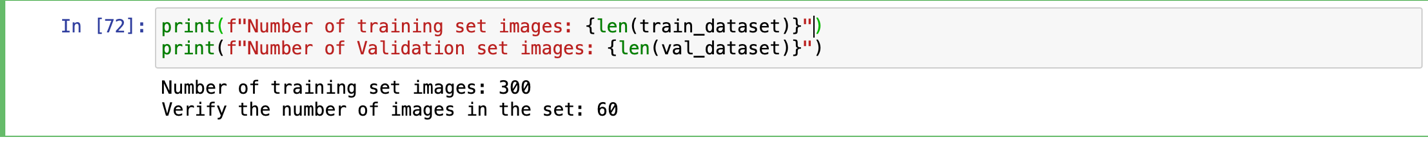
**VGG and ResNet left and right hand recognition training reported in pre-trained models**

**Datasets**  
left hand & right hand

A collage of hand gestures

Description automatically generated A collage of hands with numbers

Description automatically generated



The data are photos I took of myself, which I divided into two classes: left-handed and right-handed. A 360 photos, I made a training set of 150 left-handed photos and 150 right-handed photos, and then made a validation set of 30 left-handed photos and 30 right-handed photos.

**Pre-training model analysis**

I will mainly explain the two pre-training models I used, VGG16 and ResNet18, and analyze their model frameworks and loss and accuracy results.

Model Framework:

The core feature of VGG networks is the use of multiple consecutive small convolutional kernels (3x3) instead of large convolutional kernels (e.g., 5x5 or 7x7.) There are several versions of VGG, the common ones being VGG-16 and VGG-19, which contain 16 and 19 layers of the network, respectively. For this experiment, I used VGG-16 because training on fewer layers reduces the running burden.

ResNet introduces the concept of residual learning, which allows input information to be passed directly across layers, and due to the inclusion of residual connections, the network can be trained efficiently even when it is very deep, achieving better performance than previous deep networks. In my experiments I used resnet18 as my modeling framework.

About replacing the fully connected layers of vgg16 and resnet18:

When using pre-trained models such as VGG and ResNet, the classifier composition and functional design of each model differs, which leads to the need for different strategies when modifying these models for specific tasks (e.g., binary classification tasks).

They have differences in structural design, with VGG using multiple fully-connected layers that were used in the original ImageNet classification task to process large-scale category output (1000 classes.) ResNet, on the other hand, directly reduces the spatial dimensionality through global average pooling, with the output processed through a small, fully-connected layer, which makes modification more straightforward and efficient.

**Modifying the model to fit the binary classification problem of hand recognition**

Extract features from the model and train a linear model on top of extracted features. I use PyTorch to add a new layer onto the model, freeze all but the last layer, and train in PyTorch.

1. Classifier layers of the VGG model

The classifier of a VGG model is a sequence of several fully-connected layers (Linear layers), which are usually interspersed with ReLU activation and Dropout layers. In the original structure of VGG, the classifier is used to process the flattened feature vectors from the last convolutional layer with the aim of performing the final class classification.

When modifying the VGG model to accommodate new classification tasks (e.g., left- and right-handed binary classification problems), it is common to retain the power of the VGG model for feature learning while adapting to the new requirements by extending or simplifying the classifier part. Therefore, the code replaces the original classifier by adding a new, more complex sequence, which allows more flexibility in adjusting the model's output to match the new number of categories, as well as enhancing the model's resistance to overfitting through techniques such as Dropout.

2. Classifier Layer of the ResNet Model

In contrast to VGG, the structure of ResNet consists of a Global Average Pooling layer and a separate fully connected layer as a classifier. This design reduces the number of model parameters while making it very simple to modify the classifier-usually just replacing this one fully-connected layer is sufficient. The benefit of this design strategy is that it simplifies the final part of the model, reducing the risk of overfitting, while also reducing computational complexity.

**Analyzing loss and accuracy results**

Vgg16:

A screenshot of a computer

Description automatically generated

ResNet18:

A screenshot of a computer

Description automatically generated

Loss rate comparison:

VGG model The VGG model shows a steady improvement during the training process, and the training loss is reduced from 0.6498 to 0.1408 in the initial stage, showing good learning progress. During validation, the lowest validation loss is 0.1496 and the highest validation accuracy reaches 98.33%. However, the validation loss increased in some epochs, indicating a possible slight overfitting.

The training loss of the ResNet model decreased from 0.7434 to 0.3169, showing a slower decreasing trend. Nevertheless, the ResNet model performs better on the validation set, with the validation loss decreasing rapidly and reaching very low levels in several epochs, e.g., the validation loss in the 9th and 10th epochs is 0.1070 and 0.1105, respectively, with 100% accuracy.

Comparison of Accuracy:

The training accuracy of the VGG model increases from 58.67% to 94.67%, showing a steady improvement. The validation accuracy stays above 90% in most epochs, reaching a maximum of 98.33%. This shows that the VGG model is able to learn from data and generalize effectively.

ResNet model Improvement in training accuracy from 53.67% to 89.00%, but the growth is slower. However, the validation accuracy reaches 100% in multiple epochs, showing an extremely high generalization ability. This indicates that the ResNet model shows very high accuracy and stability on the validation set although it grows slowly during training.

**Overall Conclusion**

VGG model The performance grows faster on the training set, but fluctuates more on the validation set, showing possible overfitting issues. Nonetheless, the VGG model is able to achieve over 90% accuracy in most cases, making it a relatively stable choice.

ResNet model Although it grows slower in the early stages of training, it shows very high stability and accuracy on the validation set, especially in the later epochs, where it almost always maintains 100% accuracy. This suggests that ResNet may be a superior choice for tasks that are complex or require deeper feature abstraction.

The ResNet model performed even better in this left- and right-handed recognition task. I would recommend prioritizing the use of the ResNet model for similar image recognition tasks, especially when high accuracy performance is required. If a VGG model is to be used, model tuning may need to be considered to avoid overfitting.

**Conlusion**

The VGG16 model usually performs well in image feature extraction due to its deep and dense convolutional structure, especially in smaller details that may be better captured.

However, it has some drawbacks, in the experiments the VGG16 model has many parameters resulting in a long training time. Overfitting often occurs during training, and regularization and modification of parameters need to be added to improve the overfitting.

ResNet18, by using residual joins, can train deeper networks without causing the problem of vanishing gradients. It performed optimally in my experiments. However, for some simple tasks, ResNet18 may be a bit too complex and does not necessarily lead to better results than simple models. Although the validation set is highly accurate, the training set is not as accurate as vgg16.

Reference:

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). *Deep Residual Learning for Image Recognition*. ArXiv.org. <https://arxiv.org/abs/1512.03385>

https://www.kancloud.cn/apachecn/ml-mastery-zh/1952041

Simonyan, K., & Zisserman, A. (2015, April 10). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. ArXiv.org. <https://arxiv.org/abs/1409.1556>

<https://www.kancloud.cn/apachecn/ml-mastery-zh/1952041>

<https://blog.csdn.net/hnu_zzt/article/details/85092092>

<https://blog.csdn.net/TracelessLe/article/details/115579791>

<https://www.bilibili.com/video/BV1fU4y1E7bY?p=4&vd_source=92478abdd72edfdef79a29258b9bd716>

<https://www.bilibili.com/video/BV1jB4y1w7qB/?spm_id_from=333.337.search-card.all.click&vd_source=92478abdd72edfdef79a29258b9bd716>

<https://www.bilibili.com/video/BV1Bo4y1T7Lc/?spm_id_from=333.337.search-card.all.click&vd_source=92478abdd72edfdef79a29258b9bd716>