



**QUEEN'S
UNIVERSITY
BELFAST**

Data-Augmentation for Reducing Dataset Bias in Person Re-identification

McLaughlin, N., Martinez del Rincon, J., & Miller, P. (2015). Data-Augmentation for Reducing Dataset Bias in Person Re-identification. In *Proceedings of 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/AVSS.2015.7301739>

Published in:

Proceedings of 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Data-Augmentation for Reducing Dataset Bias in Person Re-identification

Niall McLaughlin, Jesus Martinez Del Rincon, Paul Miller
Centre for Secure Information Technologies (CSIT)
Queen’s University Belfast

{n.mclaughlin,j.martinez-del-rincon,p.miller}@qub.ac.uk

1. Abstract

In this paper we explore ways to address the issue of dataset bias in person re-identification by using data augmentation to increase the variability of the available datasets, and we introduce a novel data augmentation method for re-identification based on changing the image background. We show that use of data augmentation can improve the cross-dataset generalisation of convolutional network based re-identification systems, and that changing the image background yields further improvements.

2. Introduction

Person re-identification models the problem of recognising a person as they move through a non-overlapping camera network. In the most general case, the individual cameras will have different hardware, poor image quality, non-overlapping fields of view, will capture the person from different angles, with different pose, and with differing illumination. Due all these uncontrolled sources of variation, the current state-of-the-art rank 1 re-identification performance is around 30% [23] (or around 40% with fusion of multiple systems [24]) on the challenging Viper dataset, meaning there is still much scope for improvement in this area.

The standard method of evaluating a re-identification system’s performance is to use a publicly available dataset such as Viper [5], CAVIAR4REID [2], i-Lids [25], 3DPES [1], etc., which is split into disjoint training and testing sets. The system’s parameters are learned on the training set and performance reported on the testing set. If cross-validation is used, this procedure is repeated with different splits and the average performance over all the testing-sets reported. Although this procedure uses disjoint subsets of the dataset for training and testing, a problem arises when the dataset used is small and/or contains correlated data. In this case, a form of over-fitting may occur where the system’s parameters are overly tuned to the characteristics of a specific dataset, and hence the system does not generalise well to out-of-set examples. For instance, if the training dataset uses a very small set of cameras, always in the same



Figure 1. Name That Dataset! Can you name the above re-identification datasets? Answers shown in footnote¹ below.

locations, the re-identification system may infer the camera layout by recognising the scene backgrounds, which will not extrapolate well to other datasets or camera layouts. This problem is known as *dataset bias*, and has been reported in many areas of the computer vision literature. The root cause of dataset bias is that any finite dataset can only capture a small fraction of the variation present in the real visual world [21]. This problem can be exaggerated by capture bias, where the people building the dataset select images with a specific purpose in mind, thus further reducing the variability of the data. The problem of dataset bias is highlighted by the fact that some datasets are so visually distinctive that experienced researchers may find it easy to identify a specific dataset when presented with only a few images from it (See Fig. 1 where we recreate the *Name that Dataset!* game from [21] using re-identification datasets)¹.

It has been suggested that dataset bias could be addressed by collecting much larger datasets using the web as a source of large numbers of labelled images [13]. However, despite the creation of much larger datasets with millions of images [16], further sources of dataset bias exist, such as: *capture bias*, where humans prefer certain points of view when making photographs of objects, *label bias*, where cer-

¹Answers: 1.Viper 2.CAVIAR 3.i-Lids 4.ETHZ 5.CUHK 6.PRID

tain categories of object are poorly defined, and *negative set bias*, where the limited variety of objects in each dataset leads to poor discriminative ability in classifiers [21].

Given the inherent existence of dataset bias, even in large image collections, several approaches have been proposed to reduce its influence. The main idea behind most such methods is to combine multiple datasets while taking steps to account for differences in their data distributions. The method in [9] learns a set of weights, one for each individual dataset, and another set common to all datasets, to allow multiple training sets to be combined whilst encouraging the resultant classifier to have good generalisation ability. A related method learns an image feature representation that decomposes into two orthogonal subspaces, one common to all datasets and one for each individual dataset [20]. Recently, deep networks have been successfully applied to many vision problems, including re-identification. It has been proposed that the activations of the fully-connected layers of a deep network trained on a large object recognition problem could be used as generic image features [14]. However, it has been found [19] that these features are also susceptible to dataset bias, and that previous methods proposed for addressing dataset bias [9], did not work well when applied to deep-network based features. Finally, a neural network training method has been proposed, where the cost function encourages the network to solve the main task, while penalising the learned feature representation based on its ability to discriminate between datasets [4].

Dataset bias is starting to be recognised in the re-identification literature: The cross-dataset generalisation of a convolutional network based re-identification system was tested in [8] and found to be below that obtained when training and testing disjoint subsets of the same dataset. Improving generalisation may also become more important as re-identification is applied in more challenging conditions such as in the open-set scenario, where testing persons may or may not be present in the gallery. This means verification and identification must be carried out. In the open-set scenario, the performance of current re-identification systems was found to be insufficient for practical applications [12].

The issue of dataset bias may be a particular problem for person re-identification, as collecting a new dataset typically requires a large amount of manual labour: images of each person as seen in several cameras must be manually annotated. This can lead to small datasets (the largest re-identification datasets contain only a few thousand people [11]), or datasets without much variation in background, illumination, pose, clothing, angle of view etc.

In this paper we propose to increase the cross-dataset generalisation ability of re-identification systems, by employing data augmentation. The underlying idea of data augmentation is to increase the variability of existing datasets without the need to actually collect novel data.

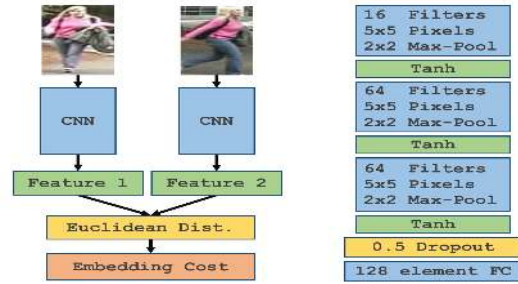


Figure 2. Left: Siamese Network architecture. Right: Structure and hyper-parameters of the convolutional networks (CNN).

Existing data are altered to better reflect the variability that can be encountered in out-of-set samples, for example, in speech-processing noise may be added to the training speech [7]. We propose to make better use of existing datasets by increasing their variability using data augmentation tailored to the re-identification problem. Our proposal will be tested using cross-dataset re-identification to evaluate its impact on dataset bias.

The contributions of this paper are: a novel data augmentation method based on changing the image background to artificially increase training-set size, and an evaluation of existing data augmentation methods in terms of their ability to improve cross-dataset re-identification performance.

3. Re-Identification System

The re-identification experiments in this paper were performed using a convolutional network [10], trained as part of a *Siamese network* architecture [6], similar to the approach used in [23]. This choice is supported by the fact that this architecture, as used by [23], provides one of the best rank one re-identification accuracies in the literature. The Siamese network architecture consists of two convolutional networks with identical weights. Each convolutional network, $G(x)$, transforms its input image, x , to a feature vector, \hat{x} . The Siamese network is trained using image pairs, (x_i, x_j) , of matched or mismatched images i.e., images belonging to the same or different persons. The embedding cost, $E(\hat{x}_i, \hat{x}_j)$, encourages the network to map pairs of matched images to feature vectors that are close, as measured by Euclidean distance, and to separate the feature vectors for pairs of mismatched images by at least a margin, m . The embedding cost function is defined as follows:

$$E(\hat{x}_i, \hat{x}_j) = \begin{cases} \frac{1}{2} \|\hat{x}_i - \hat{x}_j\|^2 & \hat{x}_i = \hat{x}_j \\ \frac{1}{2} [\max(m - \|\hat{x}_i - \hat{x}_j\|, 0)]^2 & \hat{x}_i \neq \hat{x}_j \end{cases}$$

Once trained, the convolutional network, $G(x)$, is used to map unseen person images to feature vectors, which can be compared using Euclidean distance to check if different feature vectors are likely to describe the same or dif-



Figure 3. Top Row: Data Augmentation based on linear transformations. Bottom Row: Original PRID450 image, the image’s foreground mask, and versions with simulated backgrounds.

ferent persons. In all experiments a three layer convolutional network was used, with Dropout [17] between the final convolutional, and fully connected (FC) layers. Images were represented by the activations of the 128 neurons in the FC layer. A fixed learning rate of 10^{-3} , and SGD optimisation for 500 epochs was used in all experiments. A diagram of the Siamese network architecture is shown in Fig. 2, which also illustrates the hyperparameters of the convolutional network. The hyperparameters were decided using a preliminary set of experiments on the Viper dataset, and with these settings the system achieves $\approx 33\%$ rank one accuracy (see Table 1), showing that it is a good baseline on which to validate our contributions.

4. Data Augmentation Methods

In this section we describe the specific data augmentation methods proposed (illustrated in Fig. 3). Each data augmentation method tries to simulate a type of variability present in real-world data, but which may not be well represented in the training set. By introducing these extra types of variability during training it is hoped that the network will learn to better generalise to out-of-dataset images, and hence performance in cross-dataset testing will improve.

4.1. Linear Transformations

In a realistic re-identification scenario the person may move with respect to the camera, their pose may change, and they may be viewed from a variety of angles. Additionally, small translations may occur due to errors in the person detection and extraction process, carried out prior to re-identification. Due to the small number of training examples typically available per-person, the above transformations may not be well represented by each person’s training images. We therefore augment the training set by introducing random transformations such as translation, mirroring, and rotation. Transformations may be used individually or combined together, and are applied each time a training im-

age is presented to the network during training.

Translation variability can be introduced by cropping the training image using a window with 90% of the image’s width and height, positioned uniformly at random without overlapping the image edge. Horizontal mirroring can be introduced using a Bernoulli random variable with $p = 0.5$ to indicate whether to perform mirroring. Small rotations can be applied by sampling uniformly at random from ± 5 degrees. Finally, an affine transformation combining scale, rotation and shear can be used, where scale is uniformly sampled from $\pm 5\%$ of image height, rotation uniformly sampled from ± 5 degrees, and shear uniformly sampled from ± 0.02 .

The above data augmentation approaches are computationally inexpensive and have been used in several other papers to improve the performance of convolutional network based object recognition systems [22]. However these methods have not previously been evaluated for their ability to improve generalisation for cross-dataset re-identification.

4.2. Colour

The brightness of images captured under different illuminations is easy to standardise during pre-processing, however when a person moves between different cameras, or between areas with different illumination sources, the colour of the illumination (white-balance) may vary. Accurately compensating for changing illumination colour using pre-processing is difficult. While many cameras attempt to compensate by changing their white balance, it is often not possible to completely correct for this effect, or to synchronise the information with other cameras in the network, and in any case the camera settings may be incorrect. Shifting white-balance leads to a change in the measured colour values of a person as seen from different viewpoints and/or different cameras. We simulate this effect by randomly shifting the hue of the training images by small amounts each time they are presented to the network during training. We convert the image to the HSV colour-space then modify the hue channel by adding a value, sampled uniformly at random from $[-0.1, 0.1]$, to the hue value of all pixels. The image is then converted back to RGB to allow other data augmentation methods to be applied.

4.3. Background Substitution

A further way to significantly increase the variability of the training data is by changing the image background while leaving the foreground i.e., the person, unchanged. Modifying the training images in this way should help the re-identification system to learn to discriminate between the different sources of image variability. The system should learn to focus on the parts of the images belonging to the subjects, and image features that are strongly related to

identity, while learning to ignore irrelevant sources of variability. To implement this idea the person must be accurately segmented from the image background, which can be accomplished using, for example, a deformable parts model [3] to extract the positions of the body parts, or by using background subtraction if video is available [18]. Once the person has been segmented from their original background, a simulated background can then be substituted from a corpus of realistic background images.

5. Experiments

In this section we first demonstrate experimentally that dataset bias exists. We then test the effectiveness of various data augmentation methods for improving cross-dataset generalisation performance.

5.1. Name that Dataset!

If dataset bias did not exist i.e., if all re-identification datasets contained a fair random sampling of person images from the visual world, it would not be possible to train a classifier to distinguish between images from different datasets at levels significantly different from chance. Following the approach used in [21], where the authors train a classifier to play *Name that Dataset!*, but translating it to the re-identification scenario, we show that it is possible to train a classifier to recover the parent dataset of a given image, and hence we show that dataset bias may be an issue affecting current re-identification datasets.

For this experiment, all images from the Viper, CAVIAR, i-Lids, and 3DPES datasets were combined into a single dataset. During testing and training the identity label of each image was discarded, and images were instead labelled with their parent dataset. The combined dataset was then split into 80% for training and 20% for testing, taking care to ensure there was no overlap between the persons or images included in the training and testing sets. To reduce concerns with over-fitting, the same convolutional network design and hyper-parameters were used as in all other experiments. However, instead of using a Siamese architecture, a standard feed-forward convolutional network architecture was used, with a softmax classification layer for prediction.

The results of this experiment are shown as a confusion matrix in Fig. 4, and the overall dataset identification accuracy was 96.2%. Note there is very little energy in the off-diagonal elements of the confusion matrix, showing that the re-identification datasets are very distinctive as they are not often confused by the classifier. These results indicate that dataset bias may be a significant issue, as only roughly 25% classification accuracy would be expected if the classifier was operating at chance levels, as would happen if the images from each dataset were fully unbiased samples from the visual world.

		Prediction			
		Viper	iLids	Caviar	3DPES
Test	Viper	0.96	0.02	0.02	0
	iLids	0.03	0.91	0	0.06
	Caviar	0.02	0	0.98	0
	3DPES	0	0.01	0	0.99

Figure 4. Confusion Matrix for a classifier trained to predict the parent dataset of a given re-identification image.

5.2. Cross-Dataset Generalisation

The existence of dataset bias, as demonstrated in Section 5.1, means that a re-identification system trained using only one dataset will likely not generalise well to different datasets. We show this directly by performing a cross dataset re-identification experiment. The Viper dataset was randomly split into disjoint subsets of 50% for training and 50% for testing. A re-identification system was then trained using the training subset and achieved a 33% rank one identification accuracy on the testing subset. Separate re-identification systems were then trained using 100% of the i-Lids, CAVIAR, and 3DPES datasets, and used to perform re-identification on the testing subset of viper. It can be seen from the results in Table 1 that the cross-dataset re-identification accuracies are significantly below the accuracy of the within-dataset system. This difference in accuracy is a symptom of dataset bias i.e., the features learned on the i-Lids, CAVIAR and 3DPES datasets did not generalise well to Viper.

Training Dataset	Viper	i-Lids	CAVIAR	3DPES
Viper Rank 1 (%)	33	11	12	14

Table 1. Comparing within-dataset and cross-dataset rank 1 re-identification system accuracies (%) tested on the Viper dataset.

To reduce the impact of dataset bias we propose to use data augmentation to increase the diversity of the training samples. In this experiment we quantify the effect of the various data augmentation methods on cross-dataset re-identification performance using the Viper, CAVIAR, i-Lids, and 3DPES datasets. All combinations of the different data augmentation methods were tested, however we only report a subset of the more interesting results due to space constraints. During each experiment the network was trained using all images from all persons in one dataset. Each time an image was presented to the network it was randomly augmented using one or more simultaneous data augmentation methods. After training with a particular dataset, the network was then used to perform re-identification on all the remaining datasets except the one on which it had been trained. For each combination of data augmentation

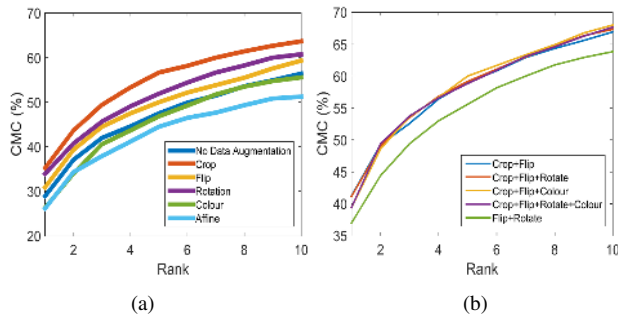


Figure 5. Data augmentation results: (a) When data augmentation methods are used alone. (b) Top five best performing combinations of data augmentation methods. (Note - Best viewed in colour).

methods, the CMC curves were averaged over all training and testing datasets, in order to allow the effect of the data augmentation methods to be separated from the specifics of their performance on particular datasets. Note that re-identification testing was conducted using all persons in each dataset (as opposed to 50% of persons as is usually the case for within-dataset scenarios, when individual datasets are split into training and testing sets). A summary of the results is shown in Fig. 5.

To understand the individual contribution of each data augmentation method, we compare performance with no data augmentation to the cases where each data augmentation method is used individually. The results, in Fig. 5(a), show that cropping gives the largest individual improvement in performance, followed by rotation and mirroring. However, both colour and affine transformations decrease performance compared to the baseline. This suggests that these augmentations, when used alone, are not representative of transformations present in realistic data.

Next we investigate which combinations give the best overall performance, as it may be the case that some methods are complimentary, or work better when used together. The results, seen in Fig. 5(b), show that very similar results are obtained by several different combinations. The highest rank one performance is obtained using cropping and mirroring, while the highest rank 10 performance is obtained using cropping, mirroring, and colour augmentation. The question of which rank is most important to optimise will be application dependent, however in this case all the top combinations achieve similar performance across a wide range of CMC values. We also note there is a large drop in performance when mirroring and colour data augmentation are used without cropping, which shows the importance of cropping for obtaining good performance improvements.

5.3. Background Substitution

As mentioned in Section 4.3 we can further increase the diversity of the training data by varying the image backgrounds to simulate a larger training-set where each person is recorded in many different environments. This experiment was only carried out using the PRID 450 [15] dataset, as it includes annotation to allow segmentation of each person from their background. A set of ten diverse empty street scene images were collected from the web for use as a background image corpus. During training, each time a person’s image was presented to the network, a background image was randomly selected from the corpus, and a window within the image was then selected uniformly at random without overlapping the edges. This background window was then substituted for the training image’s original background using the foreground mask provided with the PRID 450 dataset. See Fig. 3 for an illustration of images with simulated backgrounds.

In this experiment we compare two data augmentation scenarios: firstly, with cropping, mirroring, rotation, and colour data augmentation enabled, and secondly, with background substitution also enabled. For each scenario the network was trained using all the persons in the PRID 450 dataset, and then tested on the Viper, i-Lids, CAVIAR, and 3DPES datasets, and the CMC curve results averaged over all the testing datasets. The results in Fig. 6(a) show that use of background substitution gives an improvement in performance over using only the standard data augmentation methods. The rank one CMC, averaged over all the testing datasets, increases from 40.25% to 44.75% when simulated backgrounds are used compared to using only the standard data augmentation methods.

In the second part of this experiment we investigate how performance varies as the number of training images with artificial backgrounds is varied. For each person, a fixed number of training images with simulated backgrounds was generated offline. The system was then trained using this fixed set of images with simulated backgrounds, with cropping, mirroring, rotation and colour data augmentation enabled. The trained system was then tested on all other datasets and the average CMC curve calculated. The results in Fig. 6(b) shows how the rank one CMC accuracy varies as a function of the number of training images with simulated backgrounds. It can be seen that performance quickly improves with additional images, but starts to experience diminishing returns when around 16 additional background images per-person are used.

6. Conclusion

In this paper we have introduced a novel data augmentation method for re-identification based on changing the image background. We have shown that this can increase

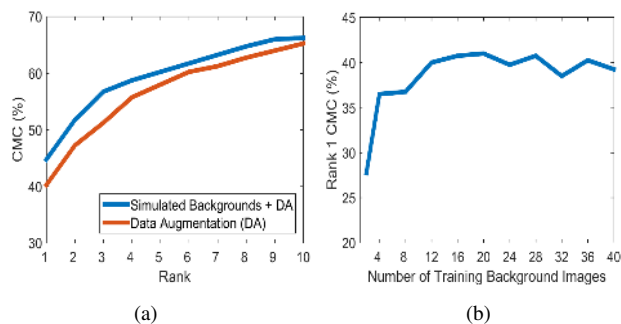


Figure 6. (a) Comparing performance with simulated background and standard data augmentation, to standard data augmentation alone. (b) Average rank-1 CMC as the number of PRID450 training images with simulated backgrounds is varied.

cross-dataset performance. We have also evaluated several existing data augmentation methods for their ability to improve cross-dataset re-identification. We find that geometric transformations, such as cropping and mirroring can significantly increase cross-dataset performance. We recommend that future work in this area should report cross-dataset re-identification accuracy as this should give a better indication of real world performance that within-dataset testing.

References

- [1] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dps: 3d people dataset for surveillance and forensics. In *Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects*, pages 59–64, Scottsdale, Arizona, USA, Nov. 2011.
- [2] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011.
- [3] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.
- [4] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.
- [5] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV 2008*, pages 262–275. Springer, 2008.
- [6] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, volume 2, pages 1735–1742. IEEE, 2006.
- [7] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, et al. Deepspeech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.
- [8] Y. Hu, D. Yi, S. Liao, Z. Lei, and S. Li. Cross dataset person re-identification. In C. V. Jawahar and S. Shan, editors, *ACCV 2014 Workshops*, volume 9010 of *Lecture Notes in Computer Science*, pages 650–664. Springer International Publishing, 2015.
- [9] A. Khosla, T. Zhou, T. Malisiewicz, A. A. Efros, and A. Torralba. Undoing the damage of dataset bias. In *ECCV 2012*, pages 158–171. Springer, 2012.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [11] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *ACCV 2012*, pages 31–44. Springer, 2013.
- [12] S. Liao, Z. Mo, Y. Hu, and S. Z. Li. Open-set person re-identification. *arXiv preprint arXiv:1408.0872*, 2014.
- [13] J. Ponce, T. L. Berg, M. Everingham, D. A. Forsyth, M. Hebert, S. Lazebnik, M. Marszalek, C. Schmid, B. C. Russell, A. Torralba, et al. Dataset issues in object recognition. In *Toward category-level object recognition*, pages 29–48. Springer, 2006.
- [14] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *CVPRW*, pages 512–519. IEEE, 2014.
- [15] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznaï, and H. Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, pages 247–267. Springer, 2014.
- [16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 2015.
- [17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, 15(1):1929–1958, 2014.
- [18] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2. IEEE, 1999.
- [19] T. Tommasi, N. Patricia, B. Caputo, and T. Tuytelaars. A deeper look at dataset bias. *arXiv preprint arXiv:1505.01257*, 2015.
- [20] T. Tommasi, N. Quadrianto, B. Caputo, and C. H. Lampert. Beyond dataset bias: Multi-task unaligned shared knowledge transfer. In *ACCV 2012*, pages 1–15. Springer, 2013.
- [21] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In *CVPR*, pages 1521–1528. IEEE, 2011.
- [22] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun. Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876*, 2015.
- [23] D. Yi, Z. Lei, and S. Z. Li. Deep metric learning for practical person re-identification. *arXiv preprint arXiv:1407.4979*, 2014.
- [24] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *Computer Vision and Pattern Recognition (CVPR)*, pages 144–151. IEEE, 2014.
- [25] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *BMVC*, volume 2, page 6, 2009.