

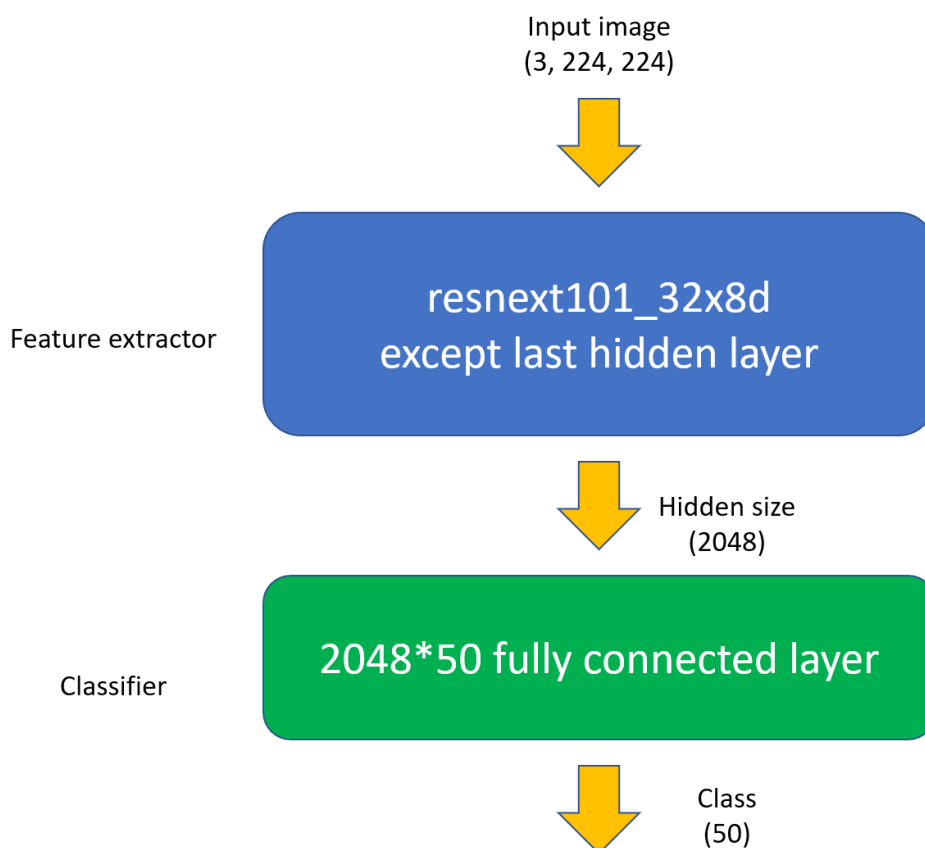
# DLCV HW1

## Problem 1

### 1. Draw the network architecture of method A or B

#### Method B:

In method B, I implemented ResNext model as my backbone model. Also, I change last fully connected layer to (2048, 50) fully connected layer considering 50 class in this task. The ResNext version is resnext101\_32x8d.



### 2. Report accuracy of your models (both A, B) on the validation set

	Model A	Model B
accuracy	0.6577	0.8820

### 3. Report your implementation details of model A

In method A, I scratch resnet101 without pretrained data. Resnet is a deep convolution network with residual block. The following picture depicts resnet model architecture.

**optimizer** : Adam

**learning rate** : 5e-5

**learning rate scheduler** : StepLR

**loss function** : cross entropy loss

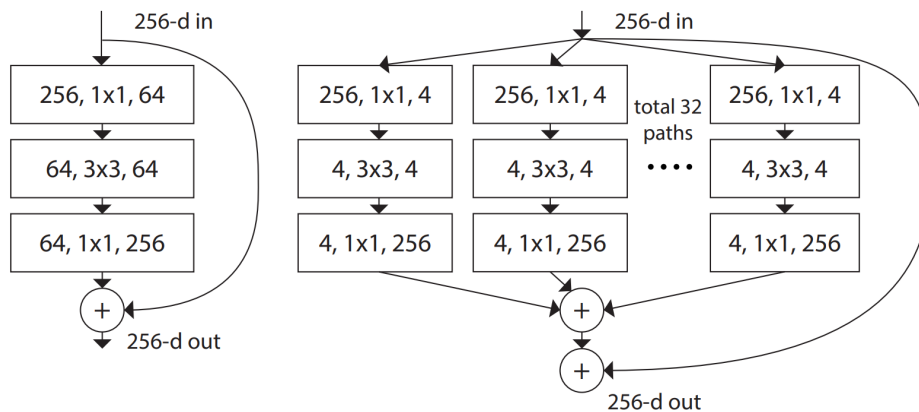
**epoch** : 18

**batch size** : 16

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

#### 4. Report your alternative model or method in B, and describe its difference from model A

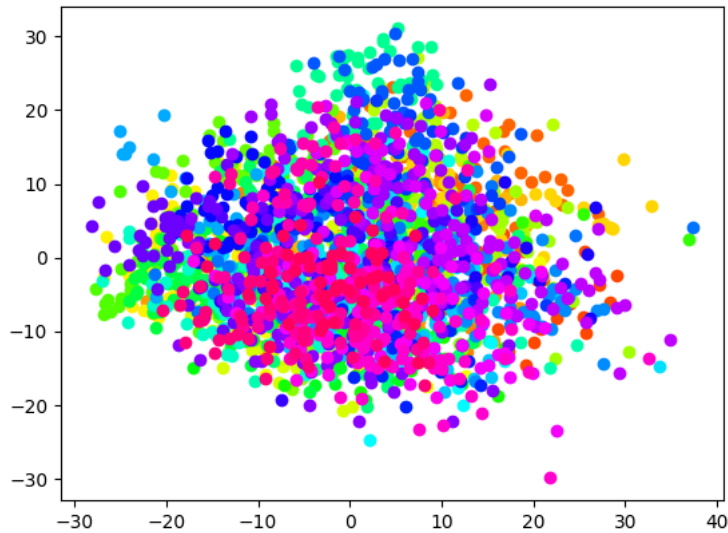
In model B, I implement resnext which is the improved version of resnet. In resnext residual block, input will pass to many convolution layer. Hence the model get wider and can handle more information in one block. The following picture is the comparison between resnet and resnext.



#### 5. PCA on model A

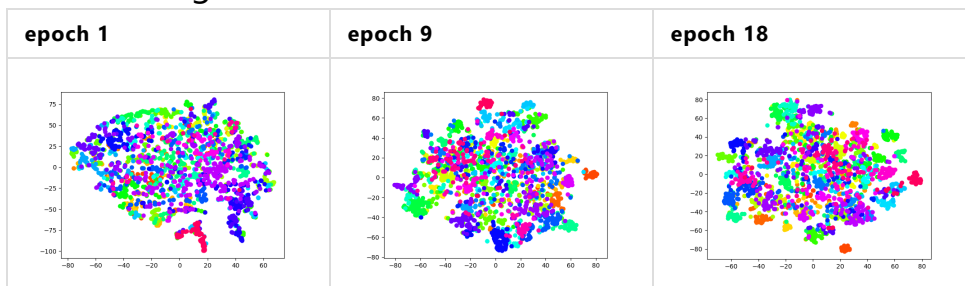
As we known about PCA. The objective of PCA is maximizing the variance of data. The picture below show this characteristic. The data scatter on 2D picture but it

doesn't maintain intra-class similarity.



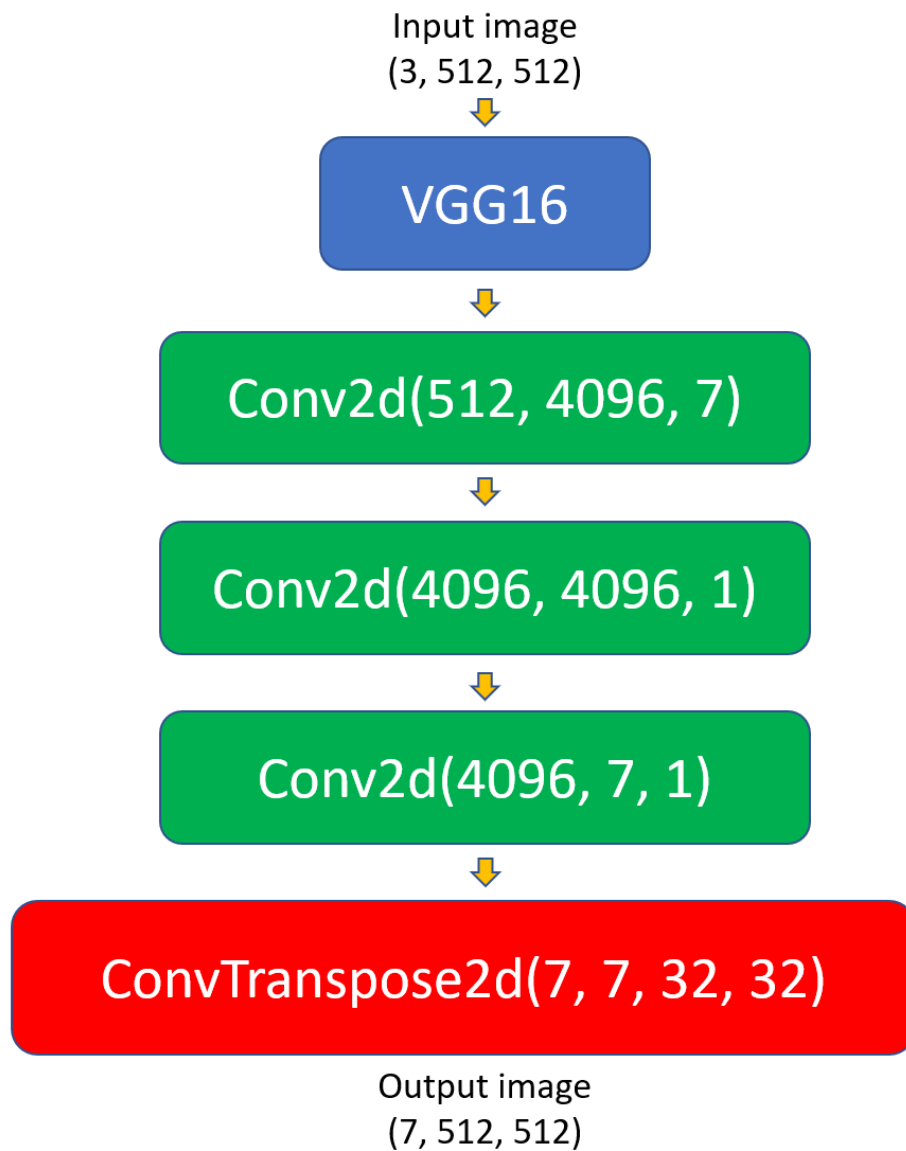
## 6. t-SNE on model A

From epoch 1 to epoch 18, we can see the datas in the same class get closer.



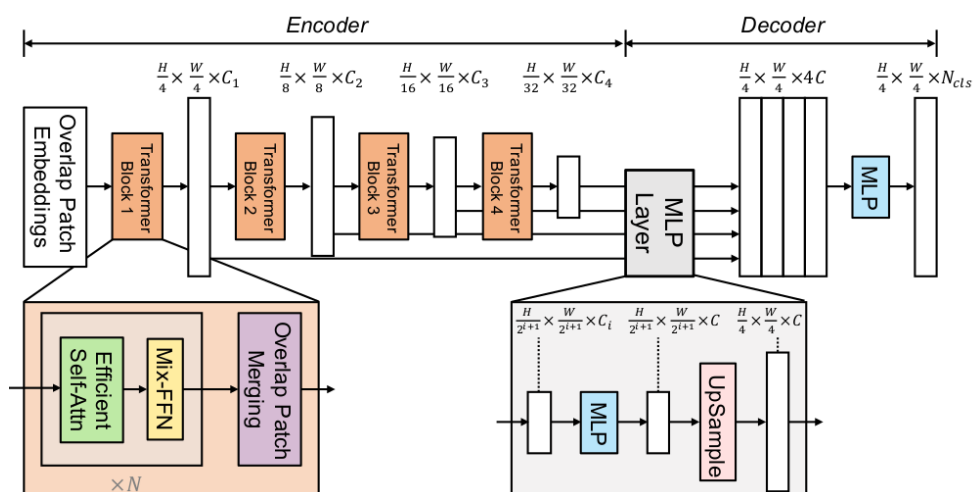
## Problem 2

### 1. Draw the network architecture of your VGG16-FCN32s model (model A)



## 2. Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model



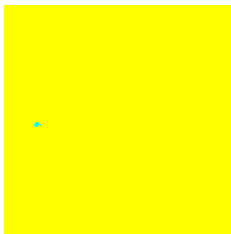
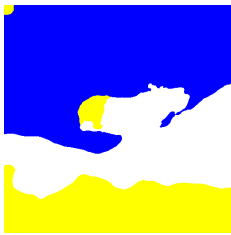




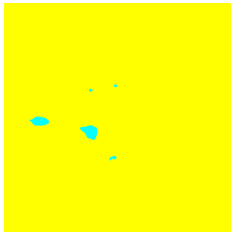
In comparison of VGG16-FCN32s which is based on CNN, segformer is based on transformer architecture which adopt attention mechanism to get more global information.



**2. Draw the network architecture of the improved model (model B) and explain it differs from your VGG16-FCN32s model**

	Method A	Method B
miou	0.6945	0.7427

**4. Show the predicted segmentation mask of “validation/0013\_sat.jpg”, “validation/0062\_sat.jpg”, “validation/0104\_sat.jpg” during the early, middle, and the final stage during the training process of the improved model**

	0013	0062	0104
epoch1			
epoch5			
epoch10			
ground truth	