Convolutional Neural Network Baseline Model Building for Person Re-Identification

Assoc. Dr. *Llahm Omar Faraj Ben Dalla
Sebha Technical University,
Department of Software Engineering, Faculty of Computer Science, Libya
Self-Lanser Group Research Center (SLGRC) Computer Department.
Tripoli/LIBYA.
E-mail: mohmdaesed@gmail.com
E-mail: selflanser@gmail.com
Phone: +218945780716

Abstract

Nowadays, there has been a wide range of critical issues that made the person re-identification task as a challenging point, these issues include human pose variation, human body occlusion, camera view variation, etc. To overcome these issues, most of the modern approaches are proposed based on deep convolutional neural network (CNN) (DevOps). This research article, sheds light on how to utilize a pre-trained CNN models which were developed for image recognition, to create a powerful CNN baseline model that could be utilized for the term of person re-identification. To build such a powerful model, this study has proposed to adjust the architecture of the CNN model by adding batch normalization and dropout layers to the classifier part of the CNN to prevent an over-fitting and re-train it with the available dataset. Then this research study has utilized cosine similarity to calculate the resemblance among the extracted features. The extensive experiments conducted on the proposed CNN baseline model utilizing the three large and well-known standard re-identification datasets to validate the performance of the proposed method, proved that the proposed approach could be associated with the state-of-the-art techniques (DevOps). Additionally, the results of this beneficial study are important for several domains such as the industrial world, the educational world as well as the scientific world in addition to researchers who aimed for some investigations outcome based on deep convolutional neural network (CNN) (DevOps).

***Keywords*:** *Computer vision, Person re-identification, Deep learning, convolutional neural network.*

INTRODUCTION

Recently, the process of re-identification person's identity has received more and more attention, having become a key point in intelligent surveillance frameworks as well as has wide application possibilities in many field practices. When you give a picture of an individual taken from single camera, the task remains to identify that an individual from the gallery taken via other multiple cameras. It's a difficult task as an outcome of critical issues of the variation on human pose and camera view, and human body occlusion. Nowadays, the advance in deep convolutional neural network (CNN) (DevOps) architecture [1-5] utilized to build efficient an individual re-identification models practices. In fact, CNN's (DevOps) main superiority is that, it can optimize the procedures of the requirements of the feature during the extraction, metric knowledge of learning as well as classification jointly in end-to-end training fashion [6].

At the current stage, most of the modern an individual reidentification approaches [1, 7, 8] were proposed based on CNN (DevOps). Generally, fine-tuning the CNN model that pre-trained on ImageNet [9] under supervision of SoftMax loss usually serves as the baseline paradigm. Investigation on building effective CNN baseline model is urgently in demand to benefit an individual re-identification research area to large extent, both from the academic and implementation frame of reference.

Regarding to this research study, the deep an individual re-identification techniques architecture designed based on the pre-trained model for ImageNet (DevOps) problem via adjusting the architecture of the CNN model and re-train it with available dataset. This study has proposed the following key changes: Adding a fully-connected layer after the adaptive pooling layer and to prevent over fitting follow it via a batch normalization layer as well as a dropout layer, then employ stochastic gradient descent (SGD) (DevOps) [10] as the optimizer for CNN training (DevOps). The feature vector that output from the adaptive pooling is utilized as the image representation..

This experimental research article remains structured as follows: Section II has reviewed some related previous studies. In section III, the definition of an individual re-identification model architecture structure in this study has described. The execution details of the experiments implementation remain provided in section IV. In Section V, this study has presented the experimental outcomes on three large and well-known an individual re-identification datasets. This study has concluded in section VI.

- **Related works**

Conventionally, hand-crafted feature designed based on color histogram come out on top in person re-identification [11-13], due to the consideration that color of clothes has good discriminability for tell the difference between an individuals. Recent researches on an individual re- identification mostly focus on building deep CNN models (DevOps) in the end-to-end learning fashion. Zheng, et. al. in [1] takes advantage of the deep convolutional models pre-trained on ImageNet [9] as well as fine-tunes it on an individual re-identification datasets utilizing softmax loss. The features that produced via the final pooling layer was utilized as an image descriptor. The learned representation achieves great performance boost against traditional hand-crafted feature. In fact, due to the favorable outcome of [1], most of the modern methods that based on deep learning technique also adopt pre-trained models as backbone network as well as have been searching other technical means to further rising the performance of re-identification framework. Therefore, for most present approaches [7, 14, 15] feature learned utilizing only softmax loss commonly be in the service of as a baseline for comparison.

Various model architectures have been explored in an individual re-identification area. Between existing an individual re-identification approaches, ResNet-50 [2] is the most commonly utilized backbone network [14, 15]. Besides, GoogLeNet [5, 16], inception networks [16] and Densenet [17] have also been chosen as backbone network via some researchers. Taking advantage of the pre-trained CNN models, via further employing metric learning methods [18], utilizing part-based CNN representation [19], otherwise carefully designing attention mechanism [20], and an individual re-identification performance could be further improved.

To prevent over-fitting for deep CNN models (DevOps) when trained on relatively small datasets, several approaches have been proposed. In a specification, random cropping [3], random flipping [4], and random erase operation [21] are commonly utilized data augmentation methods in training deep CNN model. In addition, regularization methods like weight decay is also a well-known approach for prevent over-fitting. In recent, batch normalization [22] and dropout [23] are two widely utilized tricks for training CNN and have shown benefits for preventing over-fitting. Dropout aims through the training process

randomly with a probability, to discards the output of each hidden neuron. Batch normalization aims at reducing internal co-variate shift via normalizing the output of each hidden neuron utilizing minibatch mean and variance. Since an individual re-identification dataset are relatively small, for instance, Market1501 containing only 12,936 images for training, effective means for preventing over-fitting is necessary for building high-accuracy an individual re-identification model.

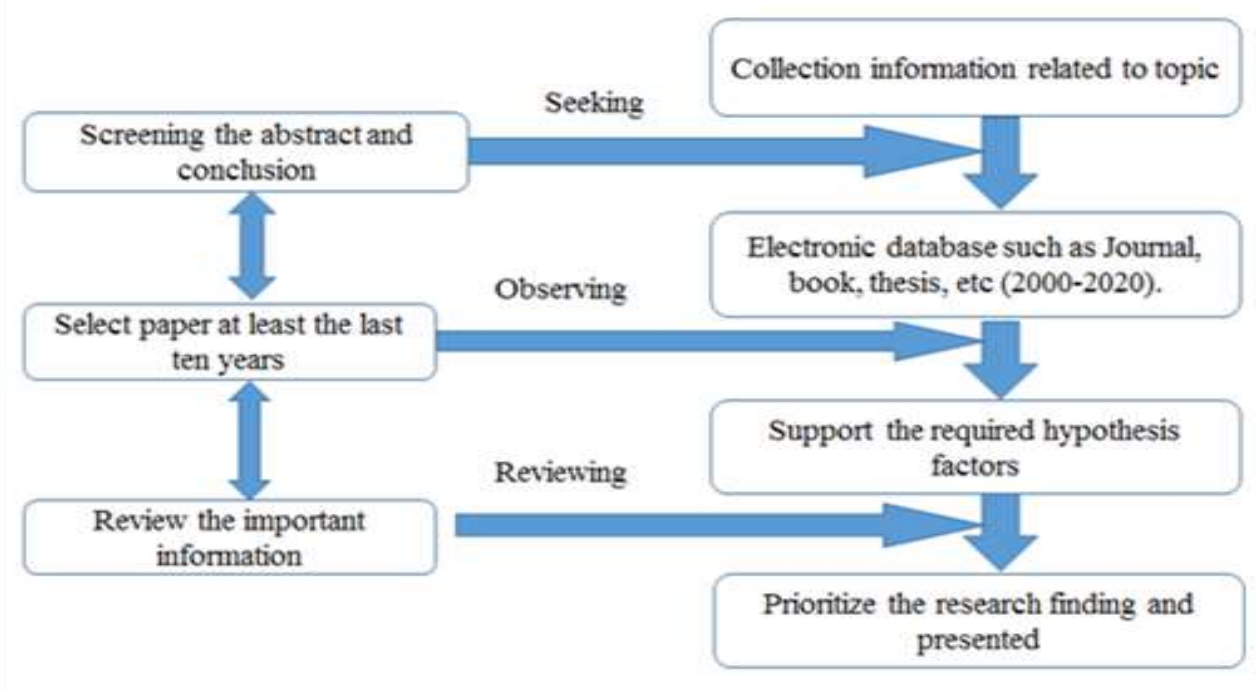


Figure.1. The required research method that used to review the main research factors and related works from the electronic database.

Zheng, et. al. in [1] takes advantage of the deep convolutional models pre-trained on ImageNet [9] as well as fine-tunes it on an individual re-identification datasets utilizing softmax loss. Furthermore, the features that produced via the final pooling layer was utilized as an image descriptor. Moreover, the learned representation achieves great performance boost against traditional hand-crafted feature. In fact, due to the favorable outcome of [1], most of the modern methods that based on deep learning technique also adopt pre-trained models as backbone network as well as have been searching other technical means to further rising the performance of re-identification framework. Therefore, for most present approaches [7, 14, 15] feature learned utilizing only softmax loss commonly be in the service of as a baseline for comparison.

Various model architectures have been explored in an individual re-identification area. Between existing an individual re-identification approaches, ResNet-50 (DevOps) [2] is the most commonly utilized backbone network [14, 15]. Besides, GoogLeNet (DevOps) [5, 16], inception networks [16] and Densenet [17] have also been chosen as backbone network via some researchers. Taking advantage of the pre-trained CNN models (DevOps), via further employing metric learning methods [18], utilizing part-based CNN representation [19], otherwise carefully designing attention mechanism [20], an individual re-identification performance could be further improved..

To prevent over-fitting for deep CNN (DevOps) models when trained on relatively small datasets, several approaches have been proposed. In a specification, random cropping [3], random flipping [4], and random erase operation [21] are commonly utilized data augmentation methods in training deep CNN model (DevOps). In addition, regularization

methods like weight decay is also a well-known approach for prevent over-fitting. In recent, batch normalization [22] and dropout [23] are two widely utilized tricks for training CNN (DevOps) and have shown benefits for preventing over-fitting. Furthermore, dropout aims through the training process randomly with a probability, to discards the output of each hidden neuron. Furthermore. batch normalization aims at reducing internal co-variate shift via normalizing the output of each hidden neuron utilizing minibatch mean and variance. In addition, since an individual re-identification dataset are relatively small, for instance, Market1501 containing only 12,936 images for training, effective means for preventing over-fitting is necessary for building high-accuracy an individual re-identification model.

The proposed Technique

According to this segment which is describes the research proposed approach towards building a powerful CNN based model that can be utilized for an individual re-identification domain. The pipeline of the proposed baseline model is exhibited in Figure.1 below. It is possible to take any CNN model designed for image classification, remove its hidden fully connected layers, and it can be utilized as the backbone, for instance, Google Inception [16] as well as ResNet [2]. As declared via this research paper utilizes the ResNet50 [2] model, taking into account its competitive performance and its comparatively concise building style. The ResNet-50 CNN model was fine-tuned utilizing the classification framework in [1] in order, to transfer its knowledge to an individual re-identification domain. As presented in Table.1 below the summarization of the ResNet-50 that is commonly utilized in an individual re-identification.

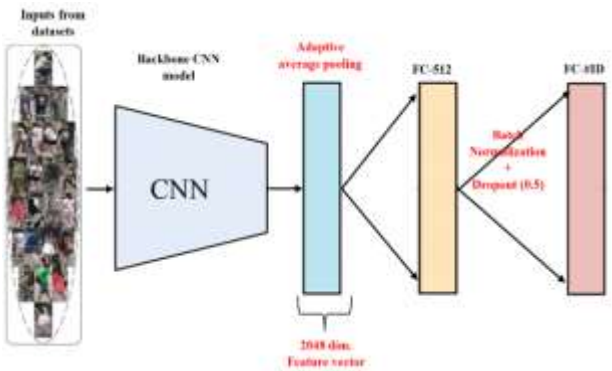


Figure.2: The pipeline of the studied CNN-based (DevOps) an individual re-identification framework.

Table.1: The ResNet-50 summary and specifications.

CNN	ResNet-50
Release year	2015
# Layers	50
top-5 error rate	3.6%
# parameters	25.5 M
Image Input Size	224- via -224

Taking ResNet-50 (DevOps) as the backbone, this study has adjusted its classifier architecture as following:

- Replace pool5 layer via adding an adaptive average pooling layer that can be utilized instead of the flatten layers that traditionally utilized in CNN models. This layer will output a 2048-dimensional requirement feature vector.
- After the pooling layer, a 512-dimensional fully connected layer is appended and then it followed via a Batch Normalization layer as well as a dropout layer sequentially. Therefore, the drop percentage of the dropout layer remains set to 0.5.
- Finally, another fully connected layer remains appended with a SoftMax loss as well as dimension based on an individual ID labels for training dataset, for instance, Market-1501 contains 751 ID labels.

• Experiments

• *Datasets Settings*

In this study, experiments conducted on the three widely utilized an individual re-identification datasets, including Market-1501 [24], as well as DukeMTMCreID [25, 26], in addition to CUHK03 [27].

Market-1501 includes 32,668 images of a total of 1501 IDs that were taken from 6 cameras positioned in front of a supermarket on campus. The Deformable Part Model (DPM) [28] was utilized for recognizing and cropping the images. The Market-1501 dataset remains divided into three portions: the first part contains 12,936 images with 751 IDs utilized for training purpose, the second part contains 19,732 images with 750 IDs dedicated for the gallery, as well as the last part contains 3,368 images with the same 750 gallery IDs utilized for the purpose of querying. 64 × 128 is the size of all images in this dataset.

DukeMTMC-reID dataset includes 36,411 images of a total of 1,812 IDs which are taken from 8 diverse standpoints. In the same way as Market-1501, this dataset remains also divided into three portions: the first term contains 16,522 pictures with 702 IDs utilized for training purpose, the second part contains 17,661 pictures with 1,110 IDs dedicated for the gallery, as well as the last part contains 2,228 images with 702 IDs utilized for the purpose of querying. In addition, images in DukeMTMC-reID dataset remain vary in size.

CUHK03 dataset includes 14,096 images of a total of 1,467 IDs. These IDs were taken via 2 cameras in the campus of CUHK University. The CUHK03 dataset offers two kinds of labeled images, one was manually labeled as well as the other detected via DPM. Experiments in this work has done on the images that detected via the DPM. Originally, the CUHK03 valuation protocol exactly include 20 train/test splits. However, for the purpose of compatibility with both Market-1501 as well as DukeMTMC-reID, [34] the protocol for training and testing that proposed in [29] remains utilized in this research study. That is 7,365 images containing 767 IDs utilized for training purpose, 5,332 images containing 700 IDs dedicated for the gallery, as well as 1,400 pictures containing the same 700 IDs as in gallery are utilized for the purpose of querying. CUHK03 dataset images are also vary in size.

The details of these three datasets are listed in Table.2.below as well as some pictures as example are exhibited in Figure.2.below.

• ***Evaluation Metrics***

For performance measurement, this study has adopted the top classification accuracy on validation data for image classification, the cumulative matching curves (CMC), as well as mean average precision (mAP). CMC is utilized to evaluate the performance of re-identification problem, this paper, reports the rank 1, 5, as well as 10 accuracy on CMC rather than plotting the actual curves for easier comparison with published outcomes. Meanwhile, following the existing works [24, 25, 34], this study also has utilized mAP to measure the

performance of individuals retrieval. For each query, the percentage precision is measured from the curve of precision-recall. mAP remains then computed as the mean value of percentage precisions across all queries. Fundamentally, CMC indicates the retrieval precision, while mAP indicates the recall.

Table.2. The details of datasets.

Datasets		Market-1501	DukeMTMC-reID	CUHK03
All	# images	32,668	36,411	14,096
	# IDs	1,501	1,812	1,467
Training	# images	12,936	16,522	7,365
	# IDs	751	702	767
Gallery	# images	19,732	17,661	5,332
	# IDs	750	1,110	700
Query	# images	3,368	2,228	1,400
	# IDs	750	702	700
# cams		6	8	2

Figure.3. The Market-1501, DukeMTMC-reID as well as CUHK03 datasets images samples.

This layer produces a 1x2048 dimensional feature requirements vector. For every dataset, this study has fed all the testing images in gallery as well as query to the trained CNN model to obtain a 2048-dimensional of an individual descriptor for each image and they are stored offline to be utilized in evaluation scenario.

Evaluation. In the evaluation phase, once the descriptors for the gallery and query images are obtained, this study compute the cosine distance amongst the query features as well as those of the gallery to measure CMC rank-1, 5, as well as 10 accuracy as well as the mAP which are utilized for re-identification evaluation.

- **Experimental Results**

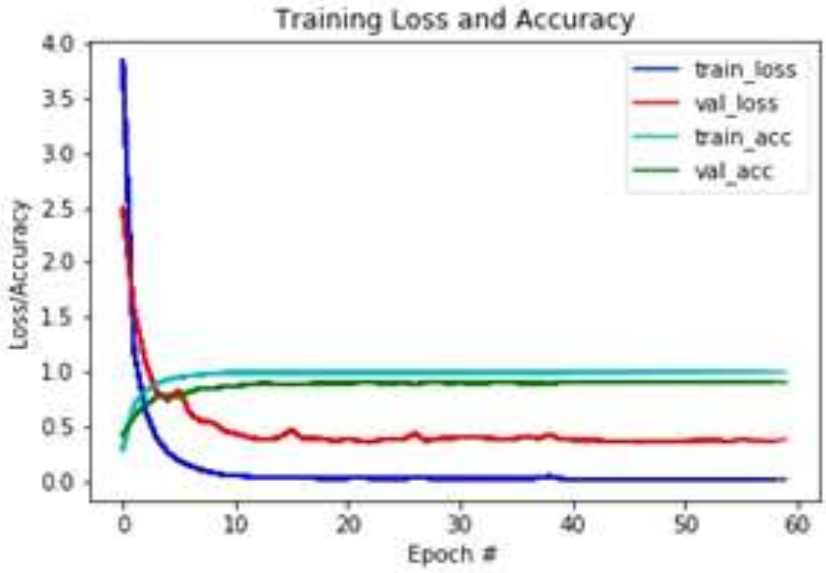
To verify the proposed model, this study has conducted experiments on three large and well-known datasets Market-1501[34], DukeMTMCreID as well as CUHK03. As an outcome, in this section this study has reported the training classification accuracy an outcome. It reported also that the CMC rank 1, 5, besides 10 accuracy as well as the mAP for the three datasets. In fact, all experiments have done utilizing a single query.

Classification Accuracy Results. The overall training accuracy along with validation accuracy and loss for each dataset are presented in figure. 3 below. As well as outcomes of the classification accuracy on validation sets are exhibited in Table.3.below.

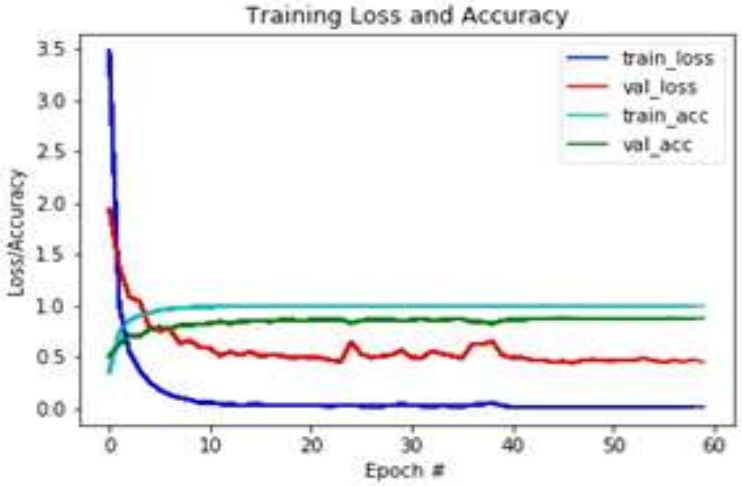
The outcomes presented that the performance of the proposed model on Market-1501, DukeMTMCreID, as well as CUHK03 datasets are remarkably good, as shown in Table.3.below, this study has obtained 91.21%, 88.32%, and 94.00% validation accuracy on the three datasets respectively. Therefore, the proposed model of this experimental study can effectively reduce the impact of over-fitting and can work well for an individual re-identification domain via producing satisfactory discriminative individual's features requirements from the image.

Re-identification Evaluation Results. To validate the effectiveness of this study approach for an individual re-identification task, we report the CMC rank 1, 5, as well as 10 accuracy as well as the mAP for all the three datasets, and compare the model that proposed in this study directly to the state-of-the-art an individual re-identification approaches that utilize the pretrained ResNet-50 CNN model as backbone model. Outcomes on the three datasets are exhibited in the Table. 4 below.

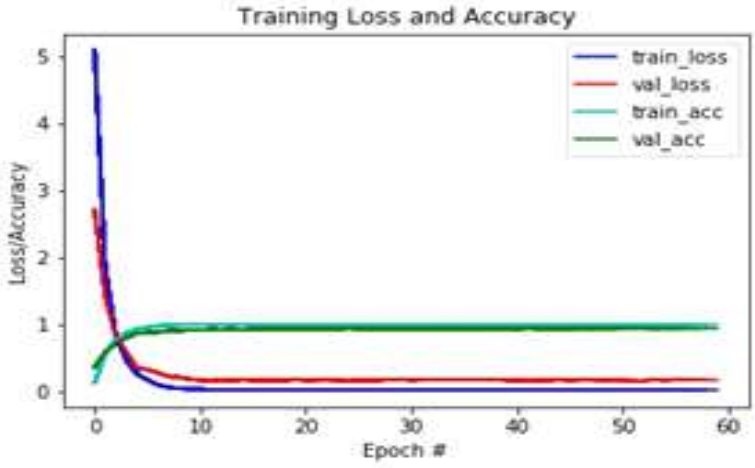
The state-of-the-art methods compression. The outcomes of the model proposed in this study were compared with the state-of-the-art techniques in terms of mAP and rank-1 accuracy on the three datasets Market1501, DukeMTMCreID, as well as CUHK03 as exhibited in Table.5.below. This study model achieved 87.7% rank-1 accuracy besides 70.5% mAP on Market1501, that remains comparable to the state-of-the-art 89.5% rank-1 accuracy as well as 71.6% mAP in [31] and exceeds the other state-of-the-art algorithms. On DukeMTMCreID, the proposed model yielded 78.1% rank-1 as well as 61.1% mAP as well as exceeds the state-of-the-art performance. On CUHK03 dataset, the proposed model yielded 43.7% rank-1 and 40.2% mAP, which is far a little bit from the state-of-the-art 54.3% rank-1 accuracy as well as 50.1% mAP [31], however, it still exceeds other state-of-the-art algorithms. This demonstrates the effectiveness as well as generality of the proposed model. To the best of this study platform, regarding to this research approach is the simplest one able to achieve the state-of-the-art performance.



(a)



(b)



(c)

Figure.4 The model accuracy and loss outcomes on (a) Market-1501, (b) DukeMTMCreID, as well as (c) CUHK03 datasets

Table. 3. The model of validation accuracy

Dataset	Validation accuracy
Market-1501	91.21%
DukeMTMCreID	88.32%
CUHK03	94.00%

Table. 4. The model re-identification evaluation outcomes

Dataset	Rank-1	Rank-5	Rank-10	mAP
Market-1501	87.7	95.4	96.9	70.5
DukeMTMCreID	78.1	88.5	92.3	61.1
CUHK03	43.7	63.6	72.6	40.2

Re-identification outcomes visualization. As presented in Figure. 4 below the exhibition of some re-identification outcomes samples on Market1501, DukeMTMCreID, as well as CUHK03 datasets. The query images have presented in the first column. The retrieved pictures remain arranged from the left to right according to the scores of similarities with the query. It could be noted that the model proposed in this study is reasonably strong to fetch the right retrieval outcomes as well as go wrong in several very hard situations that remain challenging due to that it shares one or more salient features with the query image

Table. 5. The comparison with the state-of-the-art approaches. The rank-1 accuracy as well as mAP are listed

Techniques	Market-1501		DukeMTMCreID		CUHK03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
IDE [1]	73.9	47.8	65.2	45.0	21.3	19.7
Re-rank [29]	77.1	63.6	-	-	34.7	37.4
SVDNet [14]	82.3	62.1	76.7	56.8	41.5	37.3
IDE+DaF [15]	82.3	72.4	-	-	26.4	30.0
SSM [31]	82.2	68.8	-	-	-	-
DaRe [32]	86.4	69.3	74.5	56.3	54.3	50.1
CamStyle+RE [33]	89.5	71.6	78.3	57.6	-	-
Ours	**87.7**	**70.5**	**78.1**	**61.1**	**43.7**	**40.2**

(a) Results on Market-1501 dataset
(c) Results on CUHK03 dataset

Figure. 4.1. The top-10 of an individual re-identification samples outcomes on Market-1501, DukeMTMC-reID, as well as CUHK03 datasets. The green True word and red false word on the top of each image indicate respectively the true equivalent and the false ones.

(c) Results on CUHK03 dataset

Figure. 5. The top-10 of an individual re-identification samples outcomes on Market-1501, DukeMTMC-reID, as well as CUHK03 datasets. The green True word and red False word on the top of each image indicate respectively the true equivalent and the false ones.

• Conclusion

This paper sheds the light on how to utilize a pre-trained CNN models that was developed for image recognition, to build powerful CNN baseline model that utilized for an individual re-identification domain. This study has proposed to adjust the architecture of the CNN model via adding batch normalization and dropout layers to the classifier part of the CNN model to prevent over-fitting and re-train the adjusted model with available dataset. Moreover, the experiments have exhibited that the proposed techniques regularly improved the performance of baselines as well as was very robust to features representation. Finally, the outcomes of this experimental study remain competitive with the state-of-the-art approaches outcomes on three a huge number of well-known re-identification datasets.

For further improvements, there is a plan to utilize an Adam as optimizer with the proposed model. There will be also another plane to investigate other CNN architecture that remain widely utilized for an individual re-identification practices, for instance Densenet121, and Part-based Convolutional Baseline (PCB) models.

• References

[1] Zheng, L., Y. Yang, and A.G. Hauptmann Person Re-identification: Past, Present and Future. arXiv e-prints, 2016.
[2] He, K., et al. Deep Residual Learning for Image Recognition. arXiv e-prints, 2015.
[3] Krizhevsky, A., I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems, 2012. 25.
[4] Simonyan, K. and A. Zisserman Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv e-prints, 2014.
[5] Szegedy, C., et al. Going Deeper with Convolutions. arXiv e-prints, 2014.
[6] Xiong, F., et al. Good practices on building effective CNN baseline model for person re-identification. in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2019.
[7] Chang, X., T.M. Hospedales, and T. Xiang Multi-Level Factorisation Net for Person Re-Identification. arXiv e-prints, 2018.
[8] Si, J., et al. Dual Attention Matching Network for Context-Aware Feature Sequence based Person Re-Identification. arXiv e-prints, 2018.
[9] Deng, J., et al. ImageNet: A large-scale hierarchical image database. in 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009.
[10] Ruder, S. An overview of gradient descent optimization algorithms. arXiv e-prints, 2016.
[11] Liao, S., et al. Person re-identification by Local Maximal Occurrence representation and metric learning. in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
[12] Mignon, A. and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. in 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012.
[13] Gray, D. and H. Tao. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. 2008. Berlin, Heidelberg: Springer Berlin Heidelberg.
[14] Sun, Y., et al. SVDNet for Pedestrian Retrieval. arXiv e-prints, 2017.

[15] Yu, R., et al. Divide and Fuse: A Re-ranking Approach for Person Re-identification. arXiv e-prints, 2017.
[16] Szegedy, C., et al. Rethinking the Inception Architecture for Computer Vision. arXiv e-prints, 2015.
[17] Huang, G., et al. Densely Connected Convolutional Networks. arXiv e-prints, 2016.
[18] Hermans, A., L. Beyer, and B. Leibe In Defense of the Triplet Loss for Person Re-Identification. arXiv e-prints, 2017.
[19] Zhao, L., et al. Deeply-Learned Part-Aligned Representations for Person Re-Identification. arXiv e-prints, 2017.
[20] Li, W., X. Zhu, and S. Gong, Harmonious attention network for person re-identification. CVPR, 2018. 1.
[21] Zhong, Z., et al. Random Erasing Data Augmentation. arXiv e-prints, 2017.
[22] Ioffe, S. and C. Szegedy Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv e-prints, 2015.
[23] Srivastava, N., et al., Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 2014. 15: p. 1929-1958.

[24] Zheng, L., et al. Scalable Person Re-identification: A Benchmark. in 2015 IEEE International Conference on Computer Vision (ICCV). 2015.
[25] Zheng, Z., L. Zheng, and Y. Yang Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro. arXiv e-prints, 2017.
[26] Ristani, E., et al. Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. arXiv e-prints, 2016.
[27] Li, W., et al. DeepReID: Deep filter pairing neural network for person re-identification. in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2014.
[28] Felzenszwalb, P., D. McAllester, and D. Ramanan, A discriminatively trained, multiscale, deformable part model. IEEE CVPR, 2008: p. 1-8.
[29] Zhong, Z., et al. Re-ranking Person Re-identification with k-reciprocal Encoding. arXiv e-prints, 2017.
[30] Sun, Y., et al. Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline). arXiv e-prints, 2017.
[31] Bai, S., X. Bai, and Q. Tian Scalable Person Re-identification on Supervised Smoothed Manifold. arXiv e-prints, 2017.
[32] Wang, Y., et al. Resource Aware Person Re-identification across Multiple Resolutions. arXiv e-prints, 2018.
[33] Zhong, Z., et al., Camera style adaptation for person re-identification. CVPR, 2018: p. 5157-5166.

[34] Dalla, L. O. F. B. .(2020) Dorsal Hand Vein (DHV) Verification in Terms of Deep Convolutional Neural Networks with the Linkage of Visualizing Intermediate Layer Activations Detection.