**Computer Vision HW2 Report**

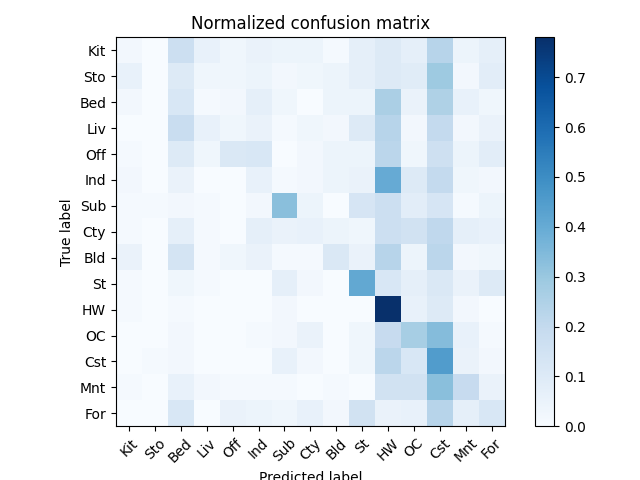
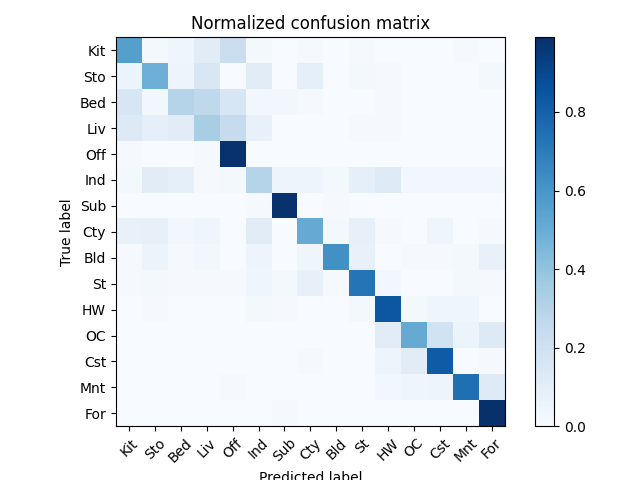
Student ID: R13921072

Name: 何家祥

**Part 1. (10%)**

**• Plot confusion matrix of two settings. (i.e. Bag of sift and tiny image) (5%)**

**Ans:**

****

Bag of shift tiny image

**• Compare the results/accuracy of both settings and explain the result. (5%)Ans:**

|  |  |
| --- | --- |
| **Accuracy** | |
| Bag of sift | 0.648 |
| Tiny image | 0.206 |

From the results, the accuracy of the Bag of SIFT method is significantly higher than that of the Tiny Image method. This is because Bag of SIFT uses more complex computations to extract local features from images, allowing it to effectively capture characteristics under variations such as translation, rotation, and lighting changes. However, it also requires more computational resources, especially in terms of memory usage.

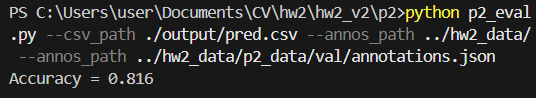
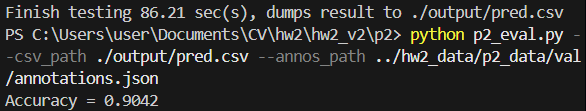
In contrast, the Tiny Image method simply uses the raw pixel values of downsampled images, which lack meaningful feature representation. Therefore, Bag of SIFT provides more discriminative and effective feature representations, making it more suitable for classification tasks.

**Part 2. (25%)**

**• Report accuracy of both models on the validation set. (2%)**

**Ans:**

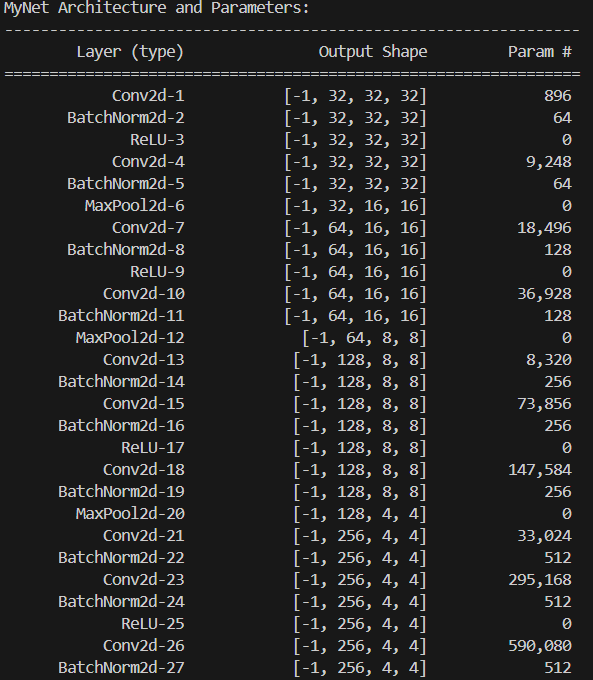
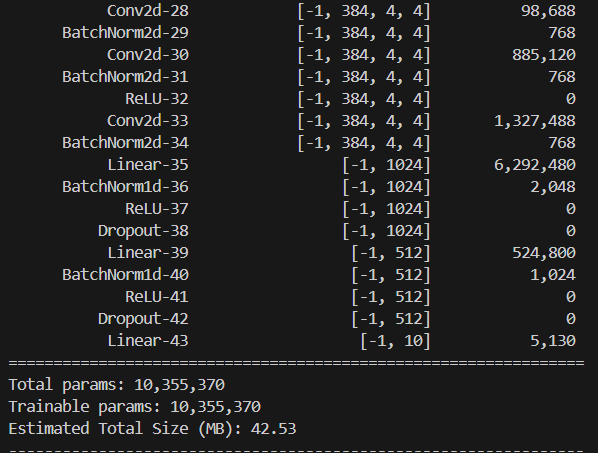
|  |  |  |
| --- | --- | --- |
|  | MyNet | ResNet |
| accuracy | 0.816 | 0.9042 |

** **

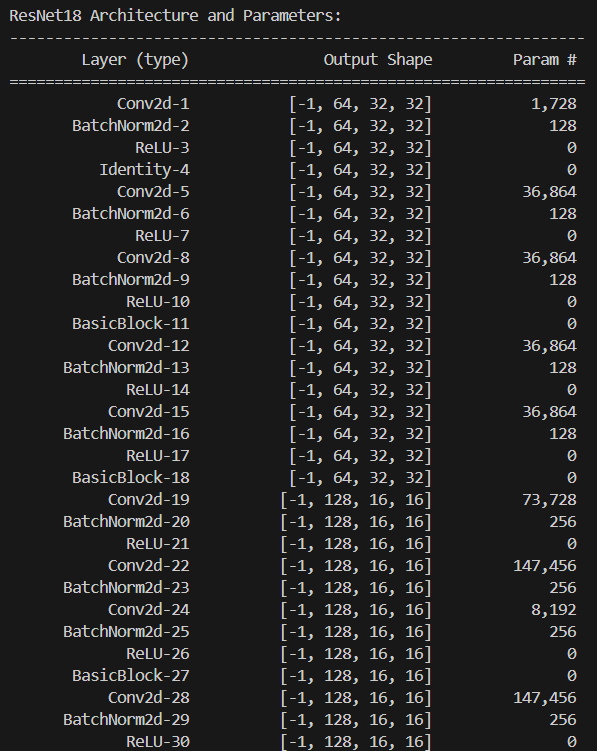
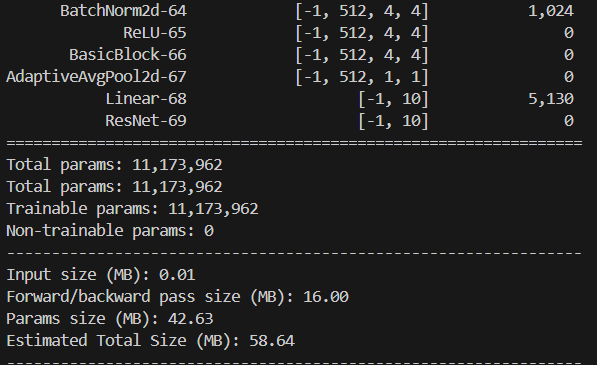
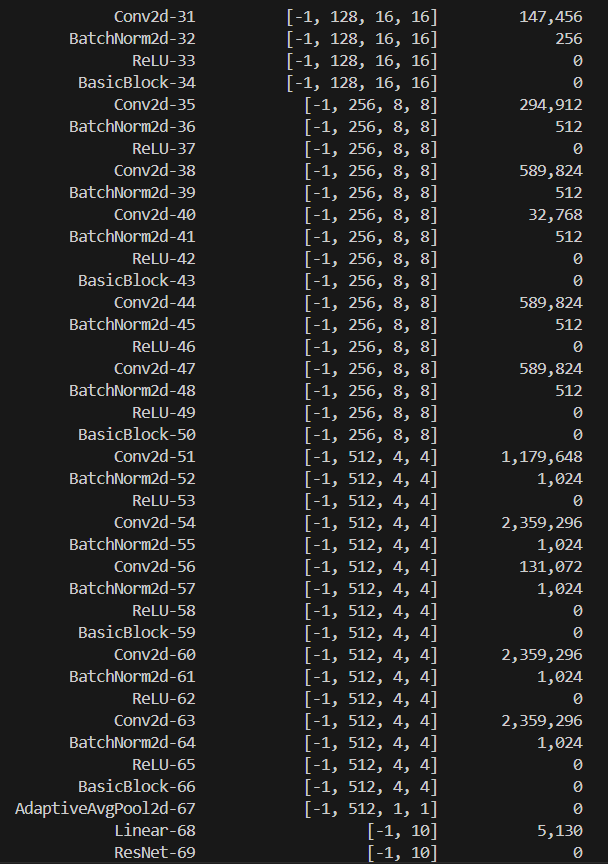
**• Print the network architecture & number of parameters of both models. What is the main difference between ResNet and other CNN architectures? (5%)**

**Ans:**

MyNet: parameters --10355370

** **

ResNet parameters --11173962

** **

The main difference between ResNet and traditional CNN architectures lies in its introduction of *residual connections*. This design allows data to bypass certain layers in the network, enabling the model to focus on learning the difference between the input and output. This mechanism effectively addresses the degradation problem that often occurs during the training of very deep networks, allowing deeper models to converge well and maintain high accuracy.

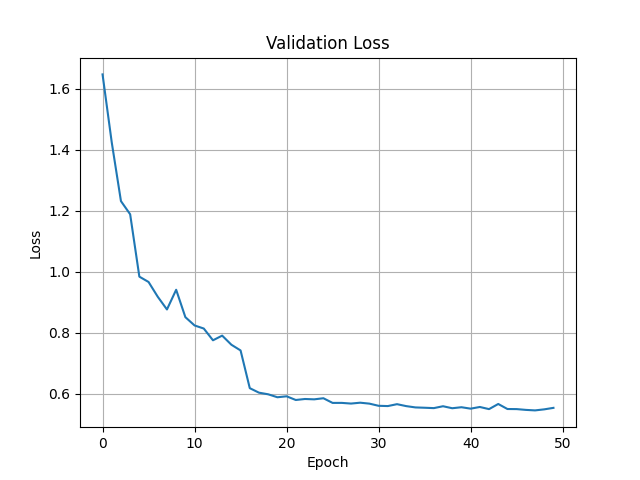
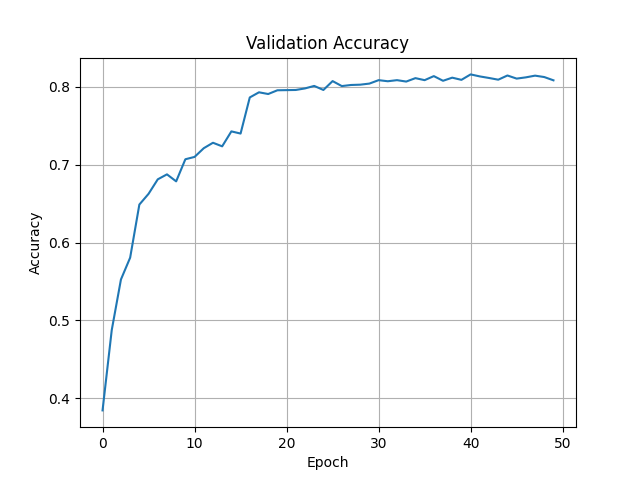
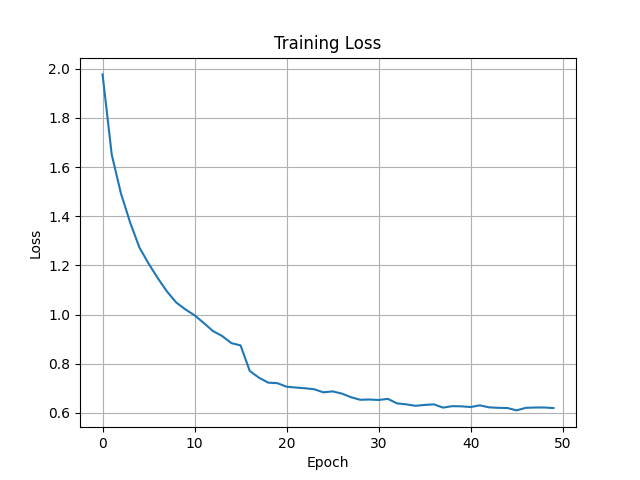
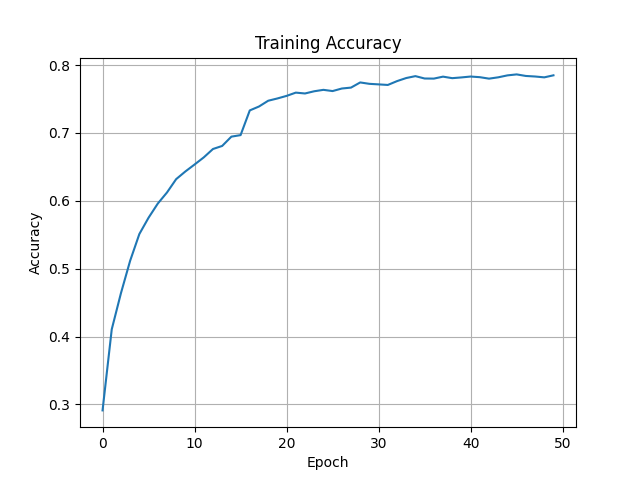
In contrast, traditional CNNs rely solely on stacking more layers to improve performance, which can lead to diminishing returns. ResNet offers a more stable and efficient solution.

Therefore, in my hand-crafted MyNet I also aimed to replicate this idea by implementing a custom residual connection function. However, this led to an increase in the number of parameters, and the performance still did not reach the level achieved by a well-trained ResNet.

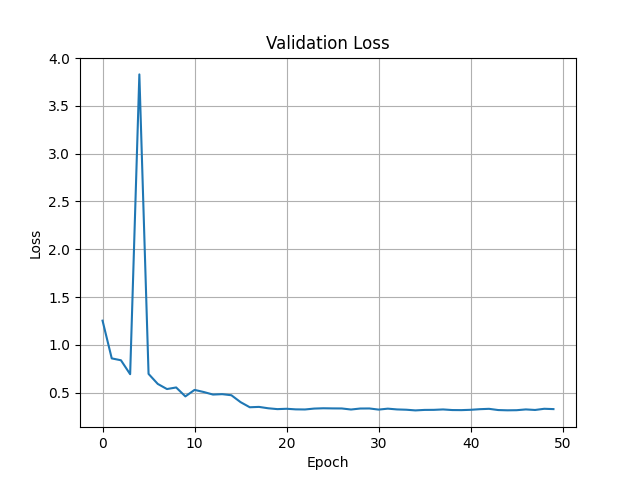
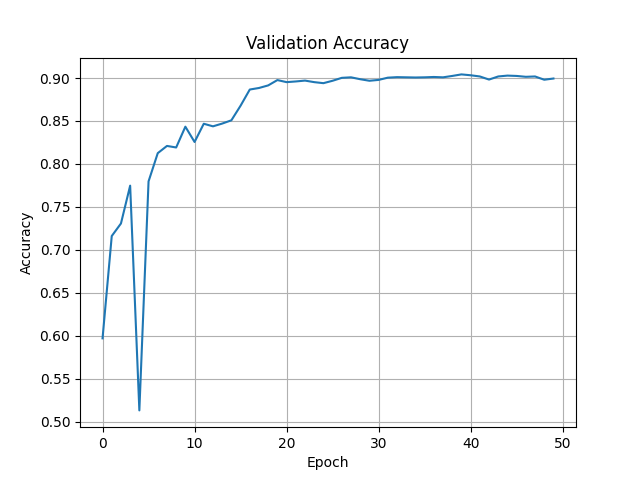
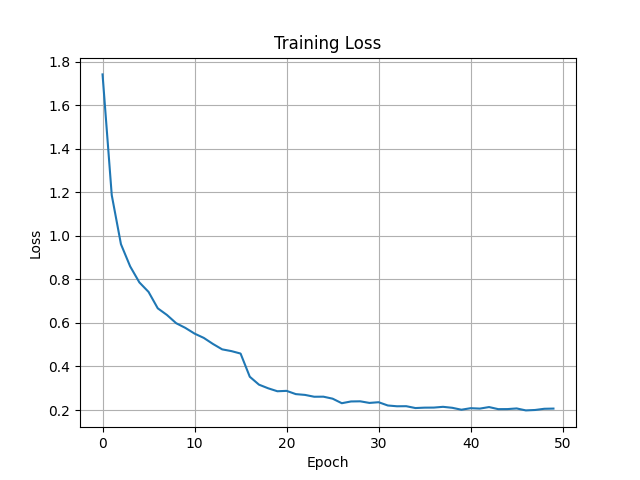
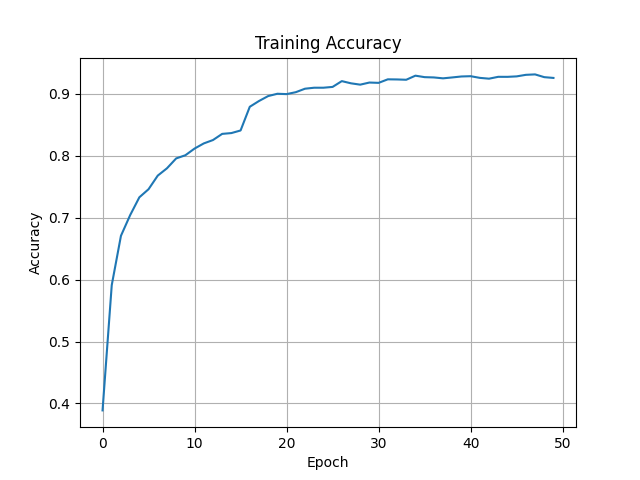
**• Plot four learning curves (loss & accuracy) of the training process (train/validation) for both models. Total 8 plots. (8%)**

**Ans:**

MyNet

****

ResNet

****

**• Briefly describe what method do you apply on your best model? (e.g. data augmentation, model architecture, loss function, etc) (10%)**

**Ans:**

1. The model uses a custom CNN inspired by ResNet, where each block includes two convolutional layers with batch normalization, ReLU, and pooling, along with shortcut connections to prevent vanishing gradients.
2. A learning rate scheduler is used to automatically adjust the learning rate during training for better stability, and data augmentation techniques like random rotation, translation, and scaling are applied to improve the model’s generalization ability.
3. The training process involves loading batches of data, performing forward and backward passes, and updating model parameters using the optimizer. The model is trained over multiple epochs with accuracy and loss calculated for both training and validation datasets.
4. Additionally, a learning rate scheduler is used to adjust the learning rate during training for better stability, and data augmentation techniques like random flipping, cropping, affine transformation, and color jittering are applied to improve the model’s generalization ability.
5. Regularization is applied by adding a Dropout layer in the fully connected layer to prevent overfitting.
6. A pretrained ResNet18 model was used, with modifications including replacing the first convolution and max-pooling layers to better suit small CIFAR-10 images, and adjusting the final fully connected layer to output 10 classes.