Proposal of Nexperia project

Qiu Chuchu
cqiuac@connect.ust.hk

Abstract

In this proposal, we fine-tuned 5 models, EfficientNet-B3, EfficientNet-B5, Resnet34, Resnet50, Resnet101 respectively. Then we did stacking, an aggregating method to do image classification and the preliminary result is that we got the accuracy of 96.35%, currently ranked the second on the Kaggle leaderboard. In order to mitigate the imbalanced distribution between the bad semi-conductors and the good ones, we adopt over-sampling, which increased the accuracy by 1%. In future work, we will try to introduce HexaGAN, HyperGAN and the incorporation of CNN model and XGBoost to bring further increasement on accuracy.

1 Introduction

This project is mainly about Semi-conductor Image Classification, currently at the first stage in Kaggle[5]. Nexperia, one of the biggest Semi-conductor company in the world, hopes the yield rate could be greatly improved by using deep learning.

According to PwC's survey[2], the annual sales of semi-conductor industry in 2019 are about US$500 billion and are most likely to continue growing in the future as artificial intelligence adds to demand. Considering the semi-conductor market's huge scale, even slight misjudgment on yield rate can lead to huge loss for company. Therefore, it is of higher importance to ensure the yield rate of semi-conductors.

According to McKinsey's report[1], for the semiconductor industry, the use of AI can lead to a reduction in yield detraction by up to 30%. To be specific, there are two approaches on adopting AI to help improve yield rate. One is to introduce AI in process flow and the other is to minimize the misclassifications on both good and bad semi-conductors. In the first stage, we focus on the latter approach because currently the yield rate is still calculated by human sampling inspection, using a microscope, human operators located and classified the defects in the system, which is not only time consuming and labor consuming, but also inaccurate. To be specific, what we need to do in the first stage of this proposal is to accurately classify the good and bad semi-conductors, avoiding classifying the good semi-conductors as the bad ones, leading to decrease in yield rate. Therefore, compared with defect review in the past, this method is novel.

2 Related work

Resnet[8] networks are commonly used in image classification and could get a relatively high accuracy. The authors designed a residual learning framework to ease the training of networks, learning residual functions instead of learning unreferenced functions, which could train a deep network better but avoid vanishing gradient problem.

However, Mingxing Tan introduced a new method called EfficientNet[15] which is 8.4 times smaller and 6.1 times faster on inference than the best existing ConvNet so we would not only use resnet to finish the project.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Besides, some researchers explore the incorporation of CNN model and XGBoost algorithm[10], because they think CNN has been recognized as the most powerful and effective mechanism for feature extraction, but traditional classifiers connected to CNN do not fully understand the extracted features. eXtreme Gradient Boosting (XGBoost) is an integrated learning algorithm based on Gradient Boosting, the principle of which is to achieve accurate classfication results through iterative computation of weak classifiers. Therefore, they integrates CNN as a trainable feature extractor to automatically obtain features from input and XGBoost as a recognizer in the top level of the network to produce better results.

Since the dataset is imbalanced, some methods should be chosen to counter imbalance simply because the classifier would get insufficiently trained and hence provide inaccurate predictions if data of the minority class is insufficient. The author of *Data Imbalance: Effects and Solutions for Classification of Large and Highly Imbalanced Data[12]* shows that the probability of getting a usable result starts to decrease with increase in the imbalance.

Oversampling is a method that can increase the size of minority class and SMOTE[6] one of the popular oversampling technique. Also, we can decrease the size of majority class but always we would like to use oversampling since the more data, the better performance. ROC and PR curves were used to measure their performance. PR curves are used to uncover information that are not available from ROC curves[7].

Iterative Metric Learning (IML)[14] conducts metric learning process until the selected samples are relatively stable and constructs a more stable and effective data space for testing data and selects the most relevant data according to the testing data. Finally, classifier such as KNN and SVM classifies the testing sample.

We could also use GAN to generating more samples. HyperGAN[13] is a approach to generate samples but different from prior GANs, HyperGAN does not require repeated samples to start with but trains directly using maximum likelihood. This significantly improve training efficiency.

3 Dataset

The photos of semi-conductor are provided by Nexperia. The amounts of bad semiconductor's photo(3000) are much smaller than the good ones(27000). Therefore, data preprocessing should be done to deal with it. This is because imbalanced dataset can discard useful information about the data. For example, if we predict all the semiconductors are good, the accurary could also be 90%. However, the goal of this stage is to correctly indentify the bad semi-conductor in order to decrease scrap rates. Thus, data processing is a must. The data could be shown in Figure 1.

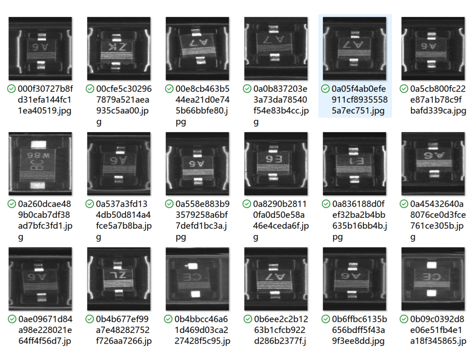

Figure 1: Sample

Table 1: Accuracy

Result	Accuracy
Testing set without oversampling	93.444
Testing set with oversampling	94.537

4 Implementation and method

4.1 Data augmentation

4.1.1 Transformation

Models trained with data augmentation would generalize better, helping introducing more varieties in the dataset. Besides, the amounts of bad semi-conductors are not enough so the transformation should be done to add more samples to the networks. A poorly trained neural network would regard pictures which are cropped and zoomed as different pictures after using transformation. Also, if all of our images are perfectly located in the center, traditional feed forward models would be confused when the image is slightly shifted to the right relative to the background. Therefore, it is of great importance to do transformation. However, transformation should be in line with reality. It is impossible to rotate images of pedestrians crossing the street to be feet up head down because on real data such a scene is incredibly unlikely. If these kinds of augmentation are be done, such augmentation of the training data could potentially hurt the results.[3] As the winner of the Kaggle competition called iMaterialist Challenge (Fashion) at FGVC5 2018 said that after experimenting with more extreme augmentation and crops but ending up using the default fastai(a python library) transformation for most of the models.[2] Default fastai transformation means we only do random flipping with probability 0.5, 10 degree rotation at 75%, 1.1x magnification with probability 0.75.

4.1.2 Methods to deal with imbalanced data

There are several ways except adding more data to combat imbalanced classes.

The first one is resampling the dataset, including over-sampling and under-sampling. Over-sampling means we should add copies of instances from the under-represented class while under-sampling means we delete instances from the over-represented class. In this project, we would like to correct the weight of sampling to own a relative balanced dataset. The dataset before re-sampling is 1:9. When we want to look closely at minority group(bad ones), we have to design the samples differently so that the enough respondents could be met in bad semi-conductors group to analyze. We do this by giving members of the minority group a higher chance. The result of classification improves about 1% after doing so in the testing set.

We have also tried weighted loss. For the minority group, we impose a large loss on them, which means if the small group is misclassified, the loss would be large. In order to minimize the loss, the neural network would take more consideration on minority group. The weight I set is that the loss on bad semi-conductors group is 9 times larger than the good ones. However, the result seems not so good.

We intend to try focal loss in the further study simply because focal loss is a loss that designs to solve the problem of imbalanced dataset. it down-weights the loss assigned to well-classified examples.[11]

The main property of focal loss is to reduce the weight of easy-to-classify samples, making the model focus on hard-to-classify samples during training. At current stage, what loss I use is the standard entropy and I guess this kind of loss would have a positive effect on this project.

4.1.3 EfficientNet

For the first stage part, I fine tune EfficientNet to do classification simply because EfficientNet significantly outperforms other ConvNets. EfficientNet-B7 achieves state-of-the-art 84.4% top-1 / 97.1% top-5 accuracy on ImageNet, while being 8.4 times smaller and 6.1 times faster on inference than the best existing ConvNet, compared with other widely used in transfer learning model such as ResnetNet50. Figure 1 shows the plot of model Size and ImageNet accuracy among different models.

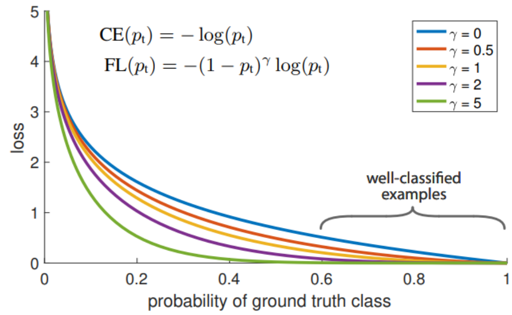

Figure 2: Focal loss[11]

Table 2: Accuracy in validation set

Model	Accuracy
Pretrained B3 model	97.33
Pretrained B5 model	97.79

EfficientNets also transfer well and achieve state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and three other transfer learning datasets, with an order of magnitude fewer parameters[15]. Therefore, I fine tune this model in the project.

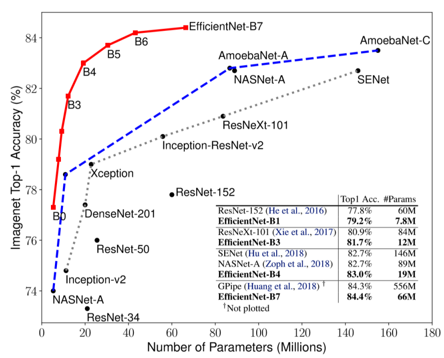

Figure 3: EfficientNet[15]

we choose EfficientNet to capture features from images of semi-conductors, adding a fully connected layer which the output is 2 features meaning which class the object belongs to. we freeze all layers but the last layer, the fully connected layer in order to keep features extracted by the network.

There are 8 pretrained models that the authors provide and we use two of them, B3 and B5 respectively. The accuracy in validation set of these two models are nearly the same.

Under this circumstance, we would not use B7 model, simply because there is not doubt that B7 model can reach a relatively good accuracy but it takes much more time in training and the accuracy may not seem significantly increases.

The learning rates we choose in these two models are around 1e-3 and 1e-5. I have done something to find a better learning rate, reaching a better accuracy more quickly. Over an epoch we begin SGD with a very low learning rate (like 1e-8) but change it by multiplying it by a certain factor for instance at each mini-batch until it reaches a very high value (like 1 or 10). we record the loss each time at each iteration and once finished, plot loss against the learning rate.[4] The optimal learning rate is the sharpest slope of the curve. Figure 4 is the loss against learning rates in pretrained B5 model.

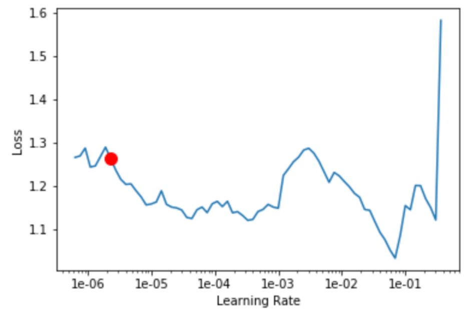

Figure 4: Finding perfect learning rate

4.1.4 Ensemble learning

Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking).

In stage 1, I would like to use stacking, a method that combine multiple classification models via a meta-classifier. There are two levels of models in this method. The base level models are trained based on a complete training set and first-level learners I choose is EfficientNet B3 and Resnet34 model. Then the meta-model is trained based on the outputs of the base level model as features. The features are the probability of the class and the output is the prediction of testing set.

4.1.5 HxeaGAN

We would like to use HxeaGAN to generate more data and do classification. It is a method that highly matches Nexperia's data, because it is designed to solve the problem of dirty data like imbalanced data in real world. HxeaGAN overcomes the class imbalance problem by training a deep generative model to follow the true data distribution, and then generates samples of minority classes for each batch. This requires conditional generation, which is regard as imputation.[9]

References

[1] Smartening up with artificial intelligence (ai) - what's in it for germany and its industrial sector?

[2] Opportunities for the global semiconductor market. `https://www.pwc.com/gx/en/industries/tmt/publications/global-tmt-semiconductor-report-2019.html`, 2019.

[3] Opportunities for the global semiconductor market. `https://www.quora.com/What-are-good-data-augmentation-techniques-for-a-small-image-data-set`, 2019.

[4] Opportunities for the global semiconductor market. `https://sgugger.github.io/how-do-you-find-a-good-learning-rate.html`, 2019.

[5] Semi-conductor image classification (first stage). `https://www.kaggle.com/c/semi-conductor-image-classification-first/data`, 2019.

[6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, Jun 1, 2002.

[7] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. ICML '06, pages 233–240. ACM, Jun 25, 2006.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, Dec 10, 2015.

[9] Uiwon Hwang, Dahuin Jung, and Sungroh Yoon. Hexagan: Generative adversarial nets for real world classification, Feb 26, 2019.

[10] Christian Kraetzer, Yun Qing Shi, Jana Dittmann, and Hyoung-Joong Kim. *Digital Forensics and Watermarking : 16th International Workshop , IWDW 2017, Magdeburg, Germany, August 23-25, 2017, Proceedings*, volume 10431. Springer International Publishing AG, Cham, Aug 9, 2017.

[11] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):1, Jul 20, 2018.

[12] J. Conference Paper. Data imbalance: Effects and solutions for classification of large and highly imbalanced data.

[13] Neale Ratzlaff and Li Fuxin. Hypergan: A generative model for diverse, performant neural networks, Jan 30, 2019.

[14] In recent years, imbalance data classification attracts more attention [Mesquita, 2016; Liu, and 2015; Sun. Iterative metric learning for imbalance data classification.

[15] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, May 28, 2019.