Joint Dataset for CNN-based Person Re-identification

Sviatlana Ihnatsyeva
Faculty of Information Technology
Polotsk State University
Novopolotsk, Belarus
ignateva604@gmail.com

Rykhard Bohush
Faculty of Information Technology
Polotsk State University
Novopolotsk, Belarus
r.bogush@psu.by

Sergey Ablameyko
United Institute of Informatics Problems of NAS of Belarus
Belarusian State University
Minsk, Belarus
ablameyko@bsu.by

Abstract. **In this paper, we propose a joint dataset for person re-identification task that includes the existing public datasets CUHK02, CUHK03, Market, Duke, LPW and our collected PolReID. We investigate the training dataset size and composition effect on the re-identification accuracy. We carried out a number of experiments with different size of dataset to solve re-identification task. The results of experiments are presented.**

Keywords: **large-scale dataset, cross domain, convolution neural network, PolReID dataset**

I. INTRODUCTION

Person re-identification (person ReID) is the process of identifying a person in another place or at different time using video surveillance systems. The ReID system extracts features of the query-image and compares them with features other persons in dataset's gallery. Convolutional neural networks (CNN) are most efficient for feature extraction.

Re-identification comes with a number of problems. People appearance may change in the course of movement, or different people may appear similar. There is also the problem of occlusion. At some points, a person part can be hidden by other people or landscape elements. Video cameras can have dissimilar resolutions, shooting at different times - different degrees of illumination, different camera positions will give different backgrounds, and this leads to the existence of such a problem as domain shift. This is of great importance when working with datasets, because each of them is a separate domain [1, 2]. Good increasing the accuracy value of the re-identification algorithm was shown by the random erasing method [3]. Random erasing is a method to increase dataset by adding images, in which an arbitrary image fragment is randomly deleted, which is filled with zero or random values. This method improves the algorithm's occlusions resistance. Currently, most re-identification systems use these augmentation methods.

The deep neural networks success makes it possible to achieve high results in the person re-identification problem [4] when the data for training and testing are independent and identically distributed. However, such models are well suited for a training set and will perform poorly in an invisible domain [5].

One of the approaches to increase the stability of the ReID system is to use a dataset that will have the maximum similarity with the data with which the re-identification algorithm will have to work. Another approach is to significantly increase the training dataset, which would include a huge number of identifiers and their images. Our paper discusses a problem of forming large dataset associations for re-identification systems.

II. EXISTED DATASETS

When training a re-identification system, a dataset is of great importance, and the more diverse the examples, the more robust the trained system will be. To increase the training set without using additional data, the simplest way is to add to the existing dataset images from the original dataset, which have undergone such manipulations as rotation, reflection vertically or horizontally, changes in brightness and contrast, color fluctuations.

In [6], a cross-domain mixup scheme is considered, and proposed scheme study is carried out, when training is carried out on the Market 1501 dataset, and testing on Duke, and then vice versa, is trained on Duke, and tested on the Market. The studies carried out have shown that the re-identification accuracy in the two considered examples is different, and it is impossible to say unambiguously how the system will behave on other data sets. A large experiments number with a different composition of training and test samples are carried out in [7], where the authors propose a new CNN framework for learn effective features, which allows to improve re-identification in the cross domain, and the authors conduct a study by training the model on one of the datasets.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Shinpuhkan2014dataset, CUHK02, CASPR, i-LIDS, PRID, and testing is performed on VIPeR, i-LIDS, Shinpuhkan2014dataset. The datasets used for training are small, which is probably one of the reasons for the low Rank1 scores. In [2], the authors propose an approach to generalize the subject area and consider training sample variants, with a different composition and number of various datasets included in its composition. Increased training set includes over 18000 IDs and almost 122000 bounding boxes by combining different datasets

In [8], authors strive to develop a universal framework for human ReID that can be generalized and work well on target domains. This work also uses an increasing strategy for the training sample at the expense of other datasets, and is conducting several cross-domain experiments, including a combined unified database that included Market, Duke, CUHK03, and MSMT17. This database includes almost 9000 identifiers and more than 220000 boxes. Dataset expansion made it possible to improve Rank1 from 33.9 when training on Duke and testing on the Market, to 82.3 when training on a combined database.

The most famous and significant in volume terms are datasets such as Market, Duke, CUHK02, CUHK03, LPW.

Market-1501 was assembled at Tsinhua University in supermarket front and includes 32668 hand-crafted bounding boxes for 1501 people. 12936 bounding boxes for 751 people are used for training, and 19732 bounding boxes in galleries for 750 people to test the re-identification algorithm. In addition, the dataset contains 2793 bounding boxes-distractors [9]. Duke MTMC-ReID is a subset of the Duke MTMC dataset acquired in March 2014 from the campus Duke University. The images were taken from 8 CCTV cameras located between the buildings, and include 36411 bounding boxes. To train the re-identification algorithm, 16522 images for 702 people are used. The remaining 17661 bounding boxes for 702 people are used for testing [10]. CUHK02 contains 1816 people and five camera views pairs. Each of them contains 971, 306, 107, 193 and 239 people, respectively. There are 4 images for each person - 2 from one camera, 2 from the another. The dataset was collected from The Chinese University of Hong Kong (CUHK) campus. The dataset contains sensitive data and the authors ask that the privacy of CUHK students be respected [11]. CUHK03 was obtained from the same campus of the Chinese University of Hong Kong, and contains 1467 people, each person has 5 images from 2 angles. This set, like CUHK02, is available only for academic research and its distribution is available only by agreement with the authors [12]. LPW (Labeled Pedestrian in the Wild) is obtained from three different scenes. The first scene includes images from 3 CCTV cameras, the other two scenes include 4 cameras. The full dataset contains 2731 people captured by at least two cameras. 7694 image sequences were generated, with an average of 77 frames per sequence, thus the total LPW dataset contains 592438 bounding boxes [13].

III. Large joint ReID dataset

We used two approaches to form a large ReID dataset. The first one was that we combined the existing datasets, which are presented in different formats and their structure is different. Our second approach involves the formation of a new images set, which is included in the joint database being created. Thus, the joint database consists of Market, Duke, CUHK02, CUHK03, LPW and PolReID.

We developed our own dataset, called PolReID [14]. For its formation, video sequences received from volunteers were used. For each person, there are from 1 to 7 video sequences with different locations, illumination levels, image quality and distance from the CCTV camera. Thus, images for most people also correspond to different domains. To extract bounding boxes from frames, the YOLOv4 detection algorithm implemented in pyTorch [15] was used. Incorrect bounding boxes were removed manually. For each person in the dataset, there are images with partial overlap, both horizontal and vertical. The same person is presented from different angles. In total, the dataset contains images for 54 people and includes 5609 bounding boxes. PolReID is split into training and testing data. For training, 30 people (3603 bounding boxes) are used, for testing - 24 people (2006 bounding boxes). Examples of images are shown in Fig. 1.

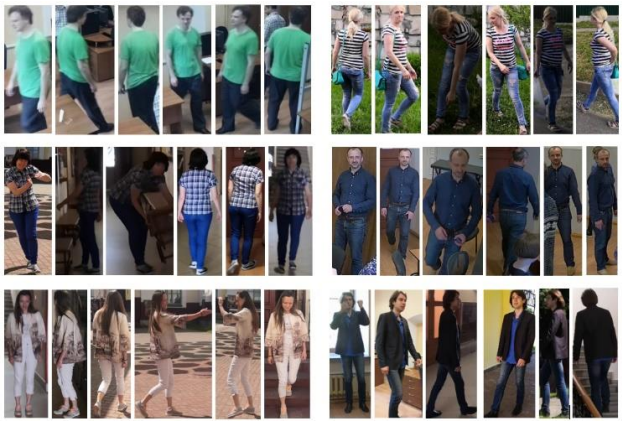

Fig. 1. Some images from PolReID dataset

When joining these datasets, considering the need to train algorithms and test them, the databases were also splinted. For Market and Duke, this task was accomplished in accordance with the protocols of the original documents.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In LPW, 666 people (141466 bounding boxes) were randomly selected for testing, the rest were added to training data. When joining sets with CUHK02, CUHK03, Market and Duke were not divided into test and training sets, as in [16, 17], and all images were used for training.

Joining dataset is challenging. This is due to the fact that different datasets have a different way of writing names, different file locations in directories hierarchy. Directory names can contain useful information such as camera number from which the image was taken, scene number. In addition, identifiers and sequence camera number values can be the same in different datasets, but they will belong to different people. To avoid such a situation, when adding each new dataset, the maximum value that was used in the existing dataset was added to the ID value and camera number.

For correct re-identification algorithm operation, the image file names were brought to a single recording format: XXXXX_cYYsZZ_AAAAAA_BB.jpg, where XXXXX is the person's identifier, YY is camera number, ZZ is the video sequence number from this camera, AAAAAA is the frame number in video sequence, BB is the different people number whose images were obtained from this frame. If the dataset did not contain any required information (usually ZZ or BB), the value was set to 0. The capital letters number in the example corresponds to digit numbers.

The joint dataset includes 8690 identifiers and 537109 images.

IV. Training and Re-Identification

A. Training model

For re-identification, the algorithm proposed in [18] was used with hyper parameters specified in Table I.

TABLE I. Hyper parameters used in model

Backbone network:	**DenseNet-121** [19], **ResNet-50** [20]
Droprate	0.5
Batch sise	32
Learning rate	0.05
Epochs	60

After epoch 40, decay learning rate by a factor of 0.1, and Fig. 2 shows that this has a positive effect on the convergence of the model.

We carried out experiments number with different consist increase the data for training the re-identification algorithm and testing with different data sets.

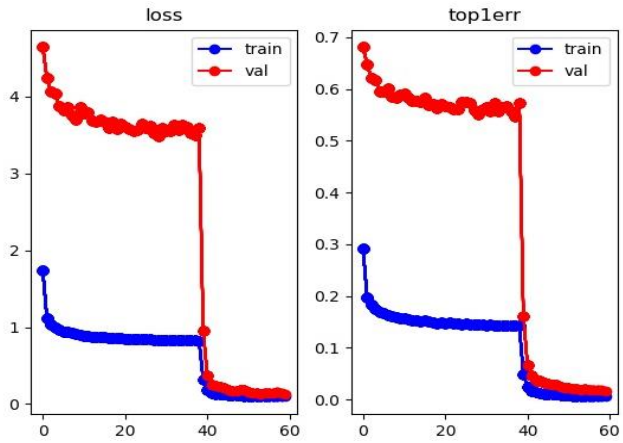

Fig. 2. Loss and top1 error graph during training re-identification model with backbone network DenseNet-121 on joint training sample

Model was trained for nine different training samples. Table II show the training sample consist and size.

This algorithm assumes that the trained neural network extracts features for each person located in the test sample gallery. Then, for each query, the all images feature table is ranked. The cosine distance is used as the similarity metric. The images obtained from the same camera as the request image are excluded from the ranked feature table. To assess the re-identification accuracy, we used metrics is Rank1, Rank5, Rank10 and mAP. The RankN metric is the ranking accuracy, i.e. the ratio of correctly obtained results to the total number of outputs among the N first issued results. mAP – this is mean Average Precision for a queries set is the mean of the average precision for each query.

TABLE II. Parametrs for training Samples

Training datasets	**Size (ID/Bboxes)**
Duke	702 / 16522
Market	751 / 12936
LPW	2064 / 448568
Market, duke	1453 / 29188
Market, duke, PolReID	1483 / 32765
Market, duke, LPW, PolReID	3547 / 481333
CUHK02, CUHK03, Market, Duke	6596 / 84964
CUHK02, CUHK03, Market, Duke, PolReID	6626 / 88541
CUHK02, CUHK03, Market, Duke, PolReID, LPW	8690 / 537109

B. Re-identification results

The experimental results are presented in Table III. Samples for training and testing do not overlap.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE III. EXPERIMENTAL RESULTS

Dataset for train \ Dataset for test		*Market*		*Duke*		*LPW*		*PolReID*	
		DenseNet	ResNet	DenseNet	ResNet	DenseNet	ResNet	DenseNet	ResNet
Market	Rank1:	89.782	87.708	39.138	31.418	32.132	27.628	65.854	60.975
	Rank5:	96.259	97.952	55.207	48.070	40.691	38.438	65.854	63.415
	Rank10:	97.298	96.675	61.715	54.533	44.294	43.234	65.854	65.854
	mAP:	73.439	70.536	21.079	16.901	19.238	17.734	58.632	57.227
Duke	Rank1:	51.456	44.151	81.688	79.623	28.679	24.775	63.415	65.854
	Rank5:	70.042	62.084	90.260	89.632	36.937	34.535	63.415	68.293
	Rank10:	76.485	69.269	92.774	92.369	40.991	39.339	63.415	68.293
	mAP:	23.536	18.590	64.029	62.001	16.306	13.835	55.038	52.840
LPW	Rank1:	63.005	56.562	41.248	36.894	79.729	71.772	**68.293**	65.854
	Rank5:	79.365	74.822	57.092	51.706	84.535	79.580	**68.293**	68.293
	Rank10:	85.273	81.799	63.600	57.900	86.036	82.733	**68.293**	68.293
	mAP:	34.663	30.413	23.449	18.808	70.069	61.809	58.180	56.584
Market, Duke	Rank1:	**92.132**	89.608	82.406	81.373	44.294	39.940	65.854	68.293
	Rank5:	97.090	95.784	91.023	89.722	55.255	48.949	65.854	68.293
	Rank10:	98.248	97.595	93.312	92.684	59.610	53.303	68.293	68.293
	mAP:	77.698	74.461	67.766	65.772	29.135	26.094	64.384	64.731
Market, Duke, PolReID	Rank1:	91.716	88.955	82.982	79.488	44.294	35.586	**68.293**	65.854
	Rank5:	96.704	95.814	91.472	88.734	55.105	46.847	**68.293**	68.293
	Rank10:	98.070	97.565	93.896	92.011	59.309	52.402	**68.293**	68.293
	mAP:	78.004	71.840	68.065	63.765	29.817	22.709	**65.872**	63.077
Market, Duke, LPW PolReID	Rank1:	**92.132**	88.717	**83.079**	78.591	80.030	75.225	**68.293**	65.854
	Rank5:	96.615	95.814	**91.607**	88.330	84.535	81.081	**68.293**	65.854
	Rank10:	97.951	97.506	**93.537**	91.158	86.687	84.084	**68.293**	65.854
	mAP:	**80.470**	74.176	**69.568**	62.842	73.638	67.154	61.603	61.004
CUHK02, CUHK03, Market, Duke	Rank1:	-	-	-	-	50.751	44.895	**68.293**	68.293
	Rank5:					60.060	54.955	**68.293**	68.293
	Rank10:					65.466	59.910	**68.293**	68.293
	mAP:					38.260	33.153	**66.800**	65.367
CUHK02, CUHK03, Market, Duke, PolReID	Rank1:	-	-	-	-	50.900	42.042	**68.293**	65.854
	Rank5:					60.661	51.051	**68.293**	65.854
	Rank10:					64.865	57.207	**68.293**	65.854
	mAP:					37.370	30.338	64.700	64.350
CUHK02, CUHK03, Market, Duke, PolReID, LPW	Rank1:	-	-	-	-	**83.934**	78.679	**68.293**	68.293
	Rank5:					**87.838**	83.484	**68.293**	68.293
	Rank10:					**89.640**	85.886	**68.293**	68.293
	mAP:					**76.286**	69.815	64.114	61.070

Horizontally, the table can be divided into three parts, each of which three lines consist. The first part includes the testing results the model when it was trained on one of the Market, Duke and LPW datasets. The best value has examples when training and testing were carried out on the same dataset, i.e. training and test samples belong to the same domain. But this result is not objective for invisible domains. The best Rank1 and mAP value corresponds to the experiment when testing invisible datasets, if the training sample was an LPW dataset, and testing was carried out on Market and Rank1 = 63.005, mAP = 34.663. When tested in invisible datasets, the LPW dataset generally showed better training ability compared to Market and Duke, which is most likely due to the significantly larger size of the LPW training set (448568 bounding boxes for 2064 IDs). If we pay attention to the example where LPW acted as a test sample, we can see that the re-identification accuracy is higher when training on the Market than on Duke, which gives us reason to assume that the different identities number is more important than the number of bounding boxes.

The second table part reflects the test results with an increase in the training sample. Here we confirm that an increase in the dataset for training the used CNN leads to an increase in the re-identification accuracy. It was found that the best test results for re-identification can be obtained when the training sample includes data belonging to the same domain as the target one. The best accuracy was achieved when combining all 4 datasets for training during testing: on Market Rank1 did not change, but the mAP increased from 77.698 to 80.470; for duke increase all parameters; for LPW, Rank1 almost doubles, and mAP more than doubles when the dataset is increased from 32765 Bbox for 1483 ID when Market and Duke are combined to 481333 bbox for 3547 ID; for PolReID we can see an increase in Rank1, Rank5, Rank10, but the mAP has

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

become smaller. The last table three rows show the test results when the approach was slightly changed when creating a training sample, i.e. the data from the Market, Duke, CUHK02 and CUHK03 sets were not divided into test and training sets, and this does not allow testing on Market and Duke. Testing on the LPW dataset showed an increase in the Rank1 accuracy to 50.900, in the case when the LPW is an invisible dataset, which is almost one and a half times higher than when using one cross dataset for training. Adding LPW to the training sample allowed us to obtain the maximum values for all estimated metrics, and Rank1 = 83.934, mAP = 76.289.

The maximum Rank1 accuracy achieved for the PolReID dataset is 68.293. The reason for this may be test sample size and composition. Some of the PolReID dataset images were obtained from only one camera, and the re-identification algorithm used cannot detect them. With the further dataset expansion, this will be taken into account.

V. Conclusion

Modern person re-identification systems use convolutional neural networks to efficiently extract features. With this approach, the training sample is of great importance. Dataset variety and size allows the re-identification system to have better generalizability and reliability. The built unified database includes 8690 identifiers and 537109 images. Such a large dataset allowed us to improve Rank 1 and / or mAP on all test sets. In further research, we plan to expand the PolReID database we have collected.

References

[1] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. "Deep Learning for Person Re-identification: A Survey and Outlook", IEEE transactions on pattern analysis and machine intelligence, 2021.

[2] J. Jieru, Q. Ruan and Timothy M. Hospedales. "Frustratingly Easy Person Re-Identification: Generalizing Person Re-ID in Practice", BMVC, 2019.

[3] Zhong, Zhun, L. Zheng, Guoliang Kang, Shaozi Li, Y. Yang. "Random Erasing Data Augmentation", 2020. URL: https://arxiv.org/pdf/1708.04896.pdf.

[4] Xiao, Tong, Hongsheng Li, Wanli Ouyang and Xiaogang Wang. "Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.1249-1258, 2016.

[5] Bąk, Sławomir and Peter Carr. "One-Shot Metric Learning for Person Re-identification", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.1571-1580, 2017.

[6] Ch. Luo, Ch. Song, Zh. Zhang, "Generalizing Person Re-Identification by Camera-Aware Invariance Learning and Cross-Domain Mixup", ECVA, 2020.

[7] Hu. Yang, Dong Yi, Shengcai Liao, Zhen Lei and S. Li. "Cross Dataset Person Re-identification", ACCV Workshops, 2014.

[8] Jin, Xin, Cuiling Lan, Wenjun Zeng, Zhibo Chen and Li Zhang. "Style Normalization and Restitution for Generalizable Person Re-Identification", IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3140-3149, 2020.

[9] L. Zheng, L. Shen, L. Tian, Sh. Wang, J. Wang, Q. Tian. "Scalable Person Re-Identification: A Benchmark", IEEE Int. Conf. on Computer Vision (ICCV), pp. 1116-1124, 2015.

[10] Zheng, Zhedong et al., "Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro", IEEE Int. Conf. on Computer Vision (ICCV), pp. 3774-3782, 2017.

[11] W. Li, Xiaogang Wang. "Locally Aligned Feature Transforms across Views", IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3594-3601, 2013.

[12] W. Li, Rui Zhao, Tong Xiao and Xiaogang Wang. "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification", IEEE Conf. on Computer Vision and Pattern Recognition, pp. 152-159, 2014.

[13] Song, Guanglu, B. Leng, Y. Liu, Congrui Hetang and Shaofan Cai. "Region-based Quality Estimation Network for Large-scale Person Re-identification", AAAI, 2018.

[14] PolReID. URL: https://github.com/SvetlanaIgn/PolReID.

[15] Pytorch-YOLOv4. URL: https://github.com/Tianxiaomo/pytorch-YOLOv4.

[16] S. Choi, T. Kim, M. Jeong, H. Park, C. Kim, "Meta Batch-Instance Normalization for Generalizable Person Re-Identification", IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3425-3435, 2021.

[17] Chen, Peixian, Pingyang Dai, Jianzhuang Liu, Feng Zheng, Q. Tian and Rongrong Ji. "Dual Distribution Alignment Network for Generalizable Person Re-Identification", AAAI , 2021.

[18] Person reID baseline pytorch. URL: https://github.com/layumi/Person_reID_baseline_pytorch.

[19] Huang, Gao, Zhuang Liu and Kilian Q. Weinberger. "Densely Connected Convolutional Networks", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2261-2269, 2017.

[20] He, Kaiming, X. Zhang, Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.