UAV-Based Real-Time Survivor Detection System in Post-Disaster Search and Rescue Operations

Jiong Dong, *Student Member, IEEE*, Kaoru Ota, *Member, IEEE*, and Mianxiong Dong, *Member, IEEE*

***Abstract*—When a natural disaster occurs, the most critical task is to search and rescue trapped people as soon as possible. In recent years, unmanned aerial vehicles (UAVs) have been widely employed because of their high durability, low cost, ease of implementation, and flexibility. In this article, we collected a new thermal image dataset captured by drones. After that, we used several different deep convolutional neural networks to train survivor detection models on our dataset, including YOLOV3, YOLOV3-MobileNetV1, and YOLOV3- MobileNetV3. Due to the limited computing power and memory of the onboard microcomputer, to balance the inference time and accuracy, we found the optimal points to prune and fine-tune the survivor detection network based on the sensitivity of the convolutional layer. We verified it on NVIDIA's Jetson TX2 and achieved a real-time performance of 26.60 frames/s (FPS). Moreover, we designed a real-time survivor detection system based on DJI Matrice 210 and Manifold 2-G to provide search and rescue services after the disaster.**

***Index Terms*—Convolutional neural networks (CNNs), search and rescue, survivor detection, thermal image, unmanned aerial vehicle (UAV).**

I. Introduction

MANY lives are lost every year through natural catastrophes (e.g., hurricanes, typhoons, and earthquakes) in every part of the world. Destructive disasters are usually accompanied by secondary disasters, such as floods, fires, and toxic gas leakage after an earthquake. Secondary disasters not only cause more deaths but also bring more difficulties to rescue work. According to research, the first 72 h after a disaster are extremely important.

In the last decades, many technologies have been developed to search for victims in disaster areas, and the most frequently used is rescue robots. Rescue robots are equipped with a thermal camera to search for survivors in the darkness or bad weather conditions. Moreover, some robots are made with sensors to detect bio-signals or audio signals from humans. Li *et al.* [1] paid attention to the 3-D robotic perception and introduced a deep learning model based on the view-invariant convolutional neural network (CNN) to understand the scene of disaster scenarios. Niroui *et al.* [2] presented a deep learning method to address the robot exploration activities in urban search and rescue (SAR) applications. Although considerable advancement has been made in the research of rescue robots for post-disaster rescue, most rescue robots do not have enough mobility to search for survivors in the disaster-affected area autonomously.

In recent years, the uses of unmanned aerial vehicles (UAVs) have overgrown because of their high durability, lower costs, easy implementation, and flexibility. UAVs are able to quickly search disaster areas from the sky to detect people in need of rescue. Therefore, an intelligent autonomous UAV enabled with image detection capabilities to detect survivors is a well-suited tool to assist SAR missions. Bejiga *et al.* [30] applied UAVs equipped with vision cameras to capture images for assisting avalanche SAR operations. In their study, they proposed a method to combine a machine learning model (support vector machine) with a pretrained CNN to determine the lost person.

The researchers apply UAV equipped with a visible light camera for SAR missions in the literature mentioned above. However, the visible light images are easily affected by smoke, fog, lightning conditions, and sometimes partially occluded by trees or buildings. It is hard for detection algorithms to detect survivors in these images. On the other hand, SAR activities usually continue until night, and standard visible lighting cameras cannot capture anything in total darkness. In this article, UAV with a thermal camera is applied to address these issues because thermal images are not influenced by light, fog, or bad weather.

The thermal camera was initially invented for military night vision objectives. It has been applied to various applications in recent decades, including fire detection, gas detection, building inspection, industrial equipment, medicine, agriculture, and surveillance. In this article, we use the Zenmuse XT2 gimbal

Manuscript received March 6, 2021; revised April 18, 2021; accepted April 19, 2021. Date of publication May 25, 2021; date of current version November 18, 2021. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP19K20250, Grant JP20F20080, and Grant JP20H04174; and in part by the Leading Initiative for Excellent Young Researchers (LEADER), MEXT, Japan. *(Corresponding author: Mianxiong Dong.)*

The authors are with the Department of Sciences and Informatics, Muroran Institute of Technology, Muroran 050-8585, Japan (e-mail: 20096511@mmm.muroran-it.ac.jp; ota@csse.muroran-it.ac.jp; mx.dong@csse.muroran-it.ac.jp).

Digital Object Identifier 10.1109/JMASS.2021.3083659

2576-3164 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

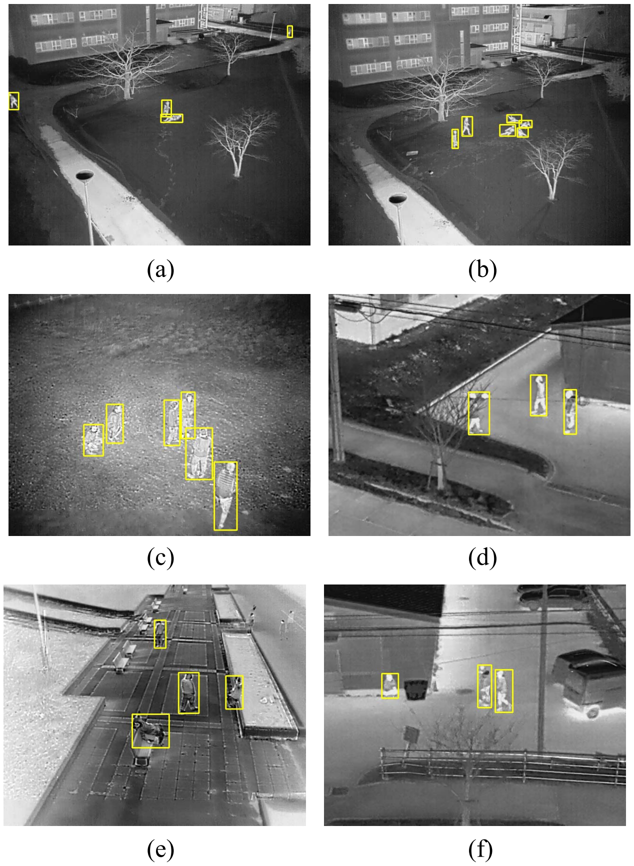


Fig. 1. Several examples of people pose in thermal images captured from DJI Matrice 210 RTK V2. (a) Walking and lying down. (b) Standing and lying down. (c) Standing and squatting. (d) Walking and occluded by the tree. (e) Standing and sitting on the ground or chair. (f) Walking and leaning on the building.

with an FLIR thermal camera for capturing thermal images; the specific details are described in Section III.

The other problem is that people show common postures like walking, standing, or cycling in many pedestrian detection datasets. However, after a disaster, the situation is complex and variable, people are in a wide range of postures like squatting or lying down and sometimes partially occluded by trees or buildings. In this article, we collect a new thermal image dataset to simulate the post-disaster scenario to address this problem. Volunteers came to our laboratory and were asked to exhibit various postures. Fig. 1 depicts various sample images in our dataset, which demonstrate the complexity of victims detection in post-disaster SAR missions.

The pedestrian detection method is essential for UAV devices in post-disaster SAR missions. Over the past 20 years, many researchers have made significant advancements in pedestrian detection, such as the well-known HOG+SVM [32] and Haar+Adaboost [33] methods. With the achievement of technology, especially the development of graphics processing units (GPUs), pedestrian detection algorithms based on the CNN become more and more popular [16]–[19]. However, deep learning methods usually require high computing power and a large runtime memory footprint to maintain good detection performance. Onboard embedded devices do not have these capabilities, and UAV platforms are also restricted by power consumption. Therefore, we need to compress the network to adapt to the UAV platform.

In this article, we use the onboard microcomputer Manifold 2-G[1] to detect survivors in UAV captured video stream; the microcomputer has 8 GB memory and its processor is NVIDIA Jetson TX2.[2] For getting real-time performance on the target platform, we go several steps further in optimizing state-of-the-art deep learning methods. The optimized model achieves three times faster than the baseline detector.

To summarize, the main contributions of this article are as follows.

1) We study the related work about SAR in post-disaster scenarios and find out that UAV cooperates with a thermal camera to solve this problem.
2) We present an original thermal images dataset designed for survivor detection in SAR missions. The dataset contains 6447 thermal images of pedestrians acquired in different scenes while standing, lying, squatting, or occluded by trees or buildings. These thermal images were captured during the day and night.
3) We find optimal values to prune the network for minimizing the model size by experiments. After that, to ensure that the detection accuracy is not lost, we apply the knowledge distillation technology to fine-tune the pruned model.
4) We design a new UAV-aided real-time survivor detection system based on DJI matrice 210 drone and Manifold 2-G to accelerate the SAR process in post-disaster scenarios.

The rest of this article is organized as follows. Related works on UAV-aided post-disaster SAR activities are described in Section II. Section III shows our designed UAV-aided SAR system, the thermal images dataset development, and the optimization detection method. Section IV illustrates the experiment and proves the feasibility of our work. Section V carries out the implementation of a real-time survivor detection system. Finally, Section VI summarizes the whole work and discusses future work.

As an extension of a conference paper [44], in this article, first, we rewrite the introduction and add much additional related work and cite more recently published papers in Sections I and II. Second, we add more details about our thermal image dataset. Third, we add the implementation of the real-time survivor detection system using DJI matrice 210 drone equipped with an FLIR thermal camera and Manifold 2-G.

II. Related Work

Post-disaster SAR, related datasets, and pedestrian detection methods are concerning research topics closely. In the following, we introduce the latest researches on these topics.

A. Post-Disaster Search and Rescue

The primary purpose of any SAR mission is to identify survivors rapidly. There are various approaches presented for

[1] https://www.dji.com/id/manifold-2

[2] https://developer.nvidia.com/embedded/jetson-tx2

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

identifying trapped victims in the literature. Each of these has advantages and disadvantages.

Of all disasters, earthquakes are among the most severe disasters, posing a fatal threat to people. Smaller tremors typically occur within hours or days of a massive earthquake and cause further destruction and death. In order to rescue the survivors as soon as possible and reduce rescue team casualties, rescue robots are used in SAR activities. Zhang *et al.* [4] developed a hybrid locomotion robot for SAR in partially collapsed buildings. Robots access these buildings to search for trapped victims and transport necessary goods, such as food, water, and communication devices. Nazarova and Zhai [5] paid attention to the multirobot SAR system, they first analyzed the rescue operation and the sequences of the rescue process based on statistical data, and then they used the intelligent optimization algorithm to optimize the search path of multirobot. Although rescue robots have been highly developed for post-disaster SAR missions, most of them still do not have sufficient mobility to explore in disaster areas.

In recent years, with the technology development of UAVs, UAVs are becoming more and more widely used in various fields because of their high durability, lower costs, easy implementation, and flexibility. UAVs also play a pivotal role in the field of post-disaster SAR. Kulkarni *et al.* [6] leveraged tools from reinforcement learning to explore the UAV-aided SAR operation in an indoor environment. Surmann *et al.* [7] presented a framework for integrating UAVs in SAR missions, consisting of UAV path planning, an intelligent image center, and a 3-D point cloud generator. Moreover, the framework was tested in several training practices in the Europe project TRADR [34]. Rohman *et al.* [42] proposed a multisensory surveillance drone system that employed an autonomous and robust UAV equipped with multiple sensors. They employ an ultra-wideband radar, microphone array, camera, and an RFID reader. For making the geolocalization of the detected survivors, they also employ laser range finder and LIDAR. However, the vision camera they employed is easily influenced by the light or weather, and this system is not tested in the real fields. Castellano *et al.* [43] presented a new dataset specifically designed for SAR operations from drones using computer vision. However, the dataset is small and is currently intended for testing and evaluation purposes only. On the other hand, the communication network may sometimes be destroyed by unpredictable disasters. And people within the range could not search for rescue. Xu *et al.* [8] proposed access point placement methods and routing to quickly connect survivors in a middle-scale post-disaster scenario. Xu *et al.* [31] focussed on network communication and design UAV-mounted mobile edge computing (MEC) task management strategies based on long range wide area networking (LoRaWAN) to achieve emergency communication. Mekikis *et al.* [41] studied the use of aerial nodes for communication recovery after a communication breakdown.

B. Related Datasets

Some of the datasets related to person detection in thermal imagery are released, such as [9]–[11]. Davis and Keck [9] collected a dataset that only consists of 284 images in 10 sequences for person detection. Then they presented another color-thermal dataset that contains 17 089 images in 6 video sequences [10]. Olmeda *et al.* [11] provided a new thermal image dataset for pedestrian detection, which was collected from driving vehicles in outdoor environments. The dataset was obtained in 13 separate video sequences, each with a different number of images. It consists of 15 224 single-channel images with dimensions of 164 × 129 pixels. The training set and test set consist of 6159 and 9065 images, respectively. In evaluations, only unoccluded pedestrians are considered. Portmann *et al.* [12] released a new dataset, which contained 4381 images, including humans and animals, such as dogs and horses. The dataset consists of 9 outdoor sequences recorded from different perspectives and at different temperatures. Wu *et al.* [13] presented the TIV dataset consisting of seven separate scenes, two of which are indoor scenes. The full resolution is 1024 × 1024. So far, the TIV dataset contains 63 782 images and records thousands of objects. The dataset is still constantly being updated. Torabi *et al.* [14] captured nine video sequences (LITIV dataset). The LITIV dataset consists of videos of various tracking situations acquired at 30 frames/s (FPS) by a visible and thermal camera, with different zoom settings and at different locations. The KAIST multispectral pedestrian dataset [15] includes images taken with varying lighting conditions under various traffic scenes (i.e., data collected both during daytime and at night). The dataset is made up of over 95 000 compatible RGB-thermal image pairs, 50 200 images are used for training, and the rest is used for testing. There are 103 128 annotations corresponding to 1182 pedestrians.

In recent years, some datasets based on drones have also been released in the computer vision area. Zhu *et al.* [38] proposed the VisDrone2018 dataset, which consists of 263 video clips and 10 209 images with rich annotations. Hsieh *et al.* [35] presented a dataset for car counting, consisting of 1448 images captured in 4 different parking lots with the drone platform, including 89 777 cars with bounding box annotations. Robicquet *et al.* [36] collected a large-scale dataset with the UAVs platform with various classes of objects on the university campus. The dataset includes more than 19 000 targets consisting of 58.95% persons, 33.68% bicyclists, 6.84% cars, a bit of skateboarders, and golf carts. Barekatain *et al.* [37] presented a new OkutamaAction dataset captured from a DJI drone at a baseball field in Okutama, Japan, for human action detection. The dataset includes 43 minute-long fully annotated sequences with 12 action classes. Detailed information about the aforementioned image datasets is illustrated in Table I.

The pedestrians in these images dataset show up limited postures, such as walking, standing, and cycling. For post-disaster SAR missions, the victims have various forms. More often, they are lying on the ground, squatting, leaning on collapsed buildings, or being buried in ruins. So, to address this problem, we collected a new thermal image dataset captured by UAV, and the person in these thermal images has different postures. Dataset development is presented in Section III-B.

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

TABLE I
THERMAL IMAGING DATASETS FOR PERSON DETECTION

Dataset	Year	Images	Resolution	Posture
OSU-T [9]	2005	284	360 * 240	Walking, Standing
OSU-CT[10]	2007	17089	320 * 240	Walking, Standing
LITIV[14]	2012	6325	320 * 240	Walking, Sitting on chair
LSI[11]	2013	15224	164 * 129	Walking, Standing
ASL-TID[12]	2014	4381	324 * 256	Walking, Standing
TIV[13]	2014	63782	1024 * 1024	Walking, Standing, Riding
Campus[36]	2016	–	1417 * 2019	Walking, Standing
CARPK[35]	2017	1448	1280 * 720	–
OkutamaAction[37]	2017	77365	3480 * 2160	Walking, Standing,Sitting
VisDrone2018[38]	2018	10209	2000 * 1500	–
Ours	2020	6447	640 * 512	Walking, Standing, Lying down, Squatting,Leaning on building

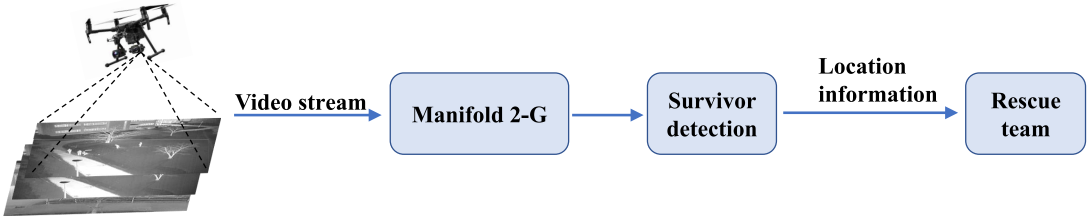


Fig. 2. Steps involved in the UAV-aided SAR system.

C. Pedestrian Detection

Pedestrian detection is a longstanding application in the computer vision area, and a lot of algorithms have been proposed over time. In the last few years, with the advancement of technology, especially the development of GPUs, pedestrian detection algorithms based on the CNN models became more and more popular. These algorithms have been proven to be more effective than any traditional geometric or statistical method. These object detectors based on the deep neural network can be simply divided into two groups: 1) single-stage and 2) two-stage detectors, the main difference is whether extra region proposal modules are required.

Two-stage detectors are all region-based, the region-based CNNs (R-CNNs) family is a typical representative of such detectors [15], [16]. The detection happens in two steps.

1) The model uses a regional proposal network to propose a set of regions of interest. Because the potential bounding box candidates can be infinite, the proposed regions are sparse.
2) Then, the region proposals are sent to the classifier for object classification.

Two-stage detectors apply region proposal modules to obtain high-quality region proposals, thus achieving a good detection accuracy. However, two-stage detectors require huge computation capacities and run-time memory footprint, thus the detection process is relatively slow.

Single-stage detectors skip the region proposal stage and run detection directly across an image via a dense sampling of potential locations. The typical representatives of such detectors are You Only Look Once (YOLO) series models [18]–[20], single shot multibox detector (SSD) [21], and RetinaNet [22]. Single-stage detectors encapsulate all computations into a single network, making it more likely to run much faster than two-stage detectors, although it maybe reaches lower accuracy rates.

However, in UAV-aided post-disaster rescue scenarios, the onboard microcomputer on UAV has a limited computing capacity. For addressing this issue and getting real-time performance, different techniques are explored in the literature. Tijtgat *et al.* [23] designed a UAV warning system and compare both the inference time and accuracy on the Jetson TX2 platform between YOLOV2 and tiny YOLOV2 neural network. He *et al.* [24] proposed an asymptotic soft filter pruning method to accelerate the inference procedure of deep CNN. Zhou *et al.* [25] paid attention to the allocation strategies and design a method to distribute the inference computation of each network layer to different devices in the embedded system.

In this article, we apply YOLOV3 [19] as base architecture and combine other backbone networks, such as MobileNetV1 [26] and MobileNetV3 [27], to test inference time and accuracy on our thermal images dataset. We employ optimization steps to prune the unnecessary filters of these models to minimize the model size. After that, in order to ensure that the detection accuracy is not lost, we use the knowledge distillation to fine-tune the pruned model. The details are described in Section III.

III. METHODOLOGY

In this part, we first present the designed UAV-aided rescue system and the dataset collection process. We then describe the deep learning model used in our rescue system and our proposed method for pruning and fine-tuning the model.

A. System Overview

The steps involved in the UAV-aided SAR system are illustrated in Fig. 2. The UAV is equipped with a thermal camera to search disaster area, the captured video stream is transferred to the onboard microcomputer Manifold 2-G.

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

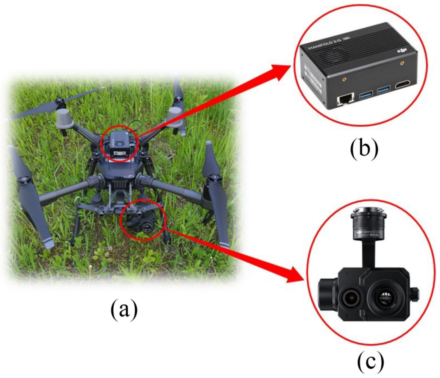


Fig. 3. (a) SAR UAV equipped with the (b) on-board embedded processing platform and manifold 2-G, and (c) Zenmuse XT2 gimbal with FLIR thermal camera.

When the survivor detection model detects victims, the location information will be sent to the rescue team.

The drone used for the SAR mission is DJI Matrice210 RTK V2 [39] (Fig. 3). DJI is the UAV manufacturing leader, has released a series of solutions for multiple disciplines and fields, such as agriculture, express transportation, disaster rescue, energy, public safety, and infrastructure. This kind of drone is designed for Industrial field, with a flight time of up to 33 min and a range of up to 5 miles. The drone has an advanced power management system and equips two batteries during a flight to provide sufficient power and enhance flight safety. It also has a battery heating system, which can significantly extend the flight time in cold temperatures. Thus the drone can be applied in many scenarios. DJI Matrice210 is compatible with Zenmuse gimbal cameras (such as Z30, X4S, XT2, etc.) and can provide high-resolution visible-light as well as thermal images and videos.

For video acquisition, a Zenmuse XT2 [40] gimbal and camera are mounted to the drone, which features a visible-lighting camera and an FLIR longwave thermal camera, simultaneously delivering both thermal and visible-light images. The FLIR longwave thermal camera provides 30 FPS or 60 FPS (depending on the camera model) high-sensitivity thermal image.

The video stream captured by Zenmuse XT2 is sent to onboard microcomputer Manifold 2-G. The Manifold 2-G is DJI's second-generation microcomputer for DJI SDK developers. Its processor is NVIDIA Jetson TX2, which has an onboard Pascal GPU with 256 CUDA cores.

B. Dataset Development

In this section, we first describe the dataset collection and the data preprocessing process. Finally, we outline the properties of our dataset.

1) Dataset Collection: Our data were collected in various locations (on our campus, the grass, or the beach) to simulate people's postures after the disaster. Moreover, participants were asked to exhibit various poses, such as walking, squatting, lying down or leaning on a building, etc. All the actors are a group of researchers in our laboratory. Our data were collected during day and night to simulate the disaster scene under different lighting. We experimented with various thermal camera angles and altitudes to capture videos to find UAVs' appropriate settings, ensuring the poses are recognizable and distinguishable. According to our experiment, we determine the camera angle of 30° or 90° and the altitude range of 15–40 m, respectively.

2) Preprocessing of Dataset: After capturing videos, there are two main steps for preprocessing the dataset: 1) frame selection and 2) pedestrian annotation.

1) *Frame Selection:* Dataset for survivor detection based on the thermal image is collected in video form. Because there are 30 frames in the video per second, to avoid duplication, we skipped every 12 frames in the captured dataset to get one frame for training and testing the survivor detection model.
2) *Pedestrian Annotation:* One of the most time-consuming processes is the annotation of objects in images in the preparation of the dataset. There are some kinds of free image annotation tools, like Labelbox[3] and LabelImg.[4] The choice of image annotation tools depends on the training method used. There is only one object to detect in our dataset, and we chose the tool LabelImg to label pedestrians in thermal images. The annotations are stored as XML files.

3) Dataset Summary and Statistics: The new UAV thermal images dataset for survivor detection contains a total of 3 video sequences and 77 365 frames in 640 × 512 resolution. We extract one image every 12 frames to avoid repetition and get a total of 6447 thermal images. 70% of these pictures were taken during the day and the rest were taken at night. Most of the pictures have more than one person, and these people show different poses. Approximately 65% of postures of people are walking or standing, 25% are lying down, 7% are squatting, and 2% are partially covered by trees or buildings, another 1% of people rely on walls or trees. Such sequences were captured using DJI Matrice210 flying at altitudes ranging from 15–40 m and with a camera angle of 30° to 90°. Table I describes the details of our dataset compared with other datasets.

C. Pedestrian Detection Methods

In this article, we apply DJI Matrice 210 with onboard microcomputer Manifold 2-G to detect survivors in post-disaster SAR missions. The target platform has limited hardware resources and still needs to get real-time performance. We selected the single-stage detector YOLOV3 series due to the excellent processing speed.

YOLO is one of the most advanced real-time object detection systems. YOLO-V3 is the third version of the YOLO algorithm [18]. YOLO-V3 uses the binary cross-entropy loss function to calculate the classification loss. Moreover, this reduces the computation complexity by avoiding the softmax function and replacing the mean square error function. Therefore, it is easy for YOLO-V3 to achieve real-time performance on a computer with a GPU. However, in embedded devices

[3] https://labelbox.com/
[4] https://github.com/tzutalin/labelImg

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

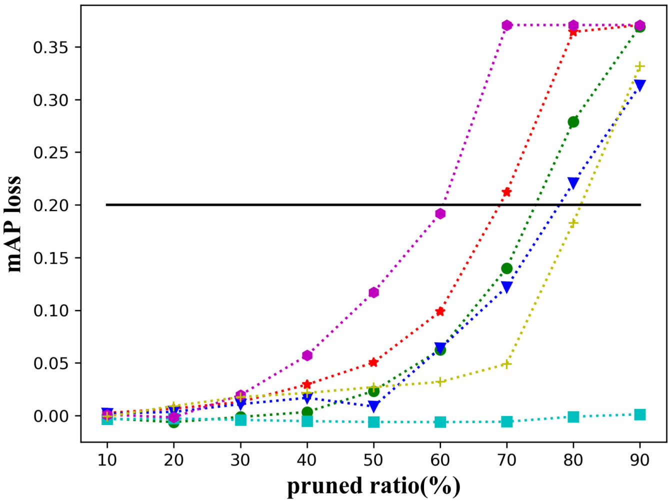


Fig. 4. Sensitives of layers on thermal images dataset.

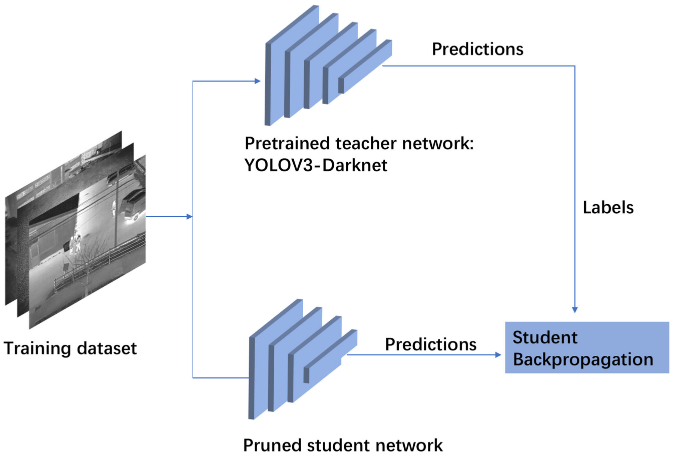


Fig. 5. Knowledge distillation process: replace the training labels of the student network with the prediction of the teacher network.

such as DJI Manifold, the YOLO-V3 model runs slowly. YOLOV3-MobileNet series used MobileNet [26] as the backbone network of YOLOV3 instead of Darknet. MobileNet uses depthwise separable convolutions to build light weight deep neural networks.

The platform Manifold 2-G has limited computation capacity, and we need to compress the network to reduce the model size and accelerate the inference time. Recent researches toward pruning the weights of various layers. However, the method reduces a significant number of parameters from the fully connected layers and causes a large loss of accuracy. In this article, we analyze the sensitivity of each layer based on the method proposed in [28] and then determine how to prune the network. Within the parameters of a convolutional layer, the filters are sorted from high to low according to l_1-norm, and the later filters are less important, these filters are preferentially pruned. When two convolutional layers are pruned filters by the same ratio, we say that the sensitivity of the layer is relatively high when the accuracy has greater impact. Therefore, according to the sensitivity of each convolution layer, different proportions of filters are pruned.

As shown in Fig. 4, the x-axis is the ratio of the filter prune, and the y-axis is the loss of accuracy. Each colored line represents a convolutional layer in the network. Each time a mean average precision (mAP) loss value is selected on the y-axis, there is a set of prune ratios on the x-axis, as shown by the solid black line. We find a set of reasonable prune ratios that meet the conditions by moving the solid black line. We prune a convolutional layer separately with different prune ratios and observe the accuracy loss on the verification dataset. The curve line rises slowly, and the corresponding convolutional layer is relatively insensitive. We give priority to prune the filters of the insensitive convolutional layer.

After pruning the network, the model size of the network is significantly reduced, but the accuracy of the network is also lost. In order to repair the recognition rate of the network, we need to fine-tune the network. In this article, we use knowledge distillation [29]. The main idea of knowledge distillation is to use a complex network with high recognition rates as a teacher model and a small network as the student. Consequently, using the teacher network to retrain the student network. With the help of the teacher network's knowledge, the accuracy of the student network can be improved.

In this article, we use the most straightforward knowledge distillation technique that replaces the labels of the student network with the prediction of the teacher network. The replacement allows the student network to learn from a network that already has its activation regions defined and makes learning easier. The process diagram is shown in Fig. 5. The trained YOLOV3-Darknet network is chosen as the teacher network, we fine-tuned the YOLOV3-MobileNetV1 and YOLOV3-MobileNetV3 networks with it.

IV. Experiment

In this chapter, we show how to adapt and evaluate the deep learning models designed to detect pedestrians with our thermal image dataset of UAV in post-disaster rescue scenarios.

Many object detection methods and CNNs were compared in detail in the previous sections. In the most advanced object detection algorithm, we apply YOLOV3-Darknet as the base architecture and compare it with YOLOV3-MobileNetV1 and YOLOV3-MobileNetV3. In order to get real-time performance on the target platform, we prune the network based on the sensitives of all layers. After that, we apply the trained YOLOV3-Darknet network as a teacher model to distill the pruned YOLOV3-MobileNetV1 and YOLOV3-MobileNetV3. All training experiments were carried out using NVIDIA GeForce GTX 1080 GPU with 8G RAM on Linux 18.04 operating system using the PaddlePaddle[5] framework and testing on the onboard microcomputer Manifold 2-G, which is equipped with the WiFi-ready NVIDIA Jetson TX2 module.

A. Parameter Settings

The parameters trained from the dataset are enormous in the object detection algorithm using deep learning. Moreover,

[5] https://github.com/paddlepaddle/paddle

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

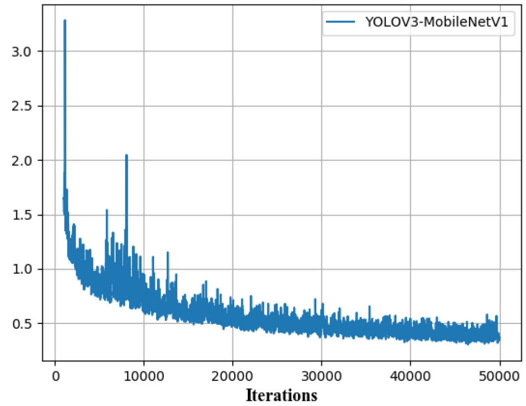


Fig. 6. Loss of YOLOV3-MobileNetV1 model.

the number of hyperparameters set by humans is large. In our experiment, the parameter setting of each algorithm is described below.

For all the three models, the maximum training iteration is set as 50 000, the initial learning rate is 0.0025, which is decreased by a factor of 10 at the iteration step of 40 000 and 45 000. We also use stochastic gradient descent (SGD) with a momentum of 0.9 and a weight decay of 0.0005. The original thermal image size is 640 × 512, which as the algorithm input is resized to 608 × 608. The batch size is set as 8 and the activation function is the ReLU function. Fig. 6 shows the convergence of YOLOV3-MobileNetV1 model.

The anchor boxes are another key factor for training the YOLO series network. Object detection models utilize anchor boxes to make bounding box predictions. To predict and localize many different objects in an image, most state-of-the-art object detection models, such as EfficientDet, and the YOLO models start with anchor boxes as a prior and adjust from there. For example, in the RetinaNet configuration, the smallest anchor box size is 32 × 32. This means that many objects smaller than this will go undetected. In the original YOLOv3, the authors applied the *k*-means algorithm to generate nine different anchor sizes on the VOC dataset. Consequently, we also applied the *k*-means algorithm to generate the corresponding anchors based on our thermal dataset to replace the default value because the object in our thermal dataset is different from the VOC dataset.

B. Training and Test Strategy

For training these CNNs, we select 2000 thermal images (1000 day images and 1000 night images) among our dataset containing 4170 instances of pedestrians. And we also select 400 images to validate these deep networks (200 day images and 200 night images) consisting of 1175 instances of pedestrians. The pedestrian distribution of each image in the training and test set is shown in Fig. 7.

During the testing period, the intersection over union (IoU) threshold of all four methods was set to 0.5. The IoU is defined as

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}. \quad (1)$$

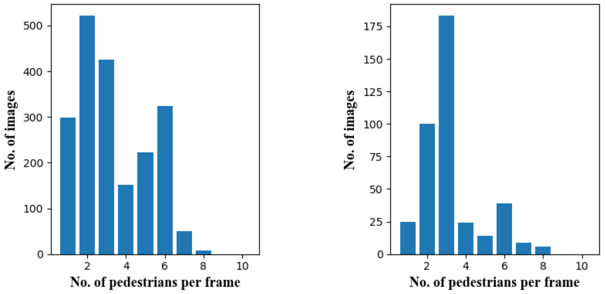


Fig. 7. Distribution of pedestrians in (left) training images and (right) test images.

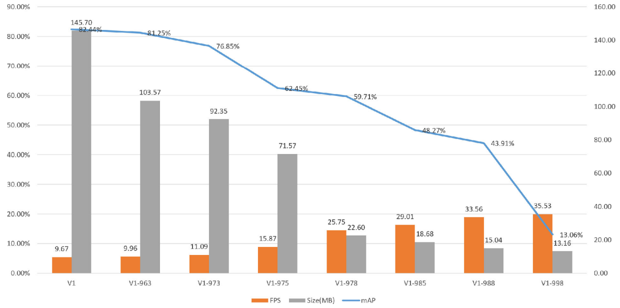


Fig. 8. Comparison of different pruning strategies.

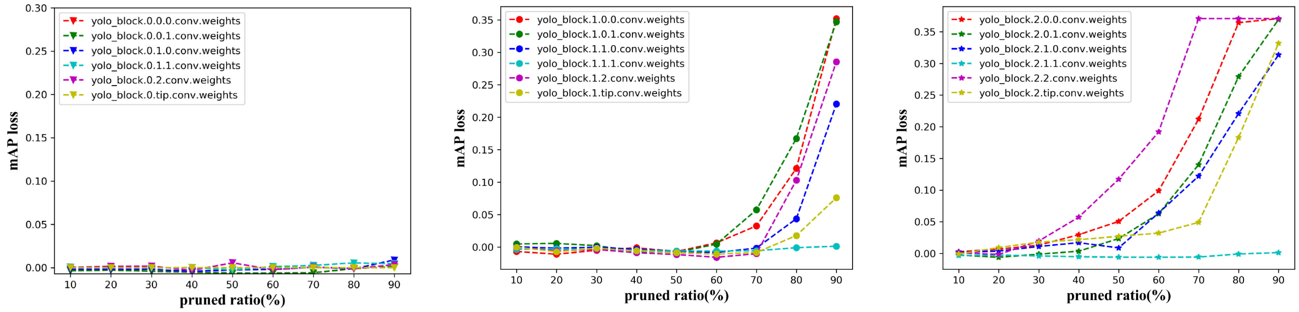
It represents the overlap between the ground truth and the detection bounding box. The larger the value of IoU, the greater the difficulty of detection, but that also means the detection bounding boxes heavily overlap with the ground truth, the detected object has a higher score.

C. Prune Strategy and Knowledge Distillation

In YOLOV3-MobileNetV1 and YOLOV3-MobileNetV3 models, a large number of parameters are mainly concentrated in the yolo-head layers, so we mainly analyze the sensitivity of each layer of the yolo-head part. The detailed sensitives of head layers in YOLOV3-MobileNetV1 are shown in Fig. 9(b). According to Fig. 9(b), it can be seen that in Fig. 9(a), these layers have very low sensitivity, the sensitives are nearly zero, so we can prune the layers with the maximum prune ratio (90%) to reduce the parameters. In Fig. 9(b) and (c), these layers are more sensitive, through a number of experiments, we found that using 70% and 80% prune ratios can minimize the model size while ensuring accuracy. We named this pruning strategy prune_978. After pruning, the inference time of the model becomes faster but the detection accuracy decreases. In order to improve the detection accuracy, we use the trained YOLOV3 network to fine-tune the pruned network. The specific steps are shown in Fig. 5.

In Fig. 8, we show the comparison of different pruning strategies using YOLOV3-MobileNetV1 as an example. For these layers in Fig. 9(a), we maintain a 90% pruning rate because of their low sensitivity. For other layers, we tried different pruning rates. As the pruning rate increases, the model size decreases rapidly and the reference speed increases, but the detection accuracy also decreases. To ensure the accuracy

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Fig. 9. Sensitives of YOLOV3-MobileNetV1 on thermal images dataset. (a) mAP loss with 90% prune ratio. (b) mAP loss with 70% prune ratio. (c) mAP loss with 80% prune ratio.

TABLE II
Performance of Deep Learning Models for UAV Thermal Image Pedestrian Detection

Model	mAP	FPS	Size(MB)	pruned FLOPS	mAP Loss
YOLOV3-Darknet	**85.03**	8.45	374.10		
YOLOV3-MobileNetV1	82.44	9.67	145.70		
YOLOV3-MobileNetV1-pruned	59.71	25.75	22.60	84.61%	27.57%
YOLOV3-MobileNetV1-pruned & fine-tune	61.98	**26.60**	22.70	84.61%	**24.82%**
YOLOV3-MobileNetV3	60.59	14.40	139.60		
YOLOV3-MobileNetV3-pruned	42.71	19.67	20.50	**86.04%**	29.51%
YOLOV3-MobileNetV3-pruned & fine-tune	43.30	20.53	**20.30**	86.04%	28.53%

TABLE III
Effect of Image Size on Accuracy and Inference Time

Image size	mAP	FPS
608 * 608	61.98	26.60
480 * 480	53.17	31.09
352 * 352	43.67	37.42
224 * 224	36.56	41.36

is above 50% and achieve real-time reference speed, the prune_978 strategy is relatively optimal.

D. Results and Discussion

When a disaster occurs, it may cause damage to the ground base station, and the UAV cannot send the captured video data back to the control center or the cloud server. The rescue system we designed uses an onboard microcomputer to process video data and detect the survivor, which solves this problem.

First, we collected a new thermal imaging dataset. Compared with normal UAV images, thermal images are not affected by bad weather and light. Table I describes the comparison of our dataset with the existing thermal imaging dataset. The pedestrians in these datasets show up limited postures, such as walking, standing, and cycling. In our dataset, the survivors have various forms, such as lying on the ground, squatting, leaning on collapsed buildings, or being buried in ruins. The new thermal imaging dataset is more suitable for rescue scenarios.

Next, we need to optimize the neural network model to achieve real-time performance. Table II describes the mAP, FPS, and model size of the selected deep learning network, and we apply YOLOV3-Darknet as the baseline. The YOLOV3-Darknet network achieves the highest accuracy, but the model size is the largest, the value of FPS is only 8.45, which means it cannot get real-time performance on NVIDIA Jetson TX2. The YOLOV3-MobileNetV1 and YOLOV3-MobileNetV3 models are using MobileNet as the backbone instead of Darknet. The model size of the two networks is significantly reduced, but comparing to YOLOV3-Darknet, the mAP values of the two smaller models reduce 3.01% and 28.74%. YOLOV3-MobileNetV3 can achieve 14.40 FPS, but during our experiment, it still has some latency. Consequently, in order to get real-time speed on embedded devices, we first prune the network with our prune_978 strategy. After pruning, the floating-point operations per second (FLOPS) of YOLOV3-MobileNetV1 reduced 84.61%, the model size decreases 93.96%, at the same time, the inference speed is three times faster than YOLOV3-Darknet, the mAP losses 27.57%. As a comparison, the FLOPS, model size, and mAP of YOLOV3-MobileNetV3 reduced 86.04%, 94.52%, and 29.51%, respectively, the FPS is 2.3 times than YOLOV3-Darknet. Second, we apply knowledge distillation technology to fine-tune the pruned network. The pretrained YOLOV3-Darknet is selected as the teacher network, replacing the training labels of the student network with the prediction of the teacher network. After distilling, the mAP and FPS of YOLOV3-MobileNetv1 increased by 3.81% and 3.30%. And these two indicators of YOLOV3-MobileNetV3 have increased by 1.38% and 4.37%.

The size of the input image will also affect the accuracy and the inference time. The initial size of the input image for all the training model is 608 × 608. We choose the pruned and fine-tuned YOLOV3-MobileNetV1 model to test the effect of changes in image size on accuracy and inference time. Table III presents the results. As the image becomes smaller, the runtime speed becomes faster, because, for smaller images, the model extracts fewer features resulting in less computation

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

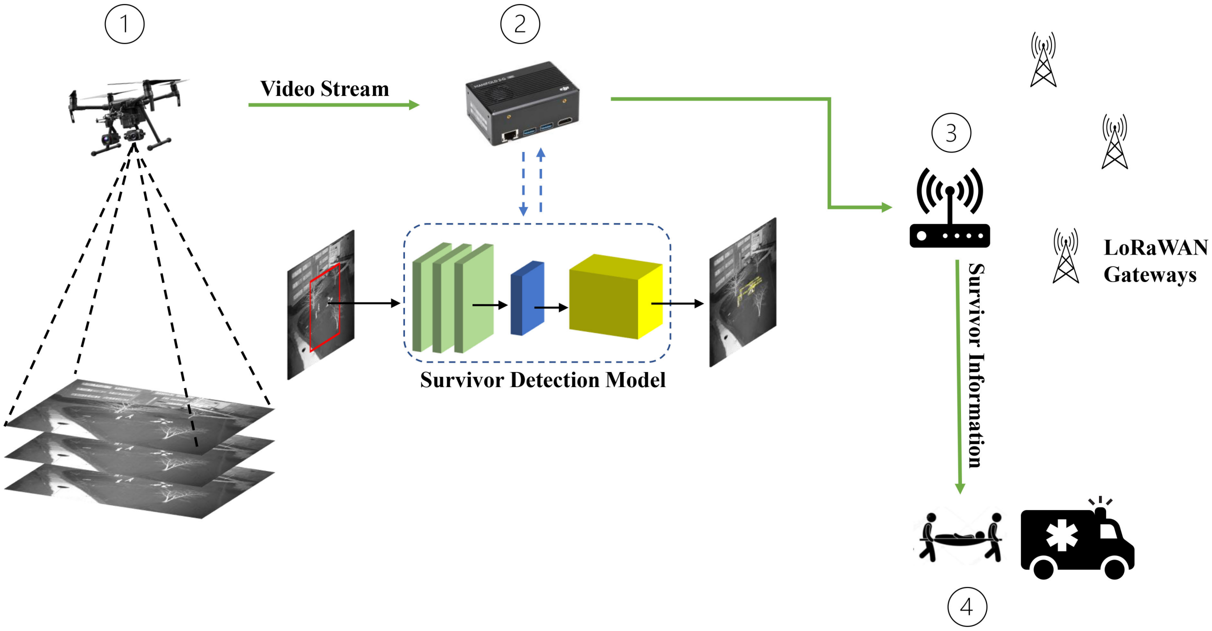


Fig. 10. Real-time UAV-aided SAR system designed in this article.

and reducing memory requirement. However, the accuracy becomes lower because of extracting fewer features to train the model. Thus, there must be a tradeoff between the inference time and accuracy for getting the best rescue results.

In the post-disaster SAR missions, the three factors of the survivor identification accuracy, the runtime speed, and the model size are equally important. Comprehensively comparing the three indicators of the four models, we believe that model YOLOV3-MobileNetV1, which is pruned with the prune_978 strategy and then fine-tuned with YOLOV3-Darknet is more suitable for this task.

In our article, we mentioned some datasets for person detection, among these datasets, some are captured only by the visible light camera, for example, the OkutamaAction and LITIV datasets. For these thermal datasets, such as the OSU-T dataset was taken by surveillance cameras, the size of people is close to our dataset, so good results can be achieved. However, for the KAIST dataset, the pictures in this dataset are all taken by the camera on the car, the postures and size of people are very different from our dataset, so the recognition effect is not good. The proposed model is trained by the thermal images captured by UAV onboard camera, so it works better for photographs taken from the drone perspective.

V. Real-Time Survivor Detection System

In this section, we focus on the implementation of the real-time survivor detection system using a DJI matrice 210 drone equipped with an FLIR thermal camera and Manifold 2-G. Fig. 10 depicts the UAV-aided SAR system. The process of the SAR activities is as follows.

① *DJI Matrice 210:* After a disaster, the drone will fly over the disaster area closely at low altitudes and search the affected areas using the FLIR thermal camera.

② *Manifold 2-G:* The video streams captured by the thermal camera will transfer to Manifold 2-G. The microcomputer will apply the survivor detection model to detect victims in the video streams. Once the victim is detected, the microcomputer will send the location information and the number of victims to LoRaWAN. The DJI Matrice 210 has a GPS module so that we can get the location information.

③ *LoRaWAN:* After a disaster, the base station maybe destroyed, The affected area will lose communication capabilities. In [31], our team workers, Xu *et al.* designed an emergency communication system by LoRaWAN. LoRa could achieve a long-range transmission up to 10 km with low power consumption. In our experiments, we apply this technology. The drone sends the information of the victims to the nearby rescue team through LoRa.

④ *Rescue Team:* Once the rescue team receives the information from the drone, they will SAR the survivors based on the location information.

In a post-disaster SAR mission, the UAV equipped with a thermal camera searches the disaster area as soon as possible. The location of the victim is locked, and the location information is sent to the rescue team through LoRa communication. As a result, with the help of the UAV-aided survivor detection system, we are able to rescue victims and reduce casualties at the same time.

VI. Conclusion

In this article, we focussed on real-time survivor SAR by UAV in post-disaster scenarios. We designed a UAV-aided rescue system based on the flexibility and easy implementation features of UAV to search affected areas quickly. To overcome the effects of bad weather and light, we applied UAV equipped

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

with a thermal camera and also collected a new thermal imaging dataset to train the state-of-the-art deep learning network. Since the onboard microcomputer has limited computing capacity and memory, we find an optimal strategy to prune the neural network by experiments and apply it to prune the YOLO-head layers of the YOLOV3-MobileNet series model, which could reduce the model size by more than 80%, and then we use the pretrained model as a teacher network to fine-tune the pruned network to improve the accuracy. The experiments show that our method can achieve 26.6 FPS real-time performance. In order to motivate further research in this area, we plan to make our dataset available to the public on the link: http://doi.org/10.5281/zenodo.4327118. This dataset can also be used for various surveillance applications for pedestrian detection and other such tasks. In future research, we are going to study the cooperation of multiple UAVs for SAR missions. Another area worthy of research is object tracking. We will verify the performance of existing tracking algorithms on our dataset and then propose new algorithms. Considering the resolution of the thermal image, this will not be trivial.

References

[1] L. Li, K. Ota, M. Dong, and W. Borjigin, “Eyes in the dark: Distributed scene understanding for disaster management,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 12, pp. 3458–3471, Dec. 2017.

[2] F. Niroui, K. Zhang, Z. Kashino, and G. Nejat, “Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments,” *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 610–617, Apr. 2019.

[3] H. Shakhatreh, A. Khreishah, and B. Ji, “UAVs to the rescue: Prolonging the lifetime of wireless devices under disaster situations,” *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 942–954, Dec. 2019.

[4] D. Zhang, Y. M. Shiguematsu, J.-Y. Lin, Y.-H. Ma, M. S. A. Maamari, and A. Takanishi, “Development of a hybrid locomotion robot for earthquake search and rescue in partially collapsed building,” in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Tianjin, China, 2019, pp. 2559–2564.

[5] A. V. Nazarova and M. Zhai, “The application of multi-agent robotic systems for earthquake rescue,” in *Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms* (Studies in Systems, Decision and Control), vol. 272, A. Kravets, Ed. Cham, Switzerland: Springer, 2020.

[6] S. Kulkarni, V. Chaphekar, M. M. U. Chowdhury F. MErden, and I. Guvenc, “UAV aided search and rescue operation using reinforcement learning,” Feb. 2020. [Online]. Available: arXiv:2002.08415.

[7] H. Surmann *et al.*, “Integration of UAVs in urban search and rescue missions,” in *Proc. IEEE Int. Symp. Safety Security Rescue Robot. (SSRR)*, Würzburg, Germany, 2019, pp. 203–209.

[8] J. Xu, K. Ota, and M. Dong, “Fast networking for disaster recovery,” *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 3, pp. 845–854, Jul.-Sep. 2020.

[9] J. W. Davis and M. A. Keck, “A two-stage template approach to person detection in thermal imagery,” in *Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, vol. 1. Breckenridge, CO, USA, 2005, pp. 364–369.

[10] J. W. Davis and V. Sharma, “Background-subtraction using contour-based fusion of thermal and visible imagery,” *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 162–182, 2007.

[11] D. Olmeda, C. Premebida, U. Nunes, J. M. Armingol, and A. de la Escalera, “Pedestrian detection in far infrared images,” *Integr. Comput.-Aided Eng.*, vol. 20, no. 4, pp. 347–360, 2013.

[12] J. Portmann, S. Lynen, M. Chli, and R. Siegwart, “People detection and tracking from aerial thermal views,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Hong Kong, 2014, pp. 1794–1800.

[13] Z. Wu, N. Fuller, D. Theriault, and M. Betke, “A thermal infrared video benchmark for visual analysis,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Columbus, OH, USA, 2014, pp. 201–208.

[14] A. Torabi, G. Massé, and G.-A. Bilodeau, “An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications,” *Comput. Vis. Image Understand.*, vol. 116, no. 2, pp. 210–221, 2012.

[15] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, “Multispectral pedestrian detection: Benchmark dataset and baseline,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1037–1045.

[16] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, 2015, pp. 1440–1448.

[17] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran, 2015, pp. 91–99.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 779–788.

[19] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018. [Online]. Available: https://arxiv.org/abs/1804.02767.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020. [Online]. Available: arXiv:2004.10934.

[21] W. Liu *et al.*, “SSD: Single shot multibox detector,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 2980–2988.

[23] N. Tijtgat, W. Van Ranst, T. Goedemé, B. Volckaert, and F. De Turck, “Embedded real-time object detection for a UAV warning system,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Venice, Italy, 2017, pp. 2110–2118.

[24] Y. He, X. Dong, G. Kang, Y. Fu, and Y. Yang, “Asymptotic soft filter pruning for deep convolutional neural networks,” 2018. [Online]. Available: arXiv:1808.07471.

[25] J. Zhou, Y. Wang, K. Ota, and M. Dong, “AAIoT: Accelerating artificial intelligence in IoT systems,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 825–828, Jun. 2019.

[26] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” 2017. [Online]. Available: arXiv:1704.04861.

[27] A. Howard *et al.*, “Searching for MobileNetV3,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.

[28] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, “Pruning filters for efficient ConvNets,” 2016. [Online]. Available: arXiv:1608.08710.

[29] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” 2015. [Online]. Available: arXiv:1503.02531.

[30] M. B. Bejiga, A. Zeggada, A. Nouffidj, and F. Melgani, “A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery,” *Remote Sens.*, vol. 9, no. 2, p. 100, 2017.

[31] J. Xu, K. Ota, and M. Dong, “Big data on the fly: UAV-mounted mobile edge computing for disaster management,” *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2620–2630, Dec. 2020.

[32] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, pp. 886–893.

[33] S. Zhang, C. Bauckhage, and A. B. Cremers, “Informed haar-like features improve pedestrian detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 947–954.

[34] *TRADR*. Accessed: Dec. 2020. [Online]. Available: http://www.tradr-project.eu/

[35] M.-R. Hsieh, Y.-L. Lin, and W. H. Hsu, “Drone-based object counting by spatially regularized regional proposal network,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 4165–4173.

[36] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning social etiquette: Human trajectory understanding in crowded scenes,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 549–565.

[37] M. Barekatain *et al.*, “Okutama-action: An aerial view video dataset for concurrent human action detection,” in *Proc. Workshops Conjunction IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 2153–2160.

[38] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, “Vision meets drones: A challenge,” 2018. [Online]. Available: arXiv:1804.07437.

[39] *DJI Matrice*. Accessed: Dec. 2020. [Online]. Available: https://www.dji.com/id/matrice-200-series-v2?site=brandsite&from=nav

[40] *Zenmuse XT2*. Accessed: Dec. 2020. [Online]. Available: https://www.dji.com/id/zenmuse-xt2?site=brandsite&from=nav

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.

[41] P.-V. Mekikis, A. Antonopoulos, E. Kartsakli, L. Alonso, and C. Verikoukis, "Communication recovery with emergency aerial networks," *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 291–299, Aug. 2017.

[42] B. P. A. Rohman, M. B. Andra, H. F. Putra, D. H. Fandiantoro, and M. Nishimoto, "Multisensory surveillance drone for survivor detection and geolocalization in complex post-disaster environment," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, 2019, pp. 9368–9371.

[43] G. Castellano, C. Castiello, C. Mencar, and G. Vessio, "Preliminary evaluation of TinyYOLO on a new dataset for search-and-rescue with drones," in *Proc. 7th Int. Conf. Soft Comput. Mach. Intell. (ISCMI)*, Stockholm, Sweden, 2020, pp. 163–166.

[44] J. Dong, K. Ota, and M. Dong, "Real-time survivor detection in UAV thermal imagery based on deep learning," in *Proc. 16th Int. Conf. Mobility Sens. Netw. (MSN)*, Tokyo, Japan, 2020, pp. 352–359.

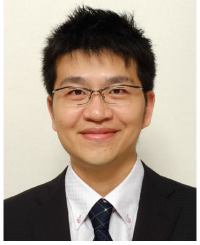
Jiong Dong (Student Member, IEEE) received the B.Eng. degree in computer science from Dalian Maritime University, Dalian, China, in 2015, and the M.Eng. degree in computer science from the École d'ingénieurs Polytechnique de l'université de Tours, Tours, France, in 2017. He is currently pursuing the Ph.D. degree in electrical engineering with the Muroran Institute of Technology, Muroran, Japan.

His main fields of research interest include computer vision and deep learning.

Kaoru Ota (Member, IEEE) was born in Aizuwakamatsu, Japan. She received the B.S. degree in computer science and engineering from the University of Aizu, Aizuwakamatsu, in 2006, the M.S. degree in computer science from Oklahoma State University, Stillwater, OK, USA, in 2008, and the Ph.D. degree in computer science and engineering from the University of Aizu, in 2012.

She is currently an Associate Professor and Ministry of Education, Culture, Sports, Science and Technology (MEXT) Excellent Young Researcher with the Department of Sciences and Informatics, Muroran Institute of Technology, Muroran, Japan. From March 2010 to March 2011, she was a Visiting Scholar with the University of Waterloo, Waterloo, ON, Canada. She was also a Japan Society of the Promotion of Science Research Fellow with Tohoku University, Sendai, Japan, from April 2012 to April 2013.

Dr. Ota is the recipient of the IEEE TCSC Early Career Award in 2017, the 13th IEEE ComSoc Asia–Pacific Young Researcher Award in 2018, and the 2020 N2Women: Rising Stars in Computer Networking and Communications. She is a Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).



Mianxiong Dong (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science and engineering from the University of Aizu, Aizuwakamatsu, Japan, in 2006, 2008, and 2013, respectively.

He is the youngest ever Vice President and a Professor of the Muroran Institute of Technology, Muroran, Japan. He was a Japan Society of the Promotion of Science (JSPS) Research Fellow with the School of Computer Science and Engineering, University of Aizu, and was a Visiting Scholar with the BBCR Group, University of Waterloo, Waterloo, ON, Canada, supported by the JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011.

Dr. Dong was selected as a Foreigner Research Fellow (a total of three recipients all over Japan) by NEC C&C Foundation in 2011. He is the recipient of the IEEE TCSC Early Career Award in 2016, the IEEE SCSTC Outstanding Young Researcher Award in 2017, the 12th IEEE ComSoc Asia–Pacific Young Researcher Award in 2017, the Funai Research Award in 2018, and the NISTEP Researcher in 2018 (one of only 11 people in Japan) in recognition of significant contributions in science and technology. He is a Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).

Authorized licensed use limited to: UNIVERSITY OF HERTFORDSHIRE. Downloaded on December 31,2024 at 11:02:34 UTC from IEEE Xplore. Restrictions apply.