Human Detection in Infrared Image Using Daytime Model-Based Transfer Learning for Military Surveillance System

1st Eun Seop Kim
Environment ICT Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
subway@etri.re.kr

2nd Whui Kim
Environment ICT Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
khsunkh@etri.re.kr

3rd Juderk Park
Environment ICT Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
zdpark@etri.re.kr

4th Kunmin Yeo
Environment ICT Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
kunmin@etri.re.kr

***Abstract*—In the field of civil and military surveillance, there has been a gradual increase in the application of deep learning-based image object detection systems to enhance accuracy. During daylight hours, the desired objects can easily be detected using cameras combined with image processing or deep learning techniques. However, during nighttime or when weather conditions prevent daylight, detecting objects through regular RGB cameras becomes challenging [1]. This paper proposes a method for detecting humans using images captured from infrared cameras, aiming to facilitate whole day surveillance systems for military bases. Typically, public image datasets almost consist of RGB images taken during the day [2]. However, gathering infrared image datasets is both essential and labor-intensive. Therefore, to minimize the collection of such infrared datasets, this study suggests using transfer learning based on daytime models to detect humans in infrared images. With a limited amount of collected infrared data, this transfer learning approach can enhance the accuracy of detection.**

***Index Terms*—Infrared image, Nighttime object detection, Transfer learning, Surveillance system**

I. Introduction

In recent years, there has been a rise in unauthorized intrusions, both in military and civilian domains, by individuals or unmanned aerial vehicles (UAVs). Often, by the time these trespasses are detected, it becomes challenging to take preventative measures due to concerns about potential damage to the protected areas. Hence, early detection from a distance before the intruders get too close is crucial.

The most straightforward surveillance method is direct visual observation by guards or using cameras. However, such manual surveillance can suffer from lapses in the observer's concentration, potentially missing critical moments [3]. To address these limitations, there's a growing trend towards using image processing and deep learning for object detection. Among these methods, traditional image processing offers relatively faster detection speeds but often lacks precision. In contrast, deep learning, known for its high accuracy, has been extensively adopted for object detection. While its computational intensity is traditionally a drawback, advancements in GPU technology have mitigated this concern [4]. However, surveillance systems relying on conventional cameras face challenges during nighttime or poor weather conditions when visibility is limited [1].

To overcome these challenges, a dual-camera system using regular RGB cameras during the day and infrared cameras at night can be deployed. While RGB cameras detect light reflected off objects from external sources, infrared cameras sense infrared radiation emitted by objects themselves, making them effective even in darkness. However, most existing datasets for object detection predominantly consist of daylight-captured RGB images. Infrared image datasets, both in volume and quality, are scarce. Gathering infrared datasets, especially for less-common objects like tanks or drones, is a cumbersome task. To address this, this study suggests maximizing the use of existing daylight image datasets while collecting a minimal amount of infrared data. By using transfer learning, one can effectively leverage the knowledge acquired from a pre-trained model on an existing dataset to facilitate the training process on a new dataset.

II. Methodology

A. YOLOv5

The deep learning model employed in this study utilizes the YOLOv5 developed based on the PyTorch framework. YOLO stands for "You Only Look Once," with the "v5" indicating its fifth version. This model has achieved state-of-the-art (SOTA) performance [5]. This framework has undergone incremental modifications with successive versions 7.0.

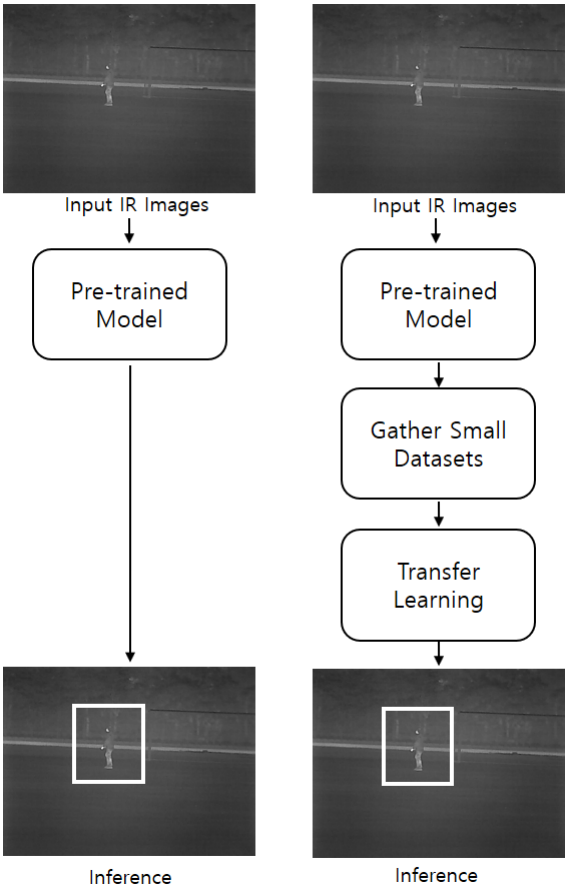

Figure 1. Proposed method

A notable change in this version is the replacement of the conventional SPP(spatial pyramid pooling) with SPPF(SPP-fast), increasing the computational throughput per hour [6]. Additionally, YOLOv5 is versatile, capable of processing images of varying sizes. The model range from the lightest, known as "nano," to the heaviest, referred to as "xlarge," with a total of five primary models. For the purpose of real-time detection in this paper, we employ the lightweight "small" model. Positioned in terms of complexity just above the "nano" model, the "small" model offers a trade-off between speed and accuracy, rendering it the most suitable choice considering its significant improvement in accuracy over the "nano".

B. Transfer Learning

The proposed method is showed in Figure 1. The left side shows the use of a pre-trained model only, while the right side shows the approach using transfer learning. Traditional deep learning based on CNN typically relies on datasets that have not undergone any previous training. In contrast, transfer learning uses a pre-trained model and further trains it, allowing the model to leverage existing knowledge. As a result, even with limited data on a new object, good performance can be achieved. Moreover, since transfer learning utilizes a smaller dataset, it reduces both training time and computational resource consumption [7]. The strategy for transfer learning can be determined based on the size or similarity of the dataset and can be categorized into three approaches: retraining the entire model, retraining without only specific layers, and using the model without further training [8].

In this paper, we capitalize on the observation that deep learning models pre-trained on daytime images can operate with considerable accuracy on infrared images intended for nighttime surveillance. However, due to differences in color representation, significant misclassifications or non-detections can occur in many images. Additionally, military surveillance targets are often challenging to obtain in high-quality or large quantities due to issues related to information scarcity and security concerns. Consequently, we aim to enhance performance by collecting a limited set of infrared images and employing transfer learning on a model pre-trained with daytime images.

C. Pre-trained Model

The YOLOv5 provides a model pre-trained on the COCO Dataset, which encompasses a vast images across 80 diverse classes. Notably, the dataset contains a substantial amount of data for humans, along with other classes such as passenger cars, trucks, quadrupeds, and birds. Given that military surveillance systems require detection of objects like tanks and drones, this pre-trained model is aptly suited for further augmentation and customization to meet such specific demands.

III. Experiment and Analysis

A. Experimental Environment

The experiment was conducted using the YOLOv5s model, pre-trained on the previously mentioned dataset. Additional data was collected from an playing field located in Daejeon, utilizing an infrared camera. Videos approximately 22 seconds long were recorded at a resolution of 640x480 with a frame rate of 28fps. From these videos, individual frames were extracted, yielding 600 images in total. Of these, 500 images were used for training, while the remaining 100 images were set aside for validation. Transfer learning and inference, both before and after training, were conducted on a PC equipped with the Intel Core i9-11900K processor and Nvidia GeForce RTX 3090 GPU.

B. Training and Inference

Training was conducted using the yolov5s model, which is identical to the pre-trained model, with an input image size set to 640, and the process ran for 300 epochs. However, training was halted at 294 epochs due to the absence of performance improvement, and for the experiments, a model trained for 184 epochs was utilized. And we employed a strategy that involves retraining the entire model without freezing any specific layers. This was performed to avoid the issue of overfitting that could arise when specific layers are frozen.

Inferences were carried out on the same infrared images using both the pre-trained and the transfer learning models. While the pre-trained model was capable of detecting humans,

Figure 2. Inference result of pre-trained model

Figure 3. Inference result of proposed method

its accuracy was relatively lower, and numerous misclassifications occurred since it was trained on RGB images. In contrast, the model post-transfer learning demonstrated a marked increase in precision in detecting humans, and the incidents of misclassification drastically decreased.

Specifically, the pre-trained model accurately detected humans without any misclassifications in 712 out of 1,100 inference images. And the transfer-learned model achieved accurate detections in 1,095 out of the 1,100 inference images. When accurately detecting humans, the confidence levels were 86% and 93% for the pre-trained and transfer-learned models, respectively. The processing time was identical for both models. The inference result of the pretrained model is shown in Figure 2. And then result of the transfer learning model is shown in Figure 3.

The unexpected efficiency of the RGB pre-trained model in detecting humans can be attributed to the fact that, when detecting objects, the morphological features might be as crucial as the color. Depending on the model, during the processing of all images, they may be converted to a binary format for inference, thereby diminishing the influence of color.

IV. Conclusion

In this paper, we utilized transfer learning based on a daytime model to detect humans in infrared image for the implementation of a military surveillance system. Pre-trained models grounded on RGB images faced challenges in detecting objects post-sunset or under unfavorable weather conditions. These models performed poorly in identifying objects within infrared images. However, the method proposed in this study overcomes these obstacles. Furthermore, by leveraging the pre-trained models as a foundation and applying transfer learning, we enhanced the accuracy using a limited dataset.

For future studies, we intend to incorporate a wider array of objects required by the military surveillance system into our proposed method. Additionally, we aim to develop a system capable of efficiently monitoring distant and small objects.

Acknowledgment

This research was supported by the Institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration and Ministry of Trade, Industry and Energy of Korean government under Grant 22-SN-EC-16

References

[1] S. Park, H. T. Kim, S. Lee, H. Joo and H. Kim, "Survey on Anti-Drone Systems: Components, Designs, and Challenges," in IEEE Access, vol. 9, pp. 42635-42659, 2021.

[2] H. Lee, M. Ra and W. -Y. Kim, "Nighttime Data Augmentation Using GAN for Improving Blind-Spot Detection," in IEEE Access, vol. 8, pp. 48049-48059, 2020.

[3] H. Garg, R. Rana and S. K. Prasad, "A Model to Detect Face Mask Using Deep Learning," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 944-949.

[4] X. Chen et al., "Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset," in IEEE Access, vol. 10, pp. 121301-121317, 2022.

[5] "Ultralytics/yolov5," GitHub, https://github.com/ultralytics/yolov5 (accessed Aug. 2, 2023).

[6] Q. Song et al., "Object detection method for grasping robot based on improved Yolov5," Micromachines, vol. 12, no. 11, p. 1273, 2021.

[7] M. Kim and S. Kim, "Robust appearance feature learning using pixel-wise discrimination for visual tracking," ETRI Journal, vol. 41, no. 4, pp. 483–493, 2019.

[8] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar and P. -A. Muller, "Transfer learning for time series classification," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 1367-1376.