Analysis and improvement of face detection based on surf cascade

Siquan Hu[1, 2], Caihong Zhang[1, 3] and Lei Liu[1, 4]

[1]School of Computer and Communication Engineering, University of Science and Technology Beijing, 100083, China
[2]Email:husiquan@ustb.edu.cn
[3]Email:s20150063@xs.ustb.edu.cn
[4]Email: liulei2776@gmail.com

Abstract. This paper aims to study limitations of the commonly employed boosting cascade framework. We focus on the factors like data, feature, weak classifier and stages. A set of novel experiments were done to show the relationship. The model contains three key points: SURF feature, weak classifier based on logistic regression and AUC-based cascade learning algorithm. This paper adds cross validation in logistic regression creatively which improves accuracy and speeds up convergence greatly. Eventually only five stages and about 100 weak classifiers are needed. The frontal face detector improves reject rate to 99% for the first three stages, decreases number of false positive greatly and achieves comparable performance among non-CNN techniques on FDDB dataset.

1. Introduction

Object detection is the basis of object recognition and tracking, and is widely used in security and other fields. The method based on machine learning consists of two steps: feature extraction and classifier design [1].Commonly used features are Hog, HAAR, Edgelet and so on. The classifier is mainly based on SVM, adaboost. Dalal proposed a pedestrian detection method based on HOG and SVM [2].Viola and Jones [3] used 3 types of 4 forms of HAAR for face detection. M.Pietikäinen [4] proposed LBP for texture feature extraction. Bo Wu [5] proposed edgelet. Dollar [6] proposed integral channel feature.

Object detection algorithms based on deep learning, like Faster RCNN [7] and YOLO [8], have achieved a high detection rate. But because of complex computing process and the requirement of GPU, it is difficult to achieve real-time demand.

2. Analysis of Boosting Cascade Model

The whole model is obtained by cascading the strong classifiers obtained by boosting a number of weak classifiers. We study all factors as below that influence model capability. We did quantities of comparison experiments to illustrate how to choose these parameters.

Data. The degree of difficulty of the samples varies greatly. Data augmentation like flip is used.

Feature. We should consider feature dimension and computational complexity. Then how to select the best feature becomes a difficult point for training.

Weak classifier. We should consider the calculation time and accuracy, and select the best method. Common methods are SVM, Logistic regression, etc.

Cascade. The number of weak classifiers and stages always influence final detection rate.

3. Details of Model

In our model, we use SURF feature, logistic regression with cross validation, AUC-based algorithm, and train in cascade structure. We will describe these below.

3.1. Feature

SURF [9], a local feature descriptor, reflects the shape and texture of feature points. It is invariant and has a great improvement in computing speed. Here, we use SURF descriptor to extract gradient information, which is robust to the rotation of faces. As shown in Figure1, We set the detection widow size 40×40, and the size of feature rectangle gradually increases from 8×8 to 40×40 and slides in the 40×40 detection window, resulting in thousands of candidate features. Each feature is divided into 2×2 or 1×4 image blocks, and we use [-1, 0, 1] to compute the gradient information of the horizontal, vertical and diagonal dimensions of the pixel. And we compute $|d_x| \pm d_x$. In this way, we get 8 dimensional information. Then, the 8 dimensional gradient information of each image block is summed $\sum(|dx| + dx), \sum(|d_x - d_x|)$. So we get 32 dimensional vectors for every feature.

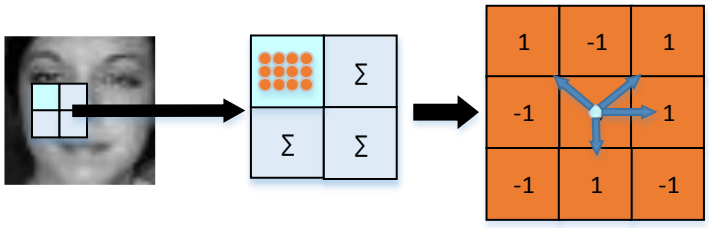

Figure 1. The flowchart of feature selection and calculation.

Compared to the 180 thousand of HAAR in 24×24 detection window, the number of candidate candidates for SURF features in this paper is greatly reduced. There is only 2143 candidate feature. There are usually three ways to select best feature from our candidate features. As in equation (1) the best feature can be found via minimizing the sum of the absolute values of the error weights of all samples, via minimizing the sum of squared errors of all samples as in equation (2).

$$\sum_{i=1}^{N} w_i |h_i - y_i| \tag{1}$$

$$\sum_{i=1}^{N} w_i (h_i - y_i)^2 \tag{2}$$

As in equation (3), the maximum value of AUC which combines the classifiers obtained in the previous training is used to find best feature.

$$J(H^{i-1} + h_j(x, w)) \tag{3}$$

3.2. Weak Classifier

The weak classifier is used to classify the result of a single feature, and the common methods are SVM, logistic regression. In this paper, we choose logistic regression, and the optimized processing method of [10], greatly improving the accuracy of weak classifier. Given samples $\{x_i, y_i\}_{i=1}^{N}$, the output of samples is define as P. The process of solving W is translated into minimization of L.

$$P = \frac{1}{1+exp(-yw^T x)} \tag{4}$$

$$L = \sum_{i=1}^{N} lg(1 + exp(-yw^T x)) + Cw^T w \tag{5}$$

The parameter C adjusts the proportion of the error and the regular term, including L1, L2.Table1 shows the great change of C. In this paper, we add cross validation to improve accuracy, and set the range of C from 0.001to 1024 and select best C which gets the minimum false positive rate. Figure 2 shows that the change of the weak classifier promotes the convergence quickly using means of cross validation.

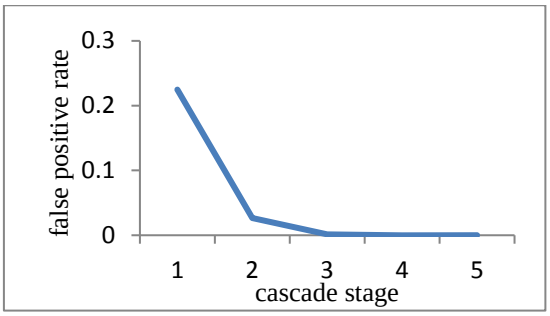

Figure 2. Accumulated false positive rate.

Table 1. Compare of parameter C.

Param.C	False positive rate
0.1	0.514
1	0.3114
10	0.2044
cross validation	0.2097

4. Experiments

The training and detection experiments were done on a personal workstation with 2.6GHZ Core-i5 and 8GB RAM. We evaluate our models on the public dataset FDDB, including 2,845 images with 5,771 faces, about 80% of which are frontal faces. We get ROC to show TPR (true positive rate) and number of FP (false positive). We only focus on four points on the ROC, the TPR of FP=0, FP=100, FP=250 and the max of FP. Frontal faces are collected from GENKI dataset, FERET dataset, totally 12,000 faces cropped and resized to 40×40 .The negative samples are mainly from images download from network by tools and Caltech dataset etc. Finally, we collect 24,000 images without faces. All experiments bellow are trained with these data and evaluated on FDDB dataset.

4.1. Data

It is impossible to collect all the samples owing to the samples diversity. Data augmentation techniques are used to improve the capabilities of the model. We get other 12,000 faces with mirror transform, and 12,000 faces by random perspective transforming face image within [-10, 10] degree. The number of false positive in the model decrease 130 but true positive rate decrease 1% compared with model without data augmentation. The model with data augmentation has more stages (5 vs 8) because of more complex data.

4.2. Feature

Figure 3 shows the features of the first three stages, which are clearly focused on the eyes, nose, mouth. The number of candidate feature is restricted by growth step size and moving step size in the detection windows, which has a great impact on memory usage and training time.Table2 shows the different results on number of candidate feature.

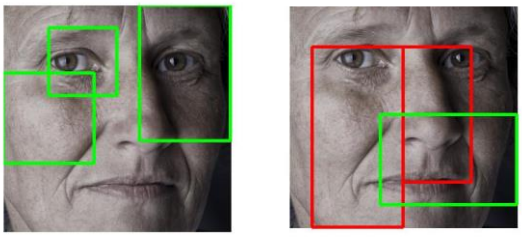

Figure 3. The feature of first three stages.

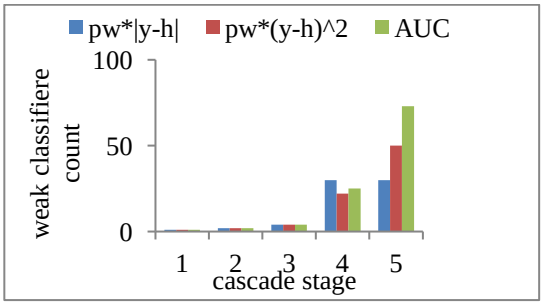

Figure 4. The number of weak classifier.

Table 2. The result of models with different number of candidate feature on FDDB.

candidate feature	TPR/FP=0	TPR/FP=100	TPR/FP=250	TPR/FP-max
482	0.392	0.705	0.742	0.762/506
1100	0.397	0.715	0.748	0.765/458
1631	0.358	0.724	0.749	0.764/414
2143	0.481	0.725	0.755	0.766/421

As the number of candidate feature increases, the true positive rate of the model is gradually improved, and the number of false positive has dropped. However, considering memory usage and training time, compared with number-2143, number-482 is a better choice. As mentioned in 3.1, there are three methods on selecting best feature. All these can rapidly converge to the FPPW 1e-6. But the AUC-based is stricter, and Figure 4 shows that the number of weak classifier of AUC-based is slightly higher. Table 3 shows that AUC-based model has smaller number of false positive. Thus we chose AUC-based method.

Table 3. The result of models with three methods of selecting best feature on FDDB.

select-feature	TPR/FP=0	TPR/FP=100	TPR/FP=250	TPR/FP-max
|pw|	0.471	0.726	0.752	0.771/606
pw^2	0.54	0.727	0.755	0.768/493
auc	0.365	0.725	0.75	0.761/ 401

4.3. Weak Classifier

Here we compare two classification algorithms, logistic regression and linear SVM with solver means with dual and original. Table 4 shows the results of SVM and LR are almost same. With smaller memory usage and faster computation speed, LR with solver means original is more suitable for the weak classifier in this paper.

Table 4. The result of models with different solver means on FDDB.

solver	TPR/FP=0	TPR/FP=100	TPR/FP=250	TPR/FP-max
L2R_LR	0.392	0.705	0.742	0.762/506
L2R_LR_DUAL	0.436	0.716	0.740	0.763/552
L2R_L1LOSS_SVC_DUAL	0.318	0.722	0.746	0.764/489

4.4. Cascade

Threshold of every stage is mainly depends on min_TPR (minimum of true positive rate). To retain more faces, we set high minimum true positive rate, resulting more weak classifiers of every stage and more stages.

Table 5. The result of models with different min_TPR on FDDB.

min_TPR	TPR/FP=0	TPR/FP=100	TPR/FP=250	TPR/FP-max
99.0%	0.159	0.687	0.743	0.746/271
99.3%	0.445	0.708	0.741	0.756/429
99.5%	0.392	0.705	0.742	0.762/506
99.6%	0.465	0.714	0.743	0.769/738
99.7%	0.609	0.724	0.749	0.779/926

Table5 shows the set of min_TPR has great influence on the TPR. Considering the final TPR and the total number of FP, we usually choose min_TPR=99.5% or 99.6%.

For an image with 450×431 pixels, we set scale to 1.1, step size to 4, and minimum of face size to 20×20, so that 200 thousand of the 40×40 detection windows need to be classified. Usually almost windows are rejected on the first few stages.

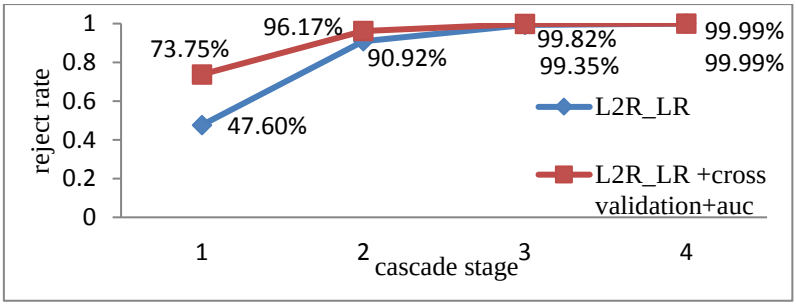

Figure 5. The accumulated reject rate of the first few stages of three different models

Table 6. The recall of models on FDDB.

Models	all2stages	all3stages	all4stages	all stages
L2R_LR	95.20%	92.88%	83.63%	75.99%
L2R_LR+crossvalidaiton+auc	95.57%	90.24%	80.58%	78.18%

Figure 5 and Table 6 show that our model with cross validation and AUC-based has higher reject rate in the first few stages and higher final TPR. The 99% reject rate shows our model has stronger learning ability.

Each parameter setting is a trade-off between detection rate and number of false positive. Compared with HAAR cascade model, as in figure 6, our model has improved a lot on the point FP=0, FP=100.And our model only has 5 stages with 105 weak classifiers. HAAR cascade model has 24stages with 2916 weak classifiers. The true positive rate of our frontal face detector is 77.9%, only missing 2% of frontal faces.

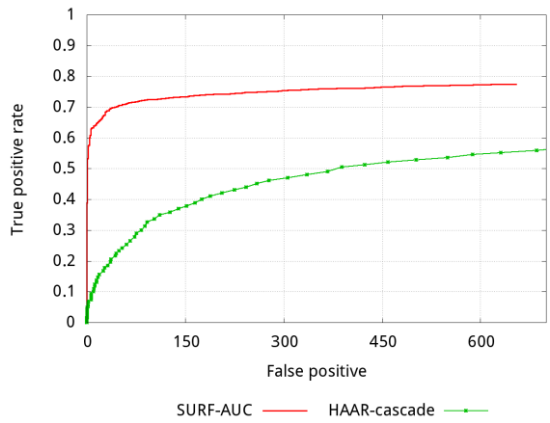

Figure 6. Compare of SURF cascade model and HAAR cascade model on FDDB dataset

5. Conclusions

This paper studies the cascade detection model based on SURF feature. The main contributions are three points. First, we analyze factors restricting the model, including data, feature, weak classifier, and cascade. And we did groups of experiments to make explicit the effect of these parameters. Second, we add cross validation in logistic regression creatively to improve accuracy, greatly speeding up convergence. Third, frontal face detector can achieve results comparable to state-of-the-art detectors using traditional machine learning means. The true positive rates are 77.9% and 60.9% when false positives are max and 0. Our detector decreases number of false positive greatly and outperforms than HAAR-cascade frontal face detector. Future work may consider combine convolutional network with cascade structure on multi-view face detection.

6. References

[1] Zhang C and Zhang Z 2010 *Technical Repor*t **66** 1-17
[2] Dalal N and Triggs B 2005 *Conf on Computer Vision and Pattern Recognition* vol1 (San Diego) pp886-893
[3] Paul V and Michael J 2001*Proc of the 2nd Int Workshop on Statistical and Computation* Theories of *Vision Modeling Learning Computing and Sampling* vol57 (Vancouver) p 87
[4] Ojala T and Pietikäinen M 2002 *Transactions on Pattern Analysis & Machine Intelligence* **24** 971-87
[5] Wu B and Ramakant N 2005 *Tenth IEEE International Conference* vol1 (Beijing) pp 90-97
[6] Dollár P et al 2009 *British Machine Vision Conf* vol 2(London) p3
[7] Ren S et al 2015*Transactions on Pattern Analysis & Machine Intelligence* 91-99
[8] Z Redmon J et al 2016 *Conf on Computer Vision and Pattern Recognition* (Las Vegas) pp 779-788.
[9] Li J et al 2012 *Int Conf on Computer Vision Workshops* vol21 pp2183-90
[10] Fan RE et al 2008 *Journal of Machine Learning Research* **9** 1871-74