International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 9 (2017), pp.1-8
http//dx.doi.org/10.14257/ijsip.2017.10.9.01

Research on Pedestrian Detection Based on PCANet

Qi Mu[1], Yikai Jia[2]* and Zhanli Li[3]

[1,2,3]College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China
[2]College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China
[1]414948139@qq.com, [2]jiayikai1018@163.com, [3]wth_xust@163.com*

Abstract

In the research of pedestrian detection, it is still a difficult task to express the pedestrian feature in the complex environment. PCANet which comprises only the very basic data processing components is a simple deep learning network. Compared with the traditional manually select features, it can extract the deeper abstract expression of the data, and has achieved a high recognition rate in the field of image recognition. In this paper, PCANet is brought in to extract pedestrian and non-pedestrian features and SVM classifier is applied for dichotomy. In the stage of detection, selective search algorithm is used to achieve pedestrian primary areas. Experimental results show that the algorithm we proposed has a higher detection rate, compared with the state of the art algorithm, in the INRIA pedestrian detection datasets.

Keywords*: pedestrian detection; deep learning; PCANet; SVM*

1. Introduction

Pedestrian detection is a fundamental problem in the computer vision and plays a key role in the application of video surveillance, intelligent vehicles and human-computer interaction[14]. Because of its importance value, many researchers have devoted themselves to it. However, it is a still challenging task caused by the intra-class variation of pedestrians in posture, clothing and the influence of complex environment such as the change of lighting, backgrounds and occlusion.

At present, one of the most popular methods in the study of pedestrian detection is based on the feature presentation. This method mainly includes feature extraction and classifier two parts. The selective feature should capture the most discriminative information of pedestrians and non-pedestrians. Many well-known traditional hand-crafted features such as HOG[1], Haar-like[2] and SIFT[3] still have some limitations on the robustness. A plausible way to remedy the limitations is learning features from the data of interest[5]. An example of such methods is learning features through deep learning networks, which achieve significant breakthrough in the area of image recognition in recent years. Compared with the hand-crafted features, deep learning network could extract higher-level features which represent more abstract semantics of the data and provide more invariance to intra-class variability. However, represented by the Convolutional Neural Network(CNN)[4], most of the deep learning networks need expertise of parameter tuning, long-time to train and have a high request on the experiment equipment. Aiming at these problems, Tsung-Han Chan with his colleagues designed a simple deep learning network PCANet which is very easy, even trivial, to train and to adapt to different data[5]. They use the PCA filters to emulate the data-adapting convolution filter bank and binary hasing for non-linear layer and block-wise histograms

Received (October 4, 2016), Review Result (December 21, 2016), Accepted (December 26, 2016)

for the feature pooling layer. Chan *et al.*, found that for almost all image classification tasks including face images, texture image, the effect of PCANet could be on par with and even better than the state-of -the-art features(hand-crafted or learned from DNNs).

Therefore, after we intensive study the problem pedestrian detection, we proposes to introduce PCANet to extract high-level feature for pedestrian detection based feature representation. Compared with other low-level feature algorithms, the experiment results show that the algorithm in this paper achieves a improving accuracy detection rate.

2. Structures of the PCANet Network(PCANet)

PCANet could be viewed as a simple convolutional neural network. it consists of three stages. The first and second stage are PCANet convolutional filters. The third stage is binary hashing(generating the nonlinear output) and block histograms. Suppose there are N training images $\{I_i\}_{i=1}^{N}$ and size of every image is m×n.

2.1. The First Stage: PCA

First of all, we use a size of $k_1 \times k_2$ patch around each pixel in every input image. Then for the image I_i, we can obtain all overlapping or non-overlapping patches of this input image and vectorize each patch, $x_{i,1}, x_{i,2}, x_{i,3},,...,x_{i,mn} \in R^{k_1k_2}$, and subtract patch mean from each patch. For all the images, putting their matrix together consisted by collected patches,we obtain:

$$X^l = [\bar{X}_1, \bar{X}_2, \bar{X}_3, ..., \bar{X}_N] \in R^{k_1k_2 \times Nmn} \tag{1}$$

Assuming that there is L_i filters in ith layer, we calculate the eigenvectors of XX^T and save the L_1 principle eigenvectors as the PCA filters of next stage. The formula as follows:

$$W_l^1 = \mathrm{mat}_{k_1,k_2}(q_l(XX^T)) \in R^{k_1 \times k_2}, \quad l = 1, 2, ..., L_1, \tag{2}$$

$\mathrm{mat}_{k_1,k_2}(v)$ will map the v∈ $R^{k_1 \times k_2}$ to a matrix, $q_l(XX^T)$ denotes the lth principal eigenvector, then every principal eigenvector captures the main variation of all the remove training patches. The first stage diagram is as follows:

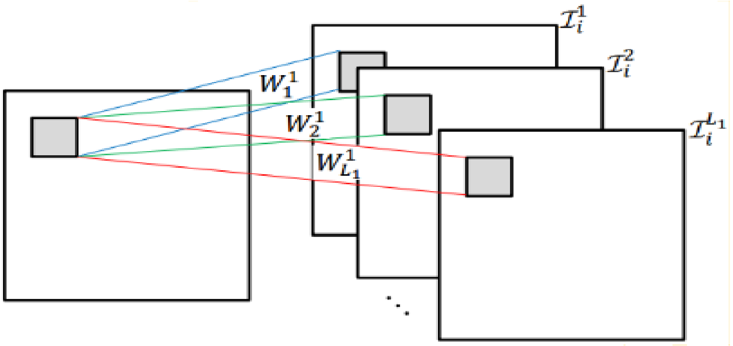

Figure 1. Structure of First Stage

2.2. The Second Stage: PCA

Almost similar with first stage, the difference is N input image I_i convolving with the lth filter output of the first stage separately, getting $L_1 \times N$ input image of second stage, the expression formula is as follows:

$$I_i^l = I_i * W_l^1, i = 1,2,...,N \tag{3}$$

Before convolution, in order to make the input image of second stage having the same size of I_i, the boundary of I_i is zero-padded.

Like the first stage, vectorize the patch of I_i^l and subtract patch mean from each patch I_i^l and expression is defined as:

$$Y^l = [\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, ..., \bar{Y}_N] \in R^{k_1 k_2 \times Nmn} \tag{4}$$

Then concatenate Y^l for all the filter outputs as:

$$Y = [Y^1, Y^2, \ldots, Y^{L_1}] \in R^{k_1 k_2 \times L_1 Nmn} \tag{5}$$

Calculate the eigenvectors of YY^T and get the L_2 largest of principal eigenvectors, the second stage expression formula is represented as:

$$W_l^2 = \mathrm{mat}_{k_1,k_2}(q_l(YY^T)) \in R^{k_1 \times k_2} \qquad l = 1, 2, \ldots, L_2, \tag{6}$$

Every input of I_i^l the second stage convoluting with W_l^2, obtain L_2 outputs of second stage as:

$$O_i^l = \{I_i^l * W_l^2\}_{l=1}^{L_2} \tag{7}$$

In the second stage, the number of outputs is L_1L_2. and is as follows:

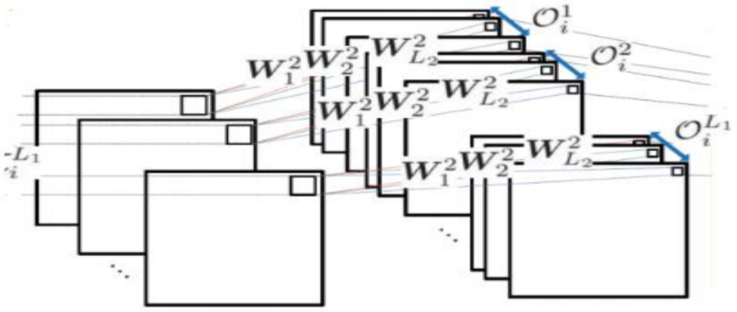

Figure 2. Structure of Second Stage

2.3. Output Stage: Hashing and Histogram

Every input image of second stage , we get:

$$T_i^l = \sum_{l=1}^{L_2} 2^{l-1} H(I_i^l * W_l^2) \qquad l = 1, 2, \ldots, L_1, \tag{8}$$

H(.) is a function that could convert a matrix to another matrix whose value is one for positive entries and zero otherwise. Then every output image multiplies a weight. This will convert output image into an integer in the range [0, 2^{L_2} -1].

Because of convolution, the dimension of feature will increase substantially, using histograms pooling could decrease feature dimension of output image. We partition every T_i^l, l=1,2,..., L_1, into B blocks and the size of every block is k$_1$k$_2$×B. Then we compute every histogram block and concatenate all of the histogram denote as Bhist(T_i^l). The final output feature could be defined as:

$$f_i = [\mathrm{Bhist}(T_i^1), \ldots, \mathrm{Bhist}(T_i^{L_1})]^T \in R^{(2^{L_2})L_1 B} \tag{9}$$

PCANet parameters mainly include the number of stage and filters in every stage, the size of filters and block size for local histograms. Because PCANet gets the filters by principal component analysis, not the gradient descent, compared with the CNN, PCANet has huge advantage on space and time complexity.

3. Pedestrian Detection Based on PCANet

In this paper, we proposed to apply PCANet for extracting feature in pedestrian detection. The flow chart of our pedestrian detection framework is shown in Figure 3.

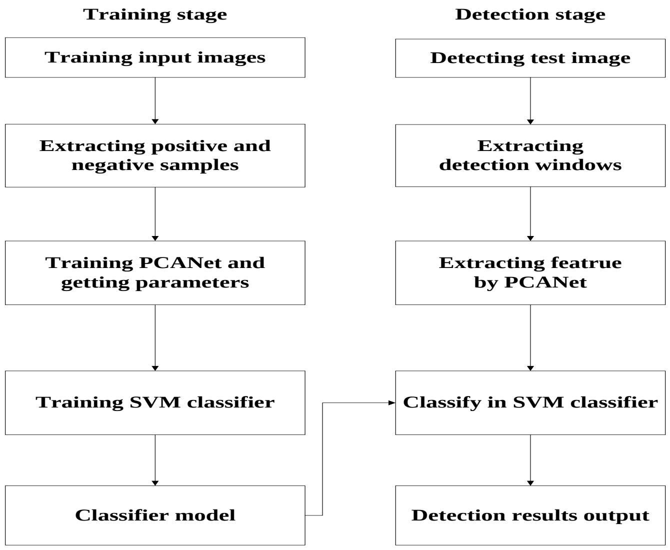
Training stage
Detection stage
Training input images
Extracting positive and negative samples
Training PCANet and getting parameters
Training SVM classifier
Classifier model
Detecting test image
Extracting detection windows
Extracting featrue by PCANet
Classify in SVM classifier
Detection results output

Figure 3. Flow Chart of Pedestrian Detection Framework Based on PCANet

Training stage: according to the ground truth files, select a certain amount of positive and negative samples from the input images. Put all of these into PCANet, train the network to get the network parameters and the features of all of the samples, then train the SVM classifier and obtain the classifier parameters.

Detection stage: first, make use of selective search[6] to generate possible pedestrian regions as candidate windows from test image. Then extract the features of these windows through trained PCANet. Next SVM classifier decides whether a candidate window shall be detected a pedestrian and making use of the approach of Non-Maximum Suppression(NMS) reduces the redundancy windows. Mark the pedestrian region using rectangle. Finally, get the final pedestrian regions and mark them on the test image.

3.1. Selective Search

The approach of selective search addresses a problem of generating possible object locations for using in object detection. This approach fully combines the advantage of exhaustive search and segmentation. Compared with traditional single strategy and exhaustive search, selective search provides multiple strategies and reduces the search space and find multi-scale regions respectively.

The selective search algorithm generates many segmentary regions based on segmentation of image algorithm[7]. Then combine the segmentary region by using hierarchical group algorithm. The process of algorithm is detailed in Table 1.

Table 1. Hierarchical Group Algorithm

Input:（colour）image
Output: a series of hypotheses *L* of object location
1. Obtain initial regions R={r_1,r_2,...r_n} by using document[7];
2. Set the initial value of similarity set S=Φ;
3. Compute similarity s(r_i,r_j) between two neighboring region r_i,r_j, and add to the similarity set S；
4. Find the two region pair r_i and r_j who have highest similarity , combine them as

a region r_t, then remove the $s(r_i, r_j)$ from similarity set and remove r_i and r_j from the region set. Calculate similarity between the new region r_t with its neighbors, meanwhile add similarity to the similarity set add r_t to the region set.
5. Extracting the bounding box of every region，this result may be the location of object L.

According different image properties, selective search present a variety of diversification strategies. This paper mainly use four similarity measure methods.

Color similarity: for each region, obtain one-demensional color histogram for each color channel by using 25 bins. So for three color channel, we could get $C_i = \{c_i^1, \ldots, c_i^n\}$ for each region with dimensionality n=75. The formula of color similarity is defined as:

$$s_{colour}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k) \tag{10}$$

After region combination, the histograms of new region is defined as:

$$C_t = \frac{\text{size}(r_i) \times C_i + \text{size}(r_j) \times C_j}{\text{size}(r_i) + \text{size}(r_j)} \tag{11}$$

Texture similarity: it adopt SIFT-Like to represent texture. We calculate the Gaussian integration in which the variance is 1 in eight orientations. We use a bin size of 10 to extract histogram for each color channel and for each orientation. For each region, we can achieve a texture histogram $T_i = \{t_i^1, \ldots, t_i^n\}$ with the dimensionality n = 240. The formula of texture similarity of is as follows:

$$s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k) \tag{12}$$

Size similarity: it encourages to merge small region early through calculating the two areas occupying the proportion of the image area. The formula is defined as:

$$s_{size}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(r_j)}{\text{size}(im)} \tag{13}$$

Where $\text{size}(im)$ denotes the size of the image in pixels.

Fill similarity: it encourages two of the region fit into each other to merge early. The formula of fill similarity is represented as:

$$fill(r_i, r_j) = 1 - \frac{\text{size}(BB_{ij}) \text{ - } \text{size}(r_i) \text{ - } \text{size}(r_i)}{\text{size}(im)} \tag{14}$$

Where BB_{ij} denotes not converted by the regions of the r_i and r_j .

The final similarity measure is combination of the above four:

$$\begin{aligned} s(r_i, r_j) &= a_1 s_{colour}(r_i, r_j) + a_2 s_{texture}(r_i, r_j) \\ &+ a_3 s_{size}(r_i, r_j) + a_4 s_{fill}(r_i, r_j) \end{aligned} \qquad a_i \in \{0,1\} \tag{15}$$

4. Experiment Results and Analysis

The experiment environment in this paper is 3.8GHZ AMD Athlon(tm) X4 760K Quad Core Processor, 8GB RAM. The software platform is MATLAB 2015.

The proposed framework is evaluated on INRIA dataset which is one of the most popular static pedestrian dataset. This dataset includes different dress and posture pedestrian, complex environments and backgrounds, which is a very challenging dataset.

All of the samples in the experiment is cropped into the size of width 64 height 128 images.

We use the evaluation methodology outlined in pedestrian detection benchmark[13], plotting miss rate versus false positives per-image(FPPI).

We compare our algorithm with the state-of-the-art domain pedestrian detection algorithms including Shapelet[8], VJ[9], HOG[1], HIKSVM[10], HogLbp[11], ChnFtrs[12]. The experiment result is shown in Figure 4.

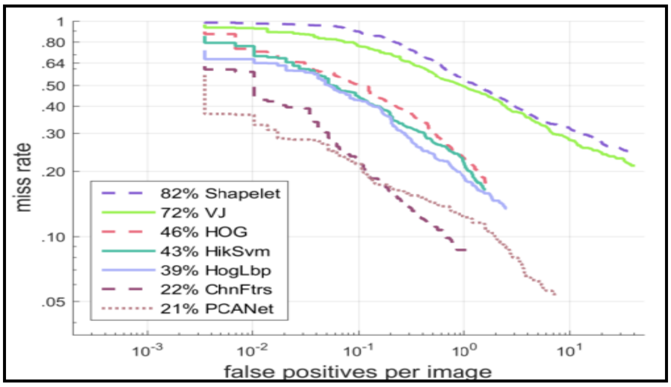

Figure 4. Experimental Comparing Results on INRIA Dataset

From Figure 4, we can conclude that the proposed algorithm in this paper, comparing with other state of art pedestrian detection algorithm, improves a lot in detection accuracy. When the false positive was 10%, our algorithm respectively achieves 25% and 22% improvement comparing with HOG and HikSvm and the detection rate is even slightly higher than some excellent pedestrian detection algorithms such as ChnFtrs algorithm.

The detection results of our algorithm evaluated in the INRIA pedestrian detection dataset is shown in Figure 5. From the detection results, we can see that in the complex environment such as the lighting variation, occlusion, our method still has good detection effect.

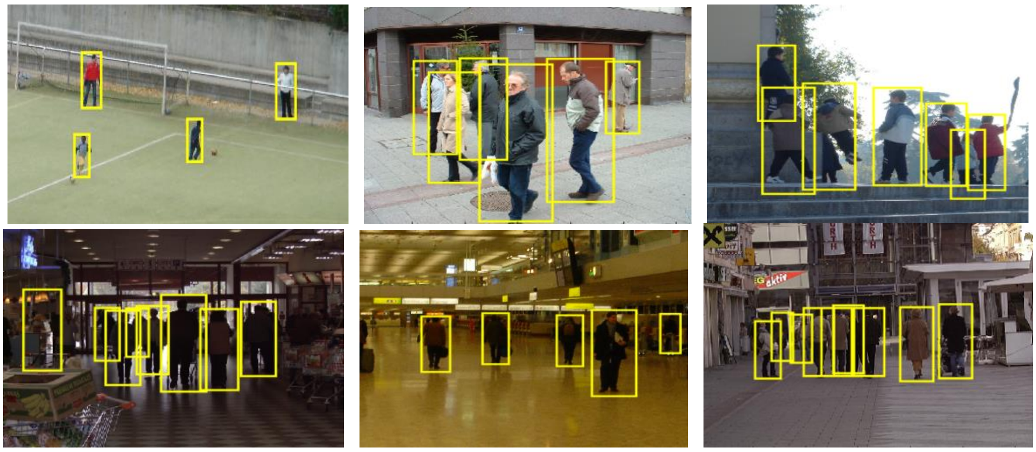

Figure 5. Experimental Result Images on the INRIA Dataset

5. Conclusion

In this paper, we propose a pedestrian detection algorithm based on PCANet, which could extract high-level feature of pedestrian. The experimental results show that the algorithm proposed in this paper significantly improves the detection rate compared with the state-of-the-art domain pedestrian detection algorithms on INRIA dataset. Our algorithm proves that feature extracted by PCANet could reflect the nature of the data, compared with the hand-crafted low-level features, and will have broad research prospects on object detection. The future work will be focused on improving the real-time efficiency.

Acknowledgments

This paper is supported by The National Natural Science Foundation of China (U1261114).

References

[1] N. Dalal and B. Triggs, "Histogram of oriented gradient for human detection", Computer Vision and Pattern Recognition, **(2005)**.

[2] C. Papageorgiou and T. Poggio, "Atrainable system for object detection", International Journal of Computer Vision, vol. 38, no. 1, **(2000)**, pp.15-33.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, **(2004)**, pp. 91-110.

[4] K. Alex, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Proceedings of Advances in Neural Information Processing Systems, Lake Tahoe, **(2012)** 748-764.

[5] T. H. Chan, K. Jia, S. Gau, J. Lu, Z. Zeng and Y. Ma, "PCANet: a simple deep learning baseline for image classification?", Eprint Arxiv, **(2014)**, pp. 1404-3606.

[6] R. Girshick, J. Donahue and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation", Computer Vision and Pattern Recognition(CVPR),2014 IEEE Conference on IEEE, **(2014)**, pp. 580-587.

[7] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation", IJCV, vol. 59, no. 59, **(2004)**, pp. 167-181.

[8] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features", CVPR, **(2007)**.

[9] P. Viola and M. Jones, "Robust real-time object detection", IJCV, vol. 57, no. 2, **(2004)**, pp. 137–154.

[10] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel SVMs is efficient", CVPR, **(2008)**.

[11] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling", ICCV, **(2009)**.

[12] P. Dollár, Z. Tu, P. Perona, and S. Belongie. "Integral channel features", BMVC, **(2009)**.

[13] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: A Benchmark", IEEE Conference of Computer Vision and Pattern Recognition, **(2009)**.

[14] X. Zeng, W. Ouyang and X. Wang, "Multi-Stage contextual deep learning for pedestrian detection", Conference: Proceedings of the 2013 IEEE International Conference on Computer Vision,**(2013)**.

Authors

Qi Mu, (1974-) was born in Xi'an city of Shanxi province of China. She received her bachelor's degree in 1996 from Xi'an University of Science and Technology and her major is computer science, and received her master's degree in 2003 from Xi'an University of Science and Technology and her major is computer applications technology. She is an Associate Professor and the Dean of Computer Science Department. Her current research interests are machine learning, computer vision. She has published more than 20 academic papers, and 4 of them were retrieved by EI. About 6 computer textbooks were issued under her general editorship, 1 of them was got the Shaanxi Province Excellent Teaching Material Award and also she has National Natural Science Foundation of China(U1261114). In addition, She had taken charge of and participate in more than 10 scientific projects and education reform projects.

Yikai Jia, She received her bachelor’s degree in Computer science and technology from Henan University, Kaifeng, China. Now, she is postgraduate in Xi’an University of Science and Technology, her research interests include image processing, machine learning, computer vision.

Zhanli Li, he was born in 1964 in Xi’an city of Shaanxi province of China and received his Doctor of Engineering degree in 1997 in Xi'an JiaoTong University. He engaged in research in Postdoctoral Program of Fudan University, the state key laboratory of Mechanical manufacturing and systems engineering and Faculty of engineering, Iwate University, Japan. He present is dean of college of Computer Science and Technology in Xi’an University of Science and Technology and his research areas are computer vision, computer graphics and so on.