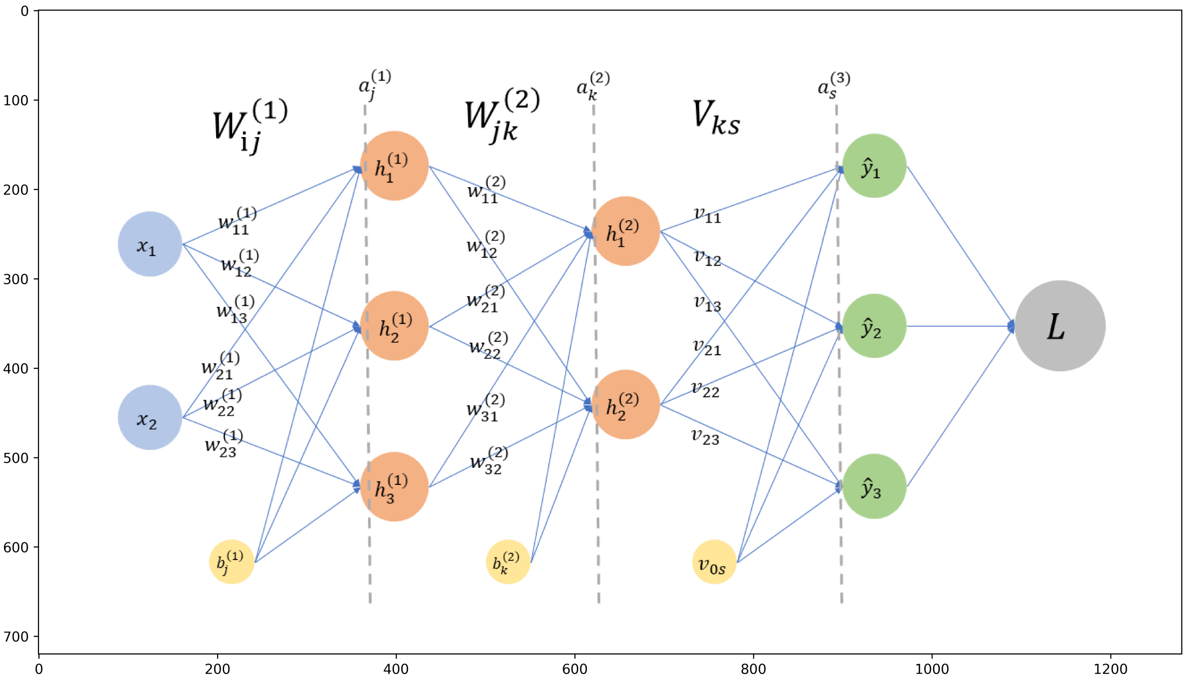
In [11]:
```
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline
import matplotlib as mpl
mpl.rcParams['figure.dpi']= 600
img=mpimg.imread("my_network.png")
plt.figure(figsize = (15,15))
plt.imshow(img)
```

Out[11]: <matplotlib.image.AxesImage at 0x2160824d780>



$$L(y,\hat{y}) = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 + y_3 \log \hat{y}_3$$

$$\frac{\partial L}{\partial V_{ks}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} \frac{\partial a_s^{(3)}}{\partial V_{ks}} \quad \text{, s=1,2,3}$$
$$= (1_{(s=true\ class)} - \hat{y}_s) h_k^{(2)}$$

$$\frac{\partial L}{\partial v_{0s}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} \frac{\partial a_s^{(3)}}{\partial v_{0s}} \quad \text{, s=1,2,3}$$
$$= \sum_s (1_{(s=true\ class)} - \hat{y}_s)$$

$$\frac{\partial L}{\partial W_{jk}^{(2)}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} \frac{\partial a_s^{(3)}}{\partial h_k^{(2)}} \frac{\partial h_k^{(2)}}{\partial a_k^{(2)}} \frac{\partial a_k^{(2)}}{\partial W_{jk}^{(2)}} \quad \text{, s=1,2,3; j = 1,2,3; k = 1, 2;}$$
$$= \sum_s (1_{(s=true\ class)} - \hat{y}_s) V_{ks} 1_{(a_k^{(2)}>0)} h_j^{(1)}$$

$$\frac{\partial L}{\partial b_k^{(2)}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} (\sum_k \frac{\partial a_s^{(3)}}{\partial h_k^{(2)}} \frac{\partial h_k^{(2)}}{\partial a_k^{(2)}} \frac{\partial a_k^{(2)}}{\partial b_k^2}) \quad \text{, s=1,2,3; j = 1,2,3; k = 1, 2;}$$
$$= \sum_s (1_{(s=true\ class)} - \hat{y}_s) (\sum_k V_{ks} 1_{(a_k^{(2)}>0)})$$

$$\frac{\partial L}{\partial W_{ij}^{(1)}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} (\sum_k \frac{\partial a_s^{(3)}}{\partial h_k^{(2)}} \frac{\partial h_k^{(2)}}{\partial a_k^{(2)}} \frac{\partial a_k^{(2)}}{\partial h_j^{(1)}} \frac{\partial h_j^{(1)}}{\partial a_j^{(1)}} \frac{\partial a_j^{(1)}}{\partial W_{ij}^{(1)}}) \quad \text{, s=1,2,3; j = 1,2,3; k = 1, 2; i = 1,2}$$
$$= \sum_s (1_{(s=true\ class)} - \hat{y}_s) (\sum_k V_{ks} 1_{(a_k^{(2)}>0)} W_{jk}^{(2)} 1_{(a_j^{(1)}>0)} x_i)$$

$$\frac{\partial L}{\partial b_j^{(1)}} = \sum_s \frac{\partial L}{\partial \hat{y}_s} \frac{\partial \hat{y}_s}{\partial a_s^{(3)}} (\sum_k \frac{\partial a_s^{(3)}}{\partial h_k^{(2)}} \frac{\partial h_k^{(2)}}{\partial a_k^{(2)}} (\sum_j \frac{\partial a_k^{(2)}}{\partial h_j^{(1)}} \frac{\partial h_j^{(1)}}{\partial a_j^{(1)}} \frac{\partial a_j^{(1)}}{\partial b_j^1})) \quad \text{, s=1,2,3; j = 1,2,3; k = 1, 2; i = 1,2}$$
$$= \sum_s (1_{(s=true\ class) - \hat{y}_s}) (\sum_k V_{ks} 1_{(a_k^{(2)}>0)} (\sum_j W_{jk}^{(2)} 1_{(a_j^{(1)}>0)}))$$