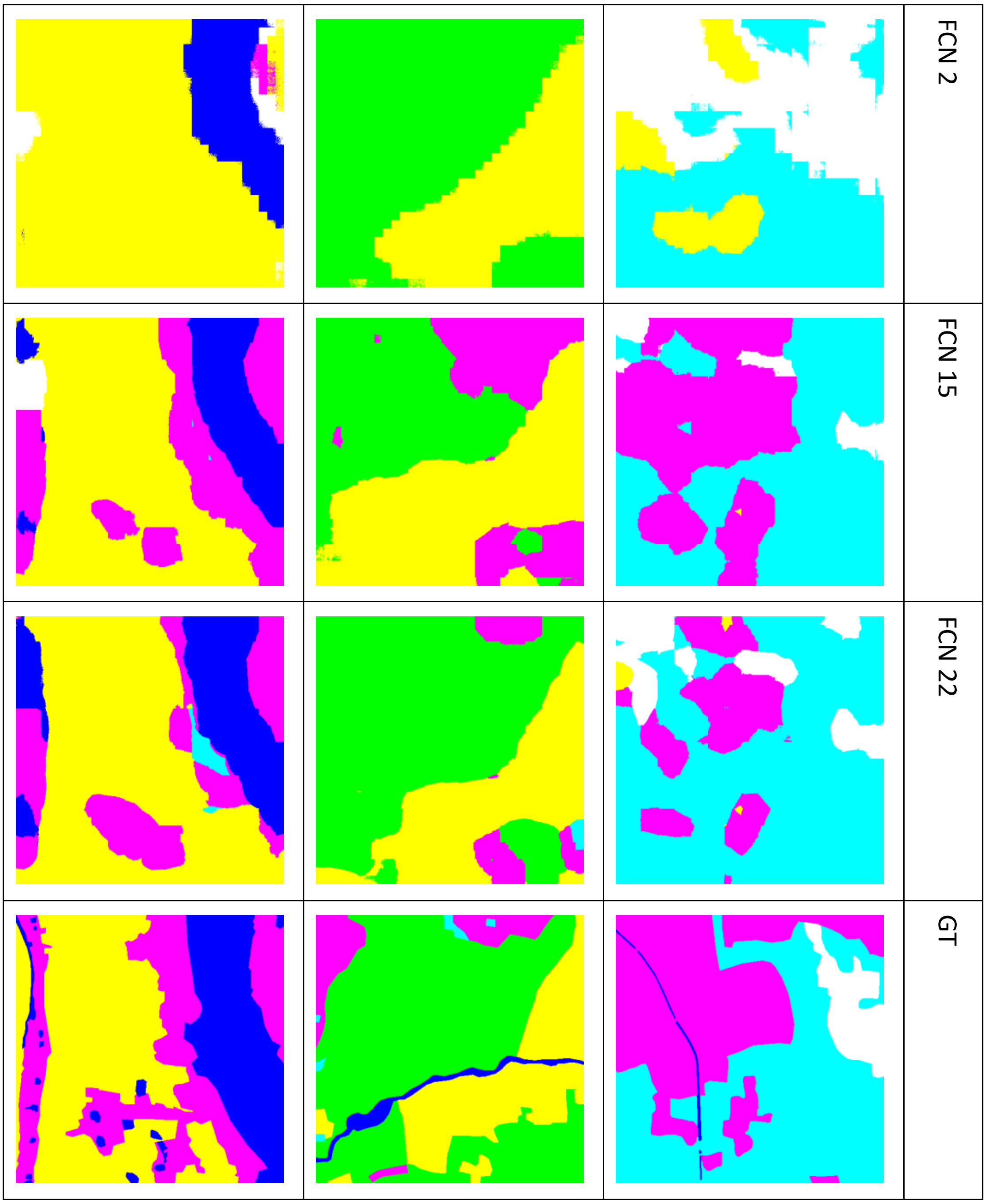
Hw3 B03901111 汪家銘

1.structure

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 512, 512, 3)	0
block1_conv1 (Conv2D)	(None, 512, 512, 64)	1792
block1_conv2 (Conv2D)	(None, 512, 512, 64)	36928
block1_pool (MaxPooling2D)	(None, 256, 256, 64)	0
block2_conv1 (Conv2D)	(None, 256, 256, 128)	73856
block2_conv2 (Conv2D)	(None, 256, 256, 128)	147584
block2_pool (MaxPooling2D)	(None, 128, 128, 128)	0
block3_conv1 (Conv2D)	(None, 128, 128, 256)	295168
block3_conv2 (Conv2D)	(None, 128, 128, 256)	590080
block3_conv3 (Conv2D)	(None, 128, 128, 256)	590080
block3_pool (MaxPooling2D)	(None, 64, 64, 256)	0
block4_conv1 (Conv2D)	(None, 64, 64, 512)	1180160
block4_conv2 (Conv2D)	(None, 64, 64, 512)	2359808
block4_conv3 (Conv2D)	(None, 64, 64, 512)	2359808
block4_pool (MaxPooling2D)	(None, 32, 32, 512)	0
block5_conv1 (Conv2D)	(None, 32, 32, 512)	2359808
block5_conv2 (Conv2D)	(None, 32, 32, 512)	2359808
block5_conv3 (Conv2D)	(None, 32, 32, 512)	2359808
block5_pool (MaxPooling2D)	(None, 16, 16, 512)	0
conv2d_1 (Conv2D)	(None, 16, 16, 4096)	102764544
dropout_1 (Dropout)	(None, 16, 16, 4096)	0
conv2d_2 (Conv2D)	(None, 16, 16, 4096)	16781312
dropout_2 (Dropout)	(None, 16, 16, 4096)	0
conv2d_3 (Conv2D)	(None, 16, 16, 7)	28679
conv2d_transpose_1 (Conv2DTr	(None, 512, 512, 7)	200704

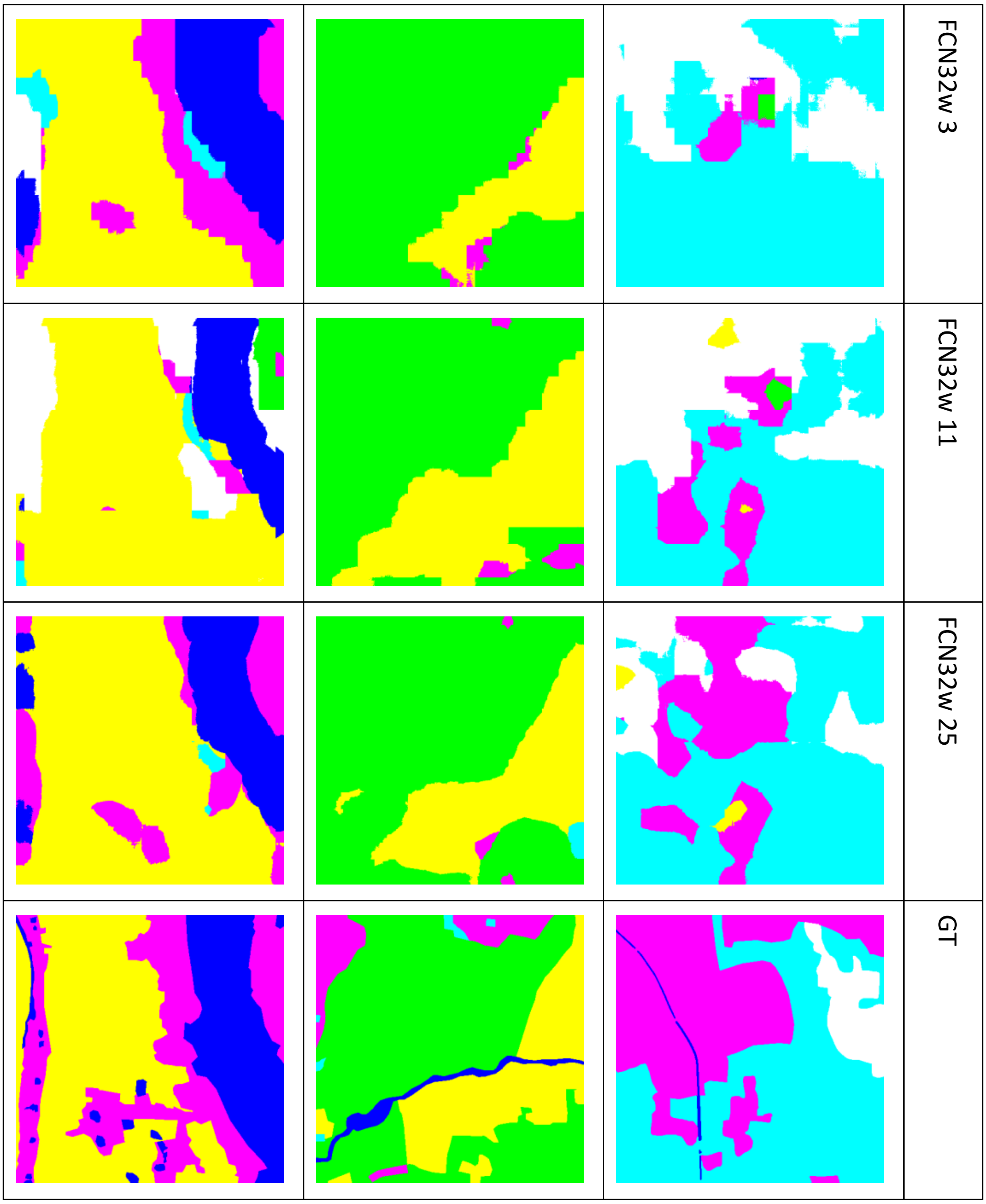
2.



3. Improved model

在本次作業中，我嘗試過了 Unet，不過效果不彰。後來我發現在本身 FCN32s 的架構上，引入 loss = weighted_categorical_crossentropy 會有部分改善，mIou 可以加高到 0.637，故架構同 1。

4.



5. 我的 weight 計算方式是：

1. 統計 0~7 各類的 pixel 總和
2. 將上值取倒數（加權少的）再取 log，得到 weight
3. 將原本架構中的 categorical_crossentropy 換成 Weighted_categorical_crossentropy(weight)（自行改寫）

我會這樣做的原因，是因為發現很經常出現 class2 的 IoU 相對偏低，並猜測這種情況是由於資料不均衡所致（黃色太多）。因此，我嘗試引入 weighted 的 loss function，並取得了較好的結果，使 FCN32 突破了原本的限制，能繼續增長。

Bonus

$$x(z^{(n)};w) = \frac{1}{1+\exp(-w \cdot z^{(n)})}$$

$$\frac{d}{dw} G(w) = \sum_n \frac{dG(w)}{dx^{(n)}} \frac{dx^{(n)}}{dw}$$

$$\frac{dx^{n}}{dw}$$

$$= -\sum_n \frac{d}{dx^{(n)}}\left[t^{(n)} \log x(z^{(n)};w) + (1-t^{(n)}) \log(1-x(z^{(n)};w))\right]$$

$$= -\sum_n \left[t^{(n)} \times \frac{1}{x^{(n)}} + (1-t^{(n)})\left(-\frac{1}{1-x^{(n)}}\right)\right] \frac{dx^{(n)}}{dw}$$

$$= -\sum_n \left[t^{n}\left(\frac{1}{x^{(n)}} + \frac{1}{1-x^{(n)}}\right) - \frac{1}{1-x^{(n)}}\right] \frac{dx^{(n)}}{dw}$$

$$= -\sum_n \left[\frac{t^{n} - x^{(n)}}{x^{(n)}(1-x^{(n)})}\right] \cdot \frac{dx^{(n)}}{dw}$$

$$= -\sum_n (t^{(n)} - x^{(n)}) \cdot \frac{[1+\exp(-wz^{(n)})]^2}{\exp(-wz^{(n)})} \cdot \frac{(-1)\cdot(-z^{(n)})\cdot \exp(-wz^{(n)})}{[1+\exp(-wz^{(n)})]^2}$$

$$= -\sum_n (t^{(n)} - x^{(n)}) \cdot z^{(n)}$$

#