OBJECT RE-IDENTIFICATION USING MULTIMODAL AERIAL IMAGERY AND CONDITIONAL ADVERSARIAL NETWORKS

V. V. Kniaz[a,b,*], P. Moshkantsev[a]

[a] State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia
vl.kniaz@gosniias.ru
[b] Moscow Institute of Physics and Technology (MIPT), 141701, 9 Institutskiy per., Dolgoprudny, Russia

Commission II, WG II/8

KEY WORDS: object re-identification, generative adversarial networks, thermal images, airborne images

ABSTRACT:

Object Re-Identification (ReID) is the task of matching a given object in the new environment with its image captured in a different environment. The input for a ReID method includes two sets of images. The probe set includes one or more images of the object that must be identified in the new environment. The gallery set includes images that may contain the object from the probe image. The ReID task’s complexity arises from the differences in the object appearance in the probe and gallery sets. Such difference may originate from changes in illumination or viewpoint locations for multiple cameras that capture images in the probe and gallery sets. This paper focuses on developing a deep learning `ThermalReID` framework for cross-modality object ReID in thermal images. Our framework aims to provide continuous object detection and re-identification while monitoring a region from a UAV. Given an input probe image captured in the visible range, our `ThermalReID` framework detects objects in a thermal image and performs the ReID. We evaluate our `ThermalReID` framework and modern baselines using various metrics. We use the IoU and mAP metrics for the object detection task. We use the cumulative matching characteristic (CMC) curves and normalized area-under-curve (nAUC) for the ReID task. The evaluation demonstrated encouraging results and proved that our `ThermalReID` framework outperforms existing baselines in the ReID accuracy. Furthermore, we demonstrated that the fusion of the semantic data with the input thermal gallery image increases the object detection and localization scores. We developed the `ThermalReID` framework for cross-modality object re-identification. We evaluated our framework and two modern baselines on the task of object ReID for four object classes. Our framework successfully performs object ReID in the thermal gallery image from the color probe image. The evaluation using real and synthetic data demonstrated that our `ThermalReID` framework increases the ReID accuracy compared to modern ReID baselines.

1. INTRODUCTION

Object Re-Identification (ReID) is the task of matching a given object in the new environment with its image captured in a different environment. The input for a ReID method includes two sets of images. The probe set includes one or more images of the object that must be identified in the new environment. The gallery set includes images that may contain the object from the probe image. The ReID task’s complexity arises from the differences in the object appearance in the probe and gallery sets. Such difference may originate from changes in illumination or viewpoint locations for multiple cameras that capture images in the probe and gallery sets.

Many ReID methods have been developed to date for the task of person ReID (Nguyen et al., 2017c, Nguyen, Park, 2016a, Ye et al., 2018b, Ye et al., 2018a). Such methods can be broadly divided into three groups: deep learning, transfer learning, and metric learning. Deep learning methods leverage neural networks to learn end-to-end models for matching objects in the probe and gallery sets. Transform learning methods learn to translate images in the probe set to match camera viewpoint and illumination conditions in the gallery set. Metric learning methods aim to develop a function that returns a distance for a given pair of samples in the probe and gallery sets. The distance is required to be small if the pair is correct and large otherwise.

While many solutions have been proposed for the ReID task for images captured in the visible range, cross-modality ReID remains challenging. Recently a new generation of neural networks has been developed focusing on generative learning. Such networks are commonly called Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). GANs are capable of learning complex image-to-image translations such as a season change or an object transfiguration. Modern research demonstrates that GANs can learn to translate probe images to different viewpoints or different illumination conditions. To the best of our knowledge, there are no results to date in the literature regarding cross-modality object ReID from airborne images.

This paper focuses on developing a deep learning `ThermalReID` framework for cross-modality object ReID in thermal images. Our framework aims to provide continuous object detection and re-identification while monitoring a region from a UAV. Given an input probe image captured in the visible range, our `ThermalReID` framework detects objects in a thermal image and performs the ReID. Our pipeline includes four significant steps. Firstly, we translate the input probe color image to the infrared range using a GAN model (Kniaz et al., 2019). After that, we perform geo-localization of the gallery image captured by an onboard infrared camera. Specifically, we generate a semantic segmentation of the gallery image and match it with a semantic map of the landscape. Next, we perform object detection in the gallery image. We use a semantic map as an additional input modality for the object detection model to improve the object detection score. Finally, we

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-131-2021

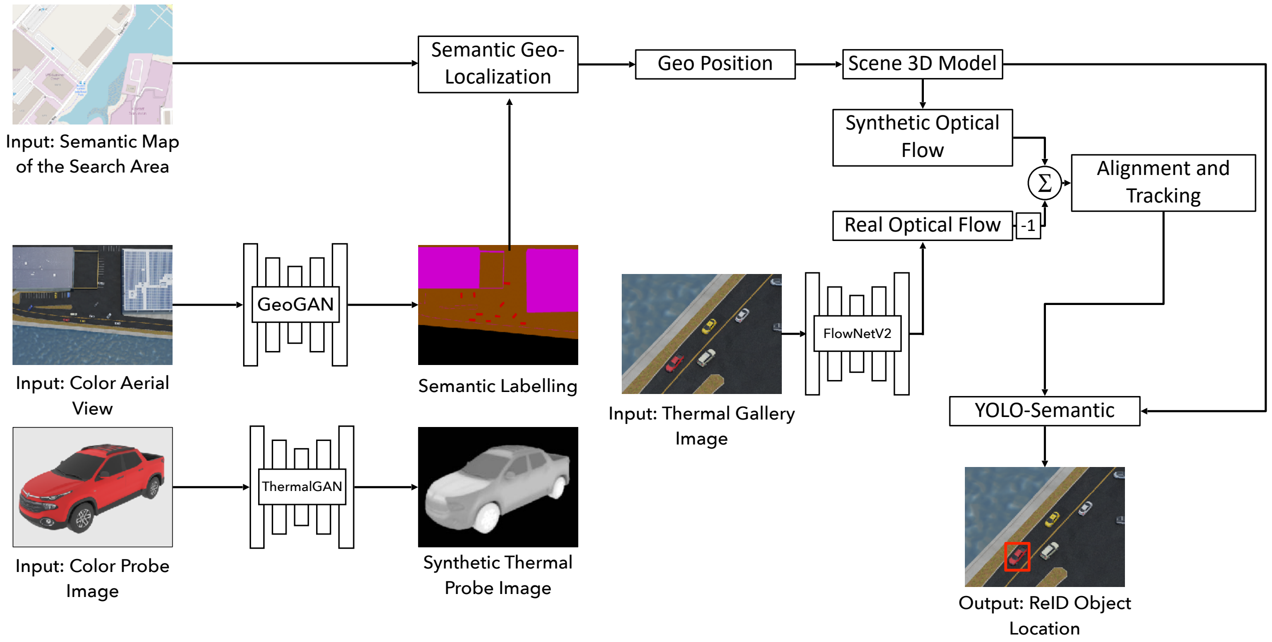

Figure 1. Overview of our proposed `ThermalReID` framework.

perform the ReID using the Bhattacharyya distance between the synthetic thermal probe image and real thermal images from the gallery set.

We developed a 3D environment to train and validate our framework. Our virtual environment includes a city scene that can be rendered in thermal and visible ranges. We include four object classes in our environment: car, person, bicycle, and dog. Using our 3D environment, we prepared a dataset consisting of 10k images divided into training and test splits. We trained our framework using the training split of the dataset and validated it using the test split and samples from the *LAERT* dataset (Knyaz, 2019).

1.1 Contributions

We present three key technical contributions:

- A unified `ThermalReID` framework for cross-modality object re-identification in thermal images.
- A new geo-localization algorithm leveraging tiled representation of the semantic map and a deep model with inverted residual blocks.
- A `YOLO-Semantic` model for object detection and localization leveraging an additional semantic labelling of the input thermal image.

2. RELATED WORK

2.1 Object Re-identification

The problem of object re-identification is important for various computer vision applications, such multi-modal image segmentation and object detection, autonomous driving, security etc. So currently it attracts attention of many researches (Farenzena et al., 2010, Gong et al., 2014, Wu et al., 2017, Wang et al., 2019, Kniaz, Knyaz, 2019, Bhuiyan et al., 2018, Prosser et al., 2008, Bhuiyan et al., 2015). New methods of re-identification allow to significantly improve the matching performance. Meanwhile, for such area as video surveillance, modern ReID systems have challenges still. Modern approaches for object re-identification can be separated into three groups (Bhuiyan et al., 2018): direct re-identification methods, metric learning methods and transform learning methods.

New transform learning-based method (Bhuiyan et al., 2018) predicts person appearance for a new camera basing on cumulative weight brightness transfer function. The method uses a robust segmentation technique to segment the human image into meaningful parts, and then matches the features extracted only from the body area. Such approach provides an improved performance of person re-identification. Multiple pedestrian detections are applied for improving the matching rate.

A specific algorithm for vehicle re-identification uses the rich annotation information available from large-scale dataset for vehicle re-identification (Wang et al., 2019). The dataset contains 137k images of 13k vehicle instances captured by cameras mounted on the board of unmanned aerial vehicle (UAV). For increasing intra-class variation, each vehicle in the dataset is captured by at least two UAVs at different locations, with diverse view-angles and flight-altitudes. The dataset contains a variety of manually labelled vehicle attributes, such as vehicle type, color, skylight, bumper, spare tire and luggage rack. In addition, the discriminative parts of each vehicle are annotated, thus contributing in distinguishing of one particular vehicle from others. The developed algorithm explicitly detects discriminative parts for each specific vehicle, and thus providing high re-identification performance, comparing with the evaluated baselines and state-of-the-art vehicle ReID approaches.

Multi-Scale and Occlusion Aware Network (Zhang et al., 2020) for UAV based imagery extracts information about vehicles in challenging conditions of arbitrary orientations, huge scale variations and partial occlusion. It consists of two parts: Multi-Scale Feature Adaptive Fusion Network (MSFAF-Net) and Regional

Attention based Triple Head Network (RATH-Net). MSFAF-Net includes a self-adaptive feature fusion module, that adaptively aggregate multi-level hierarchical feature maps, thus helping Feature Pyramid Network (FPN) to deal with the vehicle scale changes in images. The second part, Regional Attention based Triple Head Network, is used to enhance the vehicle of interest and suppress background noise caused by occlusions. Along with the developed network model, a large comprehensive vehicle dataset is collected, that contains UAV based imagery.

AI City Challenge (Chang et al., 2020) addresses to to accelerating intelligent video analysis that helps make cities smarter and safer. It is based on large city-scale real traffic data and high-quality synthetic data for evaluating developed methods. Track 2 of AI City Challenge is aimed at vehicle re-identification with real and synthetic training data. The solution (He et al., 2020) is based on a strong baseline with bag of tricks (BoT-BS) proposed in person ReID domain. It proposes a multi-domain learning method for real-world and synthetic data to train the model. The proposed Identity Mining method automatically generates pseudo labels for a part of the testing data, and performs better than the k-means clustering. Results post-processing is performed by tracklet-level re-ranking strategy with weighted features. The methods achieves 0.7322 in the mAP score on the AI City Challenge data

For overcoming the difficulties of re-identification for night-time application, additional modalities are introduced in re-identification techniques, such as infrared or long-wave infrared imagery. Using additional modalities allows to improve the robustness of matching in low-light conditions.

Exploiting of thermal camera in the field of computer vision attracts attention of many researches in re-identification field (Yilmaz et al., 2002, Davis, Keck, 2005, Knyaz, Moshkantsev, 2019). While thermal cameras serves for a significant boosting in pedestrian detection (San-Biagio et al., 2012, Xu et al., 2017) and ReID with paired color and thermal images (Nguyen et al., 2017a), cross-modality object re-identification is still a challenging task (Nguyen et al., 2017c, Nguyen, Park, 2016a, Nguyen, Park, 2016b, Nguyen et al., 2017a, Nguyen et al., 2017b). Most of problems appears from severe changes in a person appearance in color and thermal images.

To study the problem of multi-modal re-identification a set of multispectral datasets was collected in recent years (Nguyen et al., 2017a, Nguyen et al., 2017c, Nguyen, Park, 2016a, Wu et al., 2017, Ye et al., 2018a). SYSU-MM01 dataset (Wu et al., 2017) includes unpaired color and near-infrared images. RegDB dataset (Ye et al., 2018a) presents color and infrared images for evaluation of cross-modality ReID methods. Comprehensive studies of modern re-identification methods on these datasets has exposed the challenges of color-infrared matching. Simultaneously, they demonstrated the increasing performance in ReID robustness during the night-time.

Hierarchical Cross-Modality Disentanglement (Hi- CMD) method (Choi et al., 2020) automatically disentangles ID-discriminative factors and ID-excluded factors from visible-thermal images, thus reducing both intra- and cross-modality discrepancies. It uses ID-discriminative factors for robust cross-modality matching without ID-excluded factors such as pose or illumination. ID-preserving person image generation network and a hierarchical feature learning module are designed for implementing the developed approach.

Recently proposed generative adversarial networks (GAN) (Goodfellow et al., 2014) provides a background for the impressive progress in arbitrary image-to-image translation problem. We hypothesize that using a dedicated GAN framework for color-to-thermal image translation can increase color-thermal ReID performance.

2.2 Generative Adversarial Networks

Generative adversarial networks (GANs) (Goodfellow et al., 2014) exploits an antagonistic game approach, that allows to significantly increase the quality of image-to-image translation (Isola et al., 2017, Zhang et al., 2017a, Zhang et al., 2017b). pix2pix GAN framework (Isola et al., 2017) carries out arbitrary image transformations, using geometrically aligned image pairs from source and target domains. The framework successfully performs arbitrary image-to-image translations such as season change and object transfiguration. The pix2pix network model (Zhang et al., 2017a, Zhang et al., 2017b) trained to transform a thermal image of a human face to the color image allows to improve the quality of a face recognition performance in a cross-modality thermal to visible range setting.

While human face has a relatively stable temperature, color-thermal image translation for more temperature-variable objects, such as the whole human body or vehicles with an arbitrary background, is more challenging.

3. METHOD

3.1 Framework Overview

Our goal is twofold. Firstly, we would like to perform visual-based UAV geo-localization using onboard cameras. Secondly, we perform search of the probe object in thermal gallery images. Our framework works by running five deep models. Overview of the proposed ThermalReID framework is presented in Figure 1.

Our semantic geo-localization approach is inspired by tiled map representations. Our algorithm estimates the geographic coordinates (ϕ, λ) of the UAV given an input color or thermal image of the scene and rough approximate of the current geo-location. The algorithm leverages a distance learning technique. Firstly, we perform a semantic segmentation S of an input image A. Secondly, we use a deep model to estimate a distance between the generated semantic labelling S and semantic tiles from the onboard geographic dataset. We use a MobileNetV2 (Sandler et al., 2018) model for the distance estimation task.

The main object re-identification algorithm leverages the precise coordinates of the UAV estimated by the localization algorithm. It performs object re-identification in three steps. Firstly, given the estimated geo-coordinates, we generate a semantic labelling S_G of the thermal input gallery image B_G. We perform precise alignment of the semantic labelling using a differential optical flow estimation approach (Kniaz, 2018b). After that, we generate a synthetic thermal probe image B_P using image the input color probe image and a ThermalGAN conditional adversarial network (Kniaz et al., 2019). We use the thermal input gallery image B_G and its semantic labelling S_G as the input for our object detection YOLO-Semantic model. Finally, we measure distance between each candidate object detected by our YOLO-Semantic model and the synthetic thermal probe image B_P. We perform ReID by selecting the candidate object with the smallest distance (Figure 2).

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-131-2021 | © Author(s) 2021. CC BY 4.0 License.

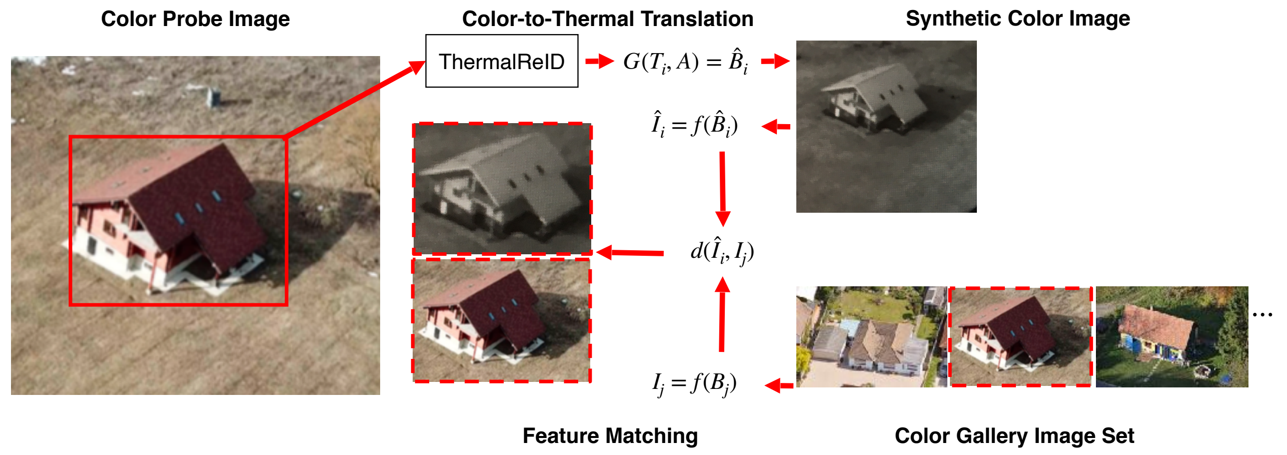

Figure 2. Cross-modality object ReID using a conditional GAN model.

3.2 Semantic Geo-Localization

We perform semantic geo-localization using two deep models and a semantic map of the search area. To optimize the search performance, we use tiled representation of the semantic map (Figure 3).

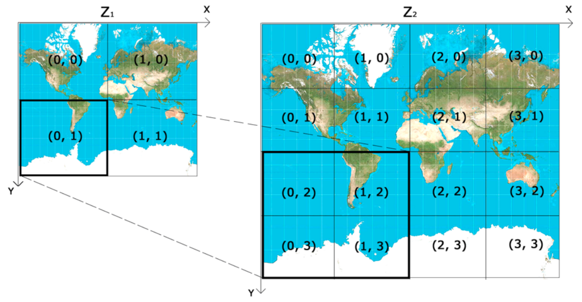

Figure 3. Tiled representation of the semantic map.

Our aim is training an algorithm that estimates the geographic coordinates (ϕ, λ) of the UAV given an input color or thermal image of the scene and rough approximate of the current geolocation. We use a `MobileNetV2` (Sandler et al., 2018) model as a staring point for our research. Our approach is twofold. We perform a semantic segmentation S of an input image A using a `GeoGAN` (Kniaz, 2018a) model. After that, we use a `MobileNetV2` model to estimate a distance between the generated semantic labelling S and semantic tiles from the onboard geographic dataset. The closest matching tile gives the current coordinates of the UAV.

3.3 YOLO-Semantic

Our `YOLO-Semantic` model is inspired by the `YOLOv3` model. We consider three domains: the thermal image domain $\mathcal{B} \in \mathbb{R}^{W \times H}$, the semantic labelling domain $\mathcal{S} \in \mathbb{R}^{K \times W \times H}$, where K is the number of semantic classes predicted by the `GeoGAN` model, and the bounding box predictions domain $\mathcal{T} \in \mathbb{R}^{(5+K) \times U \times V}$, where U, V is the number of cells in the output of our `YOLO-Semantic` model. We aim training a mapping $Y : (A_P, S_P) \rightarrow T$ from a pair on input tensors $B_P \in \mathcal{B}$ and $S_P \in \mathcal{S}$ to the bounding boxes tensor $T \in \mathcal{T}$. Details of the proposed architecture are presented in Table 1.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. The `YOLO-Semantic` architecture.

3.4 ThermalReID

We follow the general approach for thermal ReID proposed in (Kniaz et al., 2019). We use Bhattacharyya distance to compute a distance between two signatures using temperature histograms and MSER distance (Matas et al., 2002, Cheng et al., 2011)

$$d(\hat{I}_i, I_j) = \beta_H \cdot d_H(H_t(\hat{B}_i), H_t(B_j)) \\ + (1 - \beta_H) \cdot d_{\text{MSER}}(f_{\text{MSER}}(\hat{B}_i), f_{\text{MSER}}(B_j)), \quad (1)$$

where d_H is a Bhattacharyya distance, d_{MSER} is a MSER distance (Matas et al., 2002), and β_H is a calibration weight parameter. Overview of the proposed `ThermalReID` framework is presented in Figure 1.

4. EXPERIMENTS

We evaluate our ThermalReID framework and modern baselines using various metrics. We use the IoU and mAP metrics for the object detection task. We use the cumulative matching characteristic (CMC) curves and normalized area-under-curve (nAUC) for the ReID task. The evaluation demonstrated encouraging results and proved that our ThermalReID framework outperforms existing baselines in the ReID accuracy. Furthermore, we demonstrated that the fusion of the semantic data with the input thermal gallery image increases the object detection and localization scores.

4.1 Network Training

We trained our models and baselines using train split of the *LAERT* dataset. Training of the ThremalGAN model took 68 hours. We optimize network using minibatch SGD with an Adam solver. We use a learning rate of 0.0002, and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ similar to (Isola et al., 2017).

4.2 Quantitative Evaluation

We evaluate our model using *SemanticVoxels* (Kniaz et al., 2020) and *LAERT* (Knyaz, 2019) datasets. We evaluate our model and baselines quantitatively in terms of cumulative matching characteristic (CMC) curve and normalized area-under-curve (nAUC). We compare our model to two model baselines. The VRAI (Wang et al., 2019) model leverages two deep models. A joint performance of the ResNet-50 (He et al., 2016) model and the YOLOv2 (Redmon, Farhadi, 2017) model allows authors to achieve the state-of-the art in vehicle ReID in aerial images. The BoT-BS (Wang et al., 2019) model uses a strong baseline with bag of tricks commonly used for the person ReID task. The results of the evaluation are presented in Table 2.

Methods	**LAERT**					
	r = 1	**r = 5**	**r = 10**	**r = 15**	**r = 20**	**nAUC**
VRAI	12.11	22.71	32.66	39.86	41.53	33.23
BoT-BS	14.12	21.21	29.12	34.55	38.98	31.17
ThermalReID	**18.21**	**32.43**	**41.56**	**44.31**	**46.34**	**39.11**

Table 2. Experiments on *LAERT* dataset.

5. CONCLUSION

We developed the ThermalReID framework for cross-modality object re-identification. We evaluated our framework and two modern baselines on the task of object ReID for four object classes. Our framework successfully performs object ReID in the thermal gallery image from the color probe image. The evaluation using real and synthetic data demonstrated that our ThermalReID framework increases the ReID accuracy compared to modern ReID baselines.

REFERENCES

Bhuiyan, A., Perina, A., Murino, V., 2015. Person re-identification by discriminatively selecting parts and features. *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, 147–161.

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

Chang, M.-C., Chiang, C.-K., Tsai, C.-M., Chang, Y.-K., Chiang, H.-L., Wang, Y.-A., Chang, S.-Y., Li, Y.-L., Tsai, M.-S., Tseng, H.-Y., 2020. Ai city challenge 2020 - computer vision for smart transportation applications. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

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

Choi, S., Lee, S., Kim, Y., Kim, T., Kim, C., 2020. Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Davis, J. W., Keck, M. A., 2005. A two-stage template approach to person detection in thermal imagery. *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, 1, IEEE, 364–369.

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

Gong, S., Cristani, M., Yan, S., 2014. *Person Re-Identification (Advances in Computer Vision and Pattern Recognition)*. Springer London, London.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 2672–2680.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 770–778.

He, S., Luo, H., Chen, W., Zhang, M., Zhang, Y., Wang, F., Li, H., Jiang, W., 2020. Multi-domain learning and identity mining for vehicle re-identification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2485–2493.

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

Kniaz, V. V., 2018a. Conditional GANs for semantic segmentation of multispectral satellite images. L. Bruzzone, F. Bovolo (eds), *Image and Signal Processing for Remote Sensing XXIV*, 10789, International Society for Optics and Photonics, SPIE, 259 – 267.

Kniaz, V. V., 2018b. Optical flow-based filtering for effective presentation of the enhanced vision on a HUD. P. Schelkens, T. Ebrahimi, G. Cristbal (eds), *Optics, Photonics, and Digital Technologies for Imaging Applications V*, 10679, International Society for Optics and Photonics, SPIE, 162 – 171.

https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-131-2021

Kniaz, V. V., Knyaz, V. A., 2019. Chapter 6 - multispectral person re-identification using gan for color-to-thermal image translation. M. Y. Yang, B. Rosenhahn, V. Murino (eds), *Multimodal Scene Understanding*, Academic Press, 135–158.

Kniaz, V. V., Knyaz, V. A., Hladůvka, J., Kropatsch, W. G., Mizginov, V., 2019. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. L. Leal-Taixé, S. Roth (eds), *Computer Vision – ECCV 2018 Workshops*, Springer International Publishing, Cham, 606–624.

Kniaz, V. V., Knyaz, V. A., Remondino, F., Bordodymov, A., Moshkantsev, P., 2020. Image-to-voxel model translation for 3d scene reconstruction and segmentation. A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (eds), *Computer Vision – ECCV 2020*, Springer International Publishing, Cham, 105–124.

Knyaz, V., 2019. Multimodal data fusion for object recognition. E. Stella (ed.), *Multimodal Sensing: Technologies and Applications*, 11059, International Society for Optics and Photonics, SPIE, 198 – 209.

Knyaz, V. A., Moshkantsev, P. V., 2019. JOINT GEOMETRIC CALIBRATION OF COLOR AND THERMAL CAMERAS FOR SYNCHRONIZED MULTIMODAL DATASET CREATING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W18, 79–84. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W18/79/2019/.

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

Nguyen, D., Hong, H., Kim, K., Park, K., 2017a. Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras. *Sensors*, 17(3), 605–29.

Nguyen, D., Kim, K., Hong, H., Koo, J., Kim, M., Park, K., 2017b. Gender Recognition from Human-Body Images Using Visible-Light and Thermal Camera Videos Based on a Convolutional Neural Network for Image Feature Extraction. *Sensors*, 17(3), 637–22.

Nguyen, D., Park, K., 2016a. Body-Based Gender Recognition Using Images from Visible and Thermal Cameras. *Sensors*, 16(2), 156–21.

Nguyen, D., Park, K., 2016b. Enhanced Gender Recognition System Using an Improved Histogram of Oriented Gradient (HOG) Feature from Quality Assessment of Visible Light and Thermal Images of the Human Body. *Sensors*, 16(7), 1134–25.

Nguyen, D. T., Hong, H. G., Kim, K. W., Park, K. R., 2017c. Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*, 17(3), 605.

Prosser, B., Gong, S., Xiang, T., 2008. Multi-camera matching using bi-directional cumulative brightness transfer functions. *BMVC 2008 - Proceedings of the British Machine Vision Conference 2008*, Queen Mary, University of London, London, United Kingdom, British Machine Vision Association, 64.1–64.10.

Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 6517–6525.

San-Biagio, M., Ulas, A., Crocco, M., Cristani, M., Castellani, U., Murino, V., 2012. A multiple kernel learning approach to multimodal pedestrian classification. *Pattern Recognition (ICPR), 2012 21st International Conference on*, IEEE, 2412–2415.

Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., Chen, L., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, 4510–4520.

Wang, P., Jiao, B., Yang, L., Yang, Y., Zhang, S., Wei, W., Zhang, Y., 2019. Vehicle re-identification in aerial imagery: Dataset and approach. *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, 460–469.

Wu, A., Zheng, W.-S., Yu, H.-X., Gong, S., Lai, J., 2017. RGB-Infrared Cross-Modality Person Re-Identification. *The IEEE International Conference on Computer Vision (ICCV)*.

Xu, D., Ouyang, W., Ricci, E., Wang, X., Sebe, N., 2017. Learning Cross-Modal Deep Representations for Robust Pedestrian Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 4236–4244.

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

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

Yilmaz, A., Shafique, K., Shah, M., 2002. Tracking in airborne forward looking infrared imagery. 21, 623-635.

Zhang, H., Patel, V. M., Riggan, B. S., Hu, S., 2017a. Generative adversarial network-based synthesis of visible faces from polarimetrie thermal faces. *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 100–107.

Zhang, T., Wiliem, A., Yang, S., Lovell, B. C., 2017b. TV-GAN: Generative Adversarial Network Based Thermal to Visible Face Recognition.

Zhang, W., Liu, C., Chang, F., Song, Y., 2020. Multi-Scale and Occlusion Aware Network for Vehicle Detection and Segmentation on UAV Aerial Images. *Remote Sensing*, 12(11). https://www.mdpi.com/2072-4292/12/11/1760.