & \dots & (w_\ell)_{f_\ell} & 0 & & 0 \cdots 0 \\ 0 & (w_\ell)_1 & \dots & & (w_\ell)_{f_\ell} & 0 \cdots 0 \\ 0 & & \ddots & & \ddots & \vdots \\ \vdots & & & & & 0 \\ 0 & & 0 & (w_\ell)_1 & \dots & (w_\ell)_{f_\ell} \end{pmatrix}$$

(as in HW1.5)
(or Chp3 Slide 19)

(a) (cont.) $\dfrac{\partial y_L}{\partial b_\ell} = \dfrac{\partial y_L}{\partial y_\ell}\dfrac{\partial y_\ell}{\partial b_\ell}$

$$\left(\frac{\partial y_\ell}{\partial b_\ell}\right)_i = \frac{\partial}{\partial b_\ell}\left(\sigma\left(\sum_k^{n_\ell}(A_{w_\ell})_{ik}(y_{\ell-1})_k + b_\ell\right)\right)$$

$$= \sigma_i'\left(A_{w_\ell} y_{\ell-1} + b_\ell \mathbf{1}_{n_\ell}\right)$$

$$= \frac{\partial y_L}{\partial y_\ell}\,\mathrm{diag}\left(\underline{\sigma}'(A_{w_\ell} y_{\ell-1} + b_\ell \mathbf{1}_{n_\ell})\right)\mathbf{1}_{n_\ell}$$

$$= \underline{v}_\ell \mathbf{1}_{n_\ell} \quad \square$$

(b) In the forward pass, matrix vector products with A_{w_i} are used to perform convolutions (for $A_{w_i}\underline{y}_{i-1}$)
In backpropagation, calculations will require updating $\dfrac{\partial y_L}{\partial y_{\ell-1}}$ using right multiplication by A_{w_ℓ}, which should use $A_{w_i}^T$ in transpose-convolutions.
Calculations will also require $\dfrac{\partial y_L}{\partial w_\ell}$ which uses the convolutional operator (i.e. matrix) $C_{v_\ell^T}$. This should be performed using convolution since part of a matrix-vector product. Note that computing $C_{v_\ell^T}$ (by extension $\underline{v}_\ell$) will req. transpose convolutions (for $\dfrac{\partial y_L}{\partial y_\ell}$) and regular convolutions (for $A_{w_\ell} y_{\ell-1}$ in $\underline{\sigma}'$ function).

New (b) $A_{w_i}^T$ absent in both forward and backward passes; only A_{w_i} is used in forward & in $\dfrac{\partial y_L}{\partial b_\ell}, \dfrac{\partial y_L}{\partial w_\ell}$

For forward, store filter w_i and process convolutions as in HW1.
For back-prop, store v_ℓ in reverse order from $\ell = L$ to calculate $\dfrac{\partial y_L}{\partial b_\ell}$ and $\dfrac{\partial y_L}{\partial w_\ell}$ for $\ell = L, \dots, 1$

Revised 4.(a) $(\underline{y}_\ell)_i = \sigma_i\left(A_{\underline{w}_\ell}\underline{y}_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)$

$$\begin{array}{c} i=1 \\ i=2 \\ \vdots \\ i=n_\ell \end{array} \begin{pmatrix} w_1 \cdots w_{f_\ell} & 0 \cdots \\ 0\, w_1 \cdots w_{f_\ell-1}\, w_{f_\ell} & 0 \\ \vdots & \end{pmatrix} \begin{pmatrix} \underline{y}_{\ell-1} \\ \vdots \end{pmatrix}$$

convolution 'sliding' multiplication

$$= \sigma_i\left(\sum_{k=i+1}^{i+f_\ell} (\underline{w}_\ell)_{k-i} (\underline{y}_{\ell-1})_{k-1} + b_\ell\right)$$

$$\begin{array}{cccc} w_1 & w_2 & \cdots & w_{f_\ell} \\ & \times & & \\ (y_{\ell-1})_1 & (y_{\ell-1})_2 & \cdots & (y_{\ell-1})_{f_\ell} \\ 2 & 3 & & f_\ell+1 \end{array}$$

$$\left(\frac{\partial \underline{y}_\ell}{\partial \underline{w}_\ell}\right)_{ij} = \frac{\partial (\underline{y}_\ell)_i}{\partial (\underline{w}_\ell)_j} = \sigma_i'\left(A_{\underline{w}_\ell}\underline{y}_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)(\underline{y}_{\ell-1})_{i+j-1}$$

$j = k-i \Rightarrow k = i+j$

$$\Rightarrow \underbrace{\frac{\partial \underline{y}_\ell}{\partial (\underline{w}_\ell)_j}}_{\mathbb{R}^{n_\ell \times 1}} = \mathrm{diag}\left(\underline{\sigma}_i'\left(A_{\underline{w}_\ell}\underline{y}_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)\right)\begin{pmatrix} (\underline{y}_{\ell-1})_j \\ \vdots \\ (\underline{y}_{\ell-1})_{i+j-1} \\ \vdots \\ (\underline{y}_{\ell-1})_{n_\ell+j-1} \end{pmatrix}$$

$n_\ell = n_{\ell-1} - f_\ell + 1$

$j = 1, \dots, f_\ell$

so index along

$$\Rightarrow \underbrace{\frac{\partial \underline{y}_\ell}{\partial \underline{w}_\ell}}_{\mathbb{R}^{n_\ell \times f_\ell}} = \mathrm{diag}\left(\underline{\sigma}_i'\left(A_{\underline{w}_\ell}\underline{y}_{\ell-1} + b_\ell \underline{1}_{n_\ell}\right)\right) \begin{array}{c} \begin{array}{cccc} j=1 & j=2 & \cdots & j=f_\ell \end{array} \\ \begin{pmatrix} (\underline{y}_{\ell-1})_1 & (\underline{y}_{\ell-1})_2 & \cdots & (\underline{y}_{\ell-1})_{f_\ell} \\ \vdots & \vdots & & \vdots \\ (\underline{y}_{\ell-1})_i & (\underline{y}_{\ell-1})_{i+1} & \cdots & (\underline{y}_{\ell-1})_{i+f_\ell-1} \\ \vdots & \vdots & & \vdots \\ (\underline{y}_{\ell-1})_{n_\ell} & (\underline{y}_{\ell-1})_{n_\ell+1} & \cdots & (\underline{y}_{\ell-1})_{\underbrace{n_\ell+f_\ell-1}_{=n_{\ell-1}}} \end{pmatrix} \end{array}$$

By chain rule, $\underbrace{\frac{\partial y_L}{\partial \underline{w}_\ell}}_{\mathbb{R}^{1 \times f_\ell}} = \frac{\partial y_L}{\partial \underline{y}_\ell}\frac{\partial \underline{y}_\ell}{\partial \underline{w}_\ell}$

$\underline{v}_\ell \in \mathbb{R}^{1 \times n_\ell}$

$$= \underline{v}_\ell \begin{pmatrix} (\underline{y}_{\ell-1})_1 & (\underline{y}_{\ell-1})_2 & \cdots & (\underline{y}_{\ell-1})_{f_\ell} \\ \vdots & \vdots & & \vdots \\ (\underline{y}_{\ell-1})_i & (\underline{y}_{\ell-1})_{i+1} & \cdots & (\underline{y}_{\ell-1})_{i+f_\ell-1} \\ \vdots & \vdots & & \vdots \\ (\underline{y}_{\ell-1})_{n_\ell} & (\underline{y}_{\ell-1})_{n_\ell+1} & \cdots & (\underline{y}_{\ell-1})_{n_\ell+f_\ell-1} \end{pmatrix}$$

$$\left(\frac{\partial y_L}{\partial \underline{w}_\ell}\right)_{1,j} = \sum_{k=1}^{n_\ell} (\underline{v}_\ell^T)_k (\underline{y}_{\ell-1})_{j+k-1}$$

$$= \sum_{k=j+1}^{j+n_\ell} (\underline{v}_\ell^T)_{k-j} (\underline{y}_{\ell-1})_{k-1} = \left(C_{\underline{v}_\ell^T}\underline{y}_{\ell-1}\right)_{1,j} = \left(C_{\underline{v}_\ell^T}\underline{y}_{\ell-1}\right)_j^T$$

$$\Rightarrow \frac{\partial y_L}{\partial \underline{w}_\ell} = \left(C_{\underline{v}_\ell^T}\underline{y}_{\ell-1}\right)^T \text{ as required } \blacksquare$$