MATH6380P Final-Project

Rongrong GAO (20619663), Junming CHEN (20750649), Zifan SHI (20619455)

1 Introduction

Feature extraction plays a crucial role in classification task for both traditional machine learning methods and the popular deep learning methods. In this project, we compare the traditional machine learning methods with end-to-end deep learning methods on the Nexperia image classification task to verify the power of end-to-end learning. Careful data preprocessing is operated on the Nexperia dataset before the analysis of the models. Different methods of feature extraction for traditional classifiers have also been explored. Besides, class imbalance can significantly influence the performance of the model and thus will also be discussed in this study. Ensemble methods use multiple learning algorithms to obtain a better prediction results that is beneficial for classification tasks, which are also included in this report.

2 Nexperia Dataset

2.1 Dataset Overview

Nexperia image Dataset for Kaggle in-class contest is aims to classify images of semiconductor devices into two main classes, good and defect. It contains 34457 train images (27420 good and 7039 bad) and 3830 test images with similar good-to-bad ratio. As it is difficult for human workers to examine all of the products and select the good ones from the unqualified devices, we would like to use deep learning methods to help pick out as many defect devices as possible while preserving the good ones, thus improving their yield rate.

2.2 Dataset Problems

After checking the dataset, we have found the following problems.(1)There are some noisy labels, some are not semi-conductors (roughly 300 images out of 30k train images);(2)The class label is heavily imbalanced with a ratio of roughly 1:4 (detect:good); (3)Some duplicates exists in the train/test sets, all of them are tagged with "APG_ITIS" in file name. So at the beginning of the contest, we will firstly deal with the problems of the dataset and obtain a clear and accurate dataset.

3 Method

3.1 Data Preprocessing

We randomly split the dataset to training set and validation set with a ratio of 15:85.

3.1.1 Unlabeled File Detection and Removal

From the dataset, we can find that there are nearly more than 300 images doesn't have area annotations, which means they are either mislabeled or devices without any ground truth. In order to give a fair classification task for all devices, we select those files and remove them from our training/testing sets. From the detected unlabeled files, a future finding is that some of them are not devices at all, which is based on the observation that there are usually holes in the center area of the image instead of numbers existing in most cases. So in this step, we remove all non-devicess. Fig 1 show a sample of device and non-device.

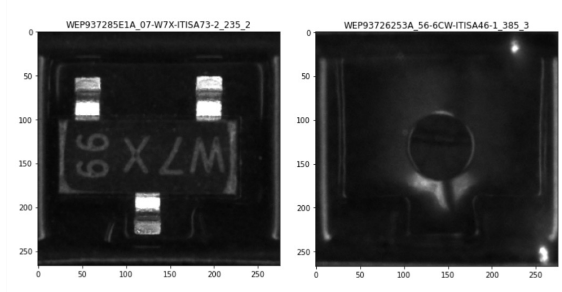

Figure 1: A sample of device and non-device

3.1.2 Class Imbalance

After the removal of unlabeled files, the proportion of good vs defect data is roughly 4:1, which means if test set are in the similar distribution, the prediction results of good devices will be 80% in accuracy. A wrong decision on defect sample will greatly affect the defection prediction rate, while it is not the case for the prediction result for good devices. And this is a common problem in most classification problem, a typical solution is that when selecting dataset from the training sets, we sample the defect cases by four times, which will make the random sampling in the training process have nearly equal change to be either good or defect. However, a better method to deal with imbalanced data is to add constraint on the loss function. We will discuss it in detail in Section 3.3.2.

3.1.3 Duplicate Detection and Removal

From the dataset, we also have some cases for validation in order to avoid over-fitting problem, which implies the phenomena that the model performs well on the training sets while not that well on the testing sets. But in our dataset, we find that there are some duplicate cases in training sets and validation sets, which will alleviate the effect of validation. So in this step, we scan the two sets and check whether there are same cases in the two sets by hash method because it is simple and fast for large dataset. The method pipeline is as follows:(1)Hash the image pixel values; (2)Check if any hash values are identical. This can effectively find out the duplicates to avoid the wrong separation of the dataset.

3.2 Feature Extraction for Traditional Method

Feature extraction plays a vital role for traditional machine learning method. We employ the ResNeXt-101 [1] for robust feature extraction. The ResNeXt-101 is either pretrained on ImageNet or fine-tuned on the Nexperia dataset. Random forest is used as the classifier, which is a well-known ensemble learning method for classification. In the random forest, a set of decision trees are constructed to fit on the various sub-samples of the dataset and averaging is adopted among all the decision tress to give the final prediction, reducing the problem of over-fitting. In this study, we explore the influence of pretrained and fine-tuned features for image classification with random forest.

3.3 End-to-End Learning

Different from traditional machine learning method, deep neural networks can leverage end-to-end learning to solve complex problems with effective feature extraction.

3.3.1 ResNeXt

We choose ResNeXt [1] as our backbone. ResNeXt is a powerful network combining the advantages of ResNet and Inception. A block of ResNeXt is shown in Figure 2. Instead of adopting the careful design for each branch in Inception, ResNeXt uses the same topology for each branch, and thus is closely related to group convolution. It adjusts the number of groups through the cardinality.

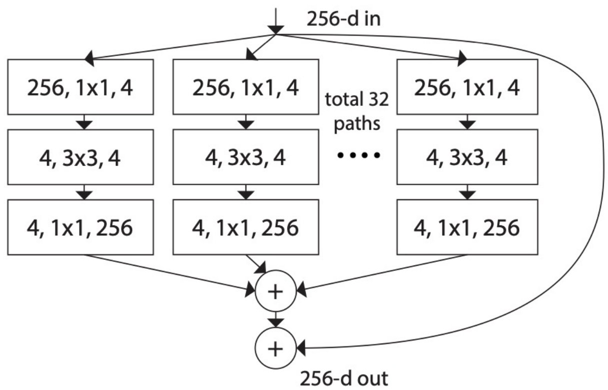

Figure 2: A block of ResNeXt with cardinality of 32.

3.3.2 Focal Loss

With a typical cross entropy loss for image classification, easily classified cases contribute to the majority of the loss and thus dominate the gradient when updating the parameters of the model. The situation is even worse when class imbalance problem is severe in the dataset. To reduce the impact of class imbalance and enhance the ability of distinguishing difficult cases, focal loss [2] is adopted for effective training. The focal loss is defined as

$$L_f(p_t) = -(1-p_t)^{\gamma} log(p_t), \tag{1}$$

where $(1-p_t)^{\gamma}$ is a modulating factor added to the original cross entropy loss and γ is a tunable parameter. The modulating factor is to reduce the contribution of the easy cases and force the model to focus on the hard examples. p_t is defined based on the estimated probability p through the model

$$p_t = \begin{cases} p & if\ class\ label\ is\ 1, \\ 1-p & otherwise. \end{cases} \tag{2}$$

3.3.3 Ensemble Learning

In the supervised learning algorithm of machine learning, our goal is to learn a stable model that performs well in all aspects, but the result is often not that ideal. Sometimes we can only get multiple models with preferences (The weakly supervised model performs better in some aspects). Ensemble learning is to combine multiple weakly-supervised models here in order to obtain a better and more comprehensive strong-supervised model. The underlying idea is that even if a certain weak classifier gets a wrong prediction, other weak classifiers can also correct the result.

In this project, we use late fusion method, which means we firstly obtain multiple scores form models trained on the full dataset and then average the scores to get the final probability of the classification.

4 Experiments

4.1 Experiments Analysis

4.1.1 Random Forest

We demonstrate the significance of properly extracted features for image classification with traditional machine learning classifier - random forest. The number of decision trees in the random forest is set to 100 for all the experiments. Three types of features are extracted from the last convolutional layer from the ResNeXt-101 pretrained on ImageNet, the ResNeXt-101 fine-tuned on Nexperia dataset with cross entropy loss, and the ResNeXt-101 fine-tuned on Nexperia dataset with focal loss, respectively.

The AUC scores for three cases are reported in the Table 1. The model pretrained on ImageNet can extract effective features for classifier to tell apart the good and defect data although there exists a domain gap between the two datasets. After fine-tuning on the target dataset, the model fits into the dataset better and extracts finer features with more confidence for the following traditional classifier. Therefore, the performance of random forest is slightly better than the one with pre-trained features. Focal loss enforces the model to focus on the difficult cases, and thus the model extracts more meaningful features, resulting in better classification performance of the random forest.

Table 1: Different feature extractions for random forest

	Pre-trained	Fine-tuned	Fine-tuned with focal loss
AUC	0.94065	0.95207	**0.95327**

4.1.2 Deep Learning

We conduct an ablation study to evaluate the effects of each component of our model. For the input, we analyze the impact of two kind of network architecture: wide-ResNet [3] and ResNeXt [1] with and without focal loss. Table 2 shows the experimental results.

Random forest achieves the worst performance with a margin of 0.04 on the Nexperia dataset compared to deep learning methods, which implies that end-to-end learning is more effective on proper feature learning and brings a significant improvement on the classification results. Focal loss helps the model to pay more attention on hard examples and therefore further improves the capability of the model to distinguish between the good and defect samples with an increase of 0.002 in AUC score. It also illustrates that class imbalanced hinders the model from learning fair and effective representations of the data.

Table 2: Results of different architectures.

	Random Forest	wide-ResNet [3]	wide-ResNet+Focal Loss	ResNeXt [1]	ResNeXt+Focal Loss
AUC	0.95327	0.99612	0.99585	0.99616	**0.99831**

Furthermore, we have researched on results of multiple ensemble methods. The ensemble method aggregates the predicted confidence scores from two methods for more robust prediction. We try the ensemble method by averaging the confidence scores of multiple models: (1)ensemble of two resNext models with and without focal loss; (2)ensemble of two resNext models and a wide-resNet without focal loss;(3)ensemble of two resNext models and two wide-resNet models. Table 3 shows the experimental results.

Table 3: Ablation study on ensemble methods.

	two ResNeXt	two ResNeXt+wide-ResNet	four models
AUC	**0.99844**	0.99820	0.99794

5 Discussion

In this study, we demonstrate that end-to-end learning extracts more meaningful features for classification and thus achieves better performance on classification task, compared with traditional classifier, e.g. random forest. Feature extraction is crucial for both the traditional machine learning methods and the deep learning methods. Class imbalance is one of the factors that influence the quality of the extracted features, and focal loss can help the model to learn from imbalanced data and hard examples effectively. Ensemble learning gathers the information learned by several models and thus produces more robust classification results. However, since we only use the late fusion method and average the confidence scores with equal importance, the improvement brought by ensemble method is limited. We believe that learning the weights for different models or combining the models in feature level could further improve the classification performance.

6 Acknowledgement

Junming CHEN: Deep learning methods, focal loss and ensemble implementation. Data pre-processing. Report writing.
Zifan SHI: Feature extraction, traditional methods implementation and experiments. Report writing.
Rongrong GAO: Report writing. Slides preparation and presentation.

Reference

[1] Saining Xie, Ross Girshick, Piotr Dolĺar, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1492–1500,2017.

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, pages 2980–2988, 2017.

[3] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. ArXiv preprint arXiv:1605.07146, 2016.