IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-2 E-ISSN: 2347-2693

Deep Learning Feature Representation Applied to Cross Dataset Pedestrian Re-identification

Hongmei Xie[1*], Yanggang Zhou[2], Qiang Liu[3]

[1*]Dept. of EST, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
[2]Dept. of EST,, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
[3]Dept. of EST,, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China

**Corresponding author: xiehm@nwpu.edu.cn Tel.: +86-029-88494613*

Available online at: www.ijcseonline.org

Received: 21/Jan//2018, Revised: 03/Feb2018, Accepted: 14/Feb/2018, Published: 28/Feb/2018

***Abstract*—**Pedestrian re-identification technology has become the current research focus due to its wide range of applications. This study conducted cross dataset pedestrian re-identification to solve the problem that the single dataset's difficulty for simulating the actual situation and its poor generalization ability. Deep learning has made remarkable achievements in the fields of machine learning recently, so the deep learning technology is integrated into cross datasets pedestrian re-identification system. Here we improved the three-layer convolutional neural network (CNN) structure proposed by Yang Hu in Asia Conference on Computer Vision (ACCV), 2014. The Batch Normalization (BN) layer has been added to reduce the over-fitting degree during training period and the adjusted cosine similarity algorithm is used for pedestrian feature match to solve the defect of cosine similarity algorithm. Finally we implemented the entire cross dataset pedestrian re-identification system and got the experimental results. The Shinpuhkan2014dataset was chosen as training set. We compared the training results before and after adding BN layer and found that test accuracy increased, test loss decreased and over-fitting phenomenon eased. The VIPeR and i_LIDS datasets were chosen as test sets. We evaluated the effects on VIPeR and i_LIDS based on the CNN model that training on Shinpuhkan2014dataset. The cumulative matching rate rank5 increased by 1.7% on VIPeR dataset compared with the current level, the rank10 and rank20 also increased. And the cumulative matching rate rank1 increased by 1.8% on i_LIDS dataset compared with the current level, the rank5 and rank10 also increased.

***Keywords*—**Cross dataset, Convolutional neural network, Batch normalization, Adjusted cosine similarity

I. INTRODUCTION

As an important and basic part of the monitoring system, pedestrian re-identification technology has many applications, such as image retrieval, behavior analysis, cross cameras recognition and tracking. However, improving the performance of the pedestrian re-identification system is a very hard problem to solve. Existing pedestrian re-identification datasets have been acquired from wide-area monitoring cameras covering a larger area. Therefore, even with high-definition cameras, the pedestrian image resolution is still low. Moreover, light changes, viewpoint, blocked between people, background, posture, camera parameters will cause the following difficulties: (1) The true distribution of each category of pedestrians cannot be generated due to the lack of valid samples; (2) The distribution within and between classes is not stable due to the diversity and ambiguity of samples; (3) Especially pedestrian re-identification datasets are inseparable, which make pedestrian re-identification a very difficult issue under different camera perspectives.

Researchers have done plentiful and meaningful work to accelerate the pedestrian re-identification development process over the past several years. A lot of important difficulties in the hard task have been solved through discriminating manual feature extraction and metric learning. However, the cross dataset pedestrian re-identification ideas has not received the attention of researchers. In the cross dataset pedestrian re-identification task, we do not know the test data and access conditions due to the training set and the test set are not the same dataset. So too many important factors have added to the difficulty of this task.

Pedestrian re-identification experiments are usually conducted on several public datasets for pedestrian re-identification. The public dataset is divided into two parts, one part is used as the training set and another part is used as the test set. It is clear that the training set and the test set come from the same public dataset (data source) in these experiments.

However, it is not easy to get the training set that is consistent with the acquisition environment of the test set under real scenarios. Therefore, many approaches based on learning are prone to over-fitting and have poor generalization ability due to similar internal structure of data and consistent acquisition environment. Even the methods based on feature design cannot ensure that the designed features are useful for new data. All in all, most pedestrian re-identification methods do not have good enough performance.

Pedestrian re-identification technology with practical value should be taken seriously by researchers due to the continuous improvement of pedestrian re-identification in single dataset. In my study, the cross dataset pedestrian re-identification problem has been solved through deep learning (DL) method. Deep learning has yielded remarkable achievements in speech, vision and sentiment analysis [1] aspects. DL is well suited for handling the large datasets. The vast majority of public pedestrian re-identification datasets are small, both in the number of categories and the number of images per category. Nevertheless, as the continuous advancement of pedestrian re-identification, more pedestrian re-identification datasets have been obtained. These datasets are large in scale but still not comparable to the dataset size in other deep learning applications [2, 3]. In this study, the relatively large pedestrian re-identification dataset Shinpukkan2014dataset is selected as the training dataset [4], the relatively small pedestrian re-identification dataset VIPeR and i-LIDS are selected as the test datasets [5,6]. This experiment will demonstrate the performance of learning pedestrian feature representation and cross-dataset pedestrian matching. Convolutional neural networks (CNN) has fewer connections and parameters and are easier to train than standard feed forward networks. Therefore, CNN was chosen to learn pedestrian feature representation in my experiment.

The main breakthrough points of this paper are as follows: (1) For the pedestrian re-identification task, combined the convolutional neural network with the Batch Normalization layer, valid features have been represented, while these features have not been emphasized before; (2) After combining the convolutional neural network with the Batch Normalization layer and using the adjusted cosine similarity algorithm, we conduct cross dataset pedestrian re-identification experiment. (3) The experimental results show that our designed cross dataset pedestrian re-identification system has good effect and can be migrated to many other pedestrian re-identification datasets.

Rest of the paper is organized as follows, Section I contains the introduction of the current application and difficulties of pedestrian re-identification, Section II contains the related work of metric learning and CNN , Section III contains the CNN architecture and essential steps of DCNN-based pedestrian re-identification and also explains the proposed Batch Normalization algorithm and improved matching algorithm with formulas, Section IV describes results and discussion and analysis of the results, Section V concludes the research work with future directions.

II. RELATED WORK

In the past several years, the recognition accuracy of pedestrian re--identification has been greatly improved. Metric learning plays a very important role in many methods. Weinberger et al. proposed the Large Margin Nearest Neighbor (LMNN) algorithm, which is mainly to learn a Mahanalobis for K nearest neighbors (KNN) [7]. And the method of semi-positive definite programming is used in this algorithm. Later, Dikmen et al. put forward an algorithm similar to LMNN, which is named Large Margin Nearest Neighbor with Rejection (LMNN-R) [8]. This method mainly added a rejection mechanism in LMNN and achieved significant improvement. Davis J V et al. proposed an algorithm called Information Theoretic Metrics Learning (ITML), its main idea is to replace the learning Mahalanobis distance method by finding the minimum differential relative entropy between two multivariate Gauss function under the limits of the distance function [9]. Zheng W S et al. considered pedestrian re-identification as the best distance learning problem to maximize the probability that a pair of truly matched images will have a smaller distance than a pair of unmatched images [10]. Koestinger M et al. put forward the KISSME algorithm that learns a distance metrics function with equivalent constraints, which is proposed from the perspective of statistical inference [11]. Li et al. put forward the local adaptive decision function (LADF) algorithm that learns the decision function of pedestrian verification, which is a combination model of local adaptive threshold regulation and distance metrics function [12]. However, almost all of these algorithms are sensitive to the choice of parameters and are easy to over-fitting. In practical application, the performance of these methods is far from satisfactory, so few people study these traditional algorithms later.

Another pedestrian re-identification method seeks to address this problem by finding the feature representation that are both sharp and stable for describing the appearance features [13,14,15,16]. Farenzena et al. proposed the Symmetry-Driven Accumulation of Local Features (SDALF) method [17]. Considering the symmetry and asymmetry property in the pedestrian image, a variety of feature combination methods were used to deal with perspective changes. Ma B et al. replaced the local feature descriptors with Fisher vectors to obtain the global representation of images [18]. Dong S C et al. used the pedestrian image structure for pedestrian re-identification [19]. The color and color position of each part of entire human image were applied to the pedestrian

matching process. Salience technology has also been used to pedestrian re-identification gradually [20,21,22]. Nevertheless, most manual features are not stable enough and have no obvious difference, and they are likely to lose their effect in poor lighting, change of perspective, blocking between people, especially the test data has migrated.

In addition to the above methods, other pedestrian re-identification methods have also been used. Gray et al. put forward to select the best features from a group of texture and color features using AdaBoost [23]. Prosser et al. considered pedestrian re-identification as a sorting question and used the Ensemble RankSVM to learn the subspaces that obtain the highest sorting in a true pedestrian match [24]. In [25], the pedestrian image pair from different perspectives are locally aligned by projecting to a common feature space and then matched with locally optimized soft assignment metrics. Liu C et al. have made significant improvements by allowing users to quickly improve their search [26]. In summary, many pedestrian re-identification methods pay more attention to the single dataset, and all of them contribute to promoting the development of pedestrian recognition. However, their performance are still not good enough under real scenarios.

Fortunately, some researchers have focused their attention on cross dataset pedestrian re-identification and some very meaningful work have been done. Ma et al. put forward the Domain Transfer Rank Support Vector Machine (DTRSVM) algorithm, which mainly used the pedestrian image pair of the source area and the unmatched (negative) pedestrian image pair of the target area to conduct pedestrian re-identification in the target area camera [27]. Although this algorithm have not been regarded as pedestrian re-identification (using the information of the target area) above the cross datasets completely, it gives us a lot of incentives and achieves exciting results.

Yi et al. put forward the Deep Metric Learning (DML) algorithm, which learns the metrics through "siamese" convolutional neural network [28]. The network has two symmetrical sub-networks which were connected by a cosine layer. In this article, the author conducted a cross dataset pedestrian re-identification experiment using the VIPeR dataset as the test set and the CUHK-campus dataset as the training set. The performance of DML on the VIPeR dataset has been greatly improved compared to DTRSVM. In addition, DML is truly the first experiment to perform cross dataset pedestrian re-identification. It facilitates the development of cross dataset pedestrian re-identification although further research need be done.

A large number of training data can be acquired and plenty of computer resources such as the development of GPU make more and more researchers begin to study the convolutional neural network (CNN). Therefore, CNN has achieved amazing success in the field of computer vision. For example, CNN has shown astonishing results in both image classification and face recognition. Inspired by DeepFace, this study mainly dealt with the cross datasets pedestrian re-identification problem by learning deep pedestrian feature representation.

III. METHODOLOGY

In recent several years, learning-based approaches have begun to outperform hand-crafted features due to more data is available and they can discover and optimize the features for particular tasks. For cross dataset pedestrian re-identification, lighting, posture changes and resolution have much worse impact. In addition, the changes of data sources will result in well-designed features performing well on one dataset and poor on the other dataset. Here, we solve the problem by learning pedestrian representations through deep neural networks in my study.

A. The model of CNN

We split the 48 × 128 size original pedestrian image into three equal-sized sections (with slight overlap), head area, body area, and leg area. The three part are all 48×48 size. And then we trained the three-part image separately and finally got three feature extraction models. At the end of pedestrian matching, we concatenated the features of the three parts and then used the fused features to conduct the pedestrian matching.

Figure 1 shows the flow chart of dividing the pedestrian image into three parts and conducting feature representation.

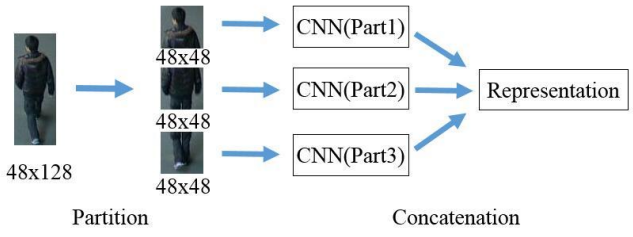


Figure 1. The flow chart of dividing the pedestrian image into three parts and concatenating the features of the three parts to conduct feature representation

The input of convolutional neural network used in this paper is a 48 × 48 split three-channel (RGB) pedestrian image. We mainly train the convolution neural network in a multi-class way. My experiment improved the three-layer convolutional neural network structure proposed by Yang Hu in ACCV,

2014 [29]. The convolution neural network model is shown in Figure 2.

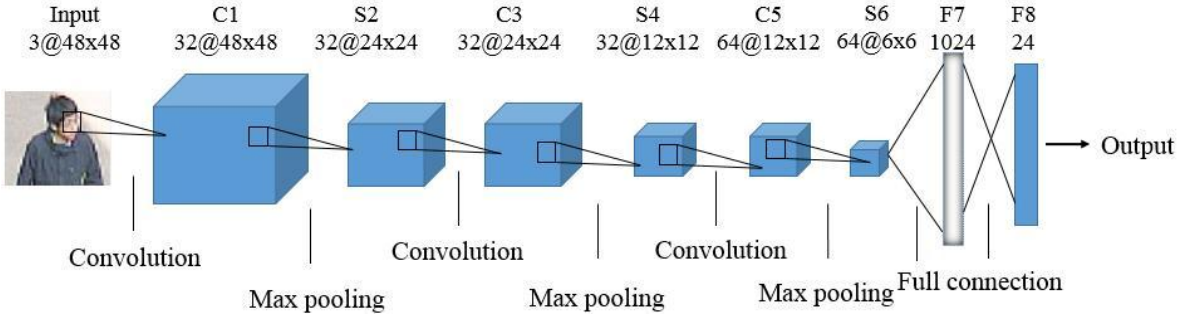


Figure 2. The structure model of CNN used in this paper

The input image is fed into convolutional layer C1 where the number of filters is 32 and the size is 5×5. And then 32 feature maps are fed to the maximum pooling layer S2 using the maximum method. In the 3×3 neighborhood of the space, we conduct the down-sampling operation for every feature map of each channel. Next S2 is followed by another convolution layer C3 with 32 filters. And the size of filter is 5×5， The first three layers can be used to extract low-level features such as simple edges and textures. The max-pooling layer plays an important role in dealing with local rotations and transformations. We attach a maximum pooling layer behind each convolution layer to make the network more robust. In some other CNN frameworks, only the maximum pooling layer is applied after the first convolutional layer of the network to avoid losing accurate location information about detailed structures and tiny textures. However, when we construct the network and take into account that pedestrian have a variety of large scope and magnitude posture changes, have no constraints and alignment, etc. The max-pooling layer is used behind each convolution layer to ensure more robust and universal learned pedestrian representation.

The subsequent layer structure is similar to the previous layer. Two convolution layers C3 and C5 have 32 filters and 64 filters respectively, and the filter size is 5×5. Some other work use the local connection layer structure to ensure that a set of filters can be learned for each position in the feature map. We choose the convolution layer instead of the local connection layer to share the weight. It can sacrifice some weight discrimination, but it is also a very safe solution that makes pedestrian representation more robust and versatile.

The last two floors F7 and F8 are fully connected. The output of fully connected F8 is fed to K-ways Softmax, which produces a distribution on K-class labels. Since the training dataset has only 24 individuals and fewer objects, we use the output of the F7 layer as the feature vector. The 1024-dimension feature vectors can be extracted from every parts of the pedestrian image, so the total feature vectors have 3072 dimensions. These feature vectors are used to conduct pedestrian feature representation and pedestrian matching.

The training of the network aims to maximize the probability of correct classification of pedestrians by maximizing multiple Logistic regression goals. We do this by minimizing the cross-entropy loss for each training sample. If k is the index of the correct category label for a given input then the loss function is as shown in (1) and (2).

$$p_k = \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} \quad (1)$$

$$L = -\log(p_k) \quad (2)$$

In Equation (1), z_k and z_j are the output value of the F8 layer. The loss is minimized by a stochastic gradient descent SGD algorithm at a batch size of 128. The gradient of the convolutional neural network is calculated and propagated through the error standard back-propagation algorithm.

Fast learning has a significant impact on the performance of the model trained in large datasets. So Rectified Linear Units (ReLU) are chosen as the activation function for each layer: max (0, x). ReLU activation function is shown in Figure 3.

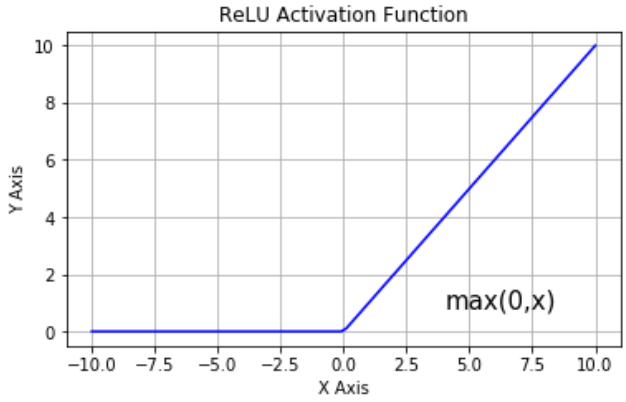


Figure 3. ReLU activation function

After getting the representation of pedestrians, we can use the similarity metric function to measure the similarity between features.

B. Batch Normalization

When using above improved convolutional neural network for training, we found a certain degree of over-fitting for the leg image training process. To reduce the degree of over-fitting. The concept of Batch Normalization (BN) has been introduced [30].

In the training process of convolutional neural network, the network data is updated in real time. Because the input data has been normalized, only the data distribution in the middle layer is considered. The data update in each layer in the middle of network will leads to the input data has been changed in the back layer. For example, the input in the second layer is calculated by the data of first layer through the weight and bias parameters. And the weights and bias parameters of first layer in the training process are constantly updated, then the input data distribution of second layer will continue to change. The data distribution change of middle layer is called Internal Covariate Shift. The Batch Normalization algorithm mainly aims to solve the situation that the data distribution changes of middle layer during the training.

Batch Normalization can also be regarded as one part of the CNN layer like convolutional layer, or pooling layer, or activation function layer, or fully connected layer. In addition to the output layer, the parameter updating of the previous layer will cause the data distribution of the latter layer to change for the other layers. The solution to this problem is to preprocess the input data of each layer. Suppose the input data of the third layer in the network is X3, and then we normalize the data to mean 0 and variance 1 before input this data to the third layer for iterative calculation. The problem of Internal Covariate Shift is solved.

The Batch Normalization layer have been placed before the activation function layer in my convolutional neural network structure, as shown in Figure 4.

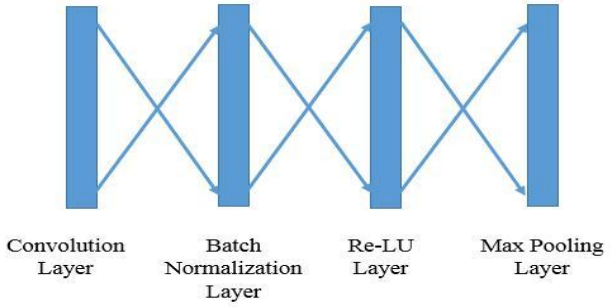


Figure 4. Improved convolution neural network unit

After adding Batch Normalization layer, the forward propagation and gradient back-propagation of convolutional neural network have changed. The forward propagation formula of BN layer is shown in (3) to (6).

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{3}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{4}$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{5}$$

$$y_i = \gamma \hat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i) \tag{6}$$

In the formulas above, μ_B is the mean of batch. σ_B^2 is the variance of batch. m is the size of batch. Formula 3-5 is the normalization operation. Formula (6) is the reconstruction transform operation and γ、β are the reconstruction parameter.

The gradient back-propagation of BN layer still follow the chain of derivation. As shown in (7) to (12).

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i}\cdot\gamma \tag{7}$$

$$\frac{\partial L}{\partial \sigma_B^2} = \sum_{i=1}^{m}\frac{\partial L}{\partial \hat{x}_i}\cdot(\hat{x}_i - \mu_B)\cdot\frac{-1}{2}(\sigma_B^2+\varepsilon)^{-3/2} \tag{8}$$

$$\frac{\partial L}{\partial \mu_B} = \sum_{i=1}^{m}\frac{\partial L}{\partial \hat{x}_i}\cdot\frac{-1}{\sqrt{\sigma_B^2+\varepsilon}} + \frac{\partial L}{\partial \sigma_B^2}\cdot\frac{\sum_{i=1}^{m}-2(x_i-\mu_B)}{m} \tag{9}$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \hat{x}_i}\cdot\frac{1}{\sqrt{\sigma_B^2+\varepsilon}} + \frac{\partial L}{\partial \sigma_B^2}\cdot\frac{2(x_i-\mu_B)}{m} + \frac{\partial L}{\partial \mu_B}\cdot\frac{1}{m} \tag{10}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m}\frac{\partial L}{\partial y_i}\cdot\hat{x}_i \tag{11}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m}\frac{\partial L}{\partial y_i} \tag{12}$$

In the above formulas, m is the batch size. And from $\frac{\partial L}{\partial y_i}$, we can get $\frac{\partial L}{\partial \gamma}$, $\frac{\partial L}{\partial \beta}$, and $\frac{\partial L}{\partial x_i}$. The back propagation process is completed.

C. Adjusted cosine similarity

Most current pedestrian re-identification methods use the cosine similarity algorithm when conducting pedestrian

matching. Although the cosine similarity algorithm has a certain effect on the measurement of sample differences, the cosine similarity algorithm only considers the similarity in the vector direction without considering the difference in the vector magnitude of each dimension. Adjusted cosine similarity algorithm is a slight improvement on the cosine similarity algorithm, and just complements this defect.

Taking the scoring system as an example, two users score (1, 2) and (4, 5) respectively for two movies. If you use cosine similarity algorithm, the two users are strikingly similar, but the actual situation does not match. The former user apparently did not like these two movies. However, the adjusted cosine similarity algorithm is mainly to correct this wrong measure method. We subtract the total mean 3 of the two scores from each score, and obtain the scores (-2, -1) and (1, 2). And then using the cosine similarity algorithm, it can be concluded that the similarity is negative and is very much in line with the actual situation.

The adjusted cosine similarity formula is shown in (13).

$$sim(i,j)=\frac{\sum_{u\in U}(R_{u,i}-\bar{R}_u)(R_{u,j}-\bar{R}_u)}{\sqrt{\sum_{u\in U}(R_{u,i}-\bar{R}_u)^2}\sqrt{\sum_{u\in U}(R_{u,j}-\bar{R}_u)^2}} \quad (13)$$

In the formula above, $R_{u,i}$ and $R_{u,j}$ are two vectors. $\bar{R}_u$ is the total mean of the two vectors.

Due to the unreasonable of cosine similarity algorithm, the adjusted cosine similarity algorithm is used to calculate the similarity between pedestrian image features in this experiment. Firstly, the head, body and leg features of the pedestrian image are obtained through the trained CNN model. And then the features are fused. Finally, the adjusted cosine similarity algorithm is used to calculate the similarity between the two pedestrian image features and to obtain the final match result.

IV. Experimental Results and Discussion

A. Datasets

The dataset used in this experiment for training the convolutional neural network is the Shinpukhan2014 pedestrian re-identification dataset, which was collected by the Japanese scholar Yasutomo Kawanishi in 2014. This dataset only has 24-class pedestrians and the images are all single pedestrian images that have been segmented from the entire surveillance image. The image size is 48 × 128 pixels. There are 16 cameras with pedestrian images from different angles. Each class pedestrian has less than about 1000 pieces. The total image number is 22,500. Due to the large number of cameras and each camera collects pedestrian images at several angles, so the dataset is more complex and can simulate most pedestrians in video surveillance under real conditions. The pedestrian image in the training set is shown in Figure 5.

Figure 5. The images in Shinpukhan2014dataset

In this experiment, cross dataset pedestrian re-identification technology was used on common VIPeR and i_LIDS pedestrian re-identification public datasets. These two datasets are widely used because the internal data distribution of the datasets is more reasonable and the collected pedestrian images are the pedestrians in the actual monitoring. There is no post-processing above the pixel level, and we just uniform the image size. Therefore, we chose to experiment on these two datasets convincingly and without any loss of fairness.

The VIPeR dataset, which is currently the most used single-frame mode standard pedestrian re-identification image library that contains pedestrian images captured by two cameras. There are 632-class pedestrians and each class has two pedestrian images. The image size is 48 × 128 pixels, and the total image number is 1264. In this dataset, the pedestrian's posture, camera's perspective, ambient light, and background occlusion vary greatly. Therefore, increasing the cumulative match rate above this dataset can be challenging. The VIPeR dataset is shown in Figure 6.

Figure 6. The images in VIPeR dataset

Then, the i_LIDS dataset is a multi-frame pedestrian re-identification dataset commonly used by researchers. There are 119-class pedestrians in this dataset, and the images are taken at an airport. The total number of pedestrian images in this dataset is 476, and averaging 4 images per person. This dataset has also pedestrian images under two cameras. However, Different from VIPeR is the existence of multi-frame images in i_LIDS. Therefore, all the pedestrian images under one camera are taken as a test set and all the pedestrian images under another camera are taken as a query image library when conducting pedestrian re-identification experiment. The size of the pedestrian image is not the same, but it is basically a complete pedestrian and the aspect ratio is more balanced. So we unified the size of all images into 48 × 128 size pedestrian image. The i_LIDS dataset is shown in Figure 7.

Figure 7. The images in i_LIDS dataset

B. Comparison result between before adding BN and after adding BN on training dataset

We added the Batch Normalization layer before each ReLU activation function layer in the convolutional neural network used in our experiments and then experimented to re-train the convolution neural network on the Shinpuhkan2014 dataset and test the classification accuracy. We trained each of the three parts of the pedestrian image to generate their own loss and accuracy curves. The comparison results between before adding BN and after adding BN on training dataset are shown in Figure 8, Figure9, and Figure 10.

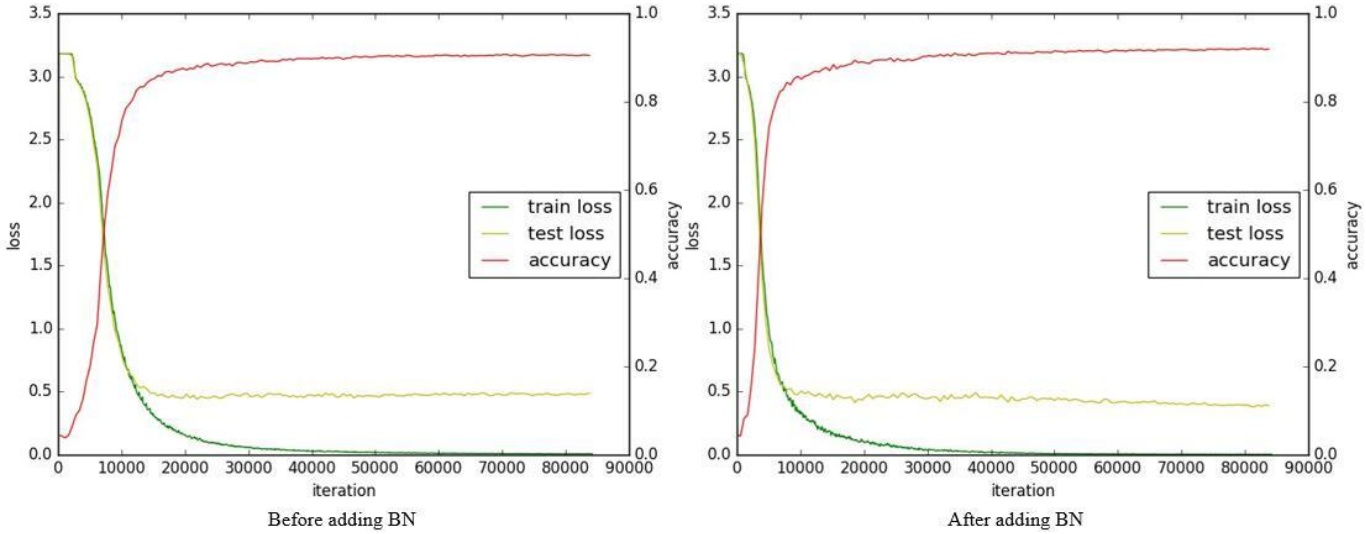


Figure 8. The result of head area

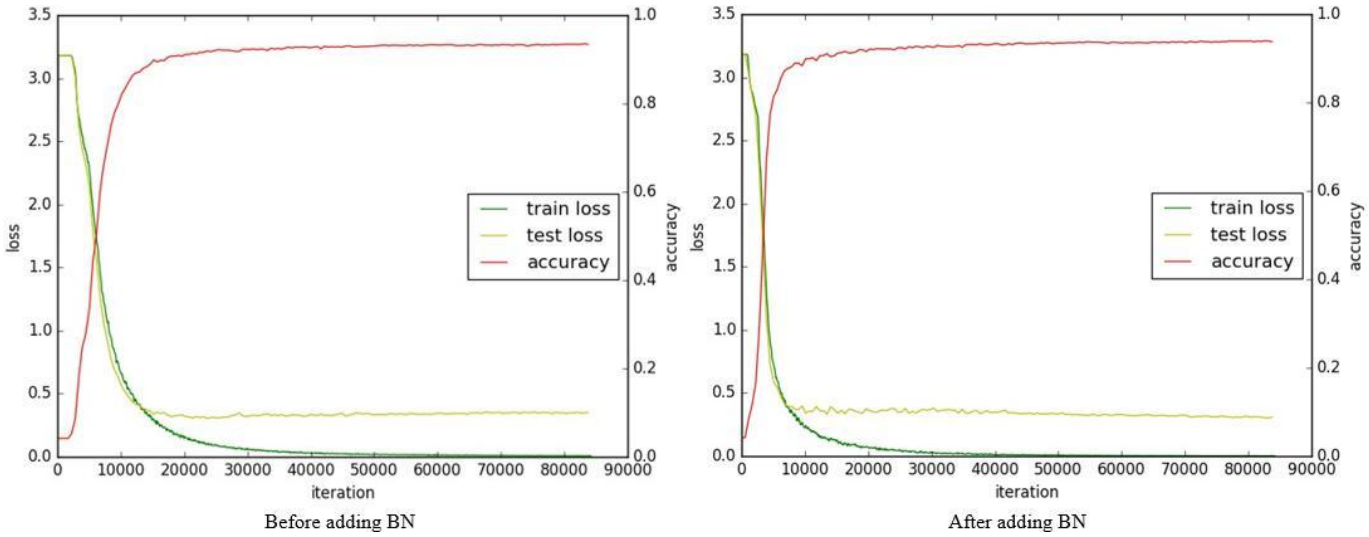


Figure 9. The result of body area

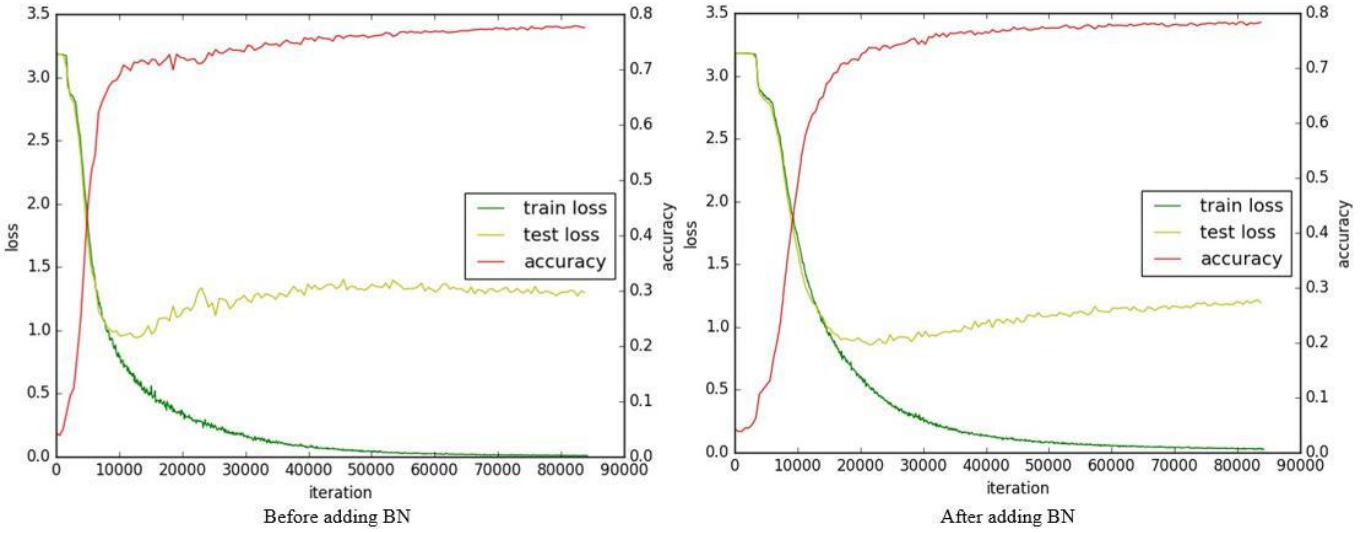


Figure 10. The result of leg area

From these result curves we can see the over-fitting in the leg area is improved. For detailed comparison data, we can refer to Table 1.

Table 1. Comparison result between before adding BN and after adding BN on training dataset

Areas	Before adding the BN layer		After adding the BN layer	
	Test loss	Accuracy	Test loss	Accuracy
head	0.50	91.01%	**0.408**	**91.7%**
body	0.38	92.7%	**0.308**	**93.7%**
leg	1.32	77.6%	**1.22**	**79.3%**

From the above table we can see the overall test loss with BN layer is smaller than without BN layer and the overall classification accuracy is improved after adding BN layer. So the training results obtained by adding BN layer are obviously better than those results without BN layer.

C. Results on VIPeR dataset

The pedestrian re-recognition results on VIPeR have been obtained, the cumulative match rate curve CMC is shown in Figure 11.

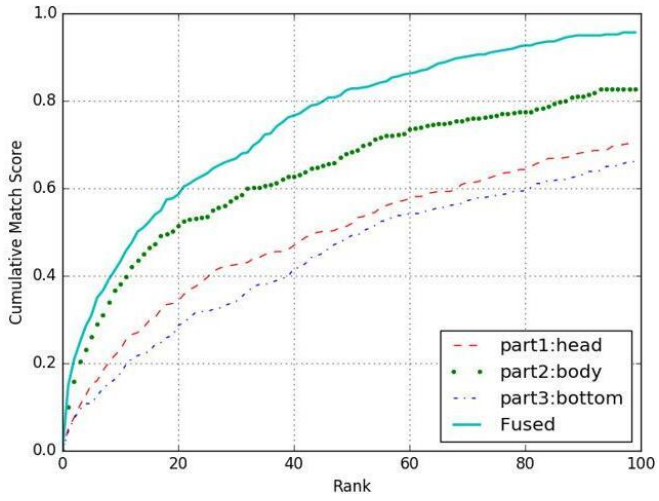


Figure 11. The cumulative match rate curve CMC on VIPeR

Using the trained models of the head, body and leg regions, the separate pedestrian matching of the three parts was carried out. It was found that the cumulative matching rate curve after fusion was higher than the cumulative matching rate curve of the three parts. The effect of the legs is worst and the effect of body area is best. So we can distinguish most of the features are concentrated in the upper body of pedestrian. The experimental results and the current better cross-dataset pedestrian re-identification results were compared, the contrast results are shown in Table 2.

Table 2. The contrast results on VIPeR

Methods	Training sets	Rank1 (%)	Rank5 (%)	Rank10 (%)	Rank20 (%)
DTRSVM	i-LIDS	8.26	31.39	44.83	53.88
DTRSVM	PRID	10.90	28.20	37.69	44.87
DML	CUHK Campus	**16.17**	45.82	57.56	64.24
Our CNN	Shinpuhkan2014	15.80	**47.59**	**58.58**	**66.81**

As can be seen from the comparison table, the experimental method far exceeds the traditional DTRSVM method. For the deep learning method DML, only the cumulative matching rate of Rank1 is slightly smaller than that of DML. The cumulative matching rate of Rank 10, Rank 20 and Rank30 are higher than the DML. So after combining the convolutional neural network with the BN algorithm and using the adjusted cosine similarity algorithm, the cross dataset pedestrian re-identification system has better performance on VIPeR dataset.

D. Results on i_LIDS dataset

For the i_LIDS dataset, there are 476 pedestrian images in this dataset. There are 119-class pedestrians. 90-class randomly selected pedestrians are tested and repeated 10 times to obtain the average cumulative matching rate. The same method as above is used to obtain the cumulative matching rate curve, as shown in Figure 12.

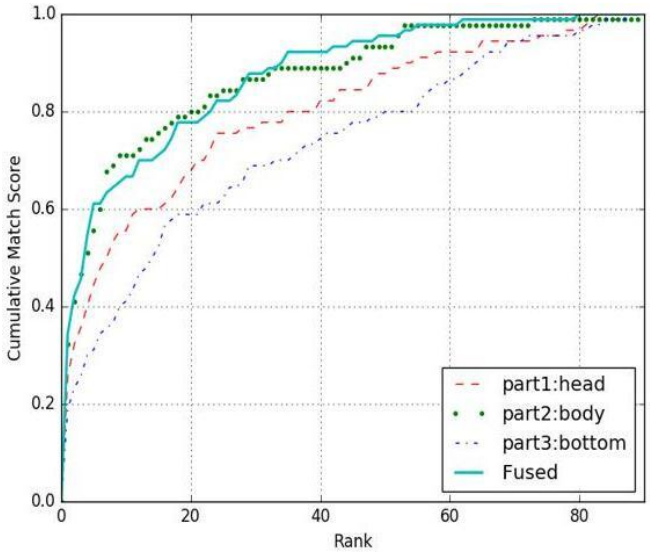


Figure 12. The cumulative match rate curve CMC on i_LIDS

As can be seen from Figure, the performance of the body area is almost equal to the fusion performance, and the performance of the legs is relatively poor. The main reason is that the i-LIDS dataset has many pedestrian shelters, and the leg area is usually blocked by a suitcase. Experimental image segmentation is not as accurate as other datasets, and only the head and body regions are relatively stable. The experimental results are compared with the various methods for pedestrian re-identification on i-LIDS dataset. As shown in Table 3.

Table 3. Compared with the various methods for pedestrian re-identification on i-LIDS dataset

Methods	Rank1 (%)	Rank5 (%)	Rank10 (%)	Rank20 (%)
MCC[31]	12.00	33.66	47.96	67.00
ITM[7]	21.67	41.80	55.12	71.31
Adaboost[32]	22.79	44.41	57.16	70.55
LMNN[5]	23.70	45.42	57.32	70.92
Xing's[33]	23.18	45.24	56.90	70.46
PRDC[8]	32,60	54.55	65.89	**78.30**
Our CNN	**34.44**	**61.10**	**66.67**	77.78

As can be seen from table 3, the cumulative matching rate Rank1, Rank5, Rank10 are better than the current level on i-LIDS dataset. So using the deep learning feature representation algorithm on i-LIDS dataset has better results than the traditional algorithms. And traditional algorithms divide the i_LIDS dataset into two parts: training and testing sub-datasets when conducting the experiment. While we only use the i_LIDS dataset as the testing set and conduct the cross-dataset (using two datasets) pedestrian re-identification experiment. And the performance of our algorithm is better than the other algorithm.

V. CONCLUSION AND FUTURE SCOPE

This study conducted cross dataset pedestrian re-identification to solve the problem that the single dataset is difficult to simulate the actual situation and the traditional pedestrian re-identification methods have the low accuracy as well as the poor generalization ability. This paper combined the 3-layer convolutional neural network with the Batch Normalization (BN) algorithm. Then the adjusted cosine similarity metric was integrated into the novel system. After that we conducted cross dataset pedestrian re-identification. The structure and training process were described in detail. Two public datasets, VIPeR and i-LIDS were tested to evaluate the performance of the learned pedestrian representation using the improved CNN model obtained from the Shinpuhkan2014 dataset. The results show that these proposed algorithm had better performance and the learned representation had good generalization capability.

ACKNOWLEDGMENT

The authors would thank for the fund support of Space Support research fund 2017KC080123.

REFERENCES

[1] Tulsi Jain, Kushagra Agarwal, Ronil Pancholia, "*Sentiment Analysis Based on a Deep Stochastic Network and Active Learning*", International Journal of Computer Sciences and Engineering (IJCSE), Vol.5, Issue.9, pp.1-6, 2017.

[2] Taigman, Y., Yang, M., Ranzato, M., Wolf, L, "*DeepFace: Closing the Gap to Human-Level Performance in Face Verification*", IEEE Conference on Computer Vision and Pattern Recognition, pp.1701-1708, 2014.

[3] Krizhevsky A, Sutskever I, Hinton G E, "*ImageNet classification with deep convolutional neural networks*", International Conference on Neural Information Processing Systems. Curran Associates Inc. pp.1097-1105, 2012.

[4] Kawanishi, Y., Yang, W., Mukunoki, M., Minoh, M., "*Shinpuhkan2014: A Multi-Camera Pedestrian Dataset for Tracking People across Multiple Cameras*", The Korea-Japan Joint Workshop on Frontiers of Computer Vision, FCV, 2014.

[5] Gray D, Brennan S, Tao H, "*Evaluating appearance models for recognition, reacquisition, and tracking*", Vol. 3, Issue. 5, pp.1-7, 2007.

[6] Zheng W S, Gong S, Xiang T, "*Person re-identification by probabilistic relative distance comparison*", In Computer Vision and Pattern Recognition(CVPR), IEEE,Vol. 42, pp.649-656,2011.

[7] Weinberger, K., Blitzer, J., Saul, L., "*Distance metric learning for large margin nearest neighbor classification*", Advances in neural information processing systems 18, pp.1473-1480, 2006.

[8] Dikmen, M., Akbas, E., Huang, T. S., Ahuja, N., "*Pedestrian recognition with a learned metric*", Lecture Notes in Computer Science, 6495 pp.501-512, 2010.

[9] Davis J V, Kulis B, Jain P, Sra S, Dhillon IS, "*Information-theoretic metric learning*", International Conference on Machine Learning, ACM, Vol. 227, pp.209-216, 2007.

[10] Zheng W S, Gong S, Xiang T, "*Person re-identification by probabilistic relative distance comparison*", In Computer Vision and Pattern Recognition(CVPR), Vol. 42, pp.649-656, 2011.

[11] Köstinger M, Hirzer M, Wohlhart P, Roth PM, Bischof H, "*Large scale metric learning from equivalence constraints*". IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp.2288-2295, 2012.

[12] Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R, "*Learning locally adaptive decision functions for person verification*", In Computer Vision and Pattern Recognition (CVPR), IEEE Conference, pp.3610-3617, 2013.

[13] Hu, Yang, Liao, Shengcai, Lei, Zhen, Yi, Dong, Li, Stan, "*Exploring Structural Information and Fusing Multiple Features for Person Re-identification*", Computer Vision and Pattern Recognition Workshops, IEEE, pp.794-799, 2013.

[14] Gheissari N, Sebastian T B, Hartley R, "*Person Re-identification Using Spatiotemporal Appearance*", Computer Vision and Pattern Recognition, Computer Society Conference on IEEE, pp.1528-1535, 2006.

[15] Hamdoun O, Moutarde F, Stanciulescu B, Steux B, "*Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences*", International Conference on Distributed Smart Cameras, IEEE, pp.1-6, 2008.

[16] Wang X, Doretto G, Sebastian T, Rittscher J, Tu P, "*Shape and Appearance Context Modeling*", International Conference on Computer Vision, IEEE, pp.1-8, 2007.

[17] Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M., "*Person re-identification by symmetry-driven accumulation of local features*", IEEE Computer Vision and Pattern Recognition, Vol.23, pp.2360-2367,2010.

[18] Ma B, Su Y, "*Local descriptors encoded by fisher vectors for person re-identification*", International Conference on Computer Vision. Springer-Verlag, pp.413-422, 2012.

[19] Dong S C, Cristani M, Stoppa M, Bazzani L, Murino V, "*Custom Pictorial Structures for Re-identification*", British Machine Vision Conference(BMVC), Vol. 68, pp.1-11, 2011.

[20] Zhao R, Ouyang W, Wang X, "*Unsupervised Salience Learning for Person Re-identification*", Computer Vision and Pattern Recognition, IEEE, pp.3586-3593, 2013.

[21] Zhao R, Ouyang W, Wang X, "*Person Re-identification by Salience Matching*", International Conference on Computer Vision, IEEE, pp.2528-2535, 2014.

[22] Liu Y, Shao Y, Sun F, "*Person re-identification based on visual saliency*", Intelligent Systems Design and Applications (ISDA) , Vol. 13,Issue. 6, pp.884-889, 2012.

[23] Gray D, Tao H, "*Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features*", Computer Vision - ECCV 2008, European Conference on Computer Vision, Marseille, France, Proceedings. DBLP, pp.262-275, October 12-18, 2008,.

[24] Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q., "*Person re-identification by support vector ranking*", In British Machine Vision Conference (BMVC), Aberystwyth, UK:BMVA Press, Vol. 2, Issue. 5, pp.1-11, 2010.

[25] Li W, Wang X, "*Locally Aligned Feature Transforms across Views*", Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, pp.3594-3601, 2013.

[26] Liu C, Chen C L, Gong S, Wang G, "*POP: Person Re-identification Post-rank Optimization*", IEEE International Conference on Computer Vision, IEEE Computer Society, pp.441-448, 2013.

[27] Ma A J, Yuen P C, Li J, "*Domain Transfer Support Vector Ranking for Person Re-identification without Target Camera Label Information*", IEEE International Conference on Computer Vision, IEEE, pp.3567-3574, 2014.

[28] Yi D, Lei Z, Liao S, Li SZ, "*Deep Metric Learning for Person Re-identification*". International Conference on Pattern Recognition, IEEE, pp.34-39, 2014.

[29] Yang Hu, Dong Yi, Shengcai Liao, Zhen Lei, Stan Z, "*Cross Dataset Person Re-identification*", Institute of Automation, Chinese Academy of Sciences (CASIA), ACCV, 9010, pp.650-664, 2014.

[30] Ioffe S, Szegedy C, "*Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*", arXiv preprint arXiv:1502.03167, pp.448-456, 2015.

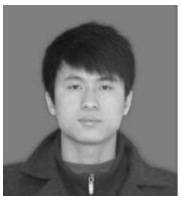
[31] Globerson A, Roweis S T, "*Metric Learning by Collapsing Classes*", Advances in Neural Information Processing Systems, Vol. 18, pp.451-458, 2006.

[32] Gray D, Tao H, "*Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features*", Computer Vision - ECCV 2008, European Conference on Computer Vision, Marseille, France , Proceedings. DBLP, pp.262-275, October 12-18, 2008.

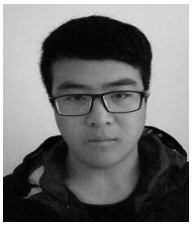
[33] Xing E P, Ng A Y, Jordan M I, Russell S, "*Distance metric learning, with application to clustering with side-information*", International Conference on Neural Information Processing Systems, MIT Press, pp.521-528, 2002.

Authors Profile

Ms.Hongmei Xie is now an associate professor on signal processing at Northwestern Polytechnical University , Xi'an,Shaanxi, China. She got her bachelor, master and Ph.D degree in 1995, 2000 and 2003 respectively all from Northwestern Polytechnical University.Her work is majorly about Electronic Circuits and System and got funds from National Science Fund of Shaanxi province. She has been a visiting scholar/researcher at University of Texas at Arlington(USA) and KULeuvin(Belgium). She has published more than 50 papers/patents in the fields of object detection and recognition in digital image or other media, signal processing and parallel algorithm design.



Mr. Yanggang Zhou is now a master student majoring at circuits and systems at Northwestern Polytechnical University, Xi'an, Shaanxi, China. He got his bachelor in 2015 from Chang'an University. His work is majorly about Artificial Intelligence and studied on image processing and deep learning. He has published several papers in the fields of object recognition and deep learning in digital image processing.



Mr. Qiang Liu is now a master student majoring at circuits and systems at Northwestern Polytechnical University, Xi'an, Shaanxi, China. He got his bachelor in 2015 from Chang'an University. His work is majorly about deep learning and FPGA programming and studied on image processing. He is working on the Space-supportive Fund in the fields of digital image processing and FPGA programming.

