



Random linear interpolation data augmentation for person re-identification

Zhi Li¹ · Jun Guo¹ · Wenli Jiao¹ · Pengfei Xu¹ · Baoying Liu¹ · Xiaowei Zhao¹

Received: 2 May 2018 / Revised: 31 October 2018 / Accepted: 11 December 2018 /

Published online: 20 December 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Person Re-Identification (person re-ID) is an image retrieval task which identifies the same person in different camera views. Generally, a good person re-ID model requires a large dataset containing over 100000 images to reduce the risk of over-fitting. Most current hand-crafted person re-ID datasets, however, are insufficient for training a learning model with high generalization ability. In addition, the lacking of images with various levels of occlusion is still remaining in most existing datasets. Motivated by these two problems, this paper proposes a new data augmentation method called Random Linear Interpolation that can enlarge the sizes of person re-ID datasets and improve the generalization ability of the learning model. The key enabler of our approach is generating fused images by interpolating pairs of original images. In other words, the innovation of the proposed approach is considering data augmentation between two random samples. Plenty of experimental results demonstrates that the proposed method is effective to improve baseline models. On Market1501 and DukeMTMC-reID datasets, our approach can achieve 92.71% and 82.19% rank-1 accuracy, respectively.

Keywords Person re-identification · Data augmentation · Linear interpolation

✉ Pengfei Xu
pfxu@nwu.edu.cn

Zhi Li
zhili@stumail.nwu.edu.cn

Jun Guo
guojun@nwu.edu.cn

Wenli Jiao
jiaowenli@stumail.nwu.edu.cn

Baoying Liu
pola.liu@nwu.edu.cn

Xiaowei Zhao
xiaoweizhao4@gmail.com

¹ School of Information Science and Technology, Northwest University, Xian, China

1 Introduction

With the rapid development of deep learning, more and more convolutional neural network models effectively deal with the problems of image classification, object detection and other tasks in computer vision. Large-scale data is required for current large convolutional networks to ensure the generalization ability of models. Person re-identification (person re-ID), identifying pedestrians among different camera views, is facing the challenge that the sizes of many person re-ID datasets are small. Although the recognition accuracy on different datasets have been increasing, the problem that some large networks can not work well on small datasets is still existing. For instance, the same CNN model has higher recognition accuracy on Market-1501 [36] with over 30,000 images than that on PRID450s [27] with less than 1000 images.

Data augmentation viewed as a data preprocessing method, generates new training samples from original datasets. It is used widely to increase data size of a dataset in image classification, object recognition and person re-identification. Moreover, data augmentation plays a key role in deep learning gradually due to the ability which alleviates over-fitting of large convolutional network models effectively. Two commonly used data augmentation methods, Random Cropping [32] and Random Flipping [29], effectively improve the generalization ability of most existing CNN models. In detail, Random Cropping decreases the influence of background in the CNN decision, and focuses more on parts of the object than on the whole object. Random Flipping enables CNN models to learn from different directions of input images. Thus, both of them concern about image processing on one image itself.

Recently, Random Erasing [41] has been proposed to implement in most existing CNN models. For person re-identification, it increases occluded images through selecting a random rectangle region in a pedestrian image and erasing its pixel with random values. Training images with different levels of occlusion are generated to reduce the risk of over-fitting. However, it only concerns a simple situation that some noise is embedded in the image.

In this paper, we propose Random Linear Interpolation (RLI), a new data augmentation method to increase images with more complicated occlusion in the person re-ID dataset. Random Linear Interpolation keeps a part of images unchanged, and generates new images by fusing pairs of images with linear interpolation [3, 4, 23]. Plenty of mixed images, looking similar to original images but not the same, are generated in training to power the generalization ability of CNN models. Furthermore, Random Linear Interpolation greatly improves the recognition accuracy in person re-ID, and makes CNN models robust to outliers and variable levels of occluded images. Examples of Random Linear Interpolation are shown in Fig. 1.

In Fig. 1, two random generated images with different interpolation strength on Market1501 [36] and DukeMTMC-reID [39] are shown. Image A is viewed as a base image, and image B is fused in it. The new generated image A^* shares both features of image A and image B. μ is a hyper-parameter that controls the strength of interpolation among pedestrian images. We can obviously observe that the similarity between image A and image A^* is determined by μ . For instance, when μ is equal to 0.3004, features of the pedestrian in image A remain a little in the generated image A on Market1501. With the increasing of μ , more original features are seemed remaining in new images. Specifically, when the value of μ is equal to 0.9212, only several features of pedestrian changes on DukeMTMC-reID.

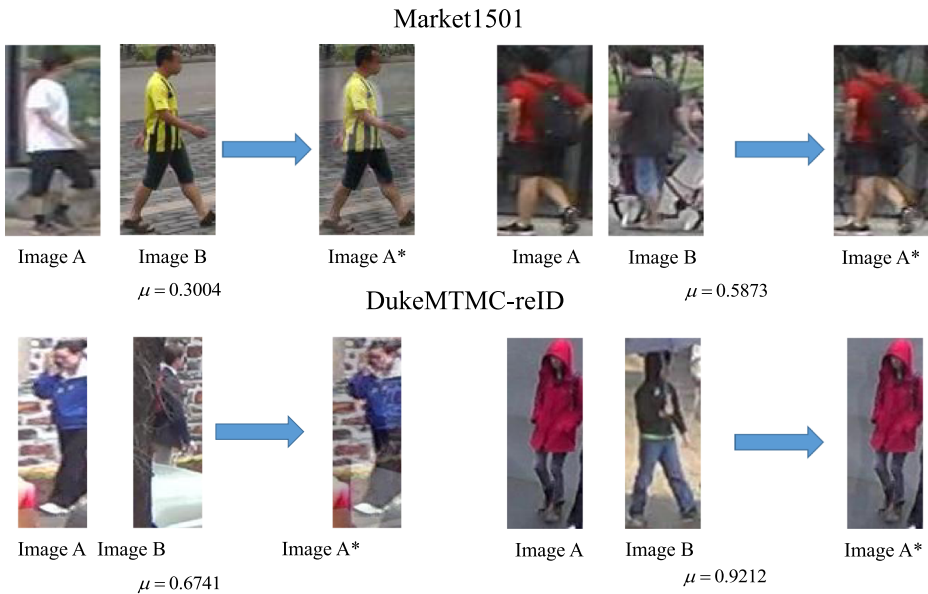


Fig. 1 Examples of Random Linear Interpolation applied on the two person re-identification datasets, i.e. Market1501, DukeMTMC-reID

In summary, this paper makes the following contributions:

- We propose a new data augmentation method - **Random Linear Interpolation (RLI)**, which is light-weighted and can implement in most existing convolutional neural network models to improve the generalization ability.
- For person re-identification, RLI can reduce the requirement of I/O and increase various levels of complicated occlusion by fusing images of mini-batches in training. Softly adjusting the proportions of interpolated samples are adopted in RLI to control the learning ability of the model for outliers.
- The proposed methods can improve the performance of baseline models including ResNet and DenseNet. Furthermore, we also augment data by exploring a new way that we consider data augmentation between two random samples rather than a sample itself.

2 Related work

2.1 Data augmentation

Data augmentation is first proposed by [31] to deal with missing value problems, as exemplified by some training samples without labels. Recently, it has been focused on once again when the convolutional neural network (CNN) developed rapidly. The size of data is vital to most existing CNN models because a CNN model with large trained samples will have a lower risk of over-fitting. Data augmentation is the technique that enlarges the number of training samples by processing original samples. In [20], random cropping was first proposed to construct a large dataset with 80 million tiny images for object and scene

recognition. Simonyan and Zisserman [29] further adopted random flipping to augment the training dataset for large-scale image recognition.

Later, to deal with the task of image classification, object detection and person re-identification, random erasing [41] is presented to randomly select a rectangle region in training images and erase its pixels with random values. The deep convolutional generative adversarial network (DCGAN) is proposed by [39]. In DCGAN, a generator network as a subnet is used to generate virtual data and a discriminator network is used to identify whether the sample is virtual or real. Different from these methods, this paper presents a new method named Random Linear Interpolation to randomly generate mixed images by fusing an image to another image and feed both generated and actual images into the CNN model, thereby enhancing robustness of the model to outliers.

2.2 Person re-identification

Person re-identification is a challenging image retrieval problem [7, 21, 42]. Due to factors such as different lighting and camera angles, the same person displays different appearance features, which poses a great challenge for recognition. In order to find the invariant features of pedestrian images in different camera views, a large number of distance metric learning methods including KISSME, KCCA, MLF, EIML, RPLM and APLM [13, 14, 22, 25, 28, 33] have been proposed. Some methods based on dictionary learning [6, 17–19] have also been used to deal with the unsupervised person re-ID problem because it can convert high-dimensional visual features into low-dimensional sparse coding.

In recent years, due to the continuous expansion of the person re-ID data sets, many convolutional neural network models have been proposed to improve the accuracy of pedestrian recognition. Zheng et al. [37] proposed the IDE method which deals with person re-ID by embedding discriminant identities. With using the triplet loss function to maximize the errors between the positive and negative samples while minimizing the errors of the positive and positive samples, TriNet [12] alleviates the problem that the scale of datasets are small and effectively works on various convolutional neural network models. SVDNet [30] optimizes the deep representation learning process in CNN training for the application of person re-identification. Zheng et al. [39] uses a deep convolutional neural network to generate unlabeled samples and proposes a label smoothing regularization method (LSRO) for integrating unlabeled outliers.

3 Our approach

In this section, we show the details of Random Linear Interpolation implementing in the convolutional neural network model for person re-identification. We first describe RLI by selecting two images to interpolate randomly in the training set. Then the theoretical basis of Random Linear Interpolation for person re-ID is analyzed. Finally, we analyze the difference among Random Linear Interpolation and other data augmentation methods.

3.1 Random linear interpolation

Random Linear Interpolation is used during training process to randomly mix two images by linear interpolation in the convolutional neural network model. Suppose there are n images in a mini-batch, then we keep k samples unchanged and perform linear interpolation operation on remaining images. In the step of data pre-processing, more and more virtual

similar samples are generated randomly. And a new image represents the fusion of two random original images, in other words, a sample in the training set owning two labels.

Algorithm 1 Random linear interpolation data augmentation.

```

Data: Input image  $I(x_a, y_a), I(x_t, y_t)$ ; interpolation strength  $\alpha$ ; interpolation possibility  $\gamma$ .
Result: Output image  $I^*(\bar{x}_a, \bar{y}_a)$ .
Initialization  $\gamma_1 \leftarrow \text{Rand}(0, 1)$ 
if  $\gamma_1 \geq \gamma$  then
     $I^*(\bar{x}_a, \bar{y}_a) \leftarrow I(x_a, y_a)$ ;
    return  $I^*(\bar{x}_a, \bar{y}_a)$ 
else
    while true do
         $\mu \leftarrow \text{Beta}(\alpha, \alpha)$ ;
         $I^*(\bar{x}_a \{R\}) \leftarrow \mu I(x_a \{R\}) + (1 - \mu)I(x_t \{R\})$ ;
         $I^*(\bar{x}_a \{G\}) \leftarrow \mu I(x_a \{G\}) + (1 - \mu)I(x_t \{G\})$ ;
         $I^*(\bar{x}_a \{B\}) \leftarrow \mu I(x_a \{B\}) + (1 - \mu)I(x_t \{B\})$ ;
         $I^*(\bar{y}_a) \leftarrow \mu I(y_a) + (1 - \mu)I(y_t)$ ;
        return  $I^*(\bar{x}_a, \bar{y}_a)$ 
    end
end

```

Two original images $I_a = (x_a, y_a)$ and $I_t = (x_t, y_t)$, where I_a represents a sample to be interpolated while I_t is a random image in the same mini-batch, are selected to mix all pixels in them. Specifically, the new virtual sample has a soft label representation which differs from the hard label representation that one data only corresponds to one label. The samples in the new dataset may be shown as $I_a^* = (\bar{x}_a, \bar{y}_a)$, where \bar{x}_a represents features of the virtual sample generated by Random Linear Interpolation and \bar{y}_a is the mixed label of (y_a, y_t) . Moreover, we keep I_a unchanged if I_a is not augmented by linear interpolation. The new generated virtual sample by Random Linear Interpolation is defined as:

$$\bar{x}_a \{R, G, B\} = \mu x_a \{R, G, B\} + (1 - \mu)x_t \{R, G, B\} \tag{1}$$

$$\bar{y}_a = \mu y_a + (1 - \mu)y_t \tag{2}$$

where R, G, B represents the pixel values of three channels of the original input image. μ is a randomly generated number through a Beta distribution $\text{Beta}(\alpha, \beta)$. For simplicity, we set β equal to α in this paper. To be noted, in training, we only use one data loader to achieve one mini-batch and generate virtual samples in the same mini-batch. This operation reduces the requirement of I/O and also decreases the confusion of the dataset. The pipeline of Random Linear Interpolation is shown as Algorithm 1.

3.2 Random linear interpolation for person re-identification

In the person re-identification problem, we focus on finding a function $f \in F$ that matches the same identities among different camera views. Most of existing methods based on convolutional neural network belong to supervised learning. In the recent convolution neural network, a large number of parameters, even more than the number of samples in the data set, may over-learn the feature of training data and have weak generalization ability to deal with outliers and the presence of occlusion. Therefore, we try to learn by enhancing the

data that approximates the original samples, which fits well with the idea of Vicinal Risk Minimization (VRM) [5].

Assume that distribution of raw data set $D = \{(x_i, y_i)\}_{i=1}^m$ is P and $(x_i, y_i) \sim P$ for all $i = 1, 2, \dots, m$. But in most practical situations, the data distribution P is unknown and \hat{P} is defined to approximate the true distribution P . VRM is one of the principles that minimize the risk between predictions and targets. The approximated distribution \hat{P} in the VRM is shown as (3):

$$\hat{P}_v(\bar{x}, \bar{y}) = \frac{1}{m} \sum_{i=1}^m v(\bar{x}, \bar{y}|x_i, y_i) \quad (3)$$

where $v(\bar{x}, \bar{y}|x_i, y_i)$ is the vicinal distribution. However, different from using gaussian kernel in VRM, we construct a virtual dataset $D_v \{(\bar{x}, \bar{y})\}_{i=1}^n$ that satisfies the principle that new samples in D_v are generated by the linear interpolation of two random samples in D . And the new vicinal distribution can be written as (4) and (5):

$$P^*(\bar{x}, \bar{y}) = \frac{1}{n} \sum_{i=1}^n \delta(\bar{x}, \bar{y}) \quad (4)$$

$$\delta(\bar{x}, \bar{y}) = \delta(\bar{x} = \mu \cdot x_j + (1 - \mu) \cdot x_i, \bar{y} = \mu \cdot y_i + (1 - \mu) \cdot y_j) \quad (5)$$

where $\delta(\bullet)$ is the Dirac distribution and (x_j, y_j) represents a random sample in D . As a hyper-parameter, μ controls the similarity between virtual and real samples.

3.3 Comparison with other data augmentation methods

We compare Random Linear Interpolation (RLI) with three effective data augmentation methods for person re-identification including random cropping, random flipping and random erasing. First of all, random cropping is a classic data augmentation method that training images is cropped randomly to meet the input size of the convolutional neural network. Parts of input pedestrian images are paid more attention to than the whole image by random cropping, and meanwhile the background information of images can be ignored. Different from random cropping, Random Linear Interpolation takes each pedestrian image as a whole, and fuses other image in it linearly. In other words, although we change all the pixel values in an image, we do not change the internal structure of the image data.

Random flipping is another data augmentation method which is commonly used in person re-identification. It enables a CNN model to learn the same image with different directions, and do not incur information loss in pedestrian images during augmentation. In comparison with random flipping, each time the model is trained by Random Linear Interpolation, the new generated image will show different values of all the pixels. Moreover, The fusion of two images can be viewed as adding various levels of noise to one of images.

Recently, random erasing has been introduced to dealing with person re-ID problems. It selects a rectangle region in an image randomly and erasing its pixels with random values. To make the CNN model robust to occlusion in person re-identification, various levels of occlusion in pedestrian images are focused on during training. Compared to random erasing, Random Linear Interpolation generates a new image by fusing a pair of images. In addition, changing the pixels of the whole image by linear interpolation, RLI expands the scope of image occlusion and improve the generalization ability of the model.

In our experiment, it is observed that all of these data augmentation methods can improve the recognition accuracy of person re-ID, and the proposed method achieves the best performance.

4 Experiments

In this section, we describe in detail the experimental performance evaluation of Random Linear Interpolation (RLI) for person re-identification. Two standard datasets, i.e., Market1501 and DukeMTMC-reID are introduced at first and then we evaluate the performance of the proposed method in some baseline models, such as ResNet18, ResNet34, ResNet50, ResNet101, ResNet152 [11] and DenseNet [15]. Finally the results that RLI compared to superior data augmentation and person re-identification methods on two benchmark datasets are analyzed.

4.1 Datasets

Two commonly used datasets including Market1501 and DukeMTMC-reID are summarized in Table 1. We take many experiments on the benchmark datasets. Market1501 is collected by six cameras in front of a supermarket in Tsinghua University. Overall, this dataset contains 32,668 bounding boxes of 1,501 pedestrian identities. Images of each identity are captured about 20 images on average. Both hand-crafted and the Deformable Part Model (DPM) [8] are used to label bounding boxes of 1501 identities. We follow the paper [36] that used 12936 images for training and another 19732 images for testing.

DukeMTMC-reID was taken from the Duke MTMC tracking dataset and has a total of 36,411 images for 1404 identities. We used 702 identities for training and the remaining identities for testing. All of images of this dataset are captured by 8 cameras, and pedestrian bounding boxes are available by hand-crafted. Follow [39], in testing, we pick one query image for each identity in each camera and put the other images in the gallery.

4.2 Settings

4.2.1 Baseline CNN model for person Re-ID

ResNet is chosen as the baseline model to evaluate the performance of the propose method. In our experiments, we take many experiments on five ResNet models including ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152. Note that we remove the full connected layer of ResNet, adding a linear layer after ResNet, and followed by Batch Normalization [16] to regularize mini-batch of input. Moreover, we use Leaky ReLU [34] as activation function with negative slope equal to 0.01 and a dropout layer is added with the possibility equal to 0.5. We use the weight parameters of ResNet model pre-trained on Image-net for fine-tuned and set bias to 0.

Table 1 Information of four person re-ID datasets in our experiments

Dataset	Market1501	DukeMTMC-reID
Year	2015	2017
IDs	1501	1812
Cameras	6	8
Bounding boxes	32668	36411
Training/Testing	12936/19732	18750/17661
Label method	Hand/DPM	Hand

In training, we divide input images into mini-batches and adopt a stochastic gradient descent algorithm proposed in [10]. Note that, we set the training batch size of the ResNet101 and ResNet152 to 16 considering the memory limit, and in other networks we set the batch size to 32. Besides, the size of input images is 256×128 . The learning rate is set to 0.1 in the full connected layer and classification layer, and 0.01 in other layers of the ResNet50 model. After 40 training epochs, we decrease the learning rate to 0.001. The momentum is 0.9 and the weight decaying is set to $5e^{-4}$. We train 60 epochs in all networks.

After augmentation for an input image in a mini-batch, there are two target labels generated for the image. Most existing loss functions can not be applied to this situation that the error between the predicted label y_{pred} and two ground-truth labels (y_a, y_b) is calculated. Thus, we adopt a novel loss function shown as (6):

$$loss = \mu \cdot (y_{pred} - y_a) + (1 - \mu) \cdot (y_{pred} - y_b) \tag{6}$$

In ResNet50, the training loss on Market1501 and DukeMTMC-reID of 10 epochs are shown in Fig. 2. It can be seen that, training loss in the model with Random Linear Interpolation is above baseline but this has no effect on the accuracy of the recognition.

4.2.2 Leaky ReLUs and dropout

Like rectified linear unit (ReLU), leaky ReLU is another commonly used activation functions to alleviate the vanishing gradient problem by identifying positive values. It is first proposed in acoustic model [26]. In comparison with ReLU, Leaky ReLU allows negative values and gives all of them with a non-zero slope. Xu et al. [34] has demonstrated that Leaky ReLU shows a better performance than ReLUs in image classification task. We used Leaky ReLU as activation function for the better performance in person re-identification.

Because of large cross-view misalignment, occlusions and pose variations in person re-ID, some patch information on the same pedestrian is likely learned incorrectly in training. We adopt the dropout strategy to overcome the influences of mismatched patches and bad neural units. When a training sample as input in each epochs, some outputs of convolutional layers are randomly set as zeros. In Dropout, we set possibility as 0.5 to reduce half of neural units in a convolutional layer randomly. This action can make the CNN model more stable in person re-ID.

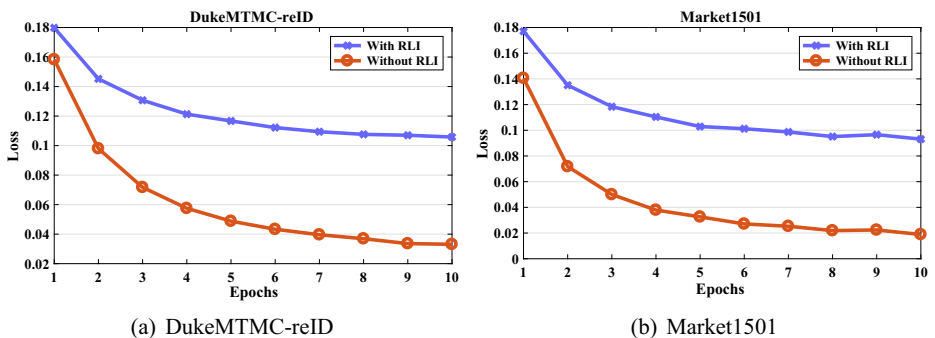


Fig. 2 Training loss of 10 epochs on DukeMTMC-reID (a) and Market1501 (b)

4.3 Parameter analysis

Two important parameters are involved with Random Linear Interpolation, i.e., interpolation strength α and interpolation possibility γ . α is the parameter in beta distribution that determines values of hyper-parameters μ . For each interpolated image, the value of μ , which controls the strength of interpolation, is randomly generated by $Beta(\alpha, \alpha)$. With $gamma$ set to 0.4, the recognition accuracy of the proposed method is shown by adjust alpha from 0.1 to 1 stepping 0.1. It can be seen that in Fig. 3b, Random Linear Interpolation with α improves over the baseline model. And rank-1 accuracy changes weakly on Market1501 when using different values of α in RLI.

γ is another parameter that determines the number of samples to be subjected to Random Linear Interpolation. When α equals to 0.1, results of rank-1 accuracy with different γ are shown in Fig. 3b. It is observed that when is set to 0.4, ResNet50 achieves the highest recognition accuracy on DukeMTMC-reID. Furthermore, with the increase of γ , the model over-learns features of generated samples and reduces attention to the original matched samples.

4.4 Performance evaluation

4.4.1 Improving different baseline models

We first verify the effectiveness of Random Linear Interpolation. γ and α are set to 0.4 and 0.1, respectively. The same parameters such as learning rate, weight decaying and dropout possibility are used in the following experiments.

Three convolution neural network models including ResNet18, ResNet34 and ResNet50 are used as baseline methods for person re-identification. Results of rank-1 accuracy and mAP on Market1501 and DukeMTMC-reID are shown in Fig. 4, respectively.

In Fig. 4a, with using RLI, the rank-1 accuracy of three models increases from 1.58% to 2.17% and the mAP increases from 1.07% to 6.23% on Market1501. Specifically, after implementing RLI in the models on DukeMTMC-reID shown in Fig. 4b, the rank-1 accuracy increased by 3.06%, 10.09% and 3.33% in ResNet18, ResNet34 and ResNet50, respectively. And mAP increased by an average of 4.13%. Results of these experiments show great performance while a baseline CNN model using Random Linear Interpolation data augmentation for person re-ID.

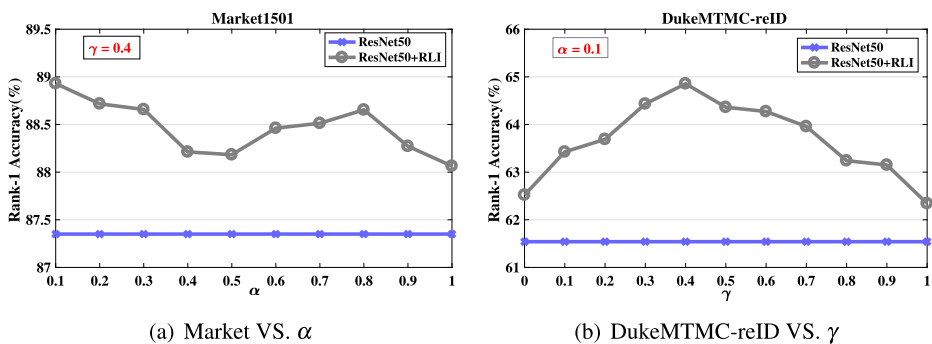


Fig. 3 Results of sensitivity analysis. In 3a, when γ is fixed, results of ResNet50 with α from 0.1 to 1. By fixing α , scores on DukeMTMC-reID with γ changed from 0 to 1

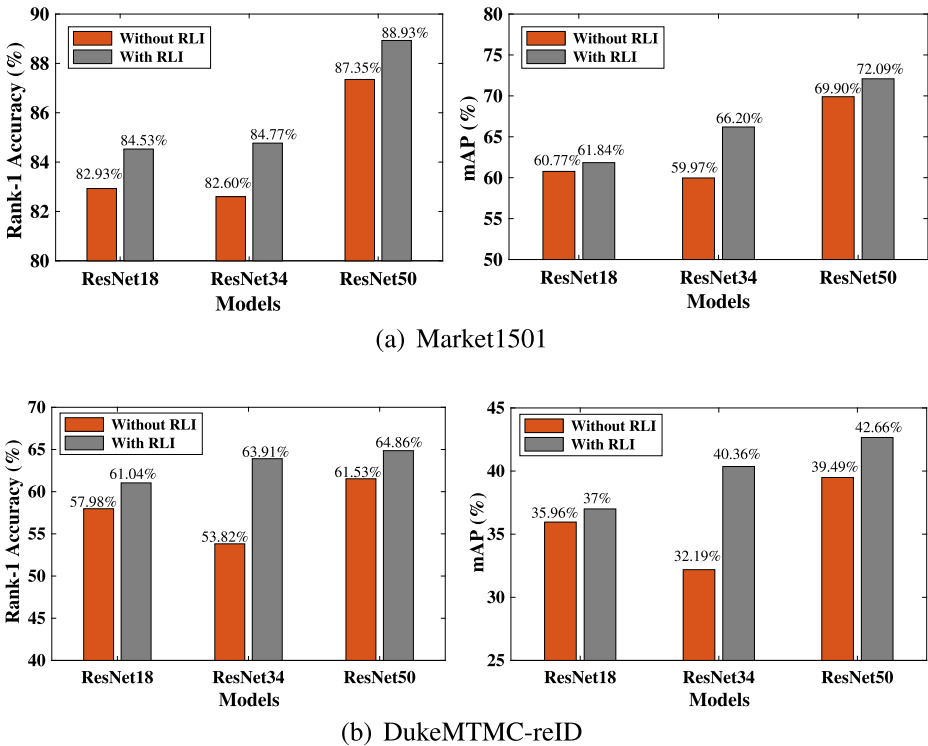


Fig. 4 We compare the performance of models with RLI and without RLI on Market1501 and DukeMTMC-reID respectively

4.4.2 Compared to superior data augmentation methods

The comparison results among different data augmentation methods including Random Cropping (RC), Random Erasing (RE) and Random Linear Interpolation (RLI) on Market1501 and DukeMTMC-re-ID, are shown in Table 2. To be fair, ResNet50 with the same structure and initialized parameters in Section 4.2.1 is used as the baseline method. Note that, random cropping is not adopted when evaluating performance of random erasing and Random Linear Interpolation. Furthermore, we combine the models with re-ranking [1, 40].

It can be observed that, all of data augmentation methods improve rank-1 accuracy and mAP of the baseline. On Market1501, both random erasing and Random Linear Interpolation show a better performance than random cropping. Specially, ResNet50 with RLI gives rank-1 accuracy of 88.93% and mAP of 72.09%. This is 1.05% higher in rank-1 accuracy and 1.31% higher in mAP than ResNet50 with RC. In addition, ResNet50 with RLI obtains 70.87% rank-1 accuracy and 63.69% mAP for DukeMTMC-reID with re-ranking. Compared to random erasing, the proposed data augmentation is 1.64% higher in rank-1 accuracy and 1.42% higher in mAP. Experimental results in Table 2 demonstrate that Random Linear Interpolation is an effective method for data augmentation and achieves a better performance than two commonly used methods, i.e., Random Cropping and Random Erasing.

Table 2 Person re-identification performance with Random Cropping (RC), Random Erasing (RE) and Random Linear Interpolation (RLI) on Market-1501 and DukeMTMC-reID based on the ResNet50 model. Random cropping is not adopted when evaluating performance of RE and RLI

Methods	Re-Ranking	Market1501				DukeMTMC-reID			
		Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
ResNet50	No	87.35	94.42	95.90	69.90	61.54	75.81	81.87	39.49
	Yes	90.53	94.74	96.23	85.09	66.56	77.96	83.39	58.41
ResNet50+RC	No	87.88	94.74	96.09	70.78	62.52	77.15	83.75	40.87
	Yes	90.59	95.49	96.58	85.86	67.95	79.80	84.82	61.08
ResNet50+RE	No	88.63	95.04	96.14	71.74	63.24	78.05	83.80	41.24
	Yes	91.00	95.64	96.97	86.67	70.02	80.79	85.49	62.97
ResNet50+RLI	No	88.93	95.16	96.53	72.09	64.86	78.46	83.98	42.66
	Yes	91.75	95.90	97.39	86.78	70.87	82.00	86.62	63.69

Furthermore, we also compare the proposed method with the Random Cropping method on three state-of-the-art baseline models including ResNet101, ResNet152 and DenseNet121. When performing experiments on ResNet101 and ResNet152, we set batch size to 16 because of memory limitations. Experiment results are shown in Table 3. The proposed method increases average rank-1 accuracy in three superior models by 2.34% and 6.32% on Market1501 and DukeMTMC-reID, respectively. In addition, it can be observed that our method has a greater performance improvement for DukeMTMC-reID dataset that is more difficult to accurately re-identify, which means that our methods can be applied to more difficult person re-identification tasks. We achieve 91.38% rank-1 accuracy and 76.89% mAP on the Market1501 dataset by DenseNet121, while achieving 74.01% rank-1 accuracy and 53.64% mAP on the DukeMTMC-reID dataset by ResNet152.

Table 3 Comparison with Random Cropping(RC) on three state-of-the-art baseline models including ResNet101, ResNet152 and DenseNet121

Methods	Re-Ranking	Market1501		DukeMTMC-reID	
		Rank-1	mAP	Rank-1	mAP
ResNet101+RC	No	82.28	61.66	66.59	48.29
	Yes	85.43	79.32	74.60	68.23
ResNet101+RLI	No	87.21	68.10	73.65	52.98
	Yes	89.92	85.49	81.07	75.88
ResNet152+RC	No	88.25	69.19	64.61	44.37
	Yes	90.99	86.59	72.62	64.33
ResNet152+RLI	No	89.06	72.04	74.01	53.64
	Yes	91.32	86.87	82.19	75.91
DenseNet121+RC	No	90.10	74.58	68.44	46.60
	Yes	91.97	86.98	76.08	69.34
DenseNet121+RLI	No	91.38	76.89	70.95	50.18
	Yes	92.71	88.98	77.97	70.99

Table 4 Comparison of our method with the state-of-the-art methods on the Market-1501 dataset. Rank-1 accuracy (%) and mAP(%) are shown

Methods	Market1501	
	Rank-1	mAP
Bow	34.4	14.09
BoW+KISSME	44.4	20.80
LOMO+XQDA	43.79	22.22
DNS	61.02	35.68
SOMAnet	73.9	47.9
PAN	82.8	63.4
Transfer	83.7	65.5
DML	87.7	68.8
IDE	72.54	46.00
SVDNet	82.3	62.1
DCGANs	83.97	66.07
TriNet	84.92	69.14
RE+SVDNet	87.08	71.31
RE+TriNet	83.94	68.67
Ours	91.38	76.89
Ours+Re-Ranking	92.71	88.98

4.4.3 Comparison with State-of-the-Art methods for person Re-ID

Many state-of-the-art methods deal with person re-identification efficiently on two benchmark datasets including Market1501 and DukeMTMC-reID. Both DukeMTMC-reID and Market1501 have been commonly used to evaluate the performance of superior methods recently. In Table 4, we show the compared results among our approach and other methods including [2, 9, 12, 24, 30, 35–39, 41] on Market1501. It can be seen that our method achieves competitive results with the state of the art. Specifically, our method obtains **rank-1 accuracy = 92.71%** and **mAP = 88.98%** for Market1501 with re-ranking. In addition,

Table 5 Comparison with prior art on DukeMTMC-reID. Rank-1 accuracy (%) and mAP (%) are shown

Methods	DukeMTMC-reID	
	Rank-1	mAP
BoW+KISSME	25.13	12.17
LOMO+XQDA	30.75	17.04
IDE	65.22	44.99
DCGANs	67.68	47.13
PAN	71.6	51.5
Ours	74.01	53.64
Ours+Re-Ranking	82.19	75.91

some superior methods including DCGANs [39], LOMO+XQDA [24], PAN [38], IDE [37] and BoW+KISSME [36] are compared to our method on DukeMTMC-reID. The detail scores of rank-1 accuracy and mAP are summarized in Table 5. In the compared methods, Random Linear Interpolation exceeds PANs [38] by 2.41% in rank-1 accuracy and 2.14% in mAP in ResNet152. Furthermore, the proposed method achieves **rank-1 accuracy = 82.19%** and **mAP = 75.91%** on DukeMTMC-reID with re-ranking.

5 Conclusion

In this paper, we proposed a novel data augmentation named Random Linear Interpolation for person re-identification. Pairs of images in training are mixed by interpolating the pixels of them. New samples with various levels of complicated occlusion are generated in training that can improve the robustness of convolutional neural network models. Experiments conducted on two benchmark datasets Market1501 and DukeMTMC-reID demonstrate that Random Linear Interpolation improves the performance of baseline CNN models for person re-identification. However, there are also many standard single-shot person re-identification datasets such as VIPeR and PRID450S with only one image per camera for each pedestrian, which poses a heavy problem for existing data augmentation methods. Therefore, we are going to optimize our method and explore a better data augmentation method based on generating mixed pedestrian background to solve the person re-identification problem of single-shot datasets in the future work.

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (Program No. 61702415, No.61502387), Natural Science Basic Research Plan in Shaanxi Province of China (Program No.2017JM6056,2016JQ6029) and Talent Support Project of Science Association in Shaanxi Province (Program No. 20180108).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bai S, Bai X (2016) Sparse contextual activation for efficient visual re-ranking. *IEEE Trans Image Process* 25(3):1056–1069
2. Barbosa IB, Cristani M, Caputo B, Rognhaugen A, Theoharis T (2017) Looking beyond appearances: synthetic training data for deep cnns in re-identification. *Computer Vision & Image Understanding*
3. Boroujeni FR, Wang S, Li Z, West N, Stantic B, Yao L, Long G (2018) Trace ratio optimization with feature correlation mining for multiclass discriminant analysis. In: *The AAAI conference on artificial intelligence*
4. Chang X, Nie F, Wang S, Yang Y, Zhou X, Zhang C (2016) Compound rank- k projections for bilinear analysis. *IEEE Transactions on Neural Networks & Learning Systems* 27(7):1502–1513
5. Chapelle O, Weston J (2011) Vicinal risk minimization. *Advances in Neural Information Processing Systems*, pp 416–422
6. Cheng D, Chang X, Liu L, Hauptmann AG, Gong Y, Zheng N (2017) Discriminative dictionary learning with ranking metric embedded for person re-identification. In: *Twenty-sixth international joint conference on artificial intelligence*, pp 964–970
7. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. *ACM Comput Surv* 40(2):1–60
8. Felzenszwalb P, Mcallester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. *Cvpr* 8:1–8

9. Geng M, Wang Y, Xiang T, Tian Y (2016) Deep transfer learning for person re-identification. *Computer Vision & Image Understanding*
10. Goodfellow IJ, Warde-Farley D, Mirza M, Courville A, Bengio Y (2013) Maxout networks. *Computer Science* pp 1319–1327
11. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. pp 770–778
12. Hermans A, Beyer L, Leibe B (2017) In defense of the triplet loss for person re-identification
13. Hirzer M, Roth PM, Bischof H (2012) Person re-identification by efficient impostor-based metric learning. In: *IEEE Ninth international conference on advanced video and signal-based surveillance*, pp 203–208
14. Hirzer M, Roth PM, Stinger M, Bischof H (2012) Relaxed pairwise learned metric for person re-identification. In: *European conference on computer vision*, pp 780–793
15. Huang G, Liu Z, Maaten LVD, Weinberger KQ (2017) Densely connected convolutional networks. In: *IEEE Conference on computer vision and pattern recognition*, pp 2261–2269
16. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. pp 448–456
17. Kodirov E, Xiang T, Fu Z, Gong S (2007) Person re-identification by unsupervised l_1 graph learning. *Hydrobiologia* 415(11):35–40
18. Kodirov E, Xiang T, Gong S (2015) Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In: *British machine vision conference*, pp 44.1–44.12
19. Kreuz-Delgado K, Murray JF, Rao BD, Engan K, Lee TW, Sejnowski TJ (2003) Dictionary learning algorithms for sparse representation. *Neural Comput* 15(2):349–396
20. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *International conference on neural information processing systems*, pp 1097–1105
21. Lei Z, Shen J, Liang X, Cheng Z (2016) Unsupervised topic hypergraph hashing for efficient mobile image retrieval. *IEEE Transactions on Cybernetics* PP(99):1–14
22. Li J, Wei Y, Liang X, Zhao F, Li J, Xu T, Feng J (2017) Deep attribute-preserving metric learning for natural language object retrieval. In: *ACM On multimedia conference*, pp 181–189
23. Li Z, Nie F, Chang X, Yang Y (2017) Beyond trace ratio: weighted harmonic mean of trace ratios for multiclass discriminant analysis. *IEEE Transactions on Knowledge & Engineering* PP(99):1–1
24. Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by local maximal occurrence representation and metric learning. In: *Computer vision and pattern recognition*, pp 2197–2206
25. Lisanti G, Masi I, Bimbo AD (2014) Matching people across camera views using kernel canonical correlation analysis. pp 1–6
26. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models
27. Roth PM, Hirzer M, K?stinger M, Beleznaï C, Bischof H (2014) Mahalanobis distance learning for person re-identification. Springer, London
28. Roth PM, Wohlhart P, Hirzer M, Kostinger M, Bischof H (2012) Large scale metric learning from equivalence constraints. In: *IEEE Conference on computer vision and pattern recognition*, pp 2288–2295
29. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Computer Science*
30. Sun Y, Zheng L, Deng W, Wang S (2017) Svdnet for pedestrian retrieval. In: *IEEE International conference on computer vision*, pp 3820–3828
31. Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *Publ Am Stat Assoc* 82(398):528–540
32. Torralba A, Fergus R, Freeman WT (2008) 80 Million Tiny Images: a Large Data Set for Nonparametric Object and Scene Recognition. *IEEE Computer Society*
33. Xiong F, Gou M, Camps O, Sznaiër M (2014) Person re-identification using kernel-based metric learning methods. In: *European conference on computer vision*, pp 1–16
34. Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network. *Computer Science*
35. Zhang L, Xiang T, Gong S (2016) Learning a discriminative null space for person re-identification. In: *Computer vision and pattern recognition*, pp 1239–1248
36. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: *IEEE international conference on computer vision*, pp 1116–1124
37. Zheng L, Yang Y, Hauptmann AG (2016) Person re-identification: past present and future
38. Zheng Z, Zheng L, Yang Y (2017) Pedestrian alignment network for large-scale person re-identification
39. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by gan improve the person re-identification baseline in vitro. pp 3774–3782
40. Zhong Z, Zheng L, Cao D, Li S (2017) Re-ranking person re-identification with k-reciprocal encoding. pp 3652–3661

41. Zhong Z, Zheng L, Kang G, Li S, Yang Y (2017) Random erasing data augmentation
42. Zhu L, Huang Z, Liu X, He X, Sun J, Zhou X (2017) Discrete multimodal hashing with canonical views for robust mobile landmark search. *IEEE Trans Multimedia* 19(9):2066–2079



Zhi Li He is currently pursuing the BS Degree in Northwest University, Xi'an, China. His main research interests include: Deep Learning and Video Analysis.



Jun Guo He is an associated Professor in the Department of Computer Science at Northwest University, China. He got his Ph.D. degree in computer science at Northwest University in 2007 and worked as a post-doctoral fellow in Tsinghua University during 2008-2010. His research interests are embedded computing, natural language processing and AI algorithm.



Wenli Jiao She is currently pursuing the doctoral degree in Northwest University, Xi'an, China. Her main research interests include: Wireless Network.



Pengfei Xu He is associated professor at School of Information Science and Technology, Northwest University in China. His main research interests include: image processing and pattern recognition.



Baoying Liu She received the B.A. degree in western culture from Xi'an Foreign Languages University, China, in 2001, the M.A. degree in western culture and history from Northwest University, Xi'an, China, in 2004, and the M.A. degree in cultural heritage preservation and tourism environment and the Ph.D. degree in cultural heritage from the University of Salento, Italy, in 2005 and 2012, respectively. She is currently with Northwest University. She is the talent introduced by Hundred-Talent Program.



Xiaowei Zhao He is currently pursuing the master's Degree in Northwest University, Xi'an, China. His main research interests include: Machine Learning and its application to multimedia and computer vision.