MULTI-DOMAIN THERMAL OBJECT DETECTION USING GENERATIVE ADVERSARIAL NETWORKS

Moataz Eltahan
Dept. of Electronics and Communications
Cairo University
Giza, Egypt
MoatazEltahan@gmail.com

Khaled Elsayed
Dept. of Electronics and Communications
Cairo University
Giza, Egypt
kelsayed@eng.cu.edu.eg

2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA) | 978-1-6654-5642-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/DICTA56598.2022.10034641

Abstract—The major recent advances in autonomous driving are mostly based on recent progress in object detection methods. Object detection broadly depends on the quality of the cameras and sensors used to generate the images. Many kinds of sensors are being utilized for the sake of improving scene perception, such as RGB, radar, and LiDAR. However, the usage of all these sensors' data leads to object detection degradation in adverse lighting conditions. Thermal imaging depicts the spatial distribution of temperature differences in a scene. This leads to the ability of a thermal camera to outperform other sensors in severe weather conditions. In this research, we propose the use of thermal imaging to improve object detection in autonomous driving. The proposed framework inherits the architecture of EfficientDet and further proposes a technique to fuse the features of two different modalities: thermal and RGB. We use GAN as an I2I translation framework to transform the thermal images into RGB images. The experimental results demonstrate that the proposed framework outperforms the detection accuracy of other state-of-the-art object detectors in thermal imagery. We propose two variants, MDFFTDet-D0 and MDFFTDet-D2 which achieve mAP of 63.15% and 77.81% respectively on the FLIR ADAS dataset.

Keywords—component, Object-detection, Infrared, Thermal Imaging, Feature Fusion, EfficientDet, GAN.

I. INTRODUCTION

Autonomous driving (AD) is undeniably gaining a substantial role in the automotive industry. AD exploits emerging advances in object detection that benefit from improvements in deep learning algorithms, especially convolutional neural networks (CNNs). Scene perception is a crucial task in self-driving vehicles. Scene perception in AD is the ability to extract relevant data from the surroundings. The efficacy of perception largely depends on the sensors and cameras used in capturing the scene. It also depends on the surrounding environmental conditions, which affect the sensors and cameras.

The main sensors utilized in perception and object detection are RGB, radar, and LiDAR. Each type has its pros and cons. RGB enables low-cost and high-resolution data capture; however, its performance is highly dependent on the surrounding lighting and weather conditions. Radar provides high resolution and accuracy and can operate in cloudy weather conditions and at night. However, it cannot produce a precise image of an object. On the other hand, LiDAR provides more accurate and precise images compared to radar and RGB, but it suffers from performance degradation in bad weather conditions, such as severe rain, fog, smoke, or snowstorm. Also, LiDAR uses light waves, which are easily affected by the medium itself. For example, moisture in the atmosphere affects the performance of a LiDAR system [1]. So, whether using RGB, radar, or LiDAR, a degradation in perception accuracy is likely to occur in adverse lighting conditions due to the lack of light or the bad weather conditions.

Thermal imaging is simply the process of sensing infrared (IR) radiation heat and converting it into visible images that depict the spatial distribution of temperature differences in a scene [2]. A thermal camera can outperform other sensors in severe lighting conditions and can also perform well in the daytime. Recently, various research endeavors have been carried out to perform sensor fusion of LiDAR and thermal cameras for autonomous driving [3,4]. This fusion can benefit from the high precision of LiDAR images and the great performance of thermal imaging in severe lighting conditions.

Object detection in the visual RGB domain is considered mature and robust and could achieve good results. This is due to the availability of large RGB datasets, such as ImageNet [5], PASCAL-VOC [6], and MS-COCO [7]. On the other hand, there is a lack of such big datasets in the thermal imagery. Therefore, there is a gap between detection results in the RGB domain compared to the thermal domain. Recently, the FLIR ADAS dataset for thermal images has been released [8]. Also, the KAIST Multi-Spectral dataset [9] is a primary dataset in thermal imagery, but it gives annotations for pedestrians only.

To tackle the problem of lacking large datasets in the thermal domain, generative adversarial networks (GAN) [10] can be used to reduce the gap between the source domain and the target domain. Multiple research papers presented the usage of GANs and domain adaptation for object detection in thermal imagery. Generally, an image to image (I2I) translation task aims to transform an image from a source domain into a target domain. Therefore, I2I could be deployed to improve object detection in the thermal domain.

In the proposed approach, an image to image (I2I) translation framework is deployed to transform the thermal images into RGB images. The proposed multi-domain framework consists of two branches. EfficientNet [11] is used on the two branches to act as the object detection backbone. UNIT GAN is used as the I2I translation framework in our work. The rest of the proposed Multi-domain feature fusion for thermal detection (MDFFTDet) blocks are employed from EfficientDet [12] object detector.

The experimental results demonstrate that using the idea of feature fusion significantly improves the performance and aids in allowing efficient fusion of the multi-domain RGB and thermal features. The proposed framework outperforms the state-of-the-art object detection methods in the thermal domain.

The rest of this paper is organized as follows. Section II discusses the literature review of object detection in thermal imagery. Section III demonstrates the proposed methodology. Section IV focuses on the carried-out experiments and their results. Section V concludes the study.

II. Related Work

Object detection and classification is a crucial techniques in computer vision applications. Recently, there is an increasing number of research works on classifying and detecting thermal imagery [13,14,15,16]. Reference [17] introduces two domain adaptation components aiming to tackle the domain shift in object detection using cross domains. Reference [13] augments thermal images with their saliency maps to serve as an attention mechanism.

Furthermore, multiple research works benefit from the recent improvements in GANs [10] to alleviate the problem of lacking publicly available large-scale thermal datasets. Generally, the idea of integrating GAN with object detectors to help in improving the detection task is tackled by multiple works [14,18,19]. Reference [20] uses CycleGAN [21] and UNIT [22] to generate RGB images from the thermal images using Faster-RCNN. Reference [15] proposes a style transfer technique using a multi-style generative network (MSGNet) to transfer the low-level features such as texture and edges from the visible spectrum to the infrared spectrum domain.

III. The Proposed Multi-Domain Feature Fusion Thermal Detection (MDFFTDet)

The key idea in the proposed multi-domain feature fusion for thermal object detection (MDFFTDet) framework is to use a feature fusion network like a bi-directional feature pyramid network (BiFPN) [12] to mix the feature vectors of thermal and RGB modalities. We argue that using BiFPN leads to the generation of high-resolution feature maps that significantly improve the performance of object detection in thermal imagery.

As summarized in Fig. 1, we propose the use of two parallel backbones of EfficientNet [11]. Each branch is responsible for processing and generating feature vectors of a specific domain, thermal or RGB. We achieve the proposed objective by deploying an image-to-image (I2I) translation framework to generate pseudo-RGB images from the input thermal images. A pre-trained UNIT GAN for transforming images from the thermal domain to the RGB domain is used as the I2I translation framework in the proposed architecture. The generated RGB images obtained from UNIT GAN are used to fine-tune one branch of the EfficientNet block that is pre-trained on the MS-COCO dataset [7]. The other EfficientNet branch is trained on thermal images, which is the FLIR ADAS dataset. The thermal and RGB weights from the two branches are concatenated and then fused using a weighted bi-directional feature pyramid network (BiFPN) as implemented in the EfficientDet model [12]. This intuition of borrowing complex features from a rich domain like the RGB domain to improve object detection in the thermal domain was first introduced in [20]. This introduced the MMTOD, which is considered one of the state-of-the-art object detection methods in thermal imagery.

The following three sections discuss some of the building blocks of the proposed scheme, and then section D discusses the details of the proposed scheme.

A. EfficientNet

EfficientNet was developed by performing neural architecture research with the goal of enhancing accuracy and reducing FLOPS. EfficientNet [11] proposed a novel model scaling method to scale up CNNs in a more structured manner using a simple compound coefficient. Conventional approaches scale network dimensions, such as width, depth, and resolution, arbitrarily. On the other hand, EfficientNet uniformly scales up all dimensions in a principled and compound way. Compound scaling scales each dimension with a fixed set of scaling coefficients. A family of models from EfficientNet-B1 to EfficientNet-B7 were obtained by applying the compound scaling method to the baseline EfficientNet-B0.

B. EfficientDet and Feature Fusion

The proposed framework is based on the EfficientDet object detection model that leverages the efficiency of the EfficientNet network [11] and uses it as its backbone network. EfficientDet also proposed a new bi-directional feature pyramid network (BiFPN) [12] to act as the feature fusion network.

The idea of feature fusion is the combination of features from different layers or branches. Feature fusion mixes the highly abstracted semantic features captured in the high-level neural network layers with the highly detailed features from the low-level layers. The key idea behind using BiFPN as proposed in EfficientDet [12] is that BiFPN allows information to flow in both the top-down and bottom-up directions, as shown in Fig. 2. The BiFPN implementation in EfficientDet [12] adds additional learnable weights to the input features and allows the network to learn the contributions of each input feature.

We leverage the success of using BiFPN to utilize and fuse the high-abstracted and low-detailed feature vectors from two different image modalities: thermal and RGB. As will be demonstrated, this fusion technique helps to improve the object detection performance in thermal imagery.

To achieve better accuracy and efficiency trade-offs under different resource constraints, EfficientDet utilized the compound scaling method first proposed in EfficientNet [11]. Compound scaling jointly scales up the resolution, width, and depth. Each EfficientDet component, i.e., EfficientNet backbone, BiFPN, and box/class prediction network has a single compound scaling factor that controls all the scaling dimensions. The baseline EfficientDet-D0 was developed by combining EfficientNet-B0 with BiFPN. A family of models from EfficientNet-D1 to EfficientNet-D7 were obtained by applying the compound scaling method with baseline EfficientNet-D0. EfficientDet-D0 and EfficientDet-D2 are utilized in the proposed framework. We refer to each implementation as MDFFTDet-D0 and MDFFTDet-D2, respectively.

C. Image to Image Translation

Due to the difficulties in finding paired thermal and RGB images, we benefit from the recent advancements in unpaired

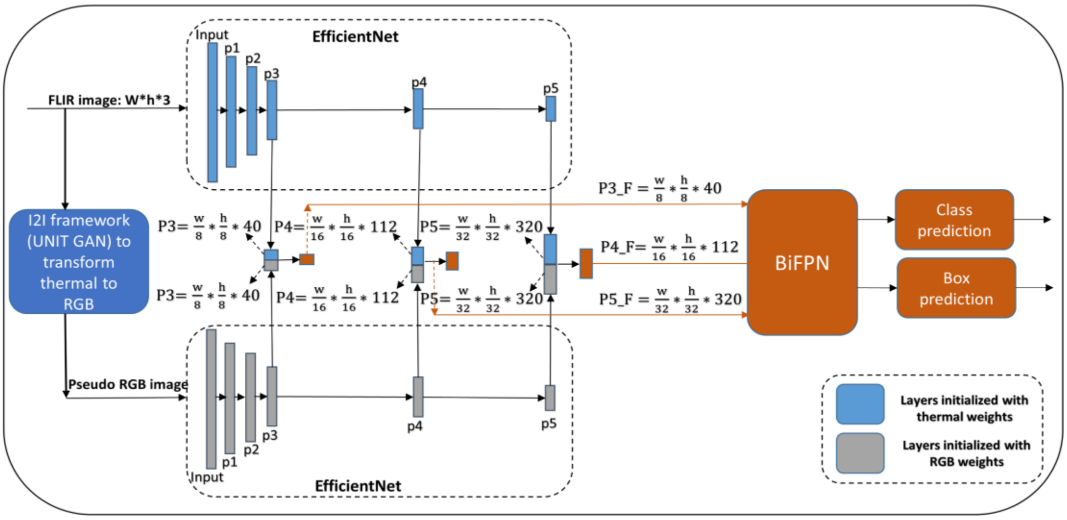

Fig 1. The proposed multi-domain feature fusion (MDFFTDet-D0) framework

image-to-image translation frameworks, especially using generative adversarial networks (GANs). UNIT GAN assumes that for any given pair of images from a domain X1 and a domain X2, there exists a shared latent space Z. Any image x1 from domain X1 or any image x2 from domain X2 could be mapped to the latent representation in a shared latent space Z.

The UNIT GAN consists of two variational autoencoders (VAE1 and VAE2) and two generative adversarial networks (GAN1 and GAN2). Each VAE consists of a pair of an encoder and a generator. Each VAE is responsible for encoding input images from a specific domain to the latent space via the encoder. The generator is then responsible for decoding and reconstructing the input image. Each GAN consists of a pair of a generator and a discriminator. The images generated by either G1 or G2 can come from two sources: the first one is the reconstruction unit (VAE), and the second is the translation unit (GAN). The discriminator is responsible for differentiating between the real images and the generated images.

D. Details of the Proposed MDFFTDet Method

We propose two variants of the proposed architecture. The first one adopts the EfficientDet-D0 [12], which adopts the EfficientNet-B0 [11]. We refer to this architecture as MDFFTDet-D0. The second one adopts the EfficientDet-D2 [12] which in turn adopts EfficientNet-B2 [11]. We refer to this architecture as MDFFTDet-D2. An EfficientDet trained on the thermal images of the training set of the FLIR ADAS dataset [8] is used as a baseline for the proposed framework.

UNIT GAN is a popular unpaired image-to-image translation framework. In the proposed framework, UNIT GAN is used to transform the thermal images into pseudo-RGB images. The obtained RGB images are used to fine-tune the RGB EfficientNet [11] backbone, which is initialized with pre-trained RGB weights of MS-COCO [7]. On the other hand, the thermal images are input to the thermal EfficientNet [11] backbone.

As shown in Fig. 1, the proposed multi-domain feature fusion for thermal object detection (MDFFTDet) framework consists of two backbone branches: the first one for thermal images and the second one for RGB. As shown in Fig. 2, the original BiFPN combines features from level P_3 to P_7 from the EfficientNet backbone network and allows the features to

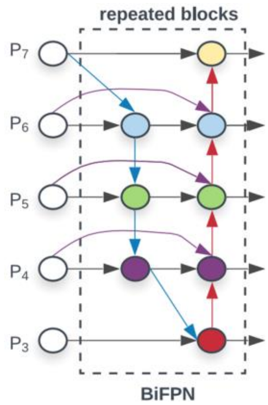

Fig 2. BiFPN uses features from level 3 to 7 (P3 - P7). [12]

repeatedly flow in both top-down and bottom-up feature fusion. We use a minimal implementation for BiFPN where it only takes three feature vectors, $P_{3\text{-}F}$, $P_{4\text{-}F}$, and $P_{5\text{-}F}$ as input from the fusing network, which fuses them using the P_3, P_4, and P_5 feature vectors from the two backbone networks. After that, BiFPN generates $P_{6\text{-}F}$ and $P_{7\text{-}F}$ internally using the fused feature vectors $P_{3\text{-}F}$, $P_{4\text{-}F}$, and $P_{5\text{-}F}$. Therefore, BiFPN combines only three feature vectors, $P_{3\text{-}F}$, $P_{4\text{-}F}$, and $P_{5\text{-}F}$ in the proposed framework as shown in Fig. 1.

The multi-fused features P_3, P_4, and P_5 are then passed through a 1x1 convolution to combine the feature vectors from the two branches and decrease the features' depth by half. As shown in Fig. 1, $P_{3\text{-}F}$, $P_{4\text{-}F}$, and $P_{5\text{-}F}$ are obtained from the 1x1 convolution layers and then fed into the BiFPN. The focal loss is the used loss function [23].

UNIT GAN provides two generators to perform the translation between source and target domains. In the proposed framework, thermal images are the source domain, and RGB images are the target domain. Therefore, only one generator is used, which is responsible for translating the thermal images to RGB. For the training of UNIT GAN, all the hyperparameters are used as stated in the original paper [22]. Also, the pre-trained weights of UNIT GAN trained on the FLIR ADAS thermal images to generate pseudo-RGB images are taken from [20]. The UNIT GAN is trained with the rest of the proposed architecture to help in optimizing the pseudo-generated RGB images for the sake of improving detection accuracy. The code and implementation for EfficientDet were taken from the PyTorch implementation [24]. More details about the implementation could be found in Section IV.

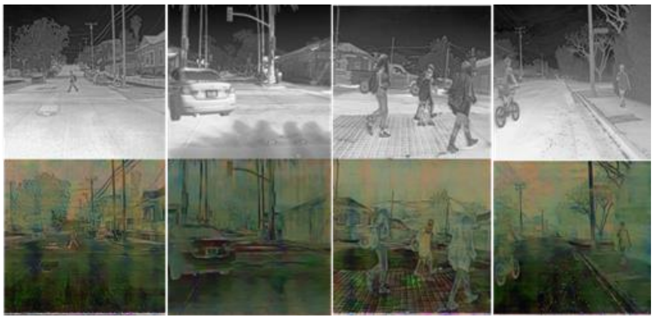

Fig 3. . Row 1: Thermal images from the FLIR ADAS dataset [8], Row 2: The corresponding pseudo generated RGB images using UNIT GAN [22].

IV. Experiments and Results

Following the literature, the proposed framework MDFFTDet is evaluated on the FLIR ADAS dataset [8] using average precision (AP) for each class at an Intersection over Union (IoU) of 0.5. The results of MDFFTDet are compared with the implementation of the EfficientDet baseline and with the recent models in the literature such as Faster R-CNN [20], MMTOD-UNIT [20], MMTOD-CG [20], ODSC models [15], and ThermalDet [16].

A. FLIR ADAS Dataset

FLIR ADAS dataset [8] consists of 10228 total images, and 9214 images are with objects annotated using a bounding box. The images are captured using a FLIR infrared Tau2 camera with a resolution of 640x512. The objects are classified into four categories, i.e., car, person, bicycle, and dog. However, the dog category has very few annotations, so it is not considered in this study. 60% of the images are captured during the daytime, and the remaining 40% are captured during the night. We use the train and test splits as provided in the dataset.

B. Baseline

An EfficientDet model [12] is trained on the thermal images of the training set of the FLIR ADAS dataset [8] and is used as a baseline to highlight the advancements brought in by the proposed framework. We followed the EfficientDet implementation in [24] for all the hyperparameters.

C. Training of the Proposed MDFFTDet Architecture

The EfficientNet [11] is trained on pseudo-generated RGB images from the UNIT GAN [22] and initialized with pre-trained RGB weights of MS-COCO [7]. The UNIT GAN [22] acts as an I2I framework to transform the thermal images into pseudo-RGB images. The pseudo-RGB images are then used to fine-tune the lower EfficientNet [11] in Fig. 1. The adoption of UNIT GAN [22] as an I2I framework helps in improving object detection in thermal imagery. However, the pseudo-generated RGB images generated from the UNIT GAN [22] are noisy compared to the real RGB domain images, as shown in Fig. 3. The UNIT GAN [22] generators are trained with the rest of the proposed framework to learn to optimize the pseudo-generated RGB images in a way that helps to reduce the overall loss of the output and therefore improve the object detection in thermal imagery.

We train two variants of the proposed architecture, MDFFTDet-D0 and MDFFTDet-D2. In the two variants, the thermal images and the corresponding pseudo-generated RGB images are passed through the two branches of EfficientNet to

TABLE I. PERFORMANCE COMPARISON OF PROPOSED METHODOLOGY (MDFFTDET-D0 AND MDFFTDET-D2) AGAINST STATE-OF-THE-ART METHODS

Method	**AP Across Each Class**			**mAP**
	Person	***Bicycle***	***Car***	
Faster R-CNN [20]	54.69	39.66	67.57	53.97
MMTOD-UNIT [20]	64.47	49.43	70.72	61.54
MMTOD-CG [20]	63.31	50.26	70.63	61.40
SSD300+ MobileNet V2 (ODSC) [15]	36.38	27.98	54.34	39.57
SSD300+ EfficientNet (ODSC) [15]	51.69	35.12	74.05	53.62
SSD512+VGG16 (ODSC) [15]	62.53	46.91	79.91	75.36
SSD300 +VGG16 (ODSC) [15]	71.01	55.53	82.33	69.62
ThermalDet [16]	78.24	60.04	85.52	74.60
EfficientDet baseline	65.24	40.99	72	59.4
MDFFTDet-D0	66.89	47.36	75.19	63.15
MDFFTDet-D2	81.29	63.34	88.82	77.81

produce a set of two feature maps, P3, P4, and P5, as shown in Fig. 1. The feature maps from each branch are then concatenated. Each concatenated feature map is passed through a 1x1 convolution layer to reduce the number of channels by half. After that, the output of each 1x1 convolution is passed as input to the bi-directional feature pyramid network (BiFPN).

The training of the proposed framework (MDFFTDet-D0) is done in two steps. In the first step, all the network blocks including the UNIT GAN are trained on the Nvidia-RTX-2080 GPU with batch size 2. The training continues until the validation dataset average precision saturates and could not

improve further. The UNIT GAN training is frozen, and the rest of the architecture continues the training process until the validation dataset average precision saturates again. The same training procedure is used for MDFFTDet-D2, but it is trained on three Nvidia-GTX-1080-Ti GPUs with a batch size of 4.

D. Results

Table 1 shows the comparison of average precision (AP) for each class and the mAP of the two frameworks, MDFFTDet-D0 and MDFFTDet-D2, compared to other state-of-the-art methods trained on the FLIR ADAS dataset. We use EfficientNet as the backbone, with an input size of 512x512. Therefore, the input FLIR thermal images are first scaled from a resolution of 640x512 to 512x512. The results show that the proposed framework MDFFTDet-D2 outperforms all the recently published detectors in thermal imagery by achieving 77.81% mAP. Notably, the two proposed MDFFTDet-D0 and MDFFTDet-D2 frameworks outperform ODSC, which uses SSD300, although it uses the same backbone as the proposed frameworks.

Table 2 shows an estimate of the models' size between the two variants proposed MDFFTDet-D0 and MDFFTDet-D2 compared to the other state-of-art methods. The comparison is based on the number of parameters in each network. Notably, the ODSC network uses SSD300 [33], which is a huge network that contains a large number of parameters 24M compared to EfficientDet, which contains 3.9M parameters for EfficientDet-D0 and 8.1 M for EfficientDet-D2. Also, the ODSC models that achieved the best results in the literature were using VGG16 as a backbone network, and VGG16 is a huge network with 138M parameters. On the other hand,

TABLE II. MODEL SIZE COMPARISON OF PROPOSED METHODOLOGY (MDFFTDET-D0 AND MDFFTDET-D2) AGAINST STATE-OF-THE-ART METHODS

Method	#Params
Faster R-CNN [26]	77M
MMTOD [26]	112M
SSD300+ Mobilenet V2 (ODSC) [18]	27.5M
SSD300+ EfficientNet (ODSC) [18]	29.3M
SSD512+VGG16 (ODSC) [18]	162.7M
SSD300 +VGG16 (ODSC) [18]	162M
ThermalDet [19]	79M
EfficientDet baseline	3.9M
MDFFTDet-D0	9.3M
MDFFTDet-D2	17.3M

EfficientNet is used as a backbone network in the proposed method MDFFTDet, and EfficientNet is a small network compared to VGG16. EfficientNet-B0 backbone used in the proposed method MDFFTDet-D0 contains 3.6M parameters and EfficientNet-B2 backbone used in the proposed method MDFFTDet-D2 contains 7.2M parameters. We use two branches of EfficientNet in each proposed method, but the two proposed frameworks MDFFTDet-D0 and MDFFTDet-D2 still use much fewer parameters. In the proposed architecture, we benefit from the BiFPN implementation, which fuses different feature maps from different levels in the backbone and gives weight to each feature, and allows the network to adapt to it.

The results clearly demonstrate that mixing the feature maps of the thermal and RGB domains leads to enhancing the effectiveness and performance of object detection in thermal imagery.

V. CONCLUSIONS

This paper focuses on improving object detection efficacy in thermal imagery, which could benefit multiple applications, such as autonomous driving, security, surveillance, and military applications, especially in severe lighting environments. This study proposes a domain adaptation of thermal and RGB images and benefits from the rich available datasets in the RGB domain. The main idea is to adopt EfficientDet and benefit from BiFPN implementation to fuse the feature maps of RGB and thermal backbones from different levels. We evaluate the proposed framework using the FLIR ADAS dataset and compare the results with recently published detectors in thermal imagery. The results demonstrate that the proposed framework achieves better performance than the state-of-the-art detectors and improves detection accuracy in thermal imagery. The future work will include using pairs of thermal and RGB inputs so that we could compare the proposed fusion technique with the other approaches that use paired images.

REFERENCES

[1] Hani, D. (2021, April 19). Lidar vs radar: Detection, tracking, and imaging. wevolver.com. Retrieved from https://www.wevolver.com/article/lidar-vs-radar-detection-tracking-and-imaging

[2] Kirk J. Havens, Edward J. Sharp, Chapter 8 - Imager Selection, Editor(s): Kirk J. Havens, Edward J. Sharp, Thermal Imaging Techniques to Survey and Monitor Animals in the Wild, Academic Press, 2016, Pages 121-141, ISBN 9780128033845, https://doi.org/10.1016/B978-0-12-803384-5.00008-7.

[3] Choi, J.D., & Kim, M.Y. (2021). A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles: its calibration and application. 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), 361-365.

[4] Azam, S., Munir, F., Sheri, A.M., Ko, Y., Hussain, I., & Jeon, M. (2019). Data fusion of lidar and thermal camera for autonomous driving.

[5] Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. CVPR.

[6] Everingham, M., Gool, L.V., Williams, C.K., Winn, J.M., & Zisserman, A. (2009). The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 88, 303-338.

[7] Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C.L. (2014). Microsoft COCO: common objects in Context. ECCV.

[8] F. A. Group. FLIR thermal dataset for algorithm training. https://www.flir.in/oem/adas/adas-dataset-form/, 2018.

[9] Hwang, S., Park, J., Kim, N., Choi, Y., & So Kweon, I. (2015). Multispectral pedestrian detection: benchmark dataset and baseline. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1037-1045).

[10] Goodfellow, I.J. (2017). NIPS 2016 Tutorial: Generative Adversarial Networks. ArXiv, abs/1701.00160.

[11] Tan, M., & Le, Q.V. (2019). EfficientNet: rethinking model scaling for convolutional neural networks. ArXiv, abs/1905.11946.

[12] Tan, M., Pang, R., & Le, Q.V. (2020). EfficientDet: scalable and efficient object detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10778-10787.

[13] Ghose, D., Desai, S.M., Bhattacharya, S., Chakraborty, D., Fiterau, M., & Rahman, T. (2019). Pedestrian detection in thermal images using saliency maps. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 988-997.

[14] Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., & Yan, S. (2017). Perceptual generative adversarial networks for small object detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1951-1959.

[15] Munir, F., Azam, S., Rafique, M. A., Sheri, A. M., & Jeon, M. (2020). Thermal object detection using domain adaptation through style consistency.

[16] Cao, Y., Zhou, T., Zhu, X., & Su, Y. (2019). Every feature counts: an improved one-stage detector in thermal imagery. *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 1965-1969.

[17] Chen, Y., Li, W., Sakaridis, C., Dai, D., & Gool, L.V. (2018). domain adaptive faster R-CNN for object detection in the Wild. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3339-3348.

[18] Wang, W., Hong, W., Wang, F., & Yu, J. (2020). GAN-knowledge distillation for one-stage object detection. *IEEE Access, 8*, 60719-60727.

[19] Zhang, Y., Bai, Y., Ding, M., & Ghanem, B. (2020). Multi-task generative adversarial network for detecting small objects in the wild. *International Journal of Computer Vision, 128*, 1810-1828.

[20] Devaguptapu, C., Akolekar, N., Sharma, M.M., & Balasubramanian, V.N. (2019). Borrow from anywhere: pseudo multi-modal object detection in thermal imagery. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1029-1038.

[21] Chu, C., Zhmoginov, A., & Sandler, M. (2017). Cyclegan, a master of steganography. arXiv preprint arXiv:1712.02950.

[22] Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. Advances in neural information processing systems, 30.

[23] Lin, T., Goyal, P., Girshick, R.B., He, K., & Dollár, P. (2020). Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, 318-327.

[24] Yet-Another-EfficientDet-Pytorch [Source code]. https://github.com/zylo117/Yet-Another-EfficientDet-Pytorch, 2020.