LONG WAVE INFRARED IMAGE COLORIZATION FOR PERSON RE-IDENTIFICATION

V. V. Kniaz[a,b,*], A. N. Bordodymov[a]

[a] State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia
(vl.kniaz, bordodymov)@gosniias.ru
[b] Moscow Institute of Physics and Technology (MIPT), Russia

Commission II, WG II/8

KEY WORDS: person Re-ID, generative adversarial networks, thermal images, image colorization

ABSTRACT:

Person re-identification (ReID) in color and thermal images require matching of the object color and its temperature. While thermal cameras increase the performance of ReID systems during the night-time, identification of corresponding features in the visible and the long-wave infrared range is challenging. The biggest challenge arises from the multimodal relationship between an object's color and its temperature. Modern ReID methods provide state-of-the-art results in person matching in the visible range. Hence, it is possible to perform multimodal matching by translation of a thermal probe image to the color domain. After that, the synthetic color probe image is matched with images from the real color gallery set. This paper is focused on the development of the `ThermalReID` multispectral person ReID framework. The framework performs matching in two steps. Firstly, it colorizes the input thermal probe image using a Generative Adversarial Network (GAN). Secondly, it matches images in the color domain using color histograms and MSCR features. We evaluate the `ThermalReID` framework using RegDB and ThermalWorld datasets. The results of the evaluation are twofold. Firstly, the developed GAN performs realistic colorization of thermal images. Secondly, the `ThermalReID` framework provides matching of persons in color and thermal images that compete with and surpass the state-of-the-art. The developed `ThermalReID` framework can be used in video surveillance systems for effective person ReID during the nighttime.

1. INTRODUCTION

Person Re-identification (ReID) is the problem of matching of the same humans captured in different images. The images can be captured using different cameras and during a different time. The only requirement is that the appearance (clothes, shoes, etc.) of a person remains unchanged across all the images. A typical ReID scenario is presented in Figure 1. Let Cam 1, Cam 2, Cam 3 be cameras located in a mall area, Let P_1, P_2 be the persons walking through the mall. Then the ReID system for a given image $I_p(i, j)$ showing a person j from a camera i should find a set of images $s = \{I_1, I_2, \ldots\}$ from other cameras, where the person P is visible. Person ReID problem received a lot of scholar attention recently (Farenzena et al., 2010, Bhuiyan et al., 2015, Bhuiyan et al., 2018). While modern ReID methods provide the-state-of-the-art accuracy in the visible range, multispectral ReID remains challenging. A robust cross-domain feature matching is required for identification of the same person in a different spectral range.

Authors participate in a research project focused on the development of a multispectral video surveillance system. The system should perform person re-identification in images captured by an outdoor thermal camera and an indoor color camera. Many robust methods for ReID in the visible range were proposed recently (Bhuiyan et al., 2018, Wei et al., 2018, Saquib Sarfraz et al., 2018). Most of the methods use either hand-crafted features (color histograms, feature descriptors) or convolutional neural networks (CNN). Most of multispectral ReID methods leverage dedicated CNN architecture to extract and match discriminative

features (Wu et al., 2017, Ye et al., 2018b, Ye et al., 2018a). Recently (Kniaz et al., 2018) proposed a new ReID approach based on a transformation of the spectral range of the probe image. The input color probe image is transformed into the thermal image domain using a dedicated Generative Adversarial Network (GAN).
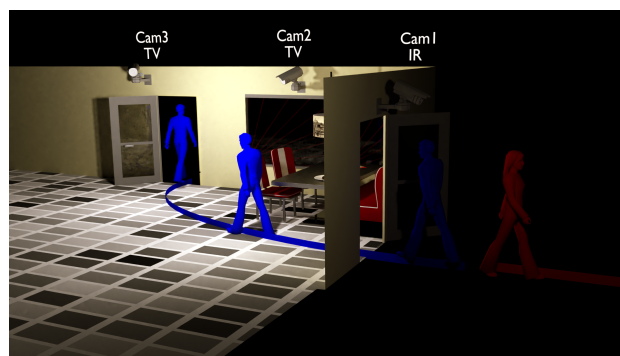

Figure 1. A typical ReID scenario.

To the best of our knowledge, there is no research in the literature regarding colorization of a thermal image for cross-domain ReID. This paper is focused on the development of a cross-modality person ReID framework `ThermalReID`. We perform re-identification in two steps. Firstly, the input thermal probe image is transformed into a synthetic color image using a GAN model. After that, we match the synthetic color probe image with a color gallery set. We evaluate our `ThermalReID` framework on ThermalWord ReID and RegDB datasets. Comparison with modern ReID models demonstrates that our `ThermalReID` framework competes with state-of-the-art ReID methods. Our main obser-

---

*Corresponding author

vation is that colorization of long-wave infrared images (LWIR) is an ill-posed problem. Realistic colorization requires an additional modality. We provide such modality as an additional 'hint' color image from a different camera. The 'hint' color image provides desired object colors and textures to our `ColorMatchGAN` model. Our `ThermalReID` framework proves to provide robust cross-modality person matching and can be used in video surveillance applications.

The rest of the paper is organized as follows. Section 2 presents a brief review of the literature regarding person ReID and cross-modality image-to-image translation. In Section 3 the developed `ThermalReID` framework is discussed. Section 4 demonstrates results of evaluation of our `ThermalReID` framework. Section 5 summarizes results and discusses the future implications of the developed method.

1.1 Contributions

We present three key technical contributions:

1. `ColorMatchGAN` model for colorization of a long-wave infrared image using a 'hint' color image.

2. `ThermalReID` cross-modality person ReID framework.

3. Evaluation of the `ThermalReID` framework and baselines on two ReID datasets.

2. RELATED WORK

Person re-identification problem received a lot of scholar attention recently (Farenzena et al., 2010, Bhuiyan et al., 2015, Bhuiyan et al., 2018). Most of modern approaches can be divided into three groups: direct methods, metric learning methods, and transform learning methods (Bhuiyan et al., 2018). The rest of this section presents a brief review of modern ReID methods organized by these groups.

2.1 Direct methods

Direct methods utilize hand-crafted or learned features to directly match person IDs. Recent works demonstrate a significant progress in continuous ReID that utilizes multiple person detections in frame sequence. In (Tang et al., 2017) a graph-based method was proposed that allows to cluster person hypotheses over time. Clustered hypotheses are solved using a minimum cost lifted multicut problem. Another approach proposed in (Bhuiyan et al., 2018) leverages cumulative weighted brightness transfer function (CWBTF) to match person images from different views.

2.2 Metric learning methods

Metric learning methods utilize the training data to learn metric spaces that allow effective ReID. In (Barman and Shah, 2017) an algorithm was proposed that allowed effective match ranking using a graph-based approach. Siamese networks prove to be efficient to learn metric spaces for person ReID (Guo and Cheung, 2018). Multilevel matching further increases the performance of the proposed method.

Kronecker Product combined with an end-to-end trainable network provide effective means for feature map matching (Shen et al., 2018). The evaluation of the method on the Market-1501 (Zheng et al., 2015), CUHK03 (Li et al., 2012), and DukeMTMC (Ristani et al., 2016) datasets proved its robustness.

2.3 Transform learning methods

Transform learning methods aim to learn a transformation from one camera to another that increases the similarity of person in different view and lightning conditions. Most of transform learning methods divide human body into parts to extract discriminative features (Zhao et al., 2017, Li et al., 2018, Qian et al., 2017, Si et al., 2018, Su et al., 2017). A deep architecture that utilizes human part cues for efficient ReID was proposed in (Su et al., 2017). In (Zhao et al., 2017) authors proposed a deep model that combines human part body decomposition and an attention based learning. A Pose-driven Deep Convolutional (PDC) model was proposed in (Su et al., 2017). The proposed model leverages a pose-driven feature weighting sub-network to effectively balance feature weights on the local and global body part levels. The proposed model prove to be robust against significant changes in person's pose.

Attention selection training (Li et al., 2018) prove to be efficient for Re-ID in uncontrolled (misaligned) images. Another attention-based approach (Si et al., 2018) learns context-aware feature sequences and performs attentive sequence comparison simultaneously. The proposed Dual ATtention Matching network (DuATM) demonstrates state-of-the-art results on various image and video ReID benchmarks.

Transform learning methods are essential for cross-modality Re-ID problem where gallery and probe images are provided in different spectral ranges (e.g., visible and infrared). Recently multiple deep learning-based methods were proposed that demonstrated the possibility of cross-modality ReID. In (Wu et al., 2017) a large cross-modality SYSU-MM01 dataset was proposed. Authors also proposed a deep learning-based method for cross-modality Re-ID that leverages a one-stream network and deep zero-padding. The evaluation of the proposed method on the SYSU-MM01 dataset proved that infrared modality can increase the matching rate in low light conditions.

Our model is based on the previous research in the field of color-to-thermal image translation for person ReID (Kniaz et al., 2018). Unlike other similar models it uses a GAN network to synthesize realistic color probe images that are matched against real color gallery images.

3. METHOD

We use the `ThermalGAN` ReID framework (Kniaz et al., 2018) as the starting point for our research. Unlike the original approach, we perform matching in the color domain. We hypothesized that a GAN network could learn the transformation from the thermal range to the visible spectrum. However, the naïve translation from to color image in RGB format thermal image was not successful. We used the assumptions of Berg et al. (Berg et al., 2018) and performed colorization of LWIR images to images in the Lab color space. Translation from thermal image to the Lab domain prove to be more efficient. However, the result still suffered from various artifacts (e.g., color uncertainty).

We hypothesized that additional modality could improve thermal image colorization. More specifically, we use a 'hint' color image as an input modality for realistic colorization. We term the resulting model as `ColorMatchGAN`. The model should be able to perform the diverse colorization for various 'hint' color image.

(a) Testing ColorMatchGAN for ReID

(b) Training ColorMatchGAN

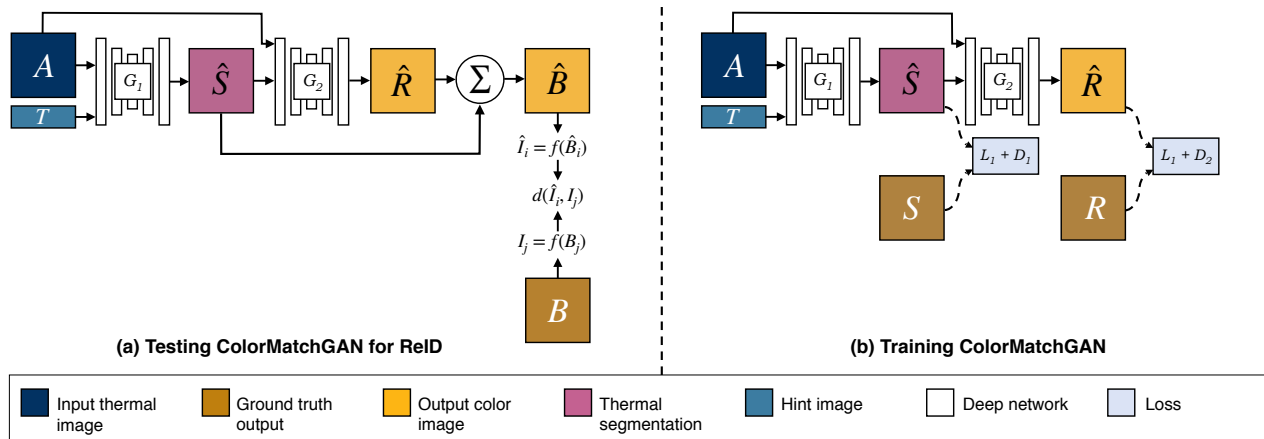| ■ Input thermal image | ■ Ground truth output | ■ Output color image | ■ Thermal segmentation | ■ Hint image | □ Deep network | □ Loss |

Figure 2. Overview of our `ColorMatchGAN` framework.

The rest of this section is organized as follows. Firstly, we describe our Color ReID framework. After that, we discuss our colorization GAN network and the architecture of the generator. A description of ReID features concludes the section.

3.1 Multimodal ReID framework

Let $V = \{v_1, v_2, v_3, \ldots, v_k, \}$ be the gallery set and I_p the thermal probe image. Where i is the index of the camera $i = \{1, 2, k\}, j_i = \{1, 2, \ldots, l\}$, j_i is the number of persons detected an image i. Our ReID framework must provide the ReID score $m(V_{i,j}, t_p)$ for a given pair of visible and thermal images. The score must be close to 1 if the image presents the same person. Otherwise, the score must be close to 0. We calculate the score using a pair of color images: a real image $v_{i,j}$ and a synthesized image $\hat{v}_p$. We use a dedicated GAN model to generate a synthetic image $\hat{v}_p$ from the real color image. Finally, we match images using a color histogram and MSCR features (Forssén, 2007).

3.2 `ColorMatchGAN` for colorization of LWIR image

Colorization of a long-wave infrared image (LWIR) is challenging. Firstly, the colorization problem requires a reconstruction of three channels (red, green, blue) from a single channel (8-14 μm). Secondly, for most of the objects, there is no correlation between an object's color texture and its infrared texture.

Our experiments with the `pix2pix` model have demonstrated that the naïve prediction of object's color from an LWIR image provide unrealistic results for RGB and Lab Color spaces. While Lab color space increases the details of the synthesized image, the result still suffers from false textures. We hypothesized that an additional modality could improve the quality of reconstruction. An example of such modality is a color image that provides examples of desired objects texture and color. We use the `ThermalGAN` framework (Kniaz et al., 2018) as a starting point for our research.

`ThermalGAN` framework is based on the `pix2pix` GAN model and the U-Net (Ronneberger et al., 2015) Generator. We made the following modifications to the `ThermalGAN`: (1) a new input tensor that includes thermal segmentation image, relative temperatures image, and a 'hint' color image, (2) output domain in Lab color space. The resulting colorization framework is presented in Figure 2.

The 'hint' image is supplied randomly from the color gallery set. We assume that the random color image $V_{i,j}$ and the probe thermal image t_p share simile textures and colors if they are located close to each other. The generator network architecture is based on the U-Net architecture. We made the following contributions to the generator: (1) two additional convolutional layers, that increase input/output dimensions, (2) we modified stride and kernel sizes, to enable filters to cover larger objects. The original U-Net architecture consists of eight convolutional and deconvolution layers. Convolutional layers consequently compress image information until it becomes a vector with shape 1×512. Deconvolution layers increase the resolution of the feature vector until it reconstructs the size of the original image.

3.3 ColorID matching

We generate feature vectors using the approach proposed by (Bhuiyan et al., 2015). We generator feature vector for real color image $v_{i,j}$ and synthesized color image $\hat{v}_p$ using Bhattacharyya distance (Matas et al., 2002, Cheng et al., 2011). The distance is given by

$$d(v_{i,j}, \hat{v}_p)) = \beta_H \cdot d_H \Big(H_c(v_{i,j}), H_c(\hat{v}_p) \Big) +$$
$$(1 - \beta_H) \cdot d_{\text{MSCR}} \Big(f_{\text{MSCR}}(v_{i,j}), f_{\text{MSCR}}(\hat{v}_p) \Big) \quad (1)$$

4. EVALUATION

4.1 Network training

The `ColorMatchGAN` framework was trained on the VOC split of the ThermalWorld dataset (Kniaz et al., 2018) using the PyTorch library (Paszke et al., 2017). The VOC split includes indoor and outdoor scenes to avoid domain shift. The training was performed using the NVIDIA 1080 Ti GPU and took 126 hours for G_1, D_1. For network optimization, we use minibatch SGD with an Adam solver. We set the learning rate to 0.0002 with momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ similar to (Isola et al., 2017).

4.2 Multimodal colorization evaluation

We evaluate multimodal colorization performance using the independent part of the VOC split of the ThermalWorld dataset. Colorizations results are presented in Figure 3. We perform a
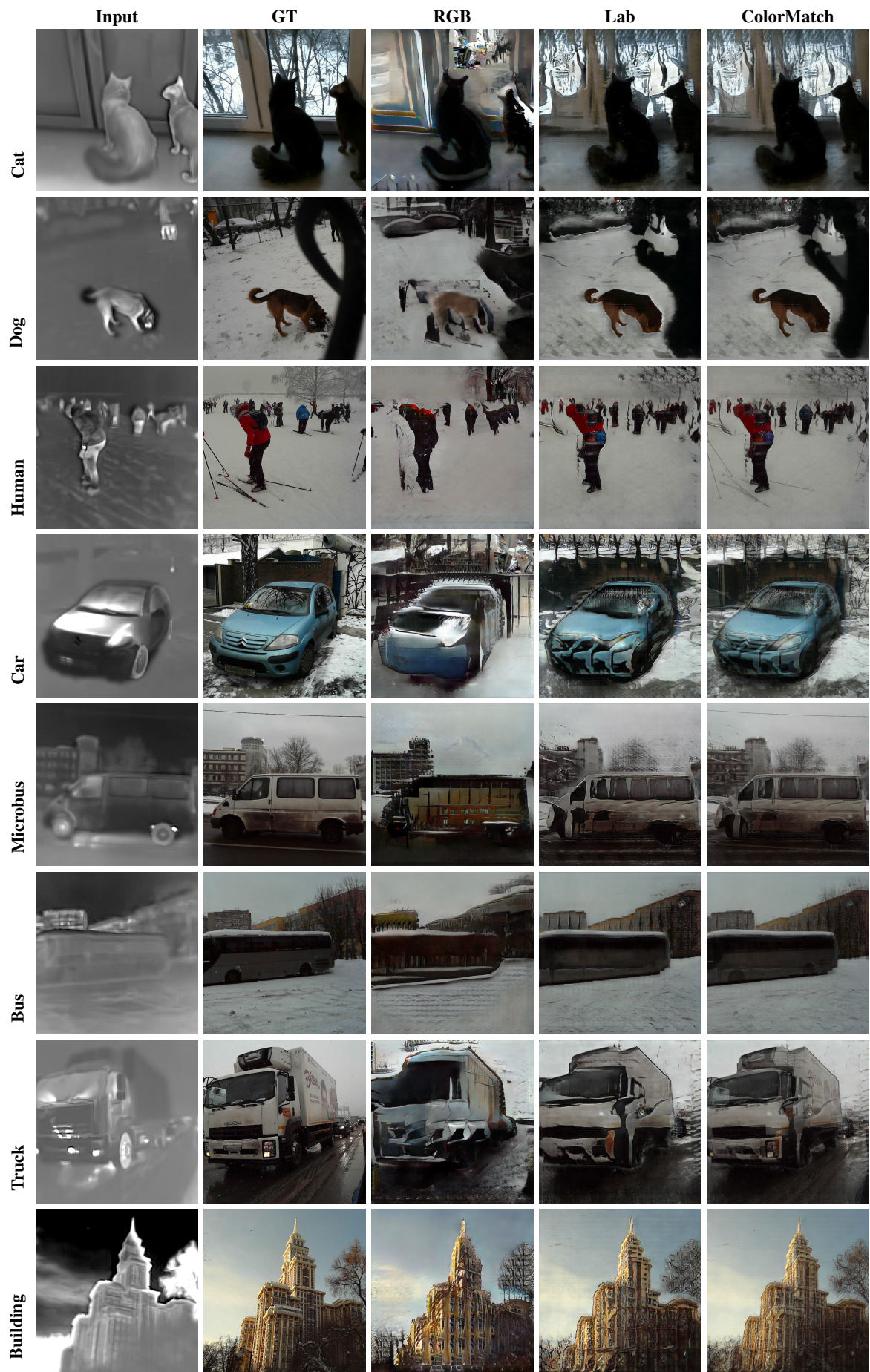
Figure 3. Qualitative colorization results for different classes and networks.

qualitative assessment of colorization results. We found that colorization results for object classes with a low diversity of surfaces colors (cats, dogs, humans) are much better than colorizations of objects with a high diversity of surface colors (cars, trucks).
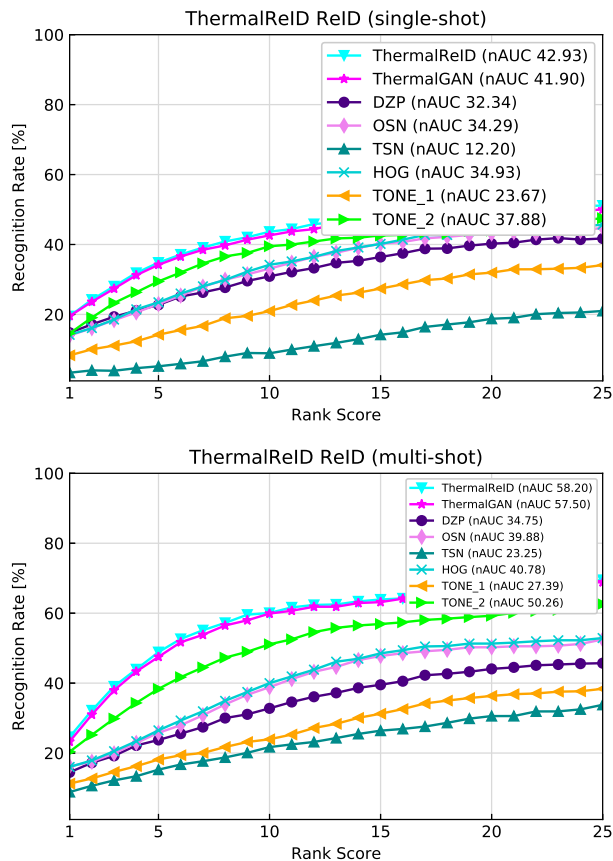
4.3 Quantitative evaluation

Figure 4. CMC plot and nAUC for evaluation of baselines and `ThermalReID` method in single-shot setting (top) and multi-shot setting (bottom).

We perform a quantitative evaluation of our `ThermalReID` framework using the ReID split of the ThermalWorld dataset. The dataset includes pairs of aligned color and thermal images for 516 person ID. We compare the performance of our `ThermalReID` framework with six baselines. We measure the performance of frameworks in terms of cumulative matching characteristic (CMC) curves and normalized area-under-curve (nAUC). The evaluation results for single-shot and multi-shot settings are presented in Figure 4. Examples of probe image colorization are shown in Figure 5.

5. CONCLUSION

We demonstrated that generative adversarial networks are capable of realistic colorization of long-wave infrared images and can reconstruct discriminative features for effective person ReID. Furthermore, our `ThermalReID` framework utilizes two-stage person re-identification for matching thermal and color images. Firstly, we supply a thermal image and an additional 'hint' color image to a GAN colorization network to generate a synthetic color probe image. The 'hint' image is sampled from the gallery set and provides the desired color distribution to the colorization network.
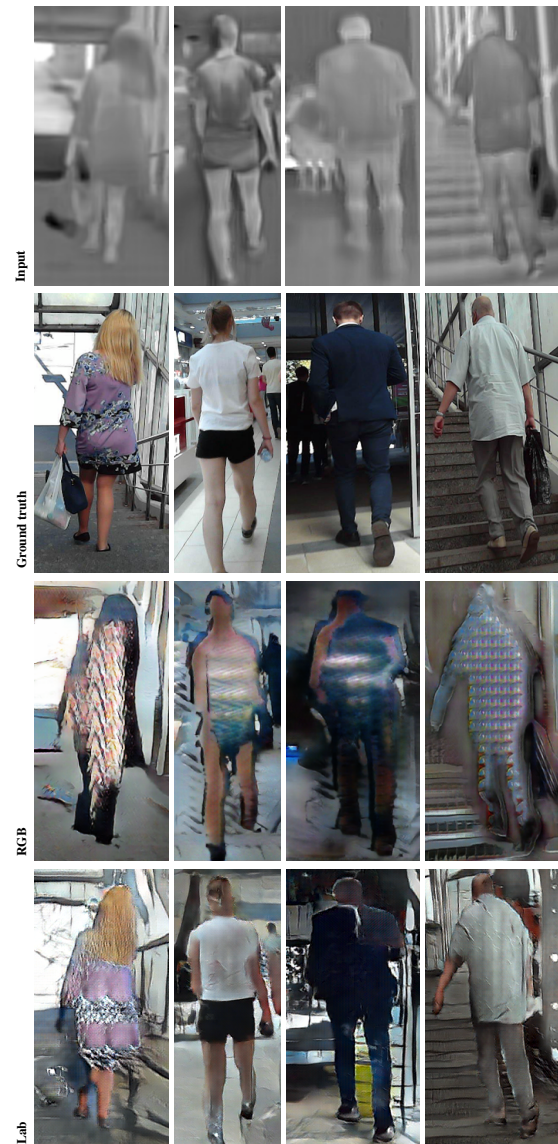

Figure 5. **Qualitative method comparison**. We compare performance of various multimodal image translation frameworks on ThermalWorld ReID dataset. We present colorization results for models trained in RGB and Lab color spaces.

Secondly, we match the resulting synthetic color probe image with a real gallery set using color histograms and MSCR features.

Our main observation is that the colorization of thermal image is an ill-posed problem and can not be performed without additional modalities. While training in Lab color space provides a significant boost in the colorization performance, the resulting images still suffer from unrealistic textures and false color proposals. Nevertheless, our `ColorMatchGAN` network can match objects in thermal and color domains and translate colors and textures from the 'hint' color image to the corresponding objects in the input thermal image.

ACKNOWLEDGEMENTS

REFERENCES

Barman, A. and Shah, S. K., 2017. Shape: A novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems. In: *The IEEE International Conference on Computer Vision (ICCV)*.

Berg, A., Ahlberg, J. and Felsberg, M., 2018. Generating visible spectrum images from thermal infrared. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Bhuiyan, A., Perina, A. and Murino, V., 2015. Person re-identification by discriminatively selecting parts and features. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Istituto Italiano di Tecnologia, Genoa, Italy, Springer International Publishing, Cham, pp. 147–161.

Bhuiyan, A., Perina, A. and Murino, V., 2018. Exploiting Multiple Detections for Person Re-Identification. *Journal of Imaging* 4(2), pp. 28.

Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L. and Murino, V., 2011. Custom pictorial structures for re-identification. In: *BMVC 2011 - Proceedings of the British Machine Vision Conference 2011*, Universita degli Studi di Verona, Verona, Italy.

Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M., 2010. Person re-identification by symmetry-driven accumulation of local features. In: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 2360–2367.

Forssén, P., 2007. Maximally stable colour regions for recognition and matching. In: *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), 18-23 June 2007, Minneapolis, Minnesota, USA*.

Guo, Y. and Cheung, N.-M., 2018. Efficient and deep person re-identification using multi-level similarity. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A., 2017. Image-to-Image Translation with Conditional Adversarial Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 5967–5976.

Kniaz, V. V., Knyaz, V. A., Hladůvka, J., Kropatsch, W. G. and Mizginov, V. A., 2018. ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-Identification in Multispectral Dataset. In: *Computer Vision – ECCV 2018 Workshops*, Springer International Publishing.

Li, W., Zhao, R. and Wang, X., 2012. Human reidentification with transferred metric learning. In: *ACCV*.

Li, W., Zhu, X. and Gong, S., 2018. Harmonious attention network for person re-identification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2002. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: *Procedings of the British Machine Vision Conference 2002*, British Machine Vision Association, pp. 36.1–36.10.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A., 2017. Automatic differentiation in pytorch.

Qian, X., Fu, Y., Jiang, Y.-G., Xiang, T. and Xue, X., 2017. Multi-scale deep learning architectures for person re-identification. In: *The IEEE International Conference on Computer Vision (ICCV)*.

Ristani, E., Solera, F., Zou, R., Cucchiara, R. and Tomasi, C., 2016. Performance measures and a data set for multi-target, multi-camera tracking. In: *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.

Saquib Sarfraz, M., Schumann, A., Eberle, A. and Stiefelhagen, R., 2018. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shen, Y., Xiao, T., Li, H., Yi, S. and Wang, X., 2018. End-to-end deep kronecker-product matching for person re-identification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Si, J., Zhang, H., Li, C.-G., Kuen, J., Kong, X., Kot, A. C. and Wang, G., 2018. Dual attention matching network for context-aware feature sequence based person re-identification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Su, C., Li, J., Zhang, S., Xing, J., Gao, W. and Tian, Q., 2017. Pose-driven deep convolutional model for person re-identification. In: *The IEEE International Conference on Computer Vision (ICCV)*.

Tang, S., Andriluka, M., Andres, B. and Schiele, B., 2017. Multiple people tracking by lifted multicut and person re-identification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wei, L., Zhang, S., Gao, W. and Tian, Q., 2018. Person transfer gan to bridge domain gap for person re-identification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wu, A., Zheng, W.-S., Yu, H.-X., Gong, S. and Lai, J., 2017. Rgb-infrared cross-modality person re-identification. In: *The IEEE International Conference on Computer Vision (ICCV)*.

Ye, M., Lan, X., Li, J. and Yuen, P. C., 2018a. Hierarchical Discriminative Learning for Visible Thermal Person Re-Identification. In: *AAAI*.

Ye, M., Wang, Z., Lan, X. and Yuen, P. C., 2018b. Visible Thermal Person Re-Identification via Dual-Constrained Top-Ranking. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, California, pp. 1092–1099.

Zhao, L., Li, X., Zhuang, Y. and Wang, J., 2017. Deeply-learned part-aligned representations for person re-identification. In: *The IEEE International Conference on Computer Vision (ICCV)*.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J. and Tian, Q., 2015. Scalable person re-identification: A benchmark. In: *The IEEE International Conference on Computer Vision (ICCV)*.