# SCALE-INVARIANT SIAMESE NETWORK FOR PERSON RE-IDENTIFICATION

*Yunzhou Zhang, Weidong Shi, Shuangwei Liu, Jining Bao, Ying Wei*

College of Information Science and Engineering, Northeastern University, Shenyang 110819, China.

## ABSTRACT

Most existing methods for person re-identification (ReID) almost match people at a single scale and ignore that people are often distinguishable at the right spatial locations and scales. Unlike previous works designing complex convolutional neural network (CNN) architecture or concatenating multi-branch scale-specific features, we aim to employ a simple network to learn scale-invariant features. Concretely, we first propose a shared two-branch framework with two-scale images from the same identity as inputs, which is beneficial for ReID network to focus on common features in different-scale images. Furthermore, we introduce a novel attention loss to enforce discriminative regions between two branches more consistent in the visual level. Finally, we conduct extensive evaluations on three large-scale datasets and report competitive performance.

## MOTIVATION


**Fig. 1.** Illustration of person bounding boxes captured at different scales(resolutions) in the public space.

Although significant progress has been made, person Re-ID still faces a huge challenge as illustrated in Fig.1, auto or manual cropping person boxes often vary more important in scale due to differences in the person-camera distance and camera deployment settings. To tackle the problems, we propose a scale-invariant siamese network (SiSNet) and introduce an attention loss called discriminative region consistency (DRC) for person ReID, which ams to improve the generalization of the model on scale.

## METHOD

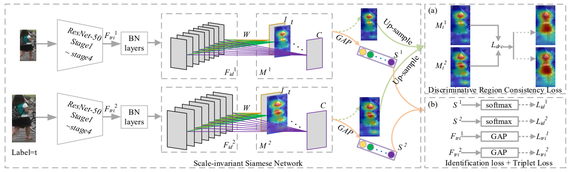


**Fig. 2.** The pipeline of our proposed framework.

As shown in Fig.2, we propose a scale-invariant siamese network (SiSNet) and introduce an attention loss called discriminative region consistency (DRC) for person ReID. Specially, SiSNet comprises two identical branches taking two-scale images from the same identity as inputs. Each branch is used to learn latent distinguishable features of single-scale images. Due to siamese network architecture that is shared with each branch, SiSNet is encouraged to exploit common features between cross-scale images. Note that the SiSNet directly adopt the existing backbone without extra complex design, so the proposed method is quite a simple architecture. Then we utilize class activation mapping (CAM) [14] to compute discriminative regions for the input image of each branch and introduce a DRC loss to enforce discriminative regions between two branches more consistent in visual level. With this design, we can only use the either-scale feature to achieve satisfactory performance, which has a low computation cost in retrieval.

## RESULTS

| Method | Publication | Backbone | Market-1501 | | DukeMTMC-ReID | | MSMT17 | |
|---|---|---|---|---|---|---|---|---|
| | | | R1 | mAP | R1 | mAP | R1 | mAP |
| **PCB [2]** | ECCV'18 | ResNet | 93.8 | 81.6 | 83.3 | 69.2 | 68.2 | 40.4 |
| **HPM [3]** | AAAI'19 | ResNet | 94.2 | 82.7 | 86.6 | 74.3 | - | - |
| **VPM [26]** | CVPR'19 | ResNet | 93.0 | 80.8 | 83.6 | 72.6 | - | - |
| **CASN [4]** | CVPR'19 | PCB | 94.4 | 82.8 | 87.7 | 73.7 | - | - |
| **AANet [5]** | CVPR'19 | ResNet | 93.9 | 83.4 | 87.7 | 74.3 | - | - |
| **IANet [6]** | CVPR'19 | ResNet | 94.4 | 83.1 | 87.1 | 73.4 | 75.5 | 46.8 |
| **DPFL [12]** | ICCV'17 | Inception | 88.9 | 73.1 | 79.2 | 60.6 | - | - |
| **MLFN [9]** | CVPR'18 | ResNeXt | 90.0 | 74.3 | 81.0 | 62.8 | - | - |
| **OSNet [11]** | ICCV'19 | OSNet | 94.8 | 84.9 | 88.6 | 73.5 | 78.7 | 52.9 |
| **Ours** | - | ResNet | **95.2** | **85.4** | **89.0** | **75.9** | **80.1** | **54.6** |

**Table 1.** Comparison with state-of-the-art methods. (Boldface denotes the best result, -: not available, R1: Rank1,hereinafter)

We compare the proposed method with some state-of-the-art methods which are mainly based on local features [2, 3, 26] (top), attention mechanism [4, 5, 6] (middle) and multi-scale features [9, 11, 12] (bottom) in table 1. The results show that the proposed approach has a clearly overwhelming performance on the three datasets.

| Loss type | Method | Market-1501 | | DukeMTMC-ReID | | MSMT17 | |
|---|---|---|---|---|---|---|---|
| | | R1 | mAP | R1 | mAP | R1 | mAP |
| **S.** | Baseline | 91.9 | 80.3 | 84.3 | 69.2 | 70.3 | 41.4 |
| | SiSNet | 93.3 | 82.3 | 86.4 | 72.0 | 73.8 | 46.1 |
| | SiSNet+DRC | **93.9** | **83.6** | **88.2** | **75.0** | **76.0** | **50.3** |
| **S.+T.** | Baseline | 94.2 | 83.2 | 86.5 | 72.5 | 76.3 | 48.4 |
| | SiSNet | 94.8 | 84.7 | 87.5 | 74.5 | 78.9 | 52.2 |
| | SiSNet+DRC | **95.2** | **85.4** | **89.0** | **75.9** | **80.1** | **54.6** |

**Table 2.** Ablation study.

From Table 2, with softmax and triplet loss, when SiSNet is employed alone, the Rank1/mAP goes up by 0.6% /1.5% on Market-1501, 1% /2% on DukeMTMC-ReID and 0.6%/3.8% on MSMT17 than the Baseline. Furthermore,with the softmax loss alone, the margin is even larger, which indicates that extracting scale-invariant features are effective.

| Market-1501 | | | | | | DukeMTMC-ReID | | | | | | MSMT17 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B. | | Si. | | Si.+D. | | B. | | Si. | | Si.+D. | | B. | | Si. | | Si.+D. | |
| R1 | m | R1 | m | R1 | m | R1 | m | R1 | m | R1 | m | R1 | m | R1 | m | R1 | m |
| 0.96 | 2.52 | 0.46 | 0.64 | 0.22 | 0.10 | 1.58 | 2.8 | 0.2 | 0.72 | 0.08 | 0.12 | 3.2 | 4.06 | 0.92 | 1.22 | 0.14 | 0.12 |
| 0.82 | 1.66 | 0.28 | 0.54 | **0.12** | **0.04** | 1.54 | 2.56 | 0.28 | 0.5 | **0.06** | **0.02** | 2.6 | 3.6 | 0.7 | 0.94 | **0.06** | **0.02** |

**Table 3.** Quantitative analysis on three datasets. (m: mAP, B.: Baseline, Si.: SiSNet, Si.+D.: SiSNet with DRC loss, the first line data are acquired under S., the second line data are acquired under S.+T.)

Under the same parameter settings, we calculate the gap of performance between non-cascading and cascading two branches of features. Table 3 shows the mean value of the gap of later five-group retrieval results.

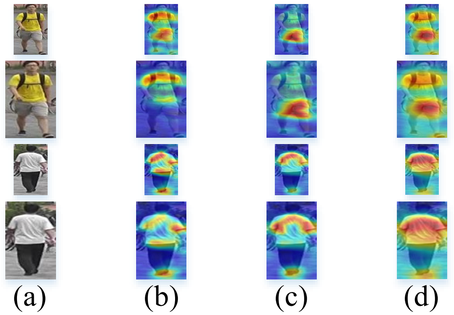


**Fig. 3.** Qualitative analysis. (a), (b), (c) and (d) are different-scale input images, the CAM produced by baseline, SiSNet, and SiSNet with DRC loss respectively.

We compare the class activation maps from the two-scale images with the same identity using the baseline and the SiSNet with or without DRC loss. Fig.3 (b) reveals that the baseline only focuses on discriminative regions, but the interest regions are inconsistent under two-scale images. When we employ SiSNet alone, class activation maps (Fig.3 (c)) show highly consistent regions, but lost some discriminative information than the baseline, e.g., the shoulder area of the first pedestrian with yellow T-shirt and the hip area of the second pedestrian with a white T-shirt. Fig.3 (d) reflects SiSNet with DRC loss produces more consistent as well as diverse regions, and DRC loss helps SiSNet to capture the lost information over again. Hence, employing DRC loss to further constraint the overall network is an effective strategy indeed.

## CONLUSION

We present a novel Scale-invariant Siamese Network (SiSNet) by aiming to discover scale-invariant information for the cross-scale matches in person ReID. In contrast to most existing approaches that employ single-scale features alone or multi-scale fusion features together, the proposed SiSNet is capable of capturing discriminative and abundant scale-invariant features of cross-scale images. Moreover, the proposed Discriminative Region Consistency (DRC) loss further promotes the capability of SiSNet learning diversity and consistency of features, which also make it possible to exploit either-scale features in the matching procedure. Extensive experiments on three large-scale datasets demonstrate the effectiveness of the proposed framework.

## ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (No. 61973066, 61471110), Foundation Project of National Key Laboratory (6142002301), and Fundamental Research Funds for the Central Universities (N172608005, N182608004).