A computer vision approach of waste sorting based on Dense Convolutional Networks

Zizhuo Wang
SIST
Shanghaitech University
wangzzh@shanghaitech.edu.cn

Suan Xia
SIST
Shanghaitech University
xiasa@shanghaitech.edu.cn

Xiner Xu
SIST
Shanghaitech University
xuxe@shanghaitech.edu.cn

Abstract

Efficiently waste sorting has become a problem which now we urgently awaits to be solved these days. In this project, we design a computer vision approach of automatic waste sorting based on Dense Convolutional Network. Waste can be sorted into more than 200 classes according to their outside visual feature. With Dense Convolutional Networks, we can make better use of features and build convolutional network more efficently.

1 Introduction

1.1 The meaning of waste sorting

Unlike the past, in the modern society, waste may contain certain chemical contents which may lead to some severe problems, such as polluting the water and soil and raising the morbidity of the residents. Though some methods of dealing waste are taken to isolate the waste from our daily life, there still exists some infiltration inevitably. Thus, proper procedures to deal waste is closely related to the waste's influence on human beings, which depends on the result of waste sorting.

The current main method of dealing waste is still landfilling, which just moves the waste from one place to another. Though it is a cheap and convenient way, it has a great cost of land resource to a country. However, waste reusage is a good way to reduce the amount of waste to be dealt with. If we do waste sorting before burning or landfilling the waste, recycled waste will be sorted out and the amount of waste will be reduced.

1.2 The difficulty of waste sorting

Though there are great benefits of waste sorting, it is not easy to do so. Current waste sorting work is mainly done manually. In most cities, waste sorting is done by workers. But human resource is so costly that every worker needs to be trained to sort waste.

In very few cities like Shanghai, waste sorting is partly done by residents in advance. However, not everyone is willing to spend time on learning waste sorting. And there will be some mistakes on the sorting result inevitably. Also, the regulation of various kinds of garbage bins costs a lot. From the observed present situation, it can be concluded that sorting waste manually is not so optimal to satisfy the goal of the published policy.

1.3 Computer vision approach of waste sorting based on machine learning

Manual work costs a lot, so we suggest using a modern way to sort waste automatically. Benefited from machine learning methods, machine can learn how to sort waste. Then waste can be sorted automatically and the cost of human resource can be greatly deducted. The workload of sorters will be reduced. Meanwhile, it will be unnecessary for residents to learn the waste sorting knowledge clearly, since they can leave the the classification work for the sorting machine.

Besides the benefits of cost saving, waste sorting based on machine learning can provide a precise result. The high accuracy classification result ensures the correct way to deal with a certain kind of waste, which helps reduce pollution and recycle some useful rubbish. Also, there is also some commercial potential of automatic waste sorting. In some developing countries, it is not easy for them to sort and recycle the waste. Due to limited government budget, it is impossible to develop professional skills for the sanitation worker, leading to low ratio of waste recycling. Instead, applying waste sorting machine is helpful to solve this situation. The initial investment on sorting machine can significantly promote the recycling ratio, which can bring economic benefits in the long term.

In short, developing the computer vision approach based on machine learning to sort waste is a future trend in this field since its benefits overweighs the initial costs of developing.

2 Methodology

2.1 Introduction to Dense Convolutional Network

Recent work has shown that convolutional networks can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. Dense Convolutional Network (DenseNet) is such a kind of convolutional networks which connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections - one between each layer and its subsequent layer - DenseNet has $L(L+1)/2$ direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. Each layer adds k feature-maps of its own to this state. The growth rate regulates how much new information each layer contributes to the global state. The global state, once written, can be accessed from everywhere within the network and, unlike in traditional network architectures, there is no need to replicate it from layer to layer. (1)

2.2 Advantages of DenseNet

DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

Counter-intuitive effect of this dense connectivity pattern is that it requires fewer parameters than traditional convolutional networks, as there is no need to relearn redundant feature maps. Traditional feed-forward architectures can be viewed as algorithms with a state, which is passed on from layer to layer. Each layer reads the state from its preceding layer and writes to the subsequent layer. It changes the state but also passes on information that needs to be preserved. DenseNet architecture explicitly differentiates between information that is added to the network and information that is preserved. DenseNet layers are very narrow (e.g., 12 feature-maps per layer), adding only a small set of feature-maps to the "collective knowledge" of the network and keep the remaining feature-maps unchanged — and the final classifier makes a decision based on all feature-maps in the network. (1)

Besides better parameter efficiency, one big advantage of DenseNets is their improved flow of information and gradients throughout the network, which makes them easy to train. Each layer has direct access to the gradients from the loss function and the original input signal, leading to an implicit deep supervision. This helps training of deeper network architectures. (1) Further, it is observed that dense connections have a regularizing effect, which reduces over- fitting on tasks with smaller training set sizes. (1)

Concatenating feature maps learned by different layers increases variation in the input of subsequent layers and improves efficiency. This constitutes a major difference between DenseNets and

ResNets (2). Compared to Inception networks (3) (4), which also concatenate features from different layers, DenseNets are simpler and more efficient.

3 Experiment

We use 2 method of DenseNet together with other several methods on a dataset of pictures of manually sorted waste.

3.1 Datasets

The datasat includes 80961 pictures of waste which are classified into 246 classes. The classification covers almost all kinds of waste in modern city life. This dataset is created by AI Studio and can be obtained here .

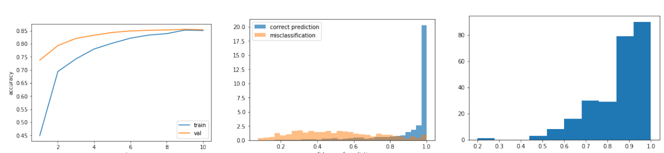
3.2 Result and Analysis

We used DenseNet of 121 layers (Figure 1) and 201 layers (Figure 2), together with ResNet (2) (Figure 3), SqueezeNet (5) (Figure 4), Binary weight network (6) (Figure 5) and Ternary weight network (7) (Figure 6) . Due to the limitaion of our PC, we only trained the models for dozens of epoches. As a result, we can see that DenseNet of 201 layers has the highest accuracy in 12 epoches.

The core of ResNet is to establish shortcuts between the front and back layers, which will facilitate the back propagation during training and enable deeper CNN networks to be trained. The basic idea of DenseNet is the same as ResNet, but it establishes dense connections between the front and back layers. Another feature of DenseNet is to achieve feature reuse through the connection of features on the channel. These features allow DenseNet to achieve better performance than ResNet with fewer parameters and calculation costs.

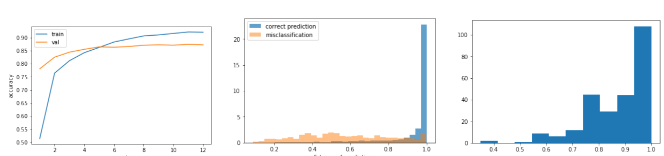
SqueezeNet proposed a new network architecture Fire Module to compress the model by reducing parameters. FireModule consists of the squeeze module(Dimension reduction using 1*1 convolution) and the expand module (Use 1*1 convolution +3*3 convolution to increase dimensions). But the trainning process of SqueezeNet is slower and it is still under-fitting after 12 epoches.

In the forward and reverse training of DNN, 1bit binary weight is used to replace floating point weight, which can simplify the multiplication operation into a simple accumulation operation which will greatly reduce the storage space. Ternary weight network quantifies weight into ternary, namely 2-bit. In fact, the Ternary weight network is better than the Binary weight network but it's still not accurate enough. It is also under-fitting after 14 epoches.



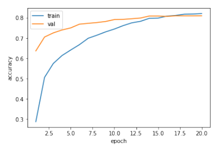
(a) epoch (b) confidence (c) accuracy

Figure 1: DenseNet 121

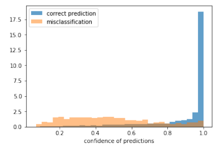


(a) epoch (b) confidence (c) accuracy

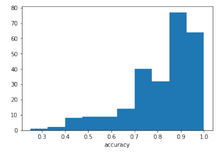
Figure 2: DenseNet 201



(a) epoch

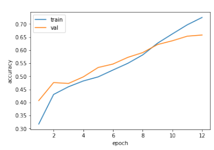


(b) confidence

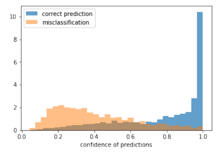


(c) accuracy

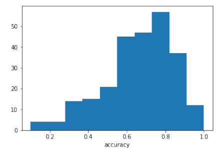
Figure 3: ResNet



(a) epoch

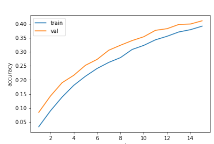


(b) confidence

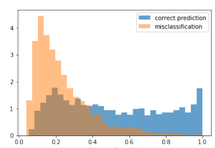


(c) accuracy

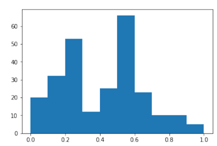
Figure 4: SqueezeNet



(a) epoch

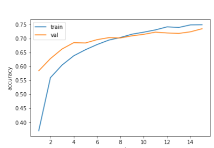


(b) confidence

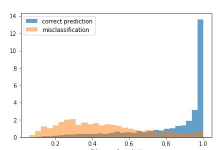


(c) accuracy

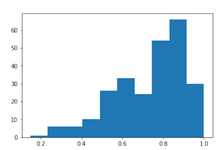
Figure 5: Binary weight network



(a) epoch



(b) confidence



(c) accuracy

Figure 6: Ternary weight network

4 Conclusion

How to sort waste efficiently has become a problem which now we urgently awaits to be solved these days. With the help of machine learning, it is possible to develop a computer vision approach of waste sorting. DenseNet is an efficent structure of convolutional networks which makes better use of features and reduces the number of parameters. This lead to a quicker and more efficent trainning process and make the convolutional networks higher accuracy after a short training.

Due to the limitaion of our PC, we only trained the models for dozens of epoches. Some of the models remains under-fitting and require more training. Also, in addition to vision features, other kinds of features like smell, weight and temperature can be very helpful in waste sorting. This might become the future works of this project.

References

[1] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017.

[2] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In *CVPR* , 2015.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR* , 2015. 2, 3

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna.Rethinking the inception architecture for computer vision. In *CVPR*, 2016.

[5] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).

[6] Rastegari, Mohammad, et al. "Xnor-net: Imagenet classification using binary convolutional neural networks." *European conference on computer vision.* Springer, Cham, 2016.

[7] Li, Fengfu, Bo Zhang, and Bin Liu. "Ternary weight networks." *arXiv preprint arXiv:1605.04711* (2016).