JD AI Fashion Challenge-Fashion style Technical Report

Ze Si
Department of automation
University of Science and Technology Beijing
Beijing, China
13472139273@163.com

Dongmei Fu
Department of automation
University of Science and Technology Beijing
Beijing, China
fdm2003@163.com

Zhicheng Huang
Department of automation
University of Science and Technology Beijing
Beijing, China
zhichenghzc@gmail.com

Yang Liu
Department of automation
University of Science and Technology Beijing
Beijing, China
ustbzdhly@163.com

***Abstract*—With the expand of fashion consume market and the pervasive technologies for AI in consumption filed,AI now is having great impact on the fashion industry.Chanese fashion consume market now is becoming individual and diverse, the fashion style now has deeper impact on consume motivation. Different designs of different clothes with different clothing matches will show more than one style those may simultaneously exist.The strong professional requirement makes fashion style recognition become one of the hardest multi label classification task. In this report,we propose a method based on the convolution neural network for the fashion style recognition.By this method,we achieved a F2-score of 0.6525,and finally got top 2 on the leaderboard.**

***Index Terms*—AI, multi-label,classification, neural network, fashion-style**

I. INTRODUCTION

Fashion style of the clothes have more and more impact on people's consume with the updating of the consumer market and the improvement of consumer demand.Fashion style reflects the social features of the times and the national tradition.There are many factors that may influnce the style of clothing,like structure, fabrics,matching between costumes, etc. One costume may have multiple styles in label space.The essence of fashion style recognition is multi-label classification task. Traditional supervised learning's research objects are often represented by single instance with single label.Recent years there are more machine learning researchers and related communities pay their attentions to multi-label learning,during 2007 to 2012,there are more than 60 papers with keyword multi-label in the title appearing in major machine learning-related conferences including ICML,ECML,ICDM,NIPS and so on[1].But when it comes to the task in the real world, there are still challenges such as evaluate metric, unbalanced sample distribution,the oversized label space ,and optimization of algorithm in practical task is still worth to study.

As the application of fashion Style recognigion, JD AI Challenge require to design a recognition algorithm to predict the style of the cloth in the image as (1):

$$s = f(I) \subseteq S \tag{1}$$

Where S represents the aggregate of thirteen styles

S={ Sport, Relax, OL,Japanese, Korean, European, England, Miden, Lady, Simple, Nature, Street and Nation.}

Each image sample usually have more than one styles. The competition organizer provided 55,000 professionally labeled image samples and their labels as training and validation set,and 10,000 images as test set to evaluation the performance of the recognition algorithm. Evaluation metric of the algrithm is F2-score (2)

$$\frac{\sum_{i=1}^{|S|} \frac{(1+2^2)Precision_i \cdot Recall_i}{2^2 Precision_i + Recall_i}}{|S|} \tag{2}$$

When we evaluate the performance of the algrithm, multi-label learning is much more complicated than the single-label learning.There are more kinds of evaluate metrics for multi-label learning include the F-score,but the metric's form is different than that in JD AI-fashion challenge.The metric proposed above pay more attention to evaluate every single label predicted by the algrithm. It may have different conclusion when we use different evaluation metrics on the same algrithm. So it is necessary to choose the appropriate optimizing way when we deal with the specific issues. Difficulties in this multi-label prediction task include the abstraction of the style recognition and unbalanced samples in each label.The style recognition is too abstract to extract features in traditional way,convolutional neural network is an effective way to solve this problem now.And we use weighted loss function and threshold-moving to relieve the problem caused by unbalanced samles.

II. MATHOD

A. Model Selection

The First step of classification problem is feature selection and extraction.For image sample it is a complex work to select the suitable feature for the task.The fashion style recognition

base on the structure, matching between costume,etc,and most importantly,it relies on the rich experience of the experts,but to select the feature we need more concrete prior knowkedge. Image classification in traditional machine learning,color feature, texture feature and shape feature tend to measure the similarity shown among the images.But for style recognition, it is higher semantic feature for expert to recognize the fashion clothes' style, more abstract feature than the color, texture and so on.Two fashion images look completely different may have the same fashion style like Fig1,this two image sample both have the label "european",however,it is difficult to classify the samples with current label by descriptive manual extraction features.

Fig.1. Two images have the same label "european"

Recent years,deep learning has gained wide attention of the researchers from different fields.Espacially, deep convolutional neural network has reached the-state-of-art in many image classification and other image processing tasks.Compared with traditional machine learning, one of the noticeable advantage of deep learning is the way of feature extraction that automatically learn features from large amount of data. Convolution neural network's layers from low-level to high- level can be seen as a series of feature extractors,low-level layers aim at extract detail features,high-level layers aim at global features and semantic information,a series of feature extractors stacked from low-level to high-level can be automatically learned from large-scale training data in an end-to-end manner[2].So,for the JD-AI Challenge task,the method of deep learning is the best choice,we choose densenet[3] as the basic network and modify the basic net according to the task.

B. Output Of The model

For the task of multi-label classification,the problem to be solved is to find the mapping relation between the sample and the labels in label space associated with the sample.For the current task,there's two possible values 0 or 1 for each label,for 13 labels the possible label sets would achieve2^{13} ,for general multi-label classification task, it can be treated as 2^{13} classification,but there is too much redundancy in this huge output space. There are only limited label sets appear in the practical samples,but the prior knowledge is not sufficient to accurately eliminate the redundant part of the label space.And for the label sets appeared in the data set, limited number of the samples for each label set and unbalanced sample quantity between different labels make it not feasible in this way.In this task,we set n-dimensional vector as the output of the model,and n is the number of labels for the current model to predict,each dimension of the output vector represents a measure of the occurrence likelihood of a corresponding label.

Conventional convolutional neural networks' classification part vectorize the feature maps of the last convolutional layer and feed it into fully connected layers,but fully connected layers with too many parameters are prone to overfitting.so Min Lin et al.proposed a method takes the average of each feature map from the last convolutional layer called global pooling[4].The model we used is densenet201 with global pooling and the vector is fed into a two layers fully connected layers,the first layer have 1024 nodes and the output layer have n nodes.For different n,multi-label net and single-label net are showed in Fig2 and Fig3.

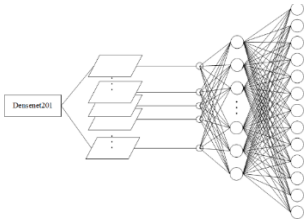

Fig.2. Multi-label net

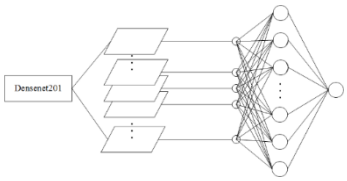

Fig.3. Single-label net

C. Loss Function

For the current task, there is serious sample imbalance problem for each label.Table 1 shows positive and negative sample ratio of each label.

LABLE	POSITIVE：NEGATIVE
Sport	1:527
Relax	1:12
OL	10:1
Japanese	1:208
Korean	1:2
European	1:5
England	1:278
Miden	1:35
Lady	9:1
Simple	1:7
Nature	1:14
Street	1:43
Nation	1:151

Table.1. Positive and negative sample ration of each label

We used weighted loss function and threshold-moving to relieve the sample imbalance problem.Weighted loss function we used based on the cross entropy,as one of the most commonly used loss function,form of cross entropy loss function showed in (3)

$$\mathrm{L} = -\big(yln(p) + (1-y)\ln(1-p)\big) \qquad (3)$$

Where y is the label of current sample an the value is 1 or 0,p is the model's output for current sample represent the probability that the sample have the label "1".Weighted cross entropy loss function add weight to each item of the polynomial,as in (4)

$$\mathrm{L} = -(Ayln(p) + B(1-y)\ln(1-p)) \qquad (4)$$

Where A and B are the weights for the positive sample and negative sample,in pratical application,we set the value of A and

B according to the positive and negative sample ratio in the training set, to give the item corresponding to smaller number of samples a larger weight.Take label “nature” as an example,set 0.5 as the decision threshold,Fig4 shows the F2-score of models trained based on these two different loss.

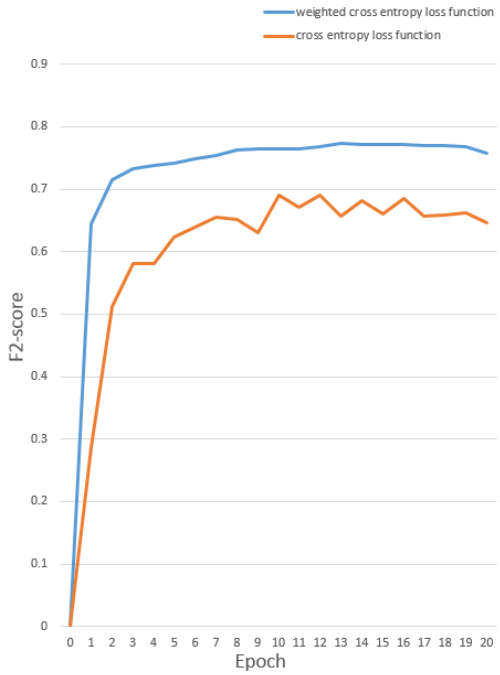

Fig.4. Training process for two different loss function

As it shown in Fig4,without threshold-moving,model trained by weighted entropy cross loss function got higher F2-score than it trained in normal cross entropy loss function.Actually,it is related to the selection of decision threshold,but according to our experiment,adjust the threshold base on the validation set for both models trained by these two different loss function,the model trained by weighted loss function still have better performance on the test set.

D. Threshold-moving

Output of the net is the predict probability of the corresponding label for the input sample.The final result is determined by whether the probability exceeds the decision threshold,so the threshold we choose will affect the F2-score of corresponding label.According to our experience,in the case of large amount of data,test set have similar data distribution to validation set.We use the trained model to predict the label of the validation set,and plot the relationship between the threshold and the F2-score,where the horizontal axis is the decision threshold and the vertical axis is F2-score.Take the threshold that maximum the F2-score for validation set as the decision threshold for the test set prediction.With limited data,the data distribution of validation set and test set will not be identical,so the best threshold may be slightly different ,but when the model trained by weighted cross entropy loss function to predict the test set label,the F2-score and threshold curve changes gently,so the F2-score is insensitive to threshold changes around around the extrem point,therefore,we can choose relatively good threshold for the test set.For example the label “nature”,Fig5 and Fig6 shows the score curves of validation set and test set.

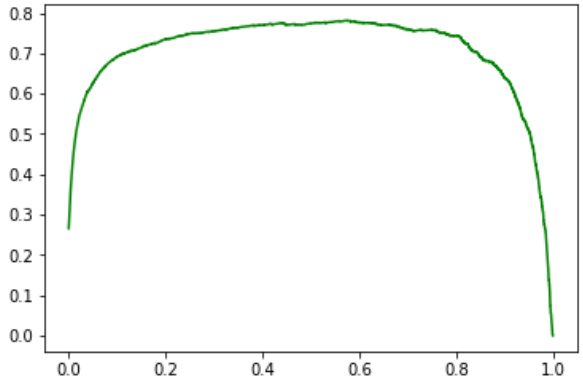

Fig.5.Validation set score curve

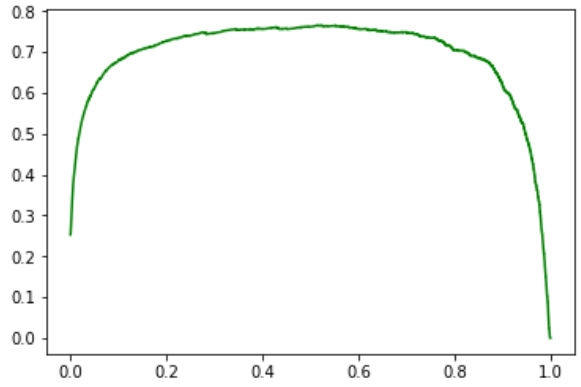

Fig.6.Test set score curve

We choose 0.5748 as the decision threshold according to the validation curve.And the F2-score on test set in this threshold is 0.7615,which is near to the maximum 0.7651.Compared to it, model trained by normal cross entropy loss function is more sensitive to threshold changes as showed in Fig7.Take threshold selected by validation set as the decision threshold of test set,the F2-score may get more divation to the highest score.

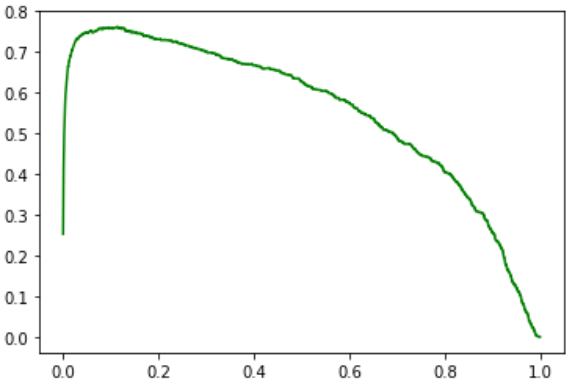

Fig.7.Validation set score curve by model trained by normal cross entropy loss function

E. *Model Training*

We trained models with two different output structure,the basic netword densenet201 have pretrained on ImageNet and fine-tune on the training set, for training process,we use two GTX1080Ti for training, choose Adam as optimization algorithm,take batch size 46 to fit the GPU's memory size.At the first epoch,we set the learning rate to 0.00001,from the second epoch we keep the learning rate or reduce it to 0.000001 according to the learning result.We trained a model to predict 13 labels and a series models for each single label,for multi-label model,we take sum of the losses of each label as the total loss.And then we compared these two different solution's performance.Fig.8 shows the muti-label training process.

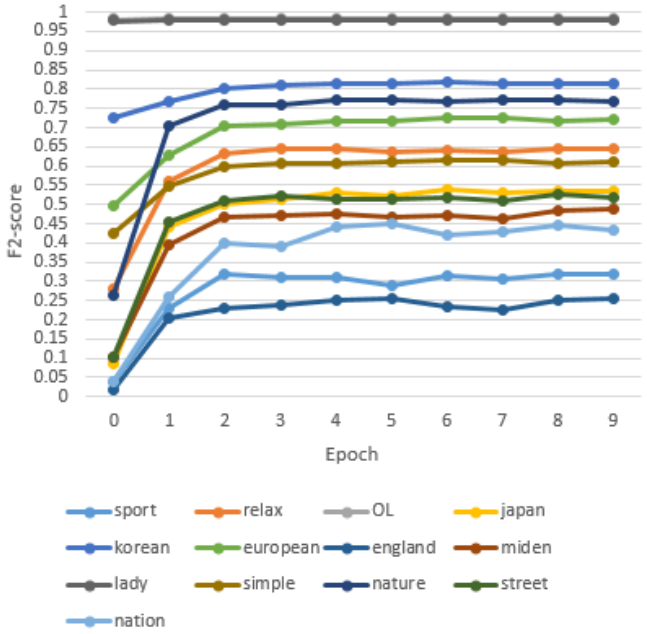

Fig.8.Multi-label network training process

The average of thirteen F2-score on the multi-label model is 0.62,but as we can see on the Fig 8,for multi-label training,we can't guarantee that all the labels get state-of-art during the training process,as the new epoch makes one label perform better but the others worse,different labels interact each other during training.The single-label models,in contrast,there are several models with same structure for each label,we can train each label independently,for different labels we can use different learning rate and take different training loops,and different labels won't affect each other.In addition,most of the single-label models get 0.02~0.03 higher F2-score than the multi-label model,single-label model's F2-score for each label shows in Table2.

LABLE	F2-SCORE
Sport	0.4
Relax	0.68
OL	0.98
Japanese	0.52
Korean	0.82
European	0.74
England	0.3
Miden	0.47
Lady	0.98
Simple	0.63
Nature	0.79
Street	0.54
Nation	0.49

Table.2. Single-label models' F2-score

The data in table 2 is the F2-score of validation,we use these models to predict the test set and submit the result as the first submission of the JD AI Fashion Challenge.Later we tried to use different basic net include resnet101 and resnet152,and fuse these models on the probability layer,and the F2-score is about 0.01 higher than before.

ACKNOWLEDGMENT

All data above are provided by the JD AI Fashion Challenge organizer.This work was partly supported by the University of Science & Technology Lab 801 and China Postdoctoral Science Fundation(No.2017M620615)

REFERENCES

[1] Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE transactions on knowledge and data engineering, 2014, 26(8): 1819-1837.

[2] Yu W, Yang K, Yao H, et al. Exploiting the complementary strengths of multi-layer CNN features for image retrieval[J]. Neurocomputing, 2017, 237: 235-241.

[3] Huang, Gao, et al. "Densely Connected Convolutional Networks." CVPR. Vol. 1. No. 2. 2017.

[4] Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.

[5] Yosinski, Jason, et al. "How transferable are features in deep neural networks?." Advances in neural information processing systems. 20