MATH 4995 Final Project: Semi-conductor Image Classification 2 mini

Liao Yi-han yliaoag@ust.hk
Department of Data Science and Technology, HKUST

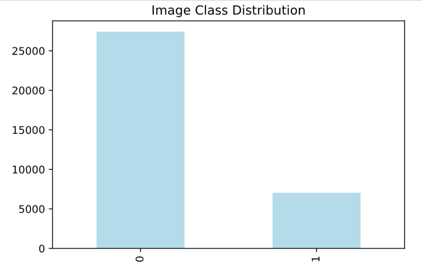
1. Introduction

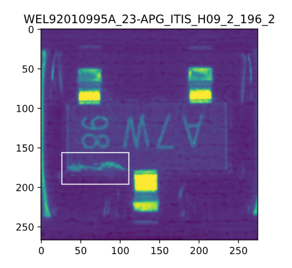
Nexperia is one of the biggest Semi-conductor manufacturers in the world. Sometimes, a lot of unqualified devices are mixed with the good ones during the production. The goal is to use deep learning methods to help Nexperia pick out the defect devices.

2. Dataset: semi-conductor images

Training Data:
- Good device: 27420
- Defect device: 7039

Testing Data: 3830





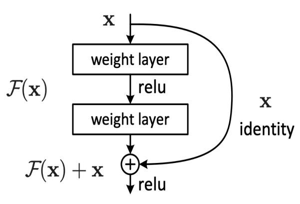

```
preprocess = transforms.Compose([
    transforms.Resize((224, 224)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
    std=[0.229, 0.224, 0.225])
])
```

3. Methodology – Transfer Learning

Considering the size of dataset is not small, it will be time-comsuming to train the model from scratch, as a result, the transfer learning technique is introduced to this project. Select ResNet and EfficientNet as the network models.

3.1 Methodology – Transfer Learning: ResNet

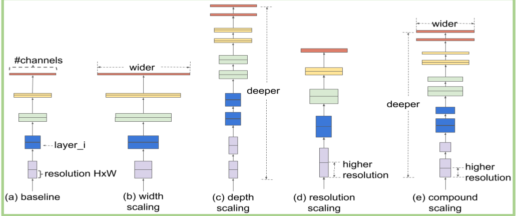
For ResNet, based on the pytorch pretrained model, the final fully-connected layer with a new fully-connected layer having the output size = number of the classes (=2, defect/good). Finetuning the convnet, and then use it to do the prediction for testing data.

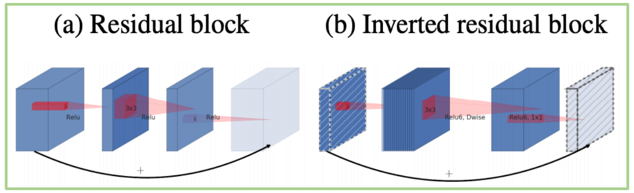


The feature for ResNet is the residual block which solves the degradation problem as stacking deeper layers. Layers in ResNet method fit a residual mapping since it is hypothesized that it's easier to optimize the residual mapping. For residual network, the gradients will be additive and that is unlikely to vanish.

3. 2 Methodology – Transfer Learning: EfficientNet

EfficientNet is aiming for an efficient model scaling for convolutional neural networks. It uniformly scales each dimension with a fixed set of scaling coefficients. This compound scaling method performs a grid search to find the relationship between different scaling dimensions of the baseline network under a constraint of fixed resources, which helps to select the appropriate scaling coefficient for each dimensions (e.g. width, depth, and image resolution). One important feature of the EfficientNet architecture is the mobile inverted bottleneck convolution (MBConv), which is a kind of residual block with an inverted structure.





Residual block: wide -> narrow -> narrow -> wide structure (#channels)
Inverted residual block: narrow -> wide -> wide -> narrow

4. Analysis

Though the dataset is quite imbalanced, the prediction results seem to be very good with the two networks. This may be because the number of image classes is only two and the training samples for each class is quite enough (27420 and 7039 respectively). The networks are able to learn the main features of each class well, and based on the transfer learning, most parameters are pretrained to a good state, only the final-layer finetuning is needed.
Both ResNet and EfficientNet are dominative neural networks, and their performances are really the state-of-the-art.

5. Results (Kaggle score)

	ResNet18	EfficientNet – B2
Training score	0.9862	0.9711
Testing score	0.99567	0.99539

6. References

K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs], Dec. 2015. arXiv: 1512.03385.
M.TanandQ.V.Le.EfficientNet:RethinkingModelScaling for Convolutional Neural Networks. arXiv:1905.11946 [cs, stat], Sept. 2020. arXiv: 1905.11946.