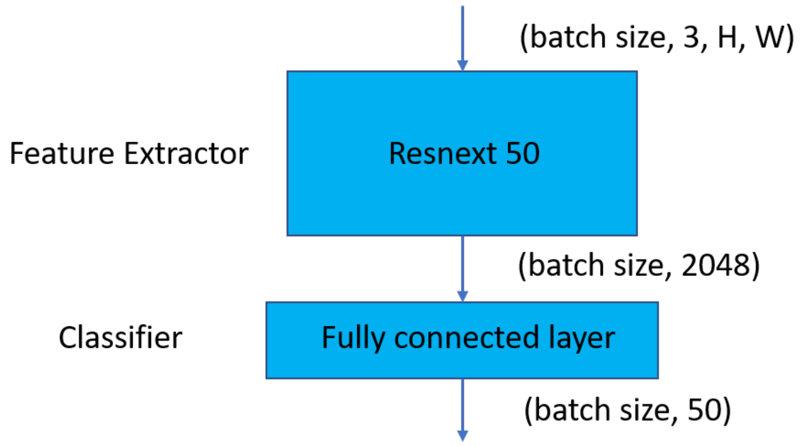
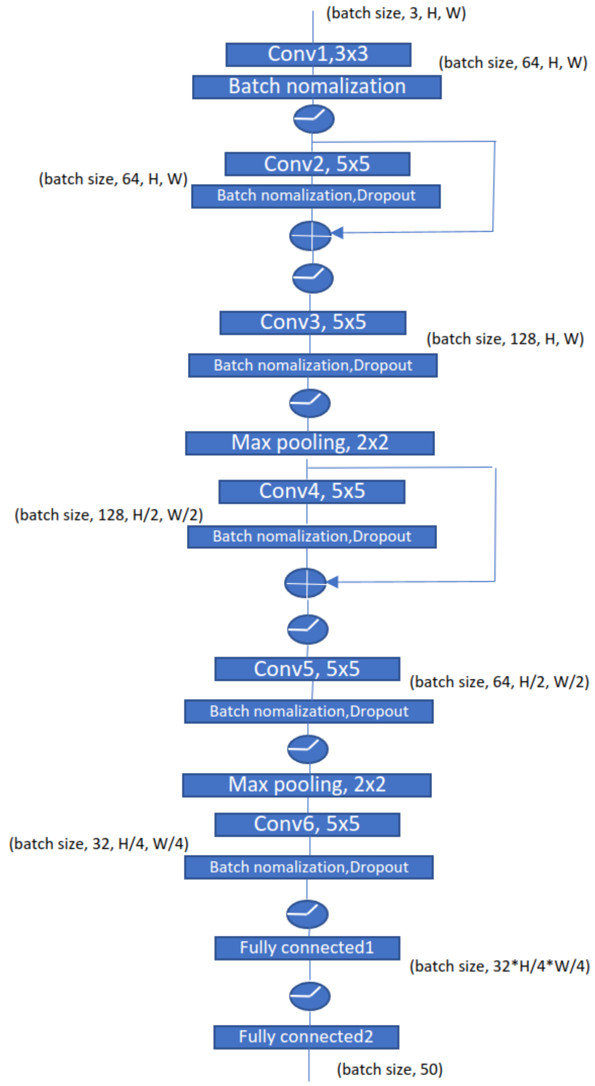
Problem 1.

(1)



(2) Model A : 51.160%, Model B : 90.440%

(3) Optimizer = Adam(all parameters are default in pytorch), loss function : cross entropy, epoch=120 data argumentation : horizontal flip, rotate, colorjitter, batch size = 16

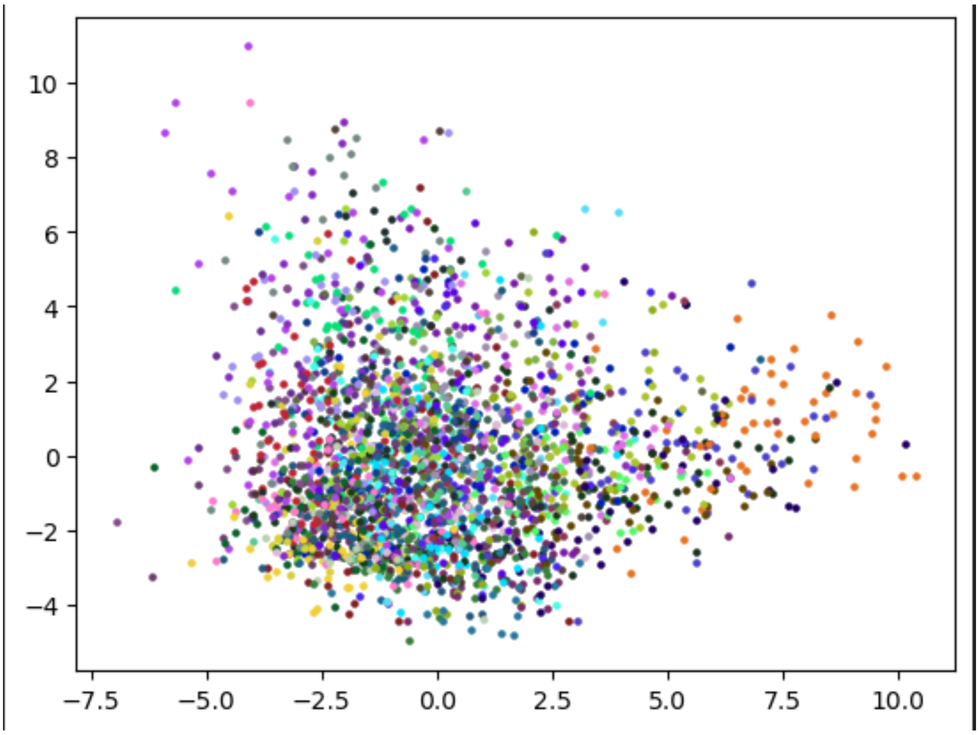


(4) Model B 使用 ResNeXt-50(32x4d)，主要差異在於 residual 的部份，Model A 僅使用 2 次單一 path 的 residual，而在 ResNeXt-50(32x4d)中，conv2 共做了 3 次合併 32 條 path 的 residual，conv3,conv4,conv5 同理，因此參數量與深度均遠大於 Model A

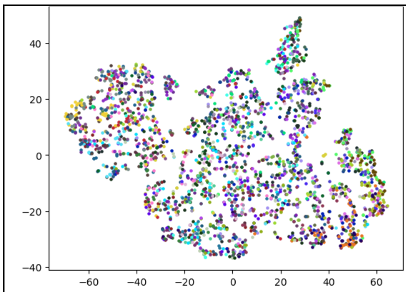
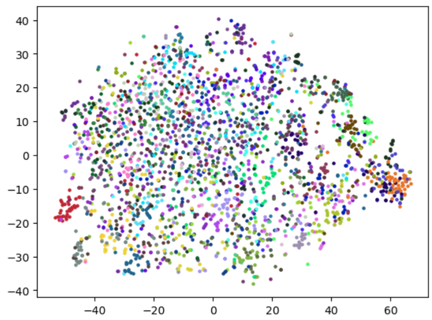
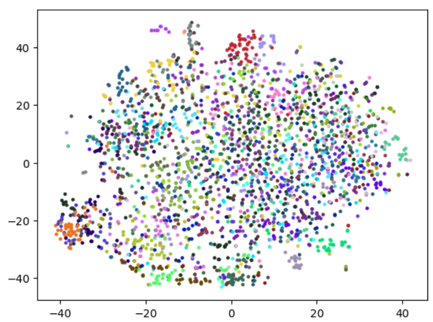
stage	output	ResNet-50	**ResNeXt-50 (32×4d)**
conv1	112×112	7×7, 64, stride 2	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2	3×3 max pool, stride 2
		$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 128\\3\times3, 128, C=32\\1\times1, 256\end{bmatrix}\times3$
conv3	28×28	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 256\\3\times3, 256, C=32\\1\times1, 512\end{bmatrix}\times4$
conv4	14×14	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 512\\3\times3, 512, C=32\\1\times1, 1024\end{bmatrix}\times6$
conv5	7×7	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 1024\\3\times3, 1024, C=32\\1\times1, 2048\end{bmatrix}\times3$
	1×1	global average pool 1000-d fc, softmax	global average pool 1000-d fc, softmax
# params.		$\mathbf{25.5}\times10^6$	$\mathbf{25.0}\times10^6$
FLOPs		$\mathbf{4.1}\times10^9$	$\mathbf{4.2}\times10^9$

XIE, Saining, et al. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p. 1492-1500.

(5) PCA 無法清楚地呈現各類別的分佈，如圖中即便是不同類別的點，依舊都集中在中左下角

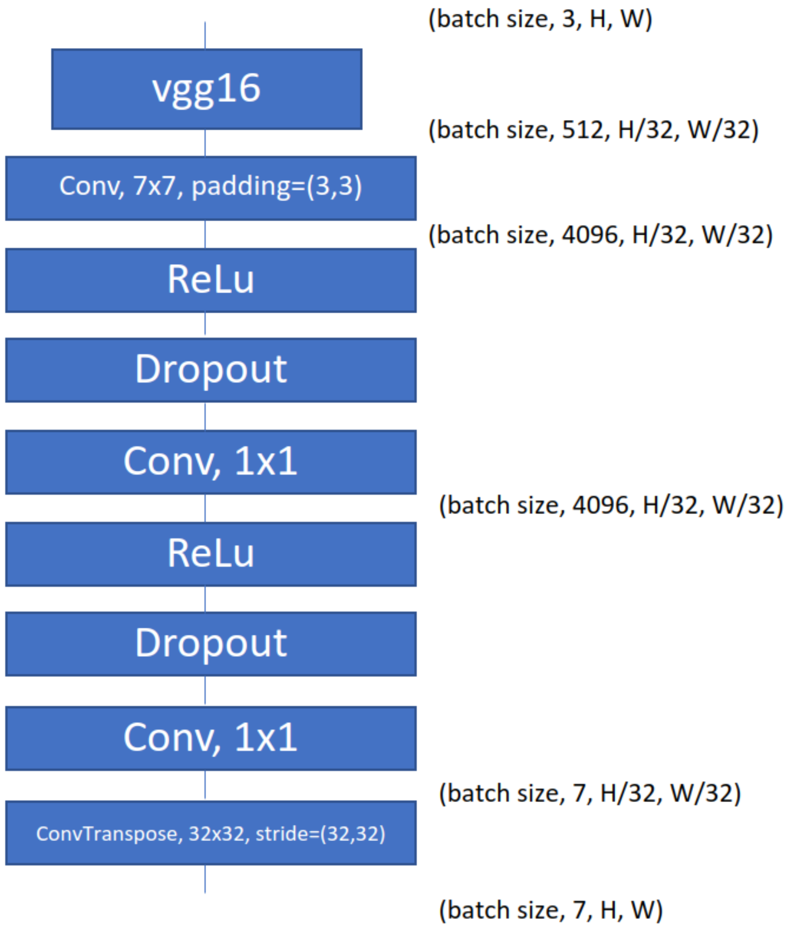


(6) t-SNE 相較於 PCA，更能觀察到分群的趨勢，仔細觀察隨著準確率上升，顏色的複雜程度愈高，這說明各群之間慢慢被區分出來，但又不像 PCA 全部擠在一團，t-SNE 在圖上有更均勻的分佈

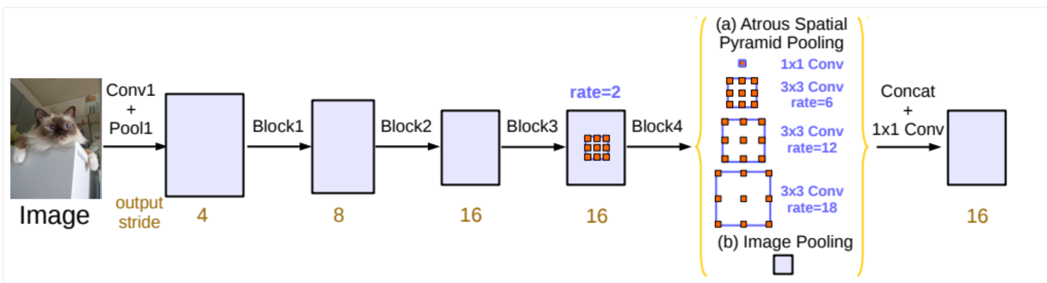
		
Epoch = 1, acc = 11.920%	Epoch = 30, acc = 44.320%	Epoch = 120, acc=50.280%

Problem 2.

(1)



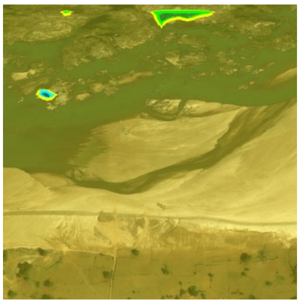
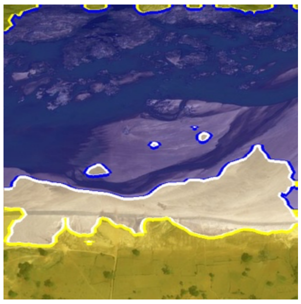
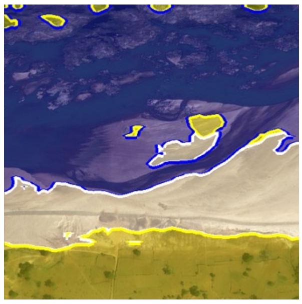
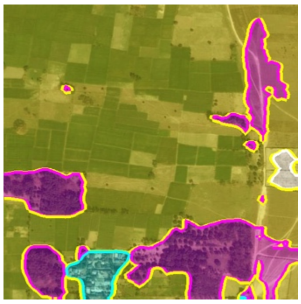
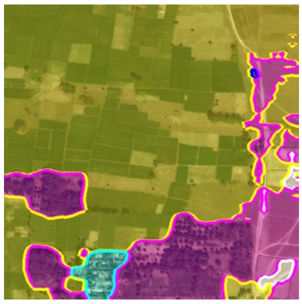
(2) Model B 為 Deeplabv3_resnet50，其使用 resnet50 作為骨架，並搭配 Atrous convolution + Spatial pyramid pooling(i.e. ASPP)。前者可擴大 receptive field，並由 Spatial pyramid pooling 組合不同 dilation rate 以獲得圖片不同區域大小的資訊，這與 ModelA 單純使用 FCN 有很大的差異



CHEN, Liang-Chieh, et al. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

(3) Model A : 66.920%, Model B : 73.888%

(4)

	Epoch 1	Epoch 50	Epoch107 (best IoU)
0013			
0062			
0104