Unsupervised object detection via LWIR/RGB translation

Rachael Abbott
ECIT
Queen's University Belfast
rabbott02@qub.ac.uk

Neil M. Robertson
ECIT
Queen's University Belfast
n.robertson@qub.ac.uk

Jesus Martinez del Rincon
ECIT
Queen's University Belfast
j.martinez-del-rincon@qub.ac.uk

Barry Connor
Thales UK
Linthouse road, Glasgow
barry.connor@uk.thalesgroup.com

Abstract

In this work, we present two new methods to overcome the lack of annotated long-wavelength infrared (LWIR) data by exploiting the abundance of similar RGB imagery. We introduce a novel unsupervised adaptation to the cycleGAN architecture for translating non-corresponding LWIR/RGB datasets. Our ultimate goal is high detection rates in the real LWIR imagery using only RGB labelled imagery for training detection algorithms. In our first experiment, we translate LWIR imagery to RGB, allowing us to use an RGB trained detection algorithm. We, thereby remove the need for labelled LWIR imagery for training detection algorithms. Experimental results show that our adaption helps to create synthetic RGB imagery with higher detection rates across two different datasets. We also find that combining the synthetic RGB and real LWIR imagery produces higher F1 scores on the RGB trained detection network. In our second experiment, we translate RGB to LWIR to fine-tune a network for detection in real LWIR imagery. This method produces the highest F1 scores out of the two methods with detection reaching up to 85.6%.

1. Introduction

Despite recent advances in detection and translation algorithms for RGB images, infra-red (IR) is still relatively understudied due to the scarcity of the data. There is a severe lack of labelled LWIR data, specifically for high performing, high-resolution thermal sensors. Detection in IR imagery is becoming increasingly important in the defence and security domain for 24-hour capability because visible light is non-optimal in extreme weather situations (e.g. fog/heavy rain/low light) and at night. Visible images captured by RGB sensors consist of *reflected* energy and provide information similar to what the human eye would process. IR images consist of *emitted* energy from objects and potential absorption/emission from the background. Long-wavelength IR (LWIR, 8 – 12 μm) imaging can be particularly useful for the observance of living beings, which is helpful in defence scenarios.

LWIR/RGB image translation could provide a solution to the lack of annotated LWIR datasets by taking advantage of a large corpus of existing labelled RGB data. One of the significant issues regarding translation research is that there exists very little work translating between modalities (i.e. IR to RGB). Also, there are no studies which focus on real-life non-corresponding (does not line up in space or time) imagery taken in cluttered urban scenes. In this paper, we present two approaches for detection in real LWIR and synthetic RGB imagery in an unsupervised manner.

Our main contributions are summarised as follows:

- We propose an unsupervised adaptation to the cycleGAN network [32]. A novel loss is introduced to minimise the appearance between the object of interest detections in the real and recovered RGB/LWIR imagery.
- We show LWIR to RGB translation is possible between non-corresponding datasets containing high resolution short-medium range targets (objects at 0-50 metres from camera) taken in real-life uncontrolled cluttered urban scenes. Figure 1 shows some examples of our synthetic RGB imagery. The synthetic RGB images can be used for detection in RGB trained algorithms. Moreover, this paper is the first to fuse synthetic RGB and real LWIR further improving F1 scores.
- We create synthetic LWIR and show, for the first time, that this imagery can be used for effective training of

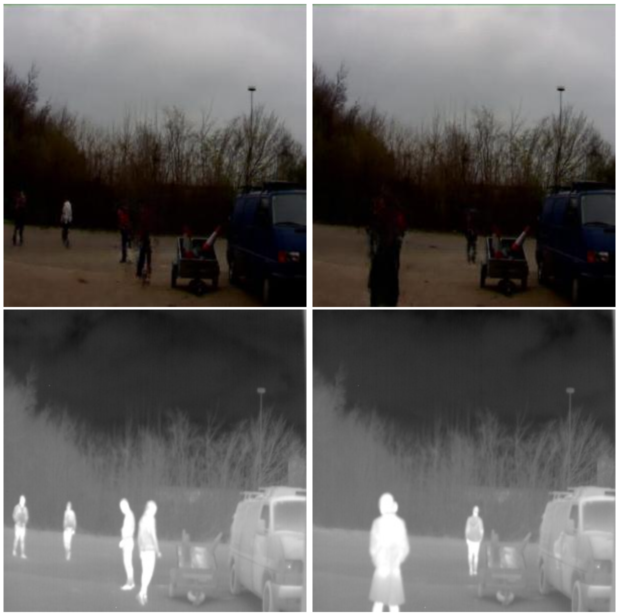

Figure 1: Thales synthetic RGB imagery (top row) translated from real LWIR imagery (bottom row) using our adapted cycleGAN network.

deep learning detectors for improved detection in real LWIR data.

This paper develops a new research area which has emerged from the need for detection in LWIR imagery and would have a real impact on the applicability of modern deep learning (DL) modelling in defence and security applications.

2. Related work

We review two main bodies of research relevant to our work here, object detection in IR imagery and image translation networks.

2.1. IR detection

When dealing with small datasets for training, transfer learning (TL) followed by fine-tuning (FT) is the most common and successful strategy [23]. TL struggles to address more radical differences between modalities meaning that detection rates are low unless FT is implemented. However, FT is not always possible due to lack of labelled data. Paper [2] propose a novel architecture for LWIR detection in an unsupervised manner for the first time. They use adaptation techniques (previously used within the RGB domain for classification tasks) for creating modality invariant features in a faster RCNN network for improving LWIR detection. The work proposed in this paper uses the mean squared error to reduce the distance between feature maps produced from RGB and LWIR imagery. This is one of only a small number of papers [4, 25, 1] which cover object detection for defence scenarios compared with other applications, given the difficulty in acquiring appropriate data. In addition to these methods, some researchers have explored fusing IR with RGB [5, 20] and IR with depth imagery to improve detection results [12, 8].

2.2. Image translation networks

For the last few years, researchers have looked at generative adversarial networks (GANs) to adapt source imagery to a different target domain [6, 9, 18, 11, 15, 7, 16]. GANs can be useful when there is only a small dataset available for training classification algorithms as we can use synthetic imagery for training purposes. These GAN architectures currently produce the best results and can be categorized into supervised and unsupervised. Supervised models use paired images, lined up in time and space, which are very difficult to collect in real scenarios. Unsupervised models alleviate the difficulty of obtaining data pairs as unpaired data is much easier to assemble. Many different architecture choices are available for unsupervised methods [32, 29, 13]. Regardless of the architecture choices, these unsupervised methods tend to struggle when significant differences between the domains need to be covered.

Another issue with translation methods is that in the past, researchers have mainly focused on translation between two domains in the RGB modality. However, recently several researchers have had success with near-infrared (NIR) to RGB translation [28, 27, 22, 17, 21, 26]. Also, CNN based translation methods have translated IR images to RGB for facial recognition [30, 24]. In addition, IR2VI [19] addresses mid-wave IR to grey-scale translation for detection of long-range targets. They introduce a structure connection to overcome the incorrect mapping problem and help translate the background more accurately. Researchers [31, 3] have had some success addressing a much harder problem of LWIR to RGB translation using the Kaist dataset [14] which consists of imagery taken in cluttered urban scenes.

Although the translation approaches described have had positive outcomes, none of them have addressed unsupervised translation between LWIR and RGB for object detection in cluttered scenes with short-medium range targets. Translating imagery for detection with short-medium range targets is a much more difficult task than for long range targets. This is because there are more discriminative details to translate, which are crucial for the subsequent localization and classification. We aim to address these gaps in research and show that unsupervised translation between LWIR and RGB for detection of short-medium range targets is possible.

In this paper, we propose to use a novel cycleGAN architecture for image translation between LWIR and RGB imagery. Our network is capable of specifically enhancing the translation of objects between LWIR and RGB in an unsupervised manner using a detection algorithm trained using only RGB labelled imagery. In addition to testing the synthetic RGB imagery with an RGB trained detection network, we also test the fusion of synthetic RGB and real LWIR imagery. Furthermore, we fine-tune the RGB trained network with synthetic LWIR imagery for detection in real LWIR.

Section 3 details our technical approach, section 4 discusses the experiments and results, and we conclude with section 5.

3. Unsupervised Object-specific CycleGAN adaption

In this section we first describe the cycleGAN network and then our novel unsupervised adaptation to enhance the object quality in adapted images.

3.1. CycleGAN network

We aim to create synthetic RGB imagery and LWIR imagery using a non-corresponding dataset. This means that the imagery is not lined up in time, viewpoint or space, as shown in Figure 2. We use the cycleGAN as our baseline architecture as it is unsupervised. The training is achieved using two sets of n images $I_i^{RGB} : i = 1, 2, .., n$ and $I_i^{IR} : i = 1, 2, .., n$ where I_i^{RGB} and I_i^{IR} are non-corresponding. To translate I_i^{IR} to its RGB version, we use a generator G and to translate I_i^{RGB} to IR we use generator F:

$$G : IR \rightarrow RGB$$

$$F : RGB \rightarrow IR$$

The synthetic RGB and IR images can be represented by $G(I_i^{IR})$ and $F(I_i^{RGB})$ respectively. Each generator has a corresponding discriminator, which seeks to tell the difference between the real and synthetic imagery:

$$D_{RGB} : \text{distinguishes } I^{RGB} \text{ from } G(I^{IR})$$

$$D_{IR} : \text{distinguishes } I^{IR} \text{ from } F(I^{RGB})$$

The training is done using the two sets of images by solving the following adversarial losses:

$$Adv_{loss}(G, D_{RGB}, I^{IR}) = \frac{1}{n}\sum_{i=1}^{n}(1 - D_{RGB}(G(I_i^{IR})))^2 \tag{1}$$

$$Adv_{loss}(F, D_{IR}, I^{RGB}) = \frac{1}{n}\sum_{i=1}^{n}(1 - D_{IR}(F(I_i^{RGB})))^2 \tag{2}$$

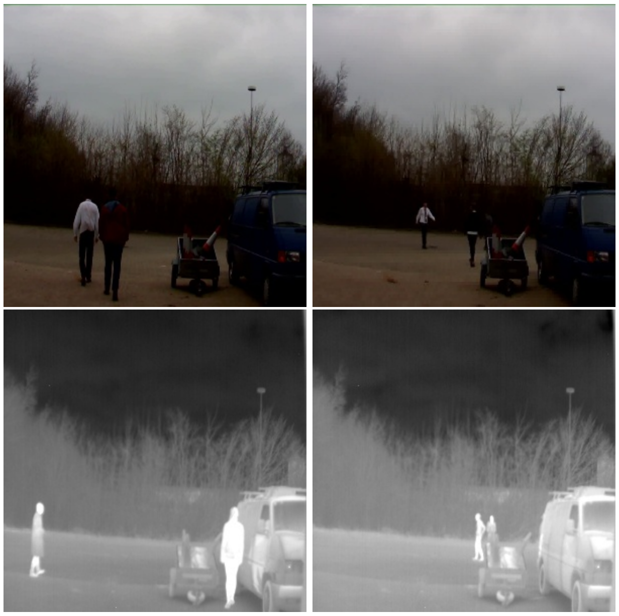

Figure 2: Examples of non-corresponding RGB (top row) and LWIR (bottom row) imagery from the Thales dataset.

These two losses alone are not sufficient to produce good quality images. The adversarial losses enforce the generated output to the appropriate modality. In other words, a generator could output an RGB image that was an excellent example of that modality, but would not have the structure or object classes of the input IR image. The cycle consistency loss addresses this issue by comparing the real image with the recovered image. The recovered image is produced when the real image is converted to one modality and then back again. The recovered image should be similar to the original image and the cycle loss enforces that $G(F(I^{RGB})) \approx I^{RGB}$ and $F(G(I^{IR})) \approx I^{IR}$. The cycle consistency loss is:

$$\begin{aligned} &Cycle_{loss}(G, F, I^{RGB}, I^{IR}) = \\ &\frac{1}{n}\sum_{i=1}^{n}[|G(F(I_i^{RGB})) - I_i^{RGB}| \\ &\quad +|F(G(I_i^{IR})) - I_i^{IR}|] \end{aligned} \tag{3}$$

The final cycleGAN loss is the adversarial losses added to a weighted cycle consistency loss by hyper-parameter α which is normally set to 10 [32].

$$CycleGAN_{loss} = TotalAdv_{loss} + \alpha Cycle_{loss} \tag{4}$$

where $TotalAdv_{loss} = Adv_{loss}(G, D_{RGB}, I^{IR}) + Adv_{loss}(F, D_{IR}, I^{RGB})$.

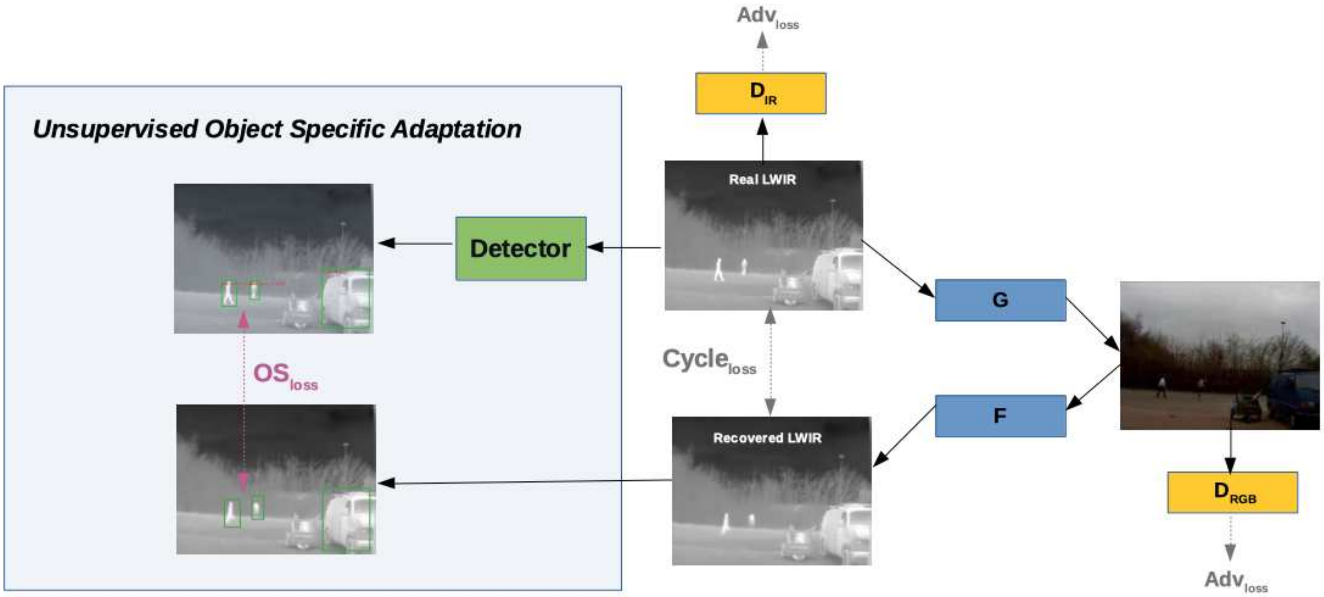

Figure 3: Overall architecture of our proposed adapted cycleGAN framework. Note this illustration needs to be duplicated for training in the cycleGAN framework with the RGB image as input.

3.2. Object-specific CycleGAN network

While the cycleGAN loses may be sufficient to adapt to similar domains and/or corresponding images, we observe that they struggle to solve non-corresponding adaptation across modalities, particularly on the fine grain details of the objects of interest. We postulate that the imagery produced using the original cycleGAN architecture does not translate the classes to a high enough standard to be detected. In this scenario, we can constrain the adaptation to focus on the objects of interest which will be later detected, since detection is our final goal. The emphasis on the object of interest regions during adaptation is made by adding a new loss function, which we call Object Specific Loss (OS Loss), and is minimised simultaneously with the other losses.

Since we aim to find where the objects are in an unsupervised manner, without requiring any label, we use the output of the detection algorithm rather than ground truth. The real LWIR and RGB images pass through an RGB-trained faster RCNN network adapted for LWIR in an unsupervised manner described in [2]. Although this network is adapted for LWIR imagery, it maintains its RGB performance, so it can be used in both modalities simultaneously. We use regions of interest (ROI) produced by faster RCNN to locate areas where objects are in the real image. We then use the ROIs to minimise the difference between these areas in the real and recovered images. This will help improve the translation of these objects to the generated image. The L1 loss is used to compare the pixel values of the bounding boxes areas for the real and recovered images. Figure 3 shows the architecture of our proposed network, which includes the original cycleGAN network plus our new unsupervised loss based on faster-RCNN network detections.

The L1 loss is multiplied by a hyperparameter λ and added to the original cycleGAN loss. The region of interest function is given as ROI(). The total OS loss is given, as shown:

$$OS_{loss} = \frac{1}{n}\sum_{i=1}^{n}|ROI(G(F(I_i^{RGB}))) - ROI(I_i^{RGB})| + \frac{1}{n}\sum_{i=1}^{n}|ROI(F(G(I_i^{IR}))) - ROI(I_i^{IR})| \tag{5}$$

Thus, the final loss function of our proposed approach is:

$$AdaptedCycleGAN_{loss} = TotalAdv_{loss} + \alpha Cycle_{loss} + \lambda OS_{loss} \tag{6}$$

4. Experimental Results

In this section, we detail the datasets used, the experimental setup and our two approaches for producing synthetic RGB/LWIR imagery for detection.

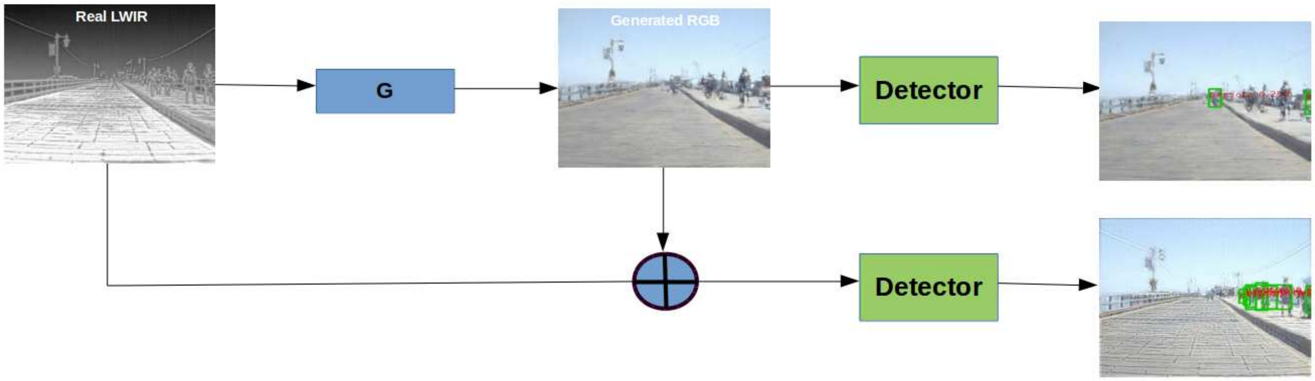

Figure 4: Overview of our LWIR to RGB translation for detection approach. The real LWIR image is translated to RGB using generator G before being passed to the detector. In addition, we combine this generated image with the original LWIR image for detection.

4.1. Datasets

Two datasets are used to validate our approach, namely, Thales and FLIR [10]. The Thales dataset contains 233 LWIR images and 233 RGB images, which are non-corresponding. The RGB camera has a resolution of 1024x768 and Thales LWIR imager has a resolution of 640x480. The background remains constant, and people move in and out of the field of view of the camera. Occlusions do occur, and there is little clutter. This dataset has only person class labels. The dataset was recorded during the day in Glasgow UK with observation distances from 1-50 metres.

In our second dataset, we choose 200 LWIR of size 640x512 and 200 RGB of size 1280x1024 from the FLIR online dataset. The RGB and LWIR images are taken with FLIR BlackFly imager and FLIR IR Tau2 imager respectively. The FLIR dataset is recorded using a dashboard camera on a vehicle driving the streets of Santa Barbara CA and, therefore has a constantly changing background with high numbers of objects per frame present and cluttered scenes. We take a selection of daytime imagery and use object classes person and vehicle. The RGB and LWIR imagery is corresponding; however, we do make use of this in our paper as we aim to replicate a realistic scenario.

4.2. Experimental setup

In this paper, we aim to achieve detection; thus, it is possible to assess our different methods by performing object detection and comparing F1 scores. We choose faster RCNN with VGG16 network trained on 70% of RGB imagery from the dataset as our baseline detection network.

We train the cycleGAN networks using 70% of the images in the dataset; the remaining 30% is used for testing. The LWIR image, which has one channel, is used three times to create a three-channel image to match the number of channels in the RGB image. We resize the images to size 300x300 and use batch size=1 for computational reasons. We use hyperparameter values $\alpha = 10$ and $\lambda = 0.1$ in equation 6 in all our experiments, which were empirically determined. All cycleGAN networks are trained using a learning rate of 0.0002 and decay epoch of 100. The method was developed using the PyTorch deep learning toolbox. We used an NVIDIA GeForce GTX 1080 Ti Founders Edition, an Intel Xeon E3-1620 v3 CPU (Quad-Core 3.5GHz) and 32 GB of Memory.

4.3. Experiment 1: LWIR to RGB translation for detection

Figure 4 shows our approach to the first experiment translating LWIR to RGB for detection. After training the RGB/LWIR translation network, we create synthetic RGB imagery using the LWIR to RGB generator network G. We then pass this imagery through an RGB trained detector to produce an F1 score. Also, we propose an additional approach where we fuse the generated RGB with its original LWIR image using a simple pixel per pixel approach. We test these synthetic RGB images and their fusion with real LWIR using our RGB trained detector to produce F1 scores. For comparison purposes, we validate the synthetic imagery produced from our adapted cycleGAN versus the synthetic imagery produced from the conventional cycleGAN and UNIT translation [13] networks, as well as the use of real LWIR imagery as input.

Figure 5 displays some of the synthetic RGB images created using the different GAN methods. Table 1 shows the F1 scores produced when using the Thales and FLIR dataset. Our cycleGAN adaptation network creates the best

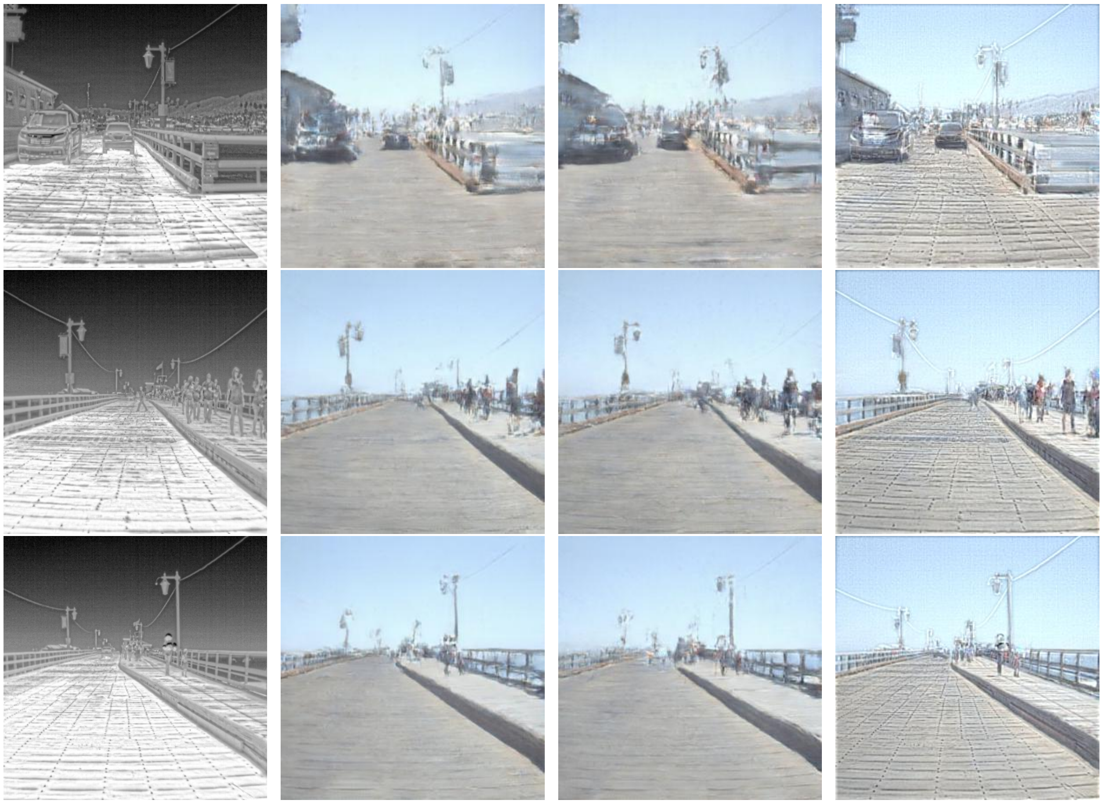

Figure 5: Figure displaying real FLIR imagery, its translated RGB image (using cycleGAN and our adaptation) and the fusion of the real LWIR and generated RGB image.

quality RGB imagery for detection and is consistent across two different datasets with respect to the F1 score. We improve the F1 score on average by 2.5% for the Thales dataset and 1.9% for the FLIR dataset with respect to the original cycleGAN network.

GAN used:	Imagery tested	Thales	FLIR
-	Real LWIR	51.3	20.0
UNIT GAN	Fake RGB	24.0	0.1
CycleGAN	Fake RGB	64.6	7.7
Ours	Fake RGB	67.1	9.5
Ours	Fake RGB + Real LWIR	**73.0**	**26.3**

Table 1: F1 scores produced from LWIR to RGB translation for detection experiment.

The Thales dataset produces the best quality synthetic RGB imagery out of the two datasets with a maximum F1 score of 67.1%. This dataset has a constant background and a maximum of 6 objects per frame which therefore helps to produce better quality imagery. The FLIR dataset produces a much lower overall F1 score of 9.5% as this dataset has many more objects, clutter, occlusions and constantly changing backgrounds. Fusing both the synthetic RGB and real LWIR image together for detection in the RGB trained network helps improve detection in the Thales dataset to 73.0% and 26.3% in the FLIR. Fusing these images together helps to maintain the structure of objects present in the image and thereby, improves F1 scores.

4.4. Experiment 2: Fine tuning with synthetic LWIR

Figure 6 explains the approach to our fine-tuning with synthetic LWIR experiment. We create synthetic LWIR imagery using the RGB to LWIR generator F. This allows us to use LWIR-wise imaginary while reusing the RGB labelling since the objects remain on the same locations after adaptation. The RGB trained detection network is fine-tuned with 70% of the synthetic LWIR with RGB ground truth labels. We train using the person class for the Thales dataset and person and vehicle for the FLIR dataset. We test 30% of the real LWIR imagery. The RGB trained network is also fine-tuned with real LWIR for comparison purposes only since

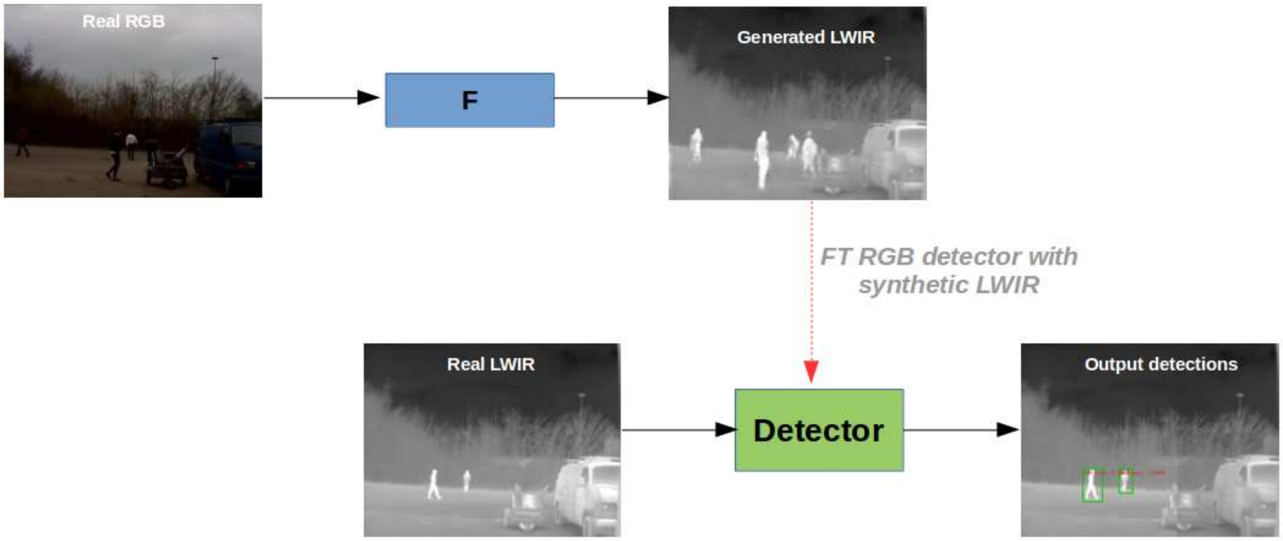

Figure 6: Overview of our fine-tuning with synthetic LWIR experiment. Real RGB is translated to synthetic LWIR using generator F. These synthetic LWIR are used for fine tuning the RGB trained detection algorithm for testing real LWIR imagery.

it requires extra manual annotation from us.

Figure 7 and 8 and show some examples of synthetic LWIR imagery produced using the Thales and FLIR datasets. The F1 scores produced using the two datasets are shown in Table 2. When testing the Thales and FLIR real LWIR imagery, we achieve F1 scores of 85.6% and 45.6% respectively. These results are comparable to the F1 scores produced when fine-tuning with real LWIR imagery but with the crucial advantage of not requiring extra annotation in LWIR. This method produces higher F1 scores than those in the LWIR to RGB translation for detection approach and shows how accurately we can detect in LWIR imagery using no real LWIR images or labels for training our detection algorithms.

Network FT with:	Imagery tested	Thales	FLIR
Synthetic LWIR	Real LWIR	**85.6**	**45.6**
Real LWIR	Real LWIR	94.2	59.8

Table 2: Fine-tuning with synthetic LWIR F1 scores.

5. Conclusion

In this paper, we achieve high detection rates in LWIR imagery in a completely unsupervised manner using only RGB labels for training detection algorithms. Firstly, we translate LWIR imagery to RGB, to make use of RGB trained detection networks. Our object-specific adapted cycleGAN produces better quality RGB imagery for detection than the original cycleGAN, producing F1 scores of up to 67.1%. In addition, we show that the fusion of the RGB generated image and the real LWIR image can further enhance detection to up to 73.0%. Secondly, we fine-tune an RGB trained detector with synthetic LWIR and test with real LWIR. This method produces the best F1 performance of up to 85.6% and confirms that our synthetic LWIR imagery is of high quality.

To summarise, for the first time, we produce synthetic RGB/LWIR imagery using our unsupervised adapted cycleGAN, which ultimately leads to high detections rates using non-corresponding and cluttered datasets. This work will be of great interest to those in the defence industry as we show 24-hour detection is possible in real-life scenarios when little or no labelled datasets are available for training.

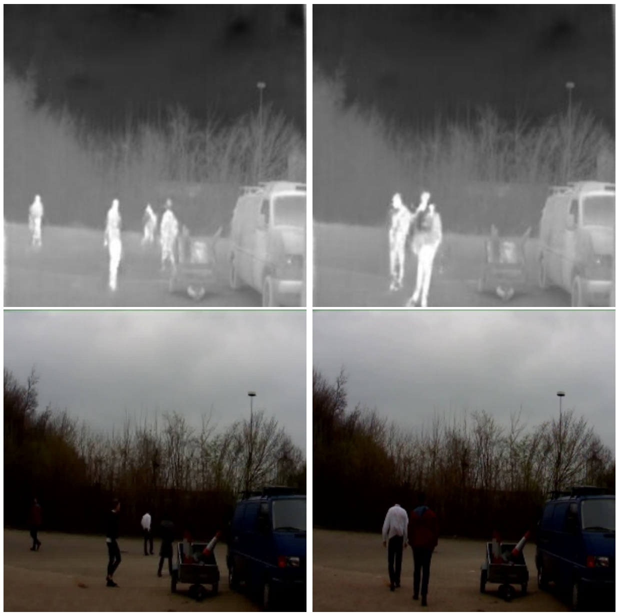

Figure 7: Thales synthetic LWIR imagery (top row) translated from RGB imagery (bottom row).

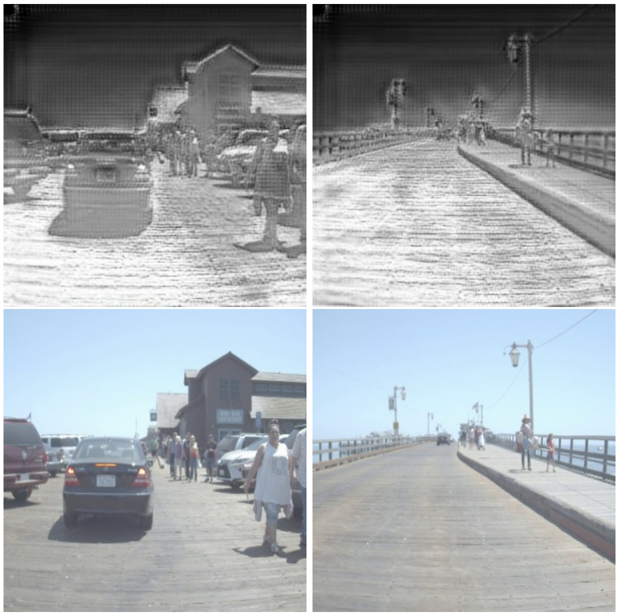

Figure 8: FLIR synthetic LWIR imagery (top row) translated from RGB imagery (bottom row).

References

[1] R. Abbott, J. M. Del Rincon, B. Connor, and N. Robertson. Deep object classification in low resolution lwir imagery via transfer learning. In *Proceedings of the 5th IMA Conference on Mathematics in Defence*, 11 2017. 2

[2] Rachael Abbott, Neil Robertson, Jesus Martinez del Rincon, and Barry Connor. Multimodal object detection using unsupervised transfer learning and adaptation techniques. In Judith Dijk, editor, *Artificial Intelligence and Machine Learning in Defense Applications*, volume 11169, pages 49 – 58. International Society for Optics and Photonics, SPIE, 2019. 2, 4

[3] Jörgen Ahlberg and Michael Felsberg. Generating visible spectrum images from thermal infrared. pages 1224–122409, 06 2018. 2

[4] Gaurav Bhatnagar and Zheng Liu. A novel image fusion framework for night-vision navigation and surveillance. *Signal, Image and Video Processing*, 9(1):165–175, Dec 2015. 2

[5] Petra Bosilj, Erchan Aptoula, Tom Duckett, and Grzegorz Cielniak. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *Journal of Field Robotics*, 0(0). 2

[6] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. *CoRR*, abs/1612.05424, 2016. 2

[7] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096, 2018. 2

[8] Zhilu Chen and Xinming Huang. Pedestrian detection for autonomous vehicle using multi-spectral cameras. *IEEE Transactions on Intelligent Vehicles*, PP:1–1, 03 2019. 2

[9] Weijian Deng, Liang Zheng, Guoliang Kang, Yi Yang, Qixiang Ye, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. *CoRR*, abs/1711.07027, 2017. 2

[10] F.A.Group. Flir thermal dataset for algorithm training. https://www.flir.co.uk/oem/adas/adas-dataset-form/. Accessed: 05-07-2019. 5

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. 2

[12] J. Hoffman, S. Gupta, J. Leong, S. Guadarrama, and T. Darrell. Cross-modal adaptation for rgb-d detection. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5032–5039, May 2016. 2

[13] Xun Huang, Ming-Yu Liu, Serge J. Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. *CoRR*, abs/1804.04732, 2018. 2, 5

[14] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baselines. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2

[15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. 2

[16] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018. 2

[17] Matthias Limmer and Hendrik P. A. Lensch. Infrared colorization using deep convolutional neural networks. *CoRR*, abs/1604.02245, 2016. 2

[18] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. *CoRR*, abs/1606.07536, 2016. 2

[19] Shuo Liu, Vijay John, Erik Blasch, Zheng Liu, and Ying Huang. IR2VI: enhanced night environmental perception by unsupervised thermal image translation. *CoRR*, abs/1806.09565, 2018. 2

[20] Shuo Liu and Zheng Liu. Multi-channel cnn-based object detection for enhanced situation awareness. *CoRR*, abs/1712.00075, 2017. 2

[21] Armin Mehri and Angel D. Sappa. Colorizing near infrared images through a cyclic adversarial approach of unpaired samples. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 2

[22] Pramuditha Perera, Mahdi Abavisani, and Vishal M. Patel. In2i : Unsupervised multi-image-to-image translation using generative adversarial networks. *CoRR*, abs/1711.09334, 2017. 2

[23] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. *CoRR*, abs/1403.6382, 2014. 2

[24] Benjamin S. Riggan, Nathaniel J. Short, and Shuowen Hu. Thermal to visible synthesis of face images using multiple regions. *CoRR*, abs/1803.07599, 2018. 2

[25] Marcel Sheeny, Andrew Wallace, Mehryar Emambakhsh, Sen Wang, and Barry Connor. POL-LWIR vehicle detection: Convolutional neural networks meet polarised infrared sensors. *CoRR*, abs/1804.02576, 2018. 2

[26] Patricia Suarez, Angel Sappa, and Boris Vintimilla. Infrared image colorization based on a triplet dcgan architecture. pages 212–217, 07 2017. 2

[27] Tian Sun, Cheolkon Jung, Qingtao Fu, and Qihui Han. Nir to rgb domain translation using asymmetric cycle generative adversarial networks. *IEEE Access*, PP:1–1, 08 2019. 2

[28] F. Ye, W. Luo, M. Dong, H. He, and W. Min. Sar image retrieval based on unsupervised domain adaptation and clustering. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5, 2019. 2

[29] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. *CoRR*, abs/1704.02510, 2017. 2

[30] He Zhang, Benjamin S. Riggan, Shuowen Hu, Nathaniel J. Short, and Vishal M. Patel. Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks. *CoRR*, abs/1812.05155, 2018. 2

[31] Lichao Zhang, Abel Gonzalez-Garcia, Joost van de Weijer, Martin Danelljan, and Fahad Shahbaz Khan. Synthetic data generation for end-to-end thermal infrared tracking. *CoRR*, abs/1806.01013, 2018. 2

[32] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017. 1, 2, 3