RGB-Infrared person re-identification (RGB-IR Re-ID) is a task aiming to retrieve and match person images between RGB images and IR images. Since most surveillance cameras capture RGB images during the day and IR images at night, RGB-IR Re-ID is helpful when checking day and night surveillance for criminal investigations. Previous related work often only extracts sharable and identity-related features in images for identification. Few researches specifically extract and make use of features that do not have the ability to distinguish identity, e.g., identity-unrelated features derived from background and modality. In this Letter, we propose a novel and concise RGB-IR Re-ID network named two-way constraint network (TWCN). Compared with traditional Re-ID networks, TWCN not only extracts and utilises identity-related features but also novelly makes full use of identity-unrelated features to improve the accuracy of the experiment. TWCN uses a reverse-triplet loss to extract identity-unrelated features, and proposes an orthogonal constraint to remove identity-unrelated information from identity-related features, which improves the purity of identity-related features. In addition, a correlation coefficient synergy and central clustering (CCSCC) loss is introduced into TWCN to extract identity-related features effectively. Extensive experiments have been conducted to prove our method is effective.

Introduction: RGB-IR Re-ID is a challenging multi-modality problem first proposed by Wu et al. [1]. In the past work [1–8], the main concern is how to extract identity-related features from images, and use identity-related features to perform similarity ranking matching for identification. However, in the RGB-IR Re-ID task, we think that not only should we extract identity-related features, but also consider the elimination of identity-unrelated information from the identity-related features. The reasons for this consideration are as follows. Compared with traditional Re-ID tasks, due to the large modality differences in the RGB-IR Re-ID task, colour and texture are difficult to use as clues to distinguish identities, which also means that the identity-related information in images has been greatly reduced. It is foreseeable that the reduction of identity-related information makes it easier for identity-unrelated information to be misjudged and appear in identity-related features. From this perspective, we try to remove the above-mentioned identity-unrelated information from identity-related features to improve the accuracy of RGB-IR Re-ID.

We make a strong assumption that in the feature extraction process, identity-related features are inevitably coupled with part of identity-unrelated information. In addition, the identity-unrelated features contain most of the identity-unrelated information. Through an orthogonal constraint between identity-related features and identity-unrelated features, the identity-unrelated information in the identity-related features can be removed. It improves the purity of the identity-related features. Besides, to solve the challenge of extracting sharable and identity-related features, a correlation coefficient synergy and central clustering (CCSCC) loss is introduced into TWCN.

It is worth mentioning that in 2020, Kansal et al. [9] proposed a spectrum-disentangled representation learning method in the RGB-IR Re-ID task, which circuitously disentangled the spectrum information in identity-related features, making the identity-related features insensitive to image modality differences. However, this method only decouples the spectrum information, which still retains other noise items (e.g., background) in identity-related features. Besides, the multi-stage training strategy increases the amount of calculation and training time. Compared with this method, our network is more concise, and can directly decouple global identity-unrelated information with an end-to-end training strategy.

In this Letters, our contributions can be listed as follows:

(1) We propose a novel and concise RGB-IR Re-ID network named two-way constraint network (TWCN), which not only extracts and utilises identity-related features but also novelly makes full use of identity-unrelated features to improve the accuracy of the experiment.

(2) TWCN uses a reverse-triplet loss to extract identity-unrelated features and proposes an orthogonal constraint to remove identity-unrelated information from identity-related features, which improves the purity of identity-related features.

(3) A CCSCC loss is introduced into TWCN to extract identity-related features effectively.

Baseline and Two-Way Constraint Network: Our baseline is a traditional Re-ID network proposed by Luo et al. [10]. Because Ye's work [11] indicates that generalised-mean pooling (GeM pooling) [12] outperforms average pooling in the RGB-IR Re-ID task, we replace average pooling with GeM pooling in the baseline. As shown in Figure 1(a), the baseline is composed of ResNet50 backbone, GeM pooling, a batch normalisation layer, and a fully connected layer.

The expression of GeM pooling [12] is as follows:

$$f_{G}^{(GeM)} = \left[ f_{c}^{(G)} | \cdots | f_{c}^{(G)} | \cdots | f_{c}^{(G)} \right]^{T},\ f_{c}^{(G)} = \left( \frac{1}{|T_{c}|} \sum_{x \in T_{c}} x^{p_{s}} \right)^{\frac{1}{p_{s}}}$$

where $p_{s}$ is a hyperparameter that will be learned in the training. With a learnable hyperparameter $p_{s}$, GeM pooling dynamically adjusts the pooling state, which retains more effective information in the deep layer of the network. More details about GeM pooling can be found in [12]. In the baseline, as shown in Figure 1(a), the initial convolution block of ResNet50 is taken out separately to perform different convolution
Compared to the original triplet loss [13]:

We use a reverse-triplet loss to extract identity-unrelated features. remove common identity-unrelated information. Finally, purer identity-related features and most of the identity-unrelated information is included in features, part of identity-unrelated information appears in identity-related features. When extracting identity-unrelated features and identity-related features, the above reasons, it is easier to extract pure identity-unrelated features.

Identity-Unrelated Information Removal: In this Letters, we make a strong assumption that in the feature extraction process, identity-related features are inevitably coupled with identity-unrelated information. And the identity-unrelated information can be removed from the identity-related features by exploiting identity-unrelated features.

Usually, we directly extract identity-related features for identification. However, as shown in Figure 2(a), in the RGB-IR Re-ID task, the identity-related features often focus on invalid areas and ignore areas of the body (coloured clothes). This is because the large modality differences make colour and texture become invalid clues, which reduce the effective identity-related information in the image and make it easier for mixing identity-unrelated information (e.g. background information) into identity-related features. If we extract identity-related features, on the contrary, identity-unrelated features focus on most of the identity-unrelated area as shown in Figure 2(b). There are two main reasons for this result: (1) Identity-unrelated information is far more than the identity-related information in an image in the RGB-IR Re-ID task. (2) Identity-unrelated information is widely distributed. Because of the above reasons, it is easier to extract pure identity-unrelated features. When extracting identity-unrelated features and identity-related features, part of identity-unrelated information appears in identity-related features and most of the identity-unrelated information is included in identity-unrelated features. Then, an orthogonal constraint is proposed between the identity-unrelated features and identity-related features to remove common identity-unrelated information. Finally, purer identity-related features are obtained for identification as shown in Figure 2(c). We use a reverse-triplet loss to extract identity-unrelated features. Compared to the original triplet loss [13]:

\[
\text{Loss}_{\text{Tri}} = \max (d_{a,p} - d_{a,n} + \alpha, 0)
\]  

(2)

The reverse-triplet loss used to extract identity-unrelated features is as follows:

\[
\text{Loss}_{\text{Tri-Reverse}} = \max (d_{a,n}^{\text{same}} - d_{a,p}^{\text{same}} + \alpha, 0) + \max (d_{a,p}^{\text{cross}} - d_{a,p}^{\text{cross}} + \alpha, 0)
\]  

(3)

where \(d_{a,n}^{\text{same}}\) and \(d_{a,p}^{\text{same}}\) are feature distances of negative and positive pairs with the same modality. \(d_{a,n}^{\text{cross}}\) and \(d_{a,p}^{\text{cross}}\) are feature distances of negative and positive pairs with different modalities. \(\alpha_1, \alpha_2\) is the margin constant of the reverse-triplet loss. TWCN calculates the reverse-triplet loss under the same modality or across modalities. The reverse-triplet loss extracts features that cannot distinguish identities, which requires features to meet: \(d_{a,n} + \alpha < d_{a,p}\). We call the features that satisfy the above formula as identity-unrelated features.

Generative adversarial nets (GAN) learning and singular vector decomposition [15] are commonly used for feature decoupling in Re-ID tasks. To decouple modality information, typical GAN learning methods [5] learn features from unified modality images that are generated from different modality images by the generator. However, the GAN methods have the disadvantages of slow convergence and a large amount of calculation. Orthogonal constraint based on singular vector decomposition [15] can perform singular value decomposition on a single matrix, so as to obtain the corresponding orthogonal matrix. In our network, identity-related features and identity-unrelated features are output in the form of two one-dimensional vectors, which is not suitable for methods based on singular vector decomposition. We try to design the loss function from the basic definition of orthogonal to achieve the purpose of orthogonal constraint. The formula for the angle between two vectors is as follows:

\[
\cos \phi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}, \quad \mathbf{a}, \mathbf{b} \text{ are vectors and } \phi \text{ is the angle between } \mathbf{a} \text{ and } \mathbf{b}. \quad \text{Let loss } = \cos \phi. \quad \text{When loss } \to 0, \cos \phi \text{ is also } \to 0. \quad \text{If } \cos \phi = 0, \text{ then } \phi = \frac{\pi}{2}, \text{ which means that the two vectors are orthogonal. We wisely use the convergence of the loss function in deep learning to design a loss function that can be used for orthogonal constraint. Based on the above considerations, a novel orthogonal constraint is proposed as follows:}

\[
\text{Loss}_{\text{orthogonal}} = \sum b \left( \frac{1}{n} \sum_{i=1}^{m} \left( \rho_{i}^{\text{unrelated}} \rho_{i}^{\text{related}} \right)^{T} \text{F}^{2} \right)
\]  

(4)

where \(b\) is the size of a mini batch, \(f_{i}^{\text{unrelated}}\) is the \(i\)th identity-unrelated feature embedding in a mini batch, and \(f_{i}^{\text{related}}\) is the corresponding identity-related feature embedding. We use the above orthogonal constraint to complete the decoupling of identity-unrelated information from identity-related features.

Correlation Coefficient Synergy and Central Clustering Loss: In addition to the removal of identity-unrelated information, we propose a CCSCC loss to extract identity-related features efficiently. The CCSCC loss consists of two items. The first item uses central features to promote cross-modality feature central clustering, and the second item uses correlation matrices [16] to synchronise feature extraction. The expression of CCSCC loss is as follows:

\[
\text{Loss}_{\text{CCSCC}} = \log \left( 1 + \lambda_1 \frac{1}{c} \sum_{i=1}^{m} \left( m_{i}^{\text{RGB}} - m_{i}^{\text{IR}} \right) \right) + \lambda_2 \frac{1}{c} \sum_{i=1}^{c} \left( R_{i}^{\text{RGB}} - R_{i}^{\text{IR}} \right)
\]  

(5)

where \(\lambda_1, \lambda_2\) are hyperparameters, \(c\) is the number of classes in a mini batch, \(m_{i}^{\text{RGB}}\) and \(m_{i}^{\text{IR}}\) are corresponding RGB/IR centre vector of the \(i\)th class in a mini batch. \(R_{i}^{\text{RGB}}\) and \(R_{i}^{\text{IR}}\) are corresponding RGB/IR feature embeddings correlation matrices of \(i\)th class in a mini batch. Given an identity-related feature embedding set \(X_{i}\), which is extracted from the \(i\)th class RGB images in a mini batch. \(X_{i} = (X_{i1}, X_{i2}, X_{i3}, \ldots)\), where \(X_{ik}\) is the \(k\)th identity-related feature embedding of the \(i\)th class in a mini batch. The calculation of \(R_{i}^{\text{RGB}}\) can be expressed as follow:

\[
R_{i}^{\text{RGB}} = \begin{bmatrix}
\rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\
\rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}
\end{bmatrix}
\]

where \(\rho_{jk} = \frac{\text{cov}(X_{ij}, X_{kj})}{\sqrt{\text{var}(X_{ij})} \sqrt{\text{var}(X_{kj})}}\) and

\[
\text{cov}(X_{ij}, X_{kj}) = E((X_{ij} - E(X_{ij})) \cdot (X_{kj} - E(X_{kj}))).
\]  

(6)
Table 1. Ablation study on the RegDB dataset (RGB-IR Re-ID).
IUIR: Identity-unrelated information removal

| Method                   | Rank-1 | mAP    |
|--------------------------|--------|--------|
| Baseline                 | 69.88  | 64.81  |
| Baseline + IUIR          | 70.83  | 65.79  |
| Baseline + IUIR + CCSCC loss (ours) | 76.87  | 72.42  |

Table 2. Ablation study on the SYSU-MM01 dataset (IR-RGB Re-ID). The SYSU-MM01 dataset has two modes (All-search mode and indoor-search mode [1]). IUIR: Identity-unrelated information removal

| Method                   | SYSU-MM01 (All-search) | SYSU-MM01 (Indoor-search) |
|--------------------------|-------------------------|---------------------------|
| Baseline                 | 44.20                   | 43.90                     |
| Baseline + IUIR          | 45.58                   | 44.95                     |
| Baseline + IUIR + CCSCC loss (ours) | 47.20 | 46.97 |

Table 3. Performance comparison with state-of-the-art methods

| Method/backbone           | RegDB (Rank-1) | RegDB (mAP) | SYSU-MM01 (All-search) | SYSU-MM01 (Indoor-search) |
|---------------------------|----------------|-------------|------------------------|---------------------------|
| Zero-pad [1] Two-stream network [1] | 17.75          | 18.00       | 14.80                  | 15.95                     |
| HCML [2] AlexNet          | 24.44          | 20.08       | 14.32                  | 16.16                     |
| e-BDTR [3] AlexNet & ResNet50 | 34.62         | 33.46       | 27.82                  | 28.42                     |
| HSME [4] AlexNet          | 50.85          | 47.00       | 20.68                  | 23.12                     |
| D^RL [5] ResNet50         | 43.40          | 44.10       | 28.90                  | 29.20                     |
| MAC [6] ResNet50          | 36.43          | 37.03       | 33.26                  | 36.22                     |
| MSR [7] DGD [17]          | 48.43          | 48.67       | 37.35                  | 38.11                     |
| AlignG [8] ResNet50       | 57.90          | 53.60       | 42.40                  | 40.70                     |
| TWCN (ours) ResNet50      | 76.87          | 72.42       | 47.20                  | 46.97                     |

The calculation of $R_i$ is similar to the above. The correlation matrices that represent the relationship between variables can be used to synchronize the convergence direction in CCSCC loss. Extensive experiments have been conducted to prove the CCSCC loss is effective.

**Experiments and Results:** In order to prove the effectiveness of our method, ablation experiments are implemented in two popular cross-modality person re-identification datasets (SYSU-MM01 [1] and RegDB dataset [14]). The SYSU-MM01 dataset contains 287,628 RGB images and 15,792 infrared images. The RegDB dataset contains 4,120 RGB images and 15,792 infrared images. All images are processed into the same size 288 x 288. In our experiments, the optimization method is the stochastic gradient descent method. The experimental results are shown in Tables 1 and 2. In the RegDB dataset, compared with the baseline, TWCN has increased by 6.99% and 6.71% in rank-1 accuracy and mAP, respectively. In the SYSU-MM01 dataset, compared with the baseline, TWCN has respectively improved in rank-1 accuracy and mAP. Table 3 shows the performance comparison between TWCN and state-of-the-art methods.

**Conclusion:** We propose a novel and concise RGB-IR Re-ID network named two-way constraint network (TWCN). TWCN not only extracts and utilizes identity-related features but also novelty makes full use of identity-unrelated features to improve the accuracy of the experiment. TWCN uses a reverse-triplet loss to extract identity-unrelated features, and proposes an orthogonal constraint to remove identity-unrelated information from identity-related features. In addition, a correlation coefficient synergy and central clustering (CCSCC) loss is introduced into TWCN to extract identity-related features effectively. Extensive experiments have proved our method is effective.

**Acknowledgement:** This work was supported in part by the Science and Technology Program of Guangdong Province under Grant 2019B010140002.

© 2021 The Authors, Electronics Letters published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 4 February 2021 Accepted: 25 April 2021
doi: 10.1049/iet-el.2021.0121

**References**

1. Wu, A.C. et al.: RGB-Infrared cross-modality person re-identification. In: IEEE International Conference on Computer Vision, pp. 5390–5399. IEEE, Piscataway, NJ (2017)
2. Ye, M. et al.: Hierarchical discriminative learning for visible thermal person re-identification. In: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 7501–7508. AAAI, Menlo Park, CA (2018)
3. Ye, M., et al.: Bi-directional center-constrained top-ranking for visible-thermal person re-identification. IEEE Trans. Inf. Forensics Secur. 15, 407–419 (2020)
4. Hao, Y., et al.: HSME: Hypersphere manifold embedding for visible-thermal person re-identification. Proceedings of the AAAI Conference on Artificial Intelligence 33, 8385–8392 (2019)
5. Wang, Z. et al.: Learning to reduce dual-level discrepancy for infrared- visible person re-identification. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 618–626. IEEE, Piscataway, NJ (2019)
6. Ye, M., Lan, X., Leng, Q.: Modality-aware collaborative learning for visible-thermal person re-identification. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 347–355. ACM, New York (2019)
7. Feng, Z., Lai, J., Xie, X.: Learning modality-specific representations for visible-infrared person re-identification. IEEE Transactions on Image Processing 29, 579–590 (2020)
8. Wang, G.A. et al.: RGB-infrared cross-modality person re-identification via joint pixel and feature alignment. In: IEEE/CVF International Conference on Computer Vision, pp. 3622–3631. IEEE, Piscataway, NJ (2019)
9. Kansal, K. et al.: SDL: Spectrum-disentangled representation learning for visible-infrared person re-identification. IEEE Transactions on Circuits and Systems for Video Technology 30(10), 3422–3432 (2020)
10. Luo, H., et al.: A strong baseline and batch normalization neck for deep person re-identification. IEEE Transactions on Multimedia 22(10), 2597–2609 (2019)
11. Ye, M. et al.: Deep learning for person re-identification: A survey and outlook. arXiv:2001.04193, (2020)
12. Radenovic, F., Tolias, G., Cham, O.: Fine-tuning CNN image retrieval with no human annotation. IEEE Trans. Pattern Anal. Mach. Intell. 41(7), 1655–1668 (2019)
13. Hermans, A., Beyer, L., Leibe, B.J.A.: In defense of the triplet loss for person re-identification. arXiv:1703.07737, (2017)
14. Nguyen, D.T. et al.: Person recognition system based on a combination of body images from visible light and thermal cameras. Sensors 17(3), (2017)
15. Sun, Y.F. et al.: SVDNet for pedestrian retrieval. In: IEEE International Conference on Computer Vision, pp. 3820–3828. IEEE, Piscataway, NJ (2017)
16. Berthold, M., Hppner, F.J.A.: On clustering time series using euclidean distance and pearson correlation. arXiv:1601.02213, (2016)
17. Xiao, T. et al.: Learning deep feature representations with domain guided dropout for person re-identification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1249-1258. IEEE, Piscataway, NJ (2016)