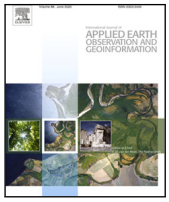
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103052

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



# Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly supervised object localization

Minsoo Park [a], Dai Quoc Tran [a], Jinyeong Bak [b], Seunghee Park [a,*]

[a] School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon, 16419, Gyeonggi-do, Republic of Korea

[b] College of Computing and Informatics, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon, 16419, Gyeonggi-do, Republic of Korea

ARTICLE INFO

*Keywords:*
Disaster response system
Decision support
Synthetic data
Weakly supervised object localization
Channel attention module

ABSTRACT

Owing to abnormal climate phenomena worldwide, forests are becoming dry and heat waves have started to occur, increasing the damage caused by wildfires. In addition to causing significant human and material damage, wildfires are also a major cause of critical pollutant emissions, in which fine dust generated by incomplete combustion pollutes the atmosphere, soil, and water. Early detection and monitoring are some of the main ways for minimizing wildfire damage, and a topic of research interest in various fields of artificial intelligence and computer vision. However, the lack of wildfire occurred image datasets is still challenge. Training deep learning model in this environment, can lead mis-detection when burning point is far from the camera or according to objects similar to flame and smoke. Our study attempted to create synthetic wildfire images in various shapes by inserting damage into a free-wildfire image using generative adversarial network (GAN) and Weakly supervised object localization (WSOL). The synthesized image can used as training data for object detection by applying the WSOL method with gradient-weighted activation map (Grad-CAM). Additionally, the YOLOv5s model was improved by adding a channel attention module; sequence-and-excitation (SE) layer and replace loss function as CIoU to address the issue of wildfire false detection in fire-like object and miss detection in small size smoke. Our proposed method, produced results as high as 7.19% in F1-score and 6.41% in average precision (AP) when compared to the existing traditional method. To use a deep learning model in practice, a lightweight model should be applied to the embedded models while maintaining high performance. The developed AI model was applied to the established drone and CCTV-based wildfire monitoring system, and a virtual experiment was conducted by generating virtual wildfires near forests in Korea.

## 1. Introduction

In the past two decades, wildfires have become more frequent and destructive because of the increased dryness and frequency of heat waves that result from global warming (Aguilera et al., 2021). A recent example is the 2019–2020 series of unprecedented wildfires in Victoria, Australia, which burned at least 10 million hectares (Jager and Coutant, 2020). In California, USA, an August Complex fire occurred in 2020, which burned 417,898 ha and was completely extinguished after nearly three months (Moreno, 2021). In the Republic of Korea, the Uljin/Samcheok wildfire of 2022 burned an area of 20,923 ha, leading to the maximum wildfire damage since 1986, according to the related statistics. Considering the severe impact of wildfires on human safety, economy, ecosystems, and infrastructure, wildfire management strategies and policies are continuously being developed worldwide. Specifically, the use of a technology that minimizes damage by detecting and extinguishing wildfires in early stages is one of the main strategies; it can be considered as the first and basic step to effectively respond to wildfire disasters.

To detect wildfires in an early stage, satellite, terrestrial, and aerial devices have been mainly used in the recent studies. With satellites, extensive wildfire detection is possible using high-temperature-sensitive short-wave infrared and thermal infrared channels (Kato et al., 2021). However, sun-synchronous satellites exhibit a high spatial resolution but low time resolution. Furthermore, it is difficult to detect wildfires in real time using sun-synchronous satellites, whereas geostationary satellites exhibit high time resolution but low spatial resolution, making it difficult to detect wildfire in early stages. Additionally, satellite data requires additional effort to eliminate errors because of interference from a wide range of clouds. Conversely, small UAVs or surveillance

* Corresponding author.
*E-mail addresses:* pms5343@skku.edu (M. Park), daitran@skku.edu (D.Q. Tran), jy.bak@skku.edu (J. Bak), shparkpc@skku.edu (S. Park).

https://doi.org/10.1016/j.jag.2022.103052
Received 10 July 2022; Received in revised form 30 September 2022; Accepted 7 October 2022
Available online 28 October 2022

cameras, such as CCTV camera, offer much lower operating costs than satellites, higher mobility, flexible perspective, and resolution, and high potential to detect wildfires early and provide on-site information (Chi et al., 2017). Given that wildfires mainly spread from low amounts of smoke, early detection from cameras installed in areas with high risk of wildfires can contribute more to securing the golden time for extinguishing wildfire, as opposed to detecting the fire with low spatial resolution over a wide range.

With the rapid development of digital cameras and image processing technologies, deep learning object detection algorithms can integrate parallel computating and graphics cards to process at real-time or near real-time speeds. There has been a significant interest worldwide in developing real-time models for detecting wildfires using common video-based surveillance systems with deep learning technology such as convolutional neural networks (CNNs) (Toulouse et al., 2015; Tran et al., 2020; Toulouse et al., 2017; Chu and Thuerey, 2017; Zhang et al., 2018b; Namozov and Im Cho, 2018; Sousa et al., 2020; Li et al., 2019; Jung et al., 2020; Park et al., 2020; Tran et al., 2022). These computer vision techniques exhibit many advantages over traditional flame and smoke detection, such as early fire detection, high accuracy, flexible system installation, and the ability to effectively detect fires in large spaces and complex building structures (Muhammad et al., 2018). Additionally, recent studies (Ren et al., 2015; Kortylewski et al., 2020) have shown that deep learning algorithms exhibit great detection performance in forest fields and can robustly detect objects in environments obscured by occlusion, lighting changes, various deposits and structures around the forest, shadows, and street trees. Recently, attempts have been made to improve the accuracy of wildfire detection models by integrating or modifying deep learning algorithms based on classifiers such as DenseNet (Huang et al., 2017) and EfficientNet (Tan and Le, 2019) and modern object detectors such as YOLOv5 and YOLOR (Wang et al., 2021b) which have been developed from idea of YOLOv4 (Bochkovskiy et al., 2020) to enhance detection accuracy.

Meanwhile, research on deep learning-based wildfire detection is challenging because of the lack of expertly annotated wildfire image datasets to train deep learning algorithms (Sousa et al., 2020; Li et al., 2019; Park et al., 2020). The lack of images is because of the following factors: (1) Accessibility of wildfire images is limited (available solely from installed fire surveillance cameras and drones that require permission from local governments to photograph) and the amount of online resources that provide wildfire images and videos is relatively sparse; (2) it is difficult to obtain sufficiently diverse image data by splitting frames of wildfire videos at the same site and using them as datasets; and (3) flame and smoke objects are difficult to label because they are amorphous and exhibit irregular shapes. Therefore, the subjective judgment of researchers cannot be excluded in annotating wildfire object detection.

Deep learning is highly dependent on data. The main problem with small datasets is that deep learning model overfits to the training examples and becomes ungeneralizable. Previous wildfire detection network studies (Park et al., 2020; Zhao et al., 2018) have mentioned the low recognition rate from small datasets.

Data augmentation (DA) is one of the standard solutions with respect to overfitting. Deep generative adversarial networks (GANs) (Foggia et al., 2015a) are algorithmic architectures that use two neural networks for creating new synthetic data instances. GANs have been used in studies pertaining to tasks, including super-resolution, I2I translation, and text-to-image translation. Specifically, the usefulness and value of GANs for I2I translation, which translates one image into another image as a form of DA, is increasing. In a recent example, GANs could quickly make up for an insufficient dataset in developing a CNN model for detecting COVID-19 infection from X-ray images. Given that this is a relatively recently discovered virus, the image dataset is insufficient, but the performance of the CNN model can be improved by supplementing this problem with the studies by Waheed et al. (2020) and Karbhari et al. (2021), as well as by many other scholars. Because it is very difficult to obtain datasets in these disease or disaster detection fields, GANs have often been used as a DA method.

In deep learning for computer vision, data labeling is the process of identifying raw data (image or video) and adding meaningful information and digital contours to objects in the images that the model can learn from. Although some relevant raw image datasets are available from several data sources, the manual annotation of images is costly in terms of time and labor. Although it may be relatively easy to label objects for classification purposes, it is very laborious to create bounding boxes for labeling localization of the object in an object detection field. Furthermore, it is difficult to guarantee objectivity in manual labeling work that inevitably involves the subjective judgment of humans.

In this study, a methodology for solving the aforementioned problems was proposed in the development of a deep learning-based computer vision model to detect wildfires. Three types of machine learning methods are used in this method: an unsupervised learning-based GAN for I2I translation, a supervised learning-based CNN model for weakly supervised method-based image annotation, and a one-stage based wildfire detector. First, a small group of non-benchmarked wildfire datasets is collected through several sources and combined with the published ordinary forest image datasets to generate synthetic images via GAN. This addresses the class imbalance between non-fire and wildfire images. Three models are used in our study, including CycleGAN (Zhu et al., 2017) for generating wildfire images and ensuring the objectivity of the experiment. Various combined datasets between the DA from the original wildfire images are used in the training of the commonly used classification networks. Best performing model from experimental results is selected via evaluation metrics that include precision, recall, F1-score, and ROC–AUC. Second, wildfire instances from generated images are localized from the selected classification model with bounding boxes, similar to object detection dataset annotation in weakly supervised methods. The overview of the proposed DA architecture is shown in Fig. 1. Finally, the labeled data generated by the weakly supervised method is trained on proposed improved YOLOv5 based detection network to prove the effectiveness of the automatic annotation method. Then, the final model is embedded on a real-time drone and CCTV wildfire monitoring system. In the proposed model, the environment for early detection of wildfires is considered.

In this process, we expect to automate the labeling task, and thereby, solve the basic problem of data shortage as well as eliminate human error. This study addresses the following challenges:

1. Can the training of a wildfire detection model with synthetic wildfire images generated by GANs improve the model performance when compared to using standard DA techniques?
2. Can weakly supervised methods with GAN be used to create datasets for wildfire object detection?
3. To what extent can the proposed wildfire early detection model be improved when compared to the existing model in terms of accuracy and device applicability?

The rest of this study is organized as follows. In Section 2, we summarize the strategy to overcome the lack of original datasets. In Section 3, we propose an image annotation method and construct a wildfire detection model. Furthermore, in this section, we define metrics for evaluating the deep learning models. In Section 4, we discuss relevant experimental results and evaluate the robustness of the model using the data generated by the proposed method. In Section 4.7 , we further discuss the experimental results. Finally, in Section 5, we conclude the study and provide suggestions for future research.

## 2. Related work

### 2.1. Address small dataset problem

Given the scarcity of wildfire-related image datasets in existing research, a wildfire detection model was developed by collecting images

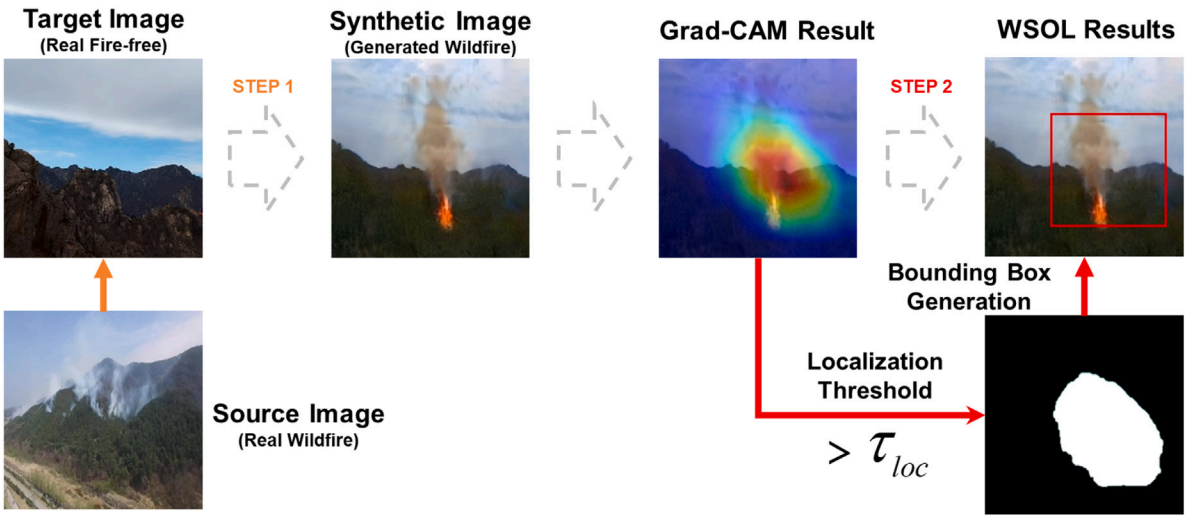



**Fig. 1.** Framework of the proposed advance wildfire detection model: Step (1) image-to-image (I2I) translation for synthesizing wildfire images. Then train wildfire image classification model (backbone CNN model for WSOL) using dataset with images created by the GAN DA method. Step (2) To insert localization information on image data, localization threshold ($\tau_{loc}$) for Grad-CAM heat map is applied. Consequently, bounding boxes are generated from the threshold. These datasets are used for training data for wildfire object detection model.

from the researchers themselves by crawling images on the Internet (Toulouse et al., 2015, 2017). Many flame detection image datasets have also been provided by researchers. Foggia et al. provided a dataset consisting of 31 different fire videos of varying resolutions (with 400 × 256-pixel resolutions on average) captured by cameras from different scenes (Foggia et al., 2015b). Steffens et al. (2015), Hüttner et al. (2017), and Chenebert et al. (2011) provided additional datasets of fire videos. However, early detection of wildfires requires data related to smoke as opposed to flames. Furthermore, even with the large quantity of datasets, the training dataset consists mostly of video frames, which include many duplicate images. This provides an insufficient association when training a model to ensure robustness with a low error rate in new environments. Moreover, these datasets do not provide any further information about each data. Hence, the researchers had manually perform pre-processing such as labeling.

Transfer learning is a method for feature representation from a pre-trained model that utilizes high-level feature information obtained via the classification task from a large dataset in advance. This method is usually trained from ImageNet, and the weights obtained from the trained model can be used for a new target model. The source domain from ImageNet and target domain differ. However, they are related such that common instances or features can be transferred between domains. This method is useful when working with similar datasets to prevent overfitting when the dataset of a new classification problem is small (Ioffe and Szegedy, 2015). Transfer learning has been used to improve the performance of wildfire detection models (Karbhari et al., 2021; Park et al., 2021) in recent research. However, given that the quantity of ImageNet data is also limited, it is not possible to learn all feature information in advance. Hence, this is not a universally applicable method. Additionally, related studies have shown that transfer learning results in poor performance when the target domain differs highly from the source domain Langnickel and Fluck (2021), Xu et al. (2017).

Large amount of relevant learning images are required for identifying high-level features (Zhao et al., 2018). In early related studies, standard DA method such as rotation, color conversion, and blurring of images were used as augmentation techniques. However, these techniques do not sufficiently change the original image in terms of color and shape, and an improved level of performance is not significantly realized (Park et al., 2020; Waheed et al., 2020). In certain studies, the performance of a model can be improved by inserting images of simulated or previously filmed smoke into the background of mountain images (Zhang et al., 2018b; Xu et al., 2017). Although this method can induce the deep learning model for detecting wildfires in various backgrounds, it still does not sufficiently change the image and requires costly manual labor for preparing the training data. Conversely, certain studies raised the possibility of quickly creating appropriate and sufficient data sets based on deep learning (Chu and Thuerey, 2017). GAN exhibit the ability to generate unique data for higher influence on the efficiency of learning model performance when compared to other general data aggregation methods. Specifically, GAN exhibited the ability to generate synthetic data that reflect various wildfire scenarios of a target forest, thereby improving the model's detection capabilities. Namozov and Im Cho (2018) used the GAN network to apply the winter and night background on the original fire image and added different seasons and times on the dataset for training their fire/non-fire classification model. However, this domain adaptation method made it difficult to detect flame–smoke with abundant morphological variations because only the background of the original image was changed and the shape of the flame–smoke is still retained during the synthetic image generation process. Park et al. (2020) attempted semantic synthesis method to generated various shape of wildfire dataset as mapping flame–smoke to non-wildfire images using Cycle-GAN. However, this method is still intended solely to train the classification model only for mapping or synthesizing images. To utilize a generated image for training the detection model, localization information, which still depends on manual labor, must be inserted.

### 2.2. Object localization from weakly supervised learning

Unlike unsupervised learning, supervised learning methods require large quantities of images with high-quality labeling (Oquab et al., 2015). Although tagging of images as an object class is relatively simple, careful annotation is required in the form of bounding boxes, which can be costly and susceptible to human error (Felzenszwalb et al., 2010). Weakly supervised object localization (WSOL) refers to the problem of finding a region of interest in a weakly labeled image. This implies that although the model was trained only with classified labeled images without localization information, this method supports the CNN localization capability from image classifier. WSOL aims to find objects during training with only image-level labels. This approach has drawn attention in computer vision research because switching from object label level to object localization level has the advantage of

**Table 1**
Description of wildfire and non-wildfire datasets. We use these datasets to generate synthetic images by GAN and train and test wildfire image classifier and wildfire detector.

| Dataset | Division | # images | GAN | Classifier | Detector | Description |
|---|---|---|---|---|---|---|
| Ours | Wildfire | 712 | ✓ | ✓ | ✓ | Drone footage; Wildfire images from Youtube |
| HPWREN (University of California San Diego, California, America, 2021) | Wildfire | 738 | ✓ | ✓ | ✓ | CCTV footage; Early wildfire |
| MLID (Center for wildfre research University of Split Faculty of Electrical Engineering et al., 2014) | Wildfire | 100 | ✓ | ✓ | ✓ | Smoke objects |
| Mendeley (Khan and Hassan, 2020) | Wildfire | 950 | – | ✓ | ✓ | Smoke objects; Flame objects |
| KTS (Jeong et al., 2019) | Non-wildfire | 900 | ✓ | ✓ | – | Mountain images from Korea |
| DNIM (Zhou et al., 2016b) | Non-wildfire | 94 | – | ✓ | ✓ | Fire-like objects |
| AI Hub (Jang, 2020) | Non-wildfire | 1,436 | – | ✓ | ✓ | Fire-like objects; Smoke-like objects |
| MIT spatial envelope dataset (Oliva and Torralba, 2001) | Non-wildfire | 612 | – | ✓ | – | Mountain and forest images |
| Ours | Non-wildfire | 2,978 | – | ✓ | – | Drone footage |

reducing the cost of generating the train label (Zhang et al., 2018c; Xue et al., 2019). Many research have shown that weak supervision learning makes errors, however despite this, it is utilized as an automatic labeling method to increase accuracy in the field of object detection and segmentation (Zhang et al., 2018d,e; Sohn et al., 2020). And a recent study tried to minimize the error of automatic labeling and generate an optimized bounding box based on the average precision of the object detection model (Li et al., 2022).

Typically, WSOL depends on localization maps from classification networks learned from small datasets to realize annotation (Lee et al., 2019) In a pioneering study, Zhou et al. (2016a) proposed adding a global average pooling (GAP) layer to classification networks. By weighting the feature map of the last convolution layer on the ConvNet and upsampling the calculated weighted map to equalize it with input image size, it is possible to localize the area of the image that corresponds to the target class. This is the basis for many weakly supervised localization problems. Some studies improve the class activation map (CAM) pipeline (Zhang et al., 2018c; Selvaraju et al., 2017). Furthermore, CAM can only be applied when the ConvNet model must include GAP. These limitations lead to a major disadvantage wherein most deep learning models use various structures at the output terminal. Therefore, a flexible method, which is not constrained by the structure of the model, was required, and Gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) was devised for this requirement. Hence, it can be applied more generally to CNN-based models because conversion of the last layer to GAP is not required. Additionally, it performs better than CAM on weakly supervised localization task.

Smoke often exhibits an amorphous and irregular shape (Tang et al., 2020). Hence, there is a possibility that noise can occur when this method is applied. In our study, we use synthesized images that are difficult to label objectively and accurately. Therefore, it aims to improve the performance of the model by performing annotation using WSOL and adjusting the localization threshold that distinguishes objects and backgrounds in the process of generating a localization map.

## 3. Materials and methods

### 3.1. Generation of wildfire images

#### 3.1.1. Image dataset

According to the proposed goal of labeling based on the classification model, datasets are classified into wildfire and non-wildfire. To the best of our knowledge, wildfire images for CNN have no public benchmarks. Hence, a small amount of data has been collected from the web to construct a dataset. In this section, the sources and types of image data used in the study are detailed. All collected images were resized to 224 × 224 pixels to fit the input size of the CNN model. 1 contains the used dataset and settings.

The proposed detection model includes photographs of wildfires in Korea acquired from unmanned aerial vehicles (UAVs) by private companies and the Korea Forest Service. To represent the forest sites of Korea as accurately as possible, images captured from wildfires in Korea were included in the training data. Fig. 2 visualizes the areas of wildfire occurrence on a map. Forested areas in Korea are heavily skewed to the east. Additionally, high-performance wireless research and education network (HPWREN) dataset (University of California San Diego, California, America, 2021), Mediterranean Landscape Image Dataset (MLID) (Center for wildfre research University of Split Faculty of Electrical Engineering et al., 2014), and Mendeley dataset (Khan and Hassan, 2020) were also included. With the addition of images obtained from the Web, a total of 2,500 wildfire images were prepared.

To augment the data by translating regular forest images provided by the Korean Tourist Spot (KTS) (Jeong et al., 2019) dataset into fire-free forest images (non-wildfire images). Moreover, wildfire-like images are similar in color and shape to smoke and flames, which can be often incorrectly identified as wildfire images, were also included in the class. Hence, day–night image matching (DNIM) (Zhou et al., 2016b) data was used to reduce the impact of illumination changes. Similar images obtained from AI Hub (Jang, 2020) managed by Ministry of Science and Technology Information and Communication in South Korea were also used. Specifically, 1436 extracted images include fog, clouds, sunlight, car lights, sunsets, and other features with visual properties similar to those of wildfires. DNIM and AI Hub dataset were used as the training data to lower false positives, and were also used to evaluate false detection rates for used testdata on similar images environment. Parts of the MIT spatial envelope dataset (Oliva and Torralba, 2001) were also classified as our non-wildfire class. Other images were collected from the web, and finally, 6,020 images were set for this class. Table 2 describes the history of wildfires photographed by drones that we used in this research.

Mountain images of KTS were synthesized from GAN as a wildfire and set as training data for classifier and detector models. Its usage is significant because it can be used to the test bed for the Korean forest fire detection system since the environmental domain are similar. GAN source data were set by combining night and day wildfire images obtained from CCTV and drones with HPWREN and MLID. Since the Mendeley dataset consists mainly of images of a wildfire in progress, which is relatively easy to collect, it has a different domain from the purpose of generating a synthetic images which is difficult to be collected (early stage of a wildfire images). Therefore, this dataset was excluded by GAN training. However, apart from the purpose of data generation, the monitoring system should be able to detect even when a wildfire is in progress or close to the recording equipment, so it is included in the training and testing of classifier and detector.

**Table 2**
Location of the wildfire in Korea with the drone capture image used in this study.

| Location | Date of occurrence | Damaged area |
|---|---|---|
| Chiak-mountain, Seongnam-ri, Sillim-myeon, Wonju-si, Gangwon-do | April 25th, 2021 | 1.5 Ha |
| 609, Geumsan-ri, Seongsan-myeon, Gangneung-si, Gangwon-do | February 18th, 2021 | 0.4 Ha |
| Gaya-mountain, 890-1, Ma-dong, Gwangyang-si, Jeollanam-do | February 10th, 2021 | 3 Ha |
| Sacheon-ri, Yangyang-eup, Yangyang-gun, Gangwon-do | February 8th, 2021 | 6.5 Ha |
| Ingeum-ri, Pungcheon-myeon, Andong-si, Gyeongsangbuk-do | April 24th, 2020 | 1944.00 Ha |
| Daebok-ri,Ungchon-myeon, Ulju-gun, Ulsan | March 19th, 2020 | 519.00 Ha |
| 30-1, Balsan-ri, Sinbuk-eup, Chuncheon-si, Gangwon-do | January 4th, 2020 | 34.26 Ha |
| Namyang-ri, Okgye-myeon, Gangneung-si, Gangwon-do | April 4th, 2019 | 1260.15 Ha |
| 394-4, Wonam-ri, Toseong-myeon, Goseong-gun, Gangwon-do | April 4th, 2019 | 1266.62 Ha |
| Songcheon-ri, Seo-myeon, Yangyang-gun, Gangwon-do | January 1st, 2019 | 97.94 Ha |
| Hajeong-ri, Miro-myeon, Samcheok-si, Gangwon-do | December 29th, 2018 | 27 Ha |
| Sikjang-mountain, Nangwol-dong, Dong-gu, Daejeon | September 9th, 2015 | 0.5 Ha |

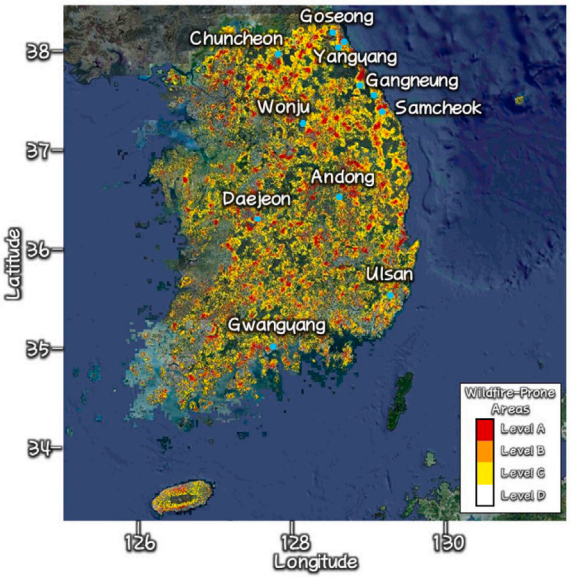



**Fig. 2.** Wildfire-prone area level of the location of wildfire where drone images were acquired; level A is the most dangerous. The vulnerability map was made by the Korea Forest Service.

### 3.1.2. Generative adversarial networks data augmentation (GAN DA)

GAN extracts the object's characteristics from the source data and maps them as target data. To overcome the phenomenon where there is no significant effect of featuring various types of damage (smoke or flame) shapes in traditional method, mapping of the damage to the non-damage map method was attempted via GAN. The I2I translation was used to diversify and increase the number of images used for binary classification model training. Specifically, 900 source images were set randomly from the collected wildfire dataset and 900 target images were set from KTS. To diversify the source set, nighttime images of wildfires were included. Wildfire image was generated using three typical I2I methods to prove that the configuration change method of our source image and target image is better than style transporter-based GAN methods. Three types of GANs were used in this study: Cycle-GAN (Zhu et al., 2017), DRIT (diverse image-to-image translation) (Lee et al., 2018), and NICE-GAN (no-independent component for encoding GAN) (Chen et al., 2020); they are marked as G1, G2, and G3, respectively, in this study.

Cycle-GAN is consistent with our experimental objective because the synthetic images are generated from different domain of datasets without using the pair-images such as pix2pix (Isola et al., 2017). The experiment was carried out under identical settings in DRIT and NICE-GAN, to strengthen the reliability of the experiment and validate the effect of GAN DA in terms of wildfire detector accuracy. DRIT is a structure that is embedded by dividing it into a shared content space between two different domains and an attribute space that represents the characteristics of each domain Lee et al. (2018). It can create flames or smoke without harming forest space, synthetic images created by receiving attributes in the fire domain and content in the forest domain can provide realistic wildfire images. Instead of creating a separate encoder for the generator, NICE-GAN uses the first layer of the discriminator as the generator encoding layer, which is a more succinct and efficient architectural framework (Chen et al., 2020). Because the encoder is trained directly on the discriminator, this structure can train color and texture of smoke successfully.

The quality of synthesized images from each GAN was evaluated using learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018a) and Fréchet inception distance (FID). The LPIPS distance measures the diversity of synthetic images by calculating the average feature distance between generated images. A higher LPIPS value indicates a greater diversity within the generated image set and a good correlation with human perceptual similarity.

**Table 3**
Number of images for training and testing wildfire image classification models.

| Data partition | Wildfire | Non-wildfire |
|---|---|---|
| Total: Original | 2,500 | 6,020 |
| Training: Original | 1,080 | 4,680 |
| Training: Brightness transformation DA | 900 | – |
| Training: Saturation transformation DA | 900 | – |
| Training: Horizontal flip DA | 900 | – |
| Training: Rotation transformation DA | 900 | – |
| Training: GAN DA | 900 | – |
| Test: Original | 1,420 | 1,340 |

## 3.2. Wildfire classifier for WSOL

### 3.2.1. Data partition

DA was needed for the Wildfire class (2500) because the balance of the number of data was insufficient when compared to the non-wildfire class (6020). Standard DA: brightness control (adjust from −10% to +10%), saturation control (adjust from −10% to +10%), and other affine transformations (Mikołajczyk and Grochowski, 2018), such as rotation transformation and horizontal transformation, were used. When the brightness or saturation was too high or low, the numerical range was adjusted in consideration of the phenomenon that the shape of the object becomes unclear during manual labeling. Additionally, DA was also performed by applying three GAN techniques mentioned in the previous chapter and existing GAN-based techniques, respectively. In the classification and detection model, the training set was classified into multiple cases according to the augmentation combination because standard DA and GAN are independent and exhibit potentially synergistic effects when used together. Bowles et al. (2018) Data was split into training, verification, and test sets. A 3:1 training and validation scheme was followed for each case. Table 3 summarizes the number of image data used for each case.

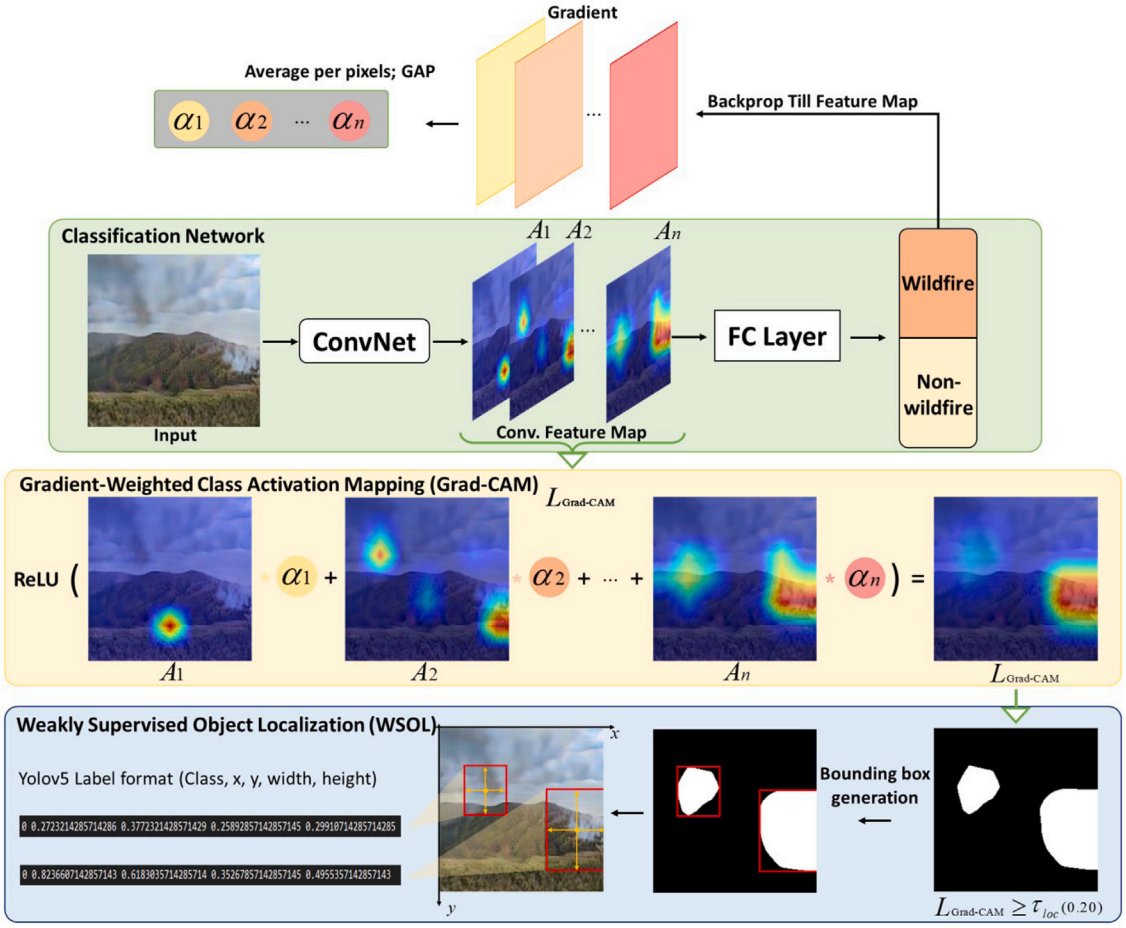



**Fig. 3.** Architecture of WSOL with Grad-CAM. Feature maps of the last convolution layer, $A_k$, are extracted during the forward propagation. Importance weight, $\alpha_k$, is extracted from global average pooling of gradient values per pixel of differential maps obtained via back propagation. In other words, if $\alpha_k$ is large, then the importance of the k-number feature map significantly contributes in determining the target class. Additionally, by multiplying $A_k$ and $\alpha_k$, the weight affecting the feature map is reflected, and only the positive value can be meaningfully determined by applying ReLU Activation; it is termed as the Grad-CAM result, $L_{Grad-CAM}$. By applying the localization threshold $\tau_x$ of $L_{Grad-CAM}$, black and white processing is performed from the boundary line. Furthermore, a bounding box is created, which can completely cover the excess value exceeding the threshold. The center point (x,y) of the created box and the width and height of the box are calculated.

### 3.2.2. Bounding box from Grad-CAM

A classification model with very high accuracy is required at this stage because if the model predicts the classification incorrectly, it also affects Grad-CAM result and labeling accuracy. Fig. 3 visualizes the process in which wildfire images generated from GAN are annotated in the form of YOLOv5 label format. YOLOv5 label format is constructed with the following classes: x, y, width, and height. Therefore, the normalized values of the width and height of the created bounding box, and the $x$ and $y$ coordinates of the center point were written and classified in the annotated label.

To annotate a synthesized image into training data for the object detection model, Grad-CAM heat map ($L_{Grad-CAM}$) was obtained from the ConvNet which exhibit the best performance among combination group of training dataset tested in binary wildfire image classification model evaluation stage. In the Grad-CAM step, the Input image obtains a feature map for each channel via the CNN Layer, and the importance of the feature map for the target wildfire class is calculated through back-propagation. The feature maps from each channel are multiplied by their respective gradient weight values according to their importance, and they are added together.

The ReLU activation function is multiplied to solely use the features that positively affect the wildfire class; further, heat map values ranging from 0 to 1 for the last convolutional layer, are generated. The expansion of this value from 0 to 255 highlights the location (flame or smoke) that influences the classification model to predict the input image as wildfire class.

We set the heatmap value threshold to differentiate between highlighted and non-highlighted places and change the heatmap to a bounding box. By drawing a bounding box around the largest segment, the model localizes annotated information on the image. Finally, to make the annotated information in yolov5 label format, the center point coordinates, width, and height of the box were extracted such that the annotation information of each image was automatically output. Fig. 3 visualizes the process in which wildfire images generated from GAN are annotated in the form of YOLOv5 label format.

## 3.3. Proposed wildfire detection method

### 3.3.1. Improved YOLOv5 with channel attention mechanism

YOLOv5 is a one-stage algorithm with high flexibility and versatility in the practical field (Wang et al., 2021a). Considering the operability by applying the model to embedded devices, YOLOv5 can expect high detection speed because it is less light model than other YOLO series. However, a disadvantage is that accuracy does not reach the requirement in recognizing the target. To secure golden time for wildfire disaster response, the detection model should consider environments with small device memory and low model performance. Hence, two strategies were applied to the basic framework YOLOv5s as follows: (1) adding a squeeze-and-excitation (SE) block (Hu et al., 2018) to the YOLOv5s of backbone network and (2) replacing the generalized intersection over union (GIoU) loss (Rezatofighi et al., 2019) as complete intersection over union (CIoU) loss.

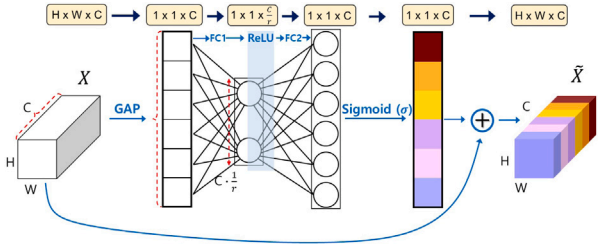



**Fig. 4.** Structure of SE block.

SE block corresponds to a module with the purpose of strengthening the expression of CNN and improving performance by modeling the interdependence of each channel. It plays a role of summarizing the information of the feature map from the existing network and recalibrating the importance of each feature map.

In Fig. 4, before the SE block operation, the input data is converted into a feature map via a series of convolution operations. In the squeeze step, a channel descriptor is created by aggregating a feature map of H × W × C size into 1-dimensional 1 × 1 × C via Global Average Pooling (GAP). The excitation operate to completely realize channel-wise dependencies from the descriptor as adjusting the fully connected layer and activation function. The excitation stage is in the form of a bottleneck structure to aid in generalization and computational efficiency. In FC1, the reduction ratio (r) decreases the number of channels to C/r, and the vector that becomes 1 × 1 × C/r passes through ReLU and FC2. In FC2, the number of channels is returned back to C, and it has a value in the range of 0 to 1 after passing through sigmoid. Finally, it is multiplied with the previous feature map to add a weight to the channel, and the output of the SE block generated in this manner can be used directly in subsequent layers. Hence, the SE layer can be flexibly applied to various existing structures without network constraints, and the model performance improvement is significant even though it slightly increases model complexity and computational cost.

The backbone of the original YOLOv5 consists of three basic modules of the convolution layer (CBS: Conv2d + Batch normalization + SiLU activation function (Elfwing et al., 2018)) with C3 (bottleneck CSP) module and the last module with spatial pyramid pooling-fast (SPPF) added to the basic module. Despite the improvement in performance values in most SE layer applications, the application of the SE layer to a place with a high feature map resolution rapidly increases the parameters of the model. This affects the detection speed (Zhan et al., 2022; Guo et al., 2022). In this study, the SE layer was not applied after the first basic module, but added after the 5th, 7th, and 10th modules of the basic YOLOv5 network with relatively low resolution of feature maps (see Fig. 5).

### 3.3.2. Bounding box regression loss function

Traditionally, models have been trained to minimize IOU loss to measure the difference between the ground truth boxes and prediction bounding boxes. GIoU (Rezatofighi et al., 2019) is used in YOLOv5, and it can compensate for the lack of a basis to determine the distance between non-overlapping IoU boxes. The metrics are defined as follows:

$$IoU = \frac{\left|A_{overlap}\right|}{\left|A_{union}\right|} \tag{1}$$

$$GIoU = \frac{\left|\frac{A_{enclose}}{A_{union}}\right|}{\left|A_{enclose}\right|} \tag{2}$$

where $A_{overlap}$ denotes the area of intersection of the predicted and ground truth bounding box, $A_{union}$ denotes the area of the union of the two bounding boxes, and $A_{enclose}$ denotes the minimum area that can cover both boxes.

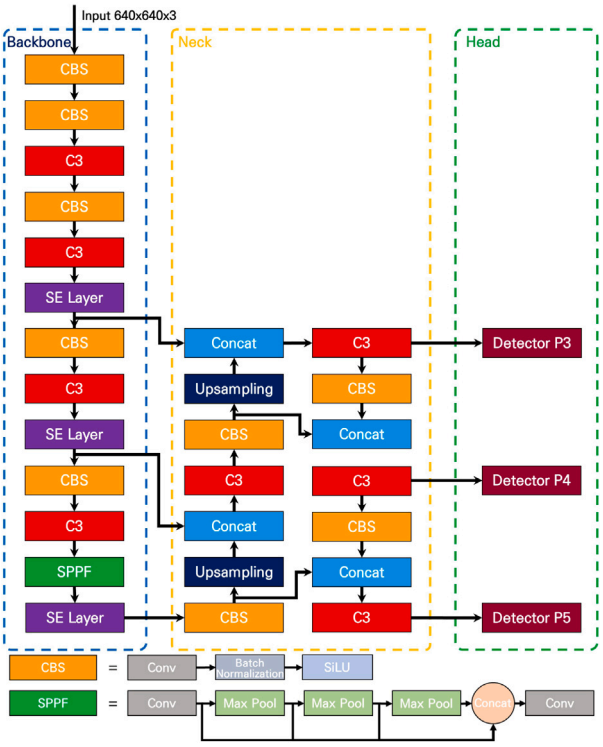



**Fig. 5.** Improved yolov5s neural network structure with SE Layer and detail of CBS and SPPF layer.

Although the GIoU loss makes up for the shortcomings of IoU loss involving the gradient vanishing problem in non-overlapping boxes, it considers only the overlapping areas of the prediction box and ground truth box, and cannot continue to optimize other positional relationships when the prediction box and ground box overlap completely. Therefore, CIoU loss function has three penalty terms: overlapping area, distance between center points, and aspect ratio (height width ratio). CIoU is used to improve the loss function to increase the positioning accuracy of small objects (Li et al., 2020), such as early smoke, and it is defined as follows:

$$CIoU = IoU - \frac{\rho^2(b^p, b^{gt})}{c^2} - \alpha\upsilon \tag{3}$$

where $\rho^2$ denotes square of the euclidean distance between center point of the predicted bounding box $b^p$ and ground truth bounding box $b^{gt}$. Furthermore, $c^2$ denotes the square of the diagonal length of the smallest box covering $b^p$ and $b^{gt}$.

$\alpha\upsilon$ is a penalty factor in CIoU. $\alpha$ represents positive trade-off parameter and the definition is shown as follows:

$$\alpha = \frac{\upsilon}{(1 - IoU) + \upsilon} \tag{4}$$

Where parameter $\upsilon$ is used for measuring the consistency of aspect ratio between $b^p$ and $b^{gt}$, and the definition is shown as follows:

$$\upsilon = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w^p}{h^p})^2 \tag{5}$$

Among them, $h^p$ and $w^p$ denote the height and width of the predicted bounding box $b^p$, respectively, and $h^{gt}$ and $w^{gt}$ denote the height and width of the ground truth bounding box $b^{gt}$, respectively. Hence, CIou loss can be defined as follows:

$$L_{CIoU} = 1 - CIoU \tag{6}$$

CIoU loss leads to faster convergence speed and accuracy on the bounding box regression problem when compared to GIoU (Bochkovskiy et al., 2020). Therefore, GIoU loss was replaced to CIoU loss as bounding box loss for improved YOLOv5s in this study.

**Table 4**
Hyper-parameter of the used GAN models.

| | Cycle-GAN(G1) | DRIT(G2) | NICE-GAN(G3) |
|---|---|---|---|
| Batch size | 12 | 10 | 4 |
| Learning rate | 0.0002 | 0.0001 | 0.0001 |
| Epoch | 1,200 | 1,000 | 1,000 |

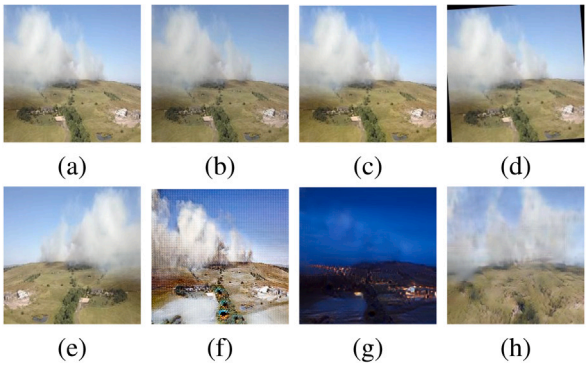



**Fig. 6.** Sample of used image augmentation methods. (a) original image. Standard augmentation: (b) brightness transformation, (c) saturation transformation, (d) random rotation, and (e) horizontal mirror. GAN method from Namozov and Im Cho (2018): (f) winter style GAN and (g) night style GAN. Proposed GAN method: (h) synthetic wildfire image from non-fire.

### 3.3.3. Object detection evaluation metrics

The evaluation index of the object detection task can be classified into two main categories : performance for average precision(AP) and F1-score, and speed for FPS and FLOPS. Specifically, AP is a method for calculating area under precision–recall curve (AUPRC). Precision–recall curve (PR-curve) shows precision and recall values at all IoU thresholds. AUPRC denotes the value of the area at the bottom of the graph. As the value becomes closer to 1, it indicates a better model.

Indicators that can calculate false alarms and miss alarms were also used because manpower was wasted by detecting non-wildfires as wildfires, or when a wildfire was not detected and delayed response caused huge damage. The statistical definition of false alarm rate (FAR) involves TN, which is typically not computed in object detection unlike classification problem. Therefore it is considered as using the false discovery rate (FDR) instead. It is tends to minimize and defined as follows:

$$FDR = \frac{FP}{TP + FP} = 1 - \frac{TP}{TP + FP} = 1 - P \tag{7}$$

where can be calculated from precision $P$, which is also known as positive predictive value. In a similar manner, the miss alarm rate can be replaced to the false negative rate (FNR). It tends to minimize and can be defined as follows:

$$FNR = \frac{FN}{TP + FN} = 1 - \frac{TP}{TP + FN} = 1 - R \tag{8}$$

where can be calculated from recall $R$. Therefore, AP, FDR, FNR, and F1-score were selected as indicators to evaluate the performance in our study.

Conversely, FPS and FLOPS, which can evaluate the speed, refer to the number of frames that can be processed per second and number of computations that measure the operating speed of the model, respectively. Hence, they can be used to verify the number of additional computations required by the improved detection model to increase performance.

## 4. Experiments and discussion

The experimental environment to construct the proposed model is an AI server with the following specifications: CentOS (Community Enterprise Operating System) Linux release 8.2.2004 with two 32 GB Nvidia Tesla V100 GPUs.

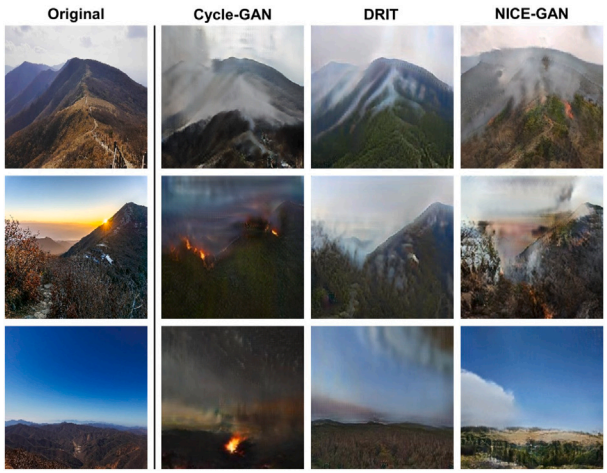



**Fig. 7.** Sample of the generated synthetic fake wildfire images from each used GAN.

**Table 5**
Quantitative comparison of I2I translation results between inputs and outputs: higher LPIPS values denote a high diversity in the dataset and lower FID values denote a high quality.

| | LPIPS ↑ | FID ↓ |
|---|---|---|
| Non-wildfire → Wildfire (Cycle-GAN) | 0.638 | 107.78 |
| Non-wildfire → Wildfire (DRIT) | 0.581 | 84.12 |
| Non-wildfire → Wildfire (NICE-GAN) | 0.582 | 101.04 |
| Wildfire → Winter style (Cycle-GAN) (Namozov and Im Cho, 2018) | 0.318 | 164.09 |
| Wildfire → Night style (Cycle-GAN) (Namozov and Im Cho, 2018) | 0.458 | 170.30 |

### 4.1. Generate synthetic wildfire image

An I2I translation with GAN methods was used to diversify and increase the number of images used for binary classification model training. To reflect more diverse scenarios, some night wildfire images were included as training data. As shown in Table 4, the models were trained with each hyper-parameters.

Samples of all DA results used in this study are shown in Fig. 6. In Standard DA, there is no significant change in the shape of the object. Further, there is no significant difference in the manner the background is changed. Conversely, GAN DA results shows the change in the shape of the object is higher than that in other DA methods, and it can make various semantic features for target object. In addition, in contrast to Fig. 6(f) and Fig. 6(g), which changed the background from the original existing object, we could see various changes in the target object in Fig. 6(h), which creates a completely new object. FID and LPIPS results in Table 5 also support the high quality and diversity of images generated by semantic synthesis based GANs (G1, G2, and G3).

Fig. 7 illustrates image samples generated from each model. Cycle-GAN results shown that reflect different wildfire scenarios, such as nighttime wildfires, even if the target is a daytime image. Conversely, despite being trained from the same dataset, DRIT results shown that largely consist of images covered in smoke, and the forest shapes are the least destroyed among the three models. NICE-GAN results shown generally exhibited evenly distributed smoke and flames, but had some unnatural background images.

### 4.2. Effectiveness of GAN DA for wildfire classification

ResNet50 and DenseNet201 models were respectively used to train the wildfire image binary classification model using a transfer learning environment from ImageNet. These two models were used to confirm the consistency of performance results. Category-cross entropy loss was also applied to train the binary classification model. By introducing the grid method, 10 random values with learning rates from 0.01

**Table 6**
F1-score and ROC–AUC values for training dataset using control variable method.

| | ResNet50 | | DenseNet201 | |
|---|---|---|---|---|
| | F1-score (%) | ROC–AUC (%) | F1-score (%) | ROC–AUC (%) |
| Dataset without augmentation | 90.84 | 91.12 | 92.89 | 93.29 |
| Dataset after augmentation | 98.48 | 98.46 | **99.08** | **99.07** |
| Remove brightness transformation | 95.88 | 96.01 | 97.67 | 97.70 |
| Remove saturation transformation | 97.21 | 97.22 | 97.17 | 97.22 |
| Remove horizontal flip method | 94.29 | 94.57 | 94.40 | 94.68 |
| Remove rotation transformation | 96.12 | 96.22 | 96.52 | 96.60 |
| Remove synthetic method from GAN | 93.88 | 94.22 | 94.41 | 94.67 |

**Table 7**
Quantitative model performance results using synthetic images as training data for the classification model. Precision refers to the ratio of actual wildfires among the images predicted by the model as wildfires, and recall refers to the ratio accuracy classified by the model among the actual wildfire test images. The best results are highlighted in bold.

| | Precision (%) | Recall (%) | F1-score (%) | ROC–AUC (%) |
|---|---|---|---|---|
| G1 | **99.64** | **98.52** | **99.08** | **99.07** |
| G2 | 99.28 | 97.25 | 98.26 | 98.25 |
| G3 | 98.31 | 98.03 | 98.17 | 98.12 |
| Winter style (Namozov and Im Cho, 2018) | 99.56 | 95.63 | 97.54 | 97.59 |
| Night style (Namozov and Im Cho, 2018) | 99.42 | 95.99 | 97.65 | 97.70 |

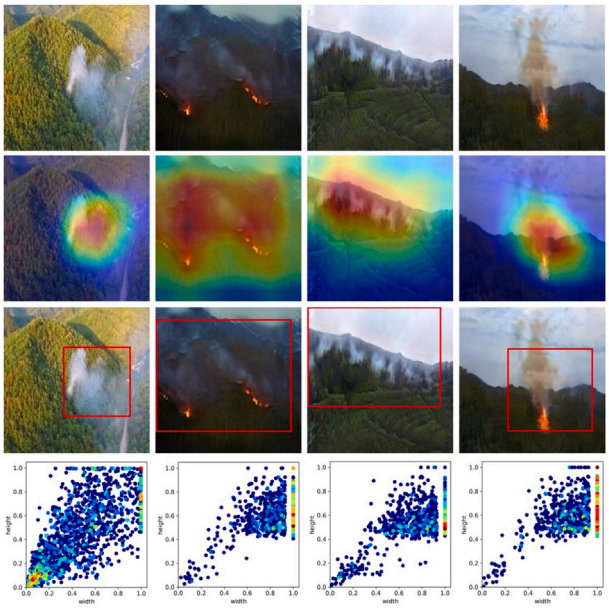


**Fig. 8.** Sample of automatic annotation result via WSOL. The first low is the original image, second low is the heat map result from the best classification model, third row corresponds to the result of generating the boundary box, and the final row is width-length ratio map of the wildfire object.

to 0.000001 were used along with three optimizer hyperparameters. The optimizers included stochastic gradient descent (SGD) (Ruder, 2016), adaptive moment estimation (ADAM) (Kingma and Ba, 2014), and Nesterov-accelerated adaptive moment (NADAM) (Dozat, 2016). These hyperparameters were trained with 200 epoch environments. Hyperparameter values were calculated with a combination that yields the highest validation accuracy.

Seven training data combinations were created from original wildfire images and augmented images. To verify the effect of DA as semantic image synthesis from GAN, the control variable method was used to remove each DA method and tested with actual images (1,420 wildfire images and 1,340 non-wildfire images).

Table 6 shows the experimental results. When a specific augmentation strategy was removed from the control variable method, the value of the metrics decreased significantly, indicating that each strategy was

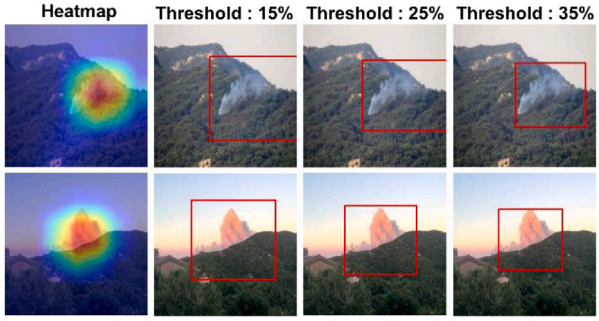



**Fig. 9.** An example sample that successfully created a boundary box while gradually increasing the threshold for Grad-CAM results.

highly influential. The results show DA strategy results in a significantly higher performance value. GAN DA effect on DenseNet201 binary classifier, F1-score for 4.67% and ROC–AUC for by 4.4%.

In the comparison results between GAN is shown in Table 7. G1, G2, and G3 results shown higher than the background style changed method in the accuracy of the wildfire classifier. This shows that misclassification can be reduced in a classifier trained with various feature environments. Additionally, all evaluation index values were high in the dataset with a high diversity, which had high LPIPS values.

### 4.3. Effectiveness of GAN DA for WSOL

This section shows the results of annotation on the dataset generated through weakly supervised learning based on the model that showed the highest performance in the experimental results of Section 4.2. The binary classification model with the highest performance was selected as weakly supervised model of Grad-CAM. A WSOL-annotated example of an original image and image generated by each image synthesis method is shown in Fig. 8. With the exception of the mis-classified results from the previous classification model, 900 images from G1, 899 images from G2, and 895 images from G3 were auto-labeled.

In the case of an inadequate selection of the backbone model among image groups, the bounding box localization identified incorrectly determined images, which were then excluded in advance. The number of cases in which the generated images were incorrectly determined from the backbone model were zero in G1 (accuracy 100%), one in G2 (accuracy 99.89%), and five in G3 (accuracy 99.44%).

### 4.4. Effectiveness of localization threshold adjustment

The bounding boxes were generated from localization threshold (annotation threshold) via Grad-CAM results. To select the optical localization threshold $\tau_{loc}$ for $L_{Grad-CAM}$ for creating a bounding box, different threshold values were used, and the box results were qualitatively evaluated on the original images. In general, 15% of the maximum heat map intensity was used as threshold to divide each pixel (Ahn et al., 2019). However, annotation was performed by switching between thresholds of 15%, 25%, and 35% to account for the ambiguous boundaries of fire and smoke objects. Fig. 9 shows an example in which the results of the annotation process are increasingly set

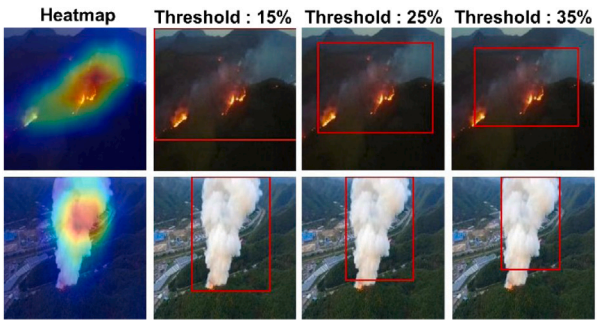



**Fig. 10.** An example sample that creates an over-tight boundary box while gradually increasing the threshold.

**Table 8**
Evaluation results obtained using adjustments to different thresholds in WSOL from DRIT(G2) synthetic images.

| | F1-score (%) | AP@50 (%) |
|---|---|---|
| Threshold: 15% (baseline) | 89.41 | 91.28 |
| Threshold: 20% | **90.00** | **92.14** |
| Threshold: 25% | 89.50 | 91.55 |

to be tight on wildfire targets while increasing the heat map threshold. The result of setting the threshold at 15% of the maximum strength was annotated as rougher than those with higher thresholds. As the processing becomes more tighter, the expected accuracy of the training results also increases. However, as shown in Fig. 10, even objects acquired at approximately 15% setting were completely included in certain image samples.

According to the results of the qualification evaluation, the candidates were classified into 15%, 20%, and 25%. Each candidate group was evaluated with an AP and F1-score performance indicator by a test dataset in our proposed model to determine the optimal value. Synthetic data as well as original images showed good results at the adjusted threshold. As shown in Table 8. F1-score was improved by 0.59%, and AP (0.5 IoU threshold) was improved by 0.88% at 20% localization threshold. Although it was not expected to dramatically improve performance, it contributed to maximizing the model performance by optimizing the labeling while converting the size of the bounding box by adjusting the localization threshold at WSOL stage for smoke and fire objects with ambiguous boundaries and shapes.

Fig. 11 shows the PR-curve in different threshold cases and AUPRC area. The area were calculated via numerical integration using trapezoidal rule; 0.92135 at the localization threshold of 20% was higher than that in other cases. Based on the aforementioned results, a threshold of 20% was finally selected for this dataset.

### 4.5. Effectiveness of the proposed wildfire detection model

The performance of the models using synthesized images as training data and models using WSOL were compared. After adjusting the hyperparameters of the object detection models, the models were trained with different training datasets. To support the effectiveness of the proposed method, three methods from GAN were used along with the manual method to compare the performance evaluation results. Based on 5210 learning images extended from 1,080 original images with standard DA, each training dataset configuration was replaced by the number of synthesized images. The training images were resized to 640 × 640 pixels. The validation scheme was applied at a ratio of 1:4 in the training data. YOLOv5s and improved YOLOv5s were trained with 0.001 initial learning rate, 32 batch size, and 300 training epochs. The manually labeled 2,427 real ground truth objects were used as the test set; synthetic images were not included because the model needed to detect actual wildfires environments.

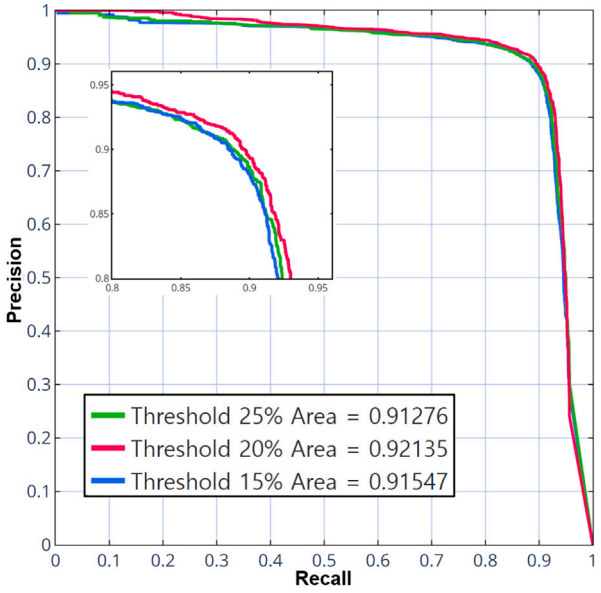



**Fig. 11.** PR-curve results of wildfire detection model using different thresholds in WSOL from DRIT(G2) synthetic images.

As shown in Table 9, Overall performance evaluation results were highly evaluated for each of the three prepared datasets in the proposed improved YOLOv5s. Specifically, in the results on the proposed model of the training dataset with the addition of synthetic images of G2 that had highest FID in experimental results, the AP increased the most among other cases to 1.72%. In G2, where only smoke was added to the synthetic fire generation result, the original model exhibited the lowest value in all evaluation indexes, as in the classification result. However, the evaluation result of the proposed model was not always proportional to the diversity of the training data.

Table 10 and Fig. 12 show the comparison of the performance evaluation results when three data set-related strategies (S1, S2, S3) and one model improvement strategy (S4) are considered to synthesize the strategies applied in this study. The dataset-related strategy consists of the following:

- S1 : Applying traditional GAN-based DA
- S2 : Using a dataset with increased feature diversity via a paired image consisting of fire-free mountain and wildfire instead of changing the background applied in previous studies
- S3 : Adjusting the localization threshold in the process of automatic labeling via WSOL
- S4 : Using proposed model (replace loss function as CIoU and add SE layer in YOLOv5s)

Specifically, M1 only adds standard DA in training dataset, M2 complements the training data to significantly improve performance in terms of F1-score and AP, M3 uses more diverse data, and M4 enhances the strength of rough annotation, allowing wildfire objects to be labeled densely as bounding boxes.

FAR is important in wildfire detection because it assigns a significant weight on model reliability and applicability in real-world scenarios (Sousa et al., 2020). It is calculated by replacing it as FDR with false positive in the denominator, in a manner similar to FAR. The overall performance improved with the addition of a strategy by significantly improving the FDR by 5.15% on M4, which uses a total of three data strategies; the performance evaluation comparison shows that this method not only saves human resources required for labeling, but also improves the accuracy of the model via data diversity and bounding box size optimization. The minimum FDR among the experimental group was 8.84% in M5, which uses the proposed data and

**Table 9**
Evaluation results obtained in IoU threshold of 0.5 using the proposed model in different datasets and a comparison with the YOLOv5s results.

| Model | Training dataset | Precision (%) | Recall (%) | F1-score (%) | AP@0.5 (%) |
|---|---|---|---|---|---|
| YOLOv5s | Add WSOL with G1 | 90.21 | 88.54 | 89.31 | 91.18 |
| | Add WSOL with G2 | 89.50 | 86.77 | 88.11 | 90.43 |
| | Add WSOL with G3 | 90.13 | 86.94 | 88.50 | 90.51 |
| Ours | Add WSOL with G1 | 91.16 | 89.65 | 90.40 | 91.97 |
| | Add WSOL with G2 | 91.20 | 88.83 | 90.00 | 92.14 |
| | Add WSOL with G3 | 91.14 | 87.76 | 89.42 | 91.52 |

**Table 10**
Evaluation results obtained in the test dataset with 0.5 IoU threshold using four strategies (use G1) and a comparison with other detection methods.

| Method | S1 | S2 | S3 | S4 | FDR↓ (%) | FNR↓ (%) | F1-score (%) | AP@0.5 (%) | #Param. | FLOPs↓ | FPS (V100) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M1 | – | – | – | – | 14.94 | 18.57 | 83.21 | 85.56 | – | – | – |
| M2 | ✓ | – | – | – | 12.86 | 15.11 | 86.00 | 89.46 | – | – | – |
| M3 | ✓ | ✓ | – | – | 12.30 | 13.02 | 87.34 | 90.48 | – | – | – |
| M4 | ✓ | ✓ | ✓ | – | 9.79 | 11.46 | 89.31 | 91.18 | 7,012,822 | 15.80G | 215.66 |
| M5 (Ours) | ✓ | ✓ | ✓ | ✓ | **8.84** | 10.35 | **90.40** | 91.97 | 7,055,830 | 15.84G | 203.09 |
| YOLOv4 (Bochkovskiy et al., 2020) | ✓ | ✓ | ✓ | – | 34.46 | 9.68 | 75.96 | 86.01 | 8,052,486 | 19.06G | 101.66 |
| YOLOR (Wang et al., 2021b) | ✓ | ✓ | ✓ | – | 11.70 | **7.66** | 90.27 | **92.48** | 36,838,416 | 81.56G | 56.81 |

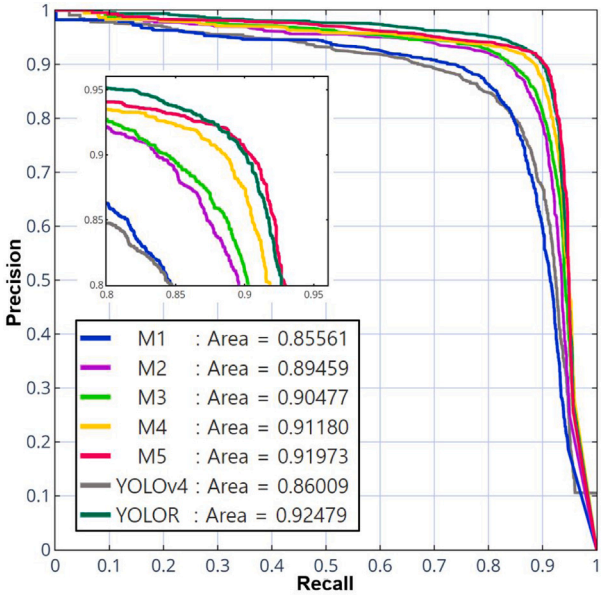



**Fig. 12.** PR-curve results of wildfire detection model using different strategy combination.

model improvement strategies; it indicates the lowest false detection rate in the test dataset.

The final experimental results exhibited an F1-score that increased from 83.21% to 90.40%, and an AP that increased from 85.56% to 91.97% with our proposed model.

The performance evaluation results of our improved detection model show that it can mitigate the aforementioned challenges. First, In the detector, model performance can be improved by synthesizing flame or smoke in the generated images in a fire-free mountain when compared to using standard DA or traditional DA from the performance evaluation results in M1, M2 and M3. Second, most of the original images were auto-labeled based on the WSOL strategy, whereas in this study, the training dataset was constructed via a class activation map to not only increase the number of data points but also simultaneously perform auto-labeling. This training dataset aided in improving the accuracy of the model, as shown in the detection model comparison results. Furthermore, by adjusting the threshold to generate a tight bounding box, useless features were minimized at the training stage (M3 and M4 results). Third, the SE layer was able to the improve the model to distinguish between important and non-important features, and the CIoU loss function was able to improve the performance on small objects by aligning the loss, which is weighted more on the center point than on the width and length of the bounding box. These two changes resulted in an improvement in overall performance metrics as shown in the results of M4 and M5.

Verification between models (M4, M5, YOLOv4, and YOLOR) was evaluated under conditions using the same data strategy. Although the number of parameters increased by about 33,000, and the FLOPS increased by only 0.04 G from M4 to M5; it improved by approximately 1% on all accuracy performance metrics. The FLOPS of this model was lower than YOLOv4 and exhibited FPS that would not be considered problematic if used for real-time object detection. Experiment results from YOLOR exhibited the lowest FNR value of 7.66% among the experimental group and the best performance in terms of detecting ground truth on wildfire objects without missing it. However, it did not exhibit good results in the false detection case. As for the PR-curve, our proposed model also performed better than YOLOR at certain confidence levels. Furthermore, for application to general equipment while using a lightweight model, our model exhibited an advantage because of the nature of the application environment that minimizes false positives. The improved model retained the advantages of YOLOv5s and showed results close to the accuracy of YOLOR, even though it was a light model with low FLOPS, and better results for false positives. Additionally, as the model has a small number of parameters, training time can be saved, which is useful for practical applications.

### 4.6. Visualization experiments

This subsection highlights examples of correct detection in the test dataset to ensure that the requirements of wildfire monitoring technology are satisfied. Additionally, it aims to draw attention to future research by analyzing common patterns of falsely detected or undetected examples.

The monitoring system is operated at day and night, and it should detect low amounts of smoke and minimize unnecessary alarms because of false detection. The test image included 112 night-time wildfire drone images. Some of these images were collected from the Korea Forest Service. Specifically, 215 small wildfire images had a ground truth

**Fig. 13.** Sample of results in night-time wildfire images: ground truth in first row, M1 (YOLOv5s) results in second row, and M5 (our) results in third row.

**Fig. 14.** Sample of results in small object: ground truth in first row, M4 (YOLOv5s with data strategy) results in second row, and M5 (Ours) results in third row.

bounding box size less than 5% of the entire image size. Additionally, 224 smoke/fire-like images were acquired.

Fig. 13 shows an example of results that are nearly consistent with the ground truth for an evening wildfire. As observed in the first and second columns, M5 exhibited robust results with respect to noise from street lights or vehicle lights at night, while M1 falsely detected the red light of the fire engine as a flame, and did not detect the smoke at night. Hence, its determination scores were generally lower than those of M5. Specifically, for wildfires occurring during fireworks display in the downtown area of the third column, only flames and smoke were accurately detected in our model; it did not detect smoke and noise generated by firecrackers as well as city lights at night.

Fig. 14 shows an example of detecting small size of smoke. Specifically, it was observed that the color and shape of the water pouring

**Fig. 15.** Sample of good results in smoke-like images (M5).

**Fig. 16.** Sample of false detection (false positive) and miss detection (false negative): ground truth on in first row and error on in second row.


**Fig. 17.** Tested drone station for monitoring wildfire.

out for evolution from the helicopter and those of smoke in the third column were similar. Furthermore, only the smoke was captured correctly despite the ambiguous boundary. Also, it was shown that fog or clouds that can be confused with smoke were correctly distinguished without a false alarm. The results of M4 show that false positives and false negatives frequently occur for small object, and that the prediction score is close to the decision threshold in an accurately detected case. Conversely, the proposed model shows results close to the bounding box of ground truth with sufficient prediction score for small smoke in fine-grained environments. Furthermore, Fig. 15 shows fog or clouds that can be confused with smoke as being correctly distinguished without a false alarm.

However, as shown in Fig. 16, there are several false and missed detections; and we will provide further insights in future research via analysis of common features of errors. First, as shown in the first column of 16, in the fire-like image, an object with a red color that is too bright or a light near a forest was falsely detected. Most false detections showed prediction values close to the decision threshold. However, in the case of night wildfires, small lights were also not detected in some cases, as shown in the second column of Fig. 16. Unlike the detection of large flame clusters, it is difficult to detect small-scale night fires from a long-range distance because smoke is barely visible at night, and thus, the detection solely relies on lights. These are very important factors for an early or residual fire treatment, but there

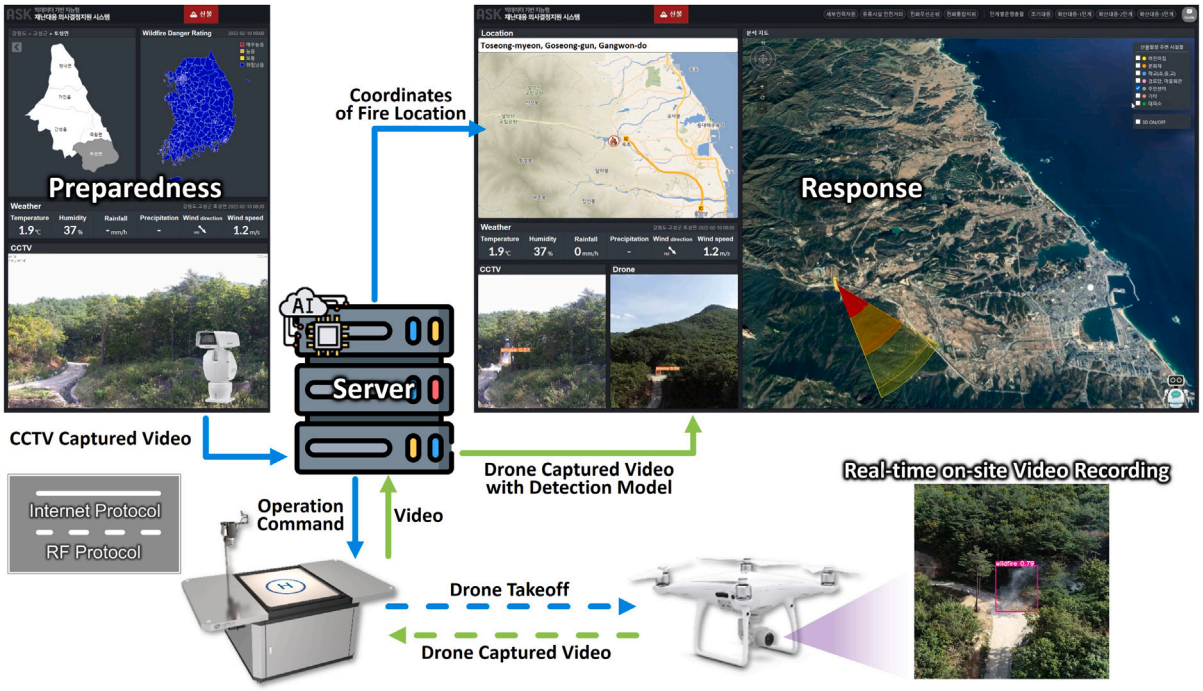



**Fig. 18.** Web-based visualization of wildfire early detection system for disaster response (ASK : AI for Safety Korea) with an example of creating a virtual forest fire via a smoke grenade by the researcher and displaying the result on the screen.

was a limitation in supplementing them because images generated from the GAN were not able to generally produce objects smaller than 1% of the total image size because of the limitations of the secured image; even localization threshold adjustment can contribute to reducing the bounding box size.

The second common error characteristic corresponds to an issue with the tilt angle and altitude of wildfire detection equipment. This is not a problem in CCTV footage, which is almost horizontal to the ground. In the case of a drone footage, given the risk of flames and smoke, it can be taken at various angles and altitudes; however, most drone footages of wildfires consist of horizontal images acquired. Additionally, given that most of them use horizontal flips in the DA stage and do not rotate more than 90 degrees, a gap can occur between learning data in the form of smoke that mostly spreads to the sky and drone test images. Hence, it is difficult to learn the features of a wildfire image shot by a drone. For this reason, a low amount of smoke in a vertical area cannot be detected and CCTV images have been mainly used for early detection of wildfires. However, if the foundation for securing early wildfire images by a drone is provided, the detection accuracy is expected to improve as it is possible to generate a synthetic smoke image that considers the drone footage characteristics from a GAN. In spite of the fact that SE Layer can assist in emphasizing important features while suppressing less important features, neural networks cannot be generalized to untrained situations. Therefore, in future studies, these limitations should be considered in the data construction stage.

### *4.7. Testing applications of wildfire early detection system for disaster response*

This subsection describes the application and testing of the proposed model to early wildfire detection for a disaster response system. Models used as monitoring equipment include TNU-6321 positioning CCTV cameras capable of shooting up to HD (1920 × 1080 pixels) videos and Phantom4 Pro v2.0 commercial drones and charging equipment capable of shooting up to 4K (3840 × 2160 pixels) 60 fps videos. The CCTV exhibits a tilt range of −40 degrees to 90 degrees and can rotate 360 degrees. Hence, it plays an important role in real-time forest detection. The drone is dispatched to identify the exact location of the fire and check the residual fire in the process of responding to a wildfire. Fig. 17 shows the equipment installed to test the system for wildfires and shows examples of images captured by each equipment. The drone operating environment was differentiated according to its purpose. To quickly detect fires or understand the direction of the fire, the altitude is set to be high in the range from 100 m to 150 m, and the tilt angle is set to be low in the range from 10 degree to 30 degree such that a wide range can be photographed at one shot. Conversely, to identify small-scale fires or fire sites in specific areas, the altitude was set to be relatively low in the range from 70 m to 100 m, and the tilt angle was set in the range from 45 degrees to 60 degrees.

In the event of a wildfire, the disaster response system operation scenario is shown in Fig. 18. In a disaster preparedness condition, the system displays the CCTV monitoring screen from the drone station, and weather data such as rainfall, wind direction, wind speed, temperature, and humidity of the installed equipment are visualized. Furthermore, the wildfire danger rating is visualized in the system. When a wildfire occurs and is detected by the proposed model, the coordinates of the fire location are transmitted to the system, which then switches to the disaster response screen. The system sends a driving command to the drone station closest to the reporting location via an Internet protocol, and the ground control system sets the drone mission path. Then, the drone takes off, acquires video, and transmits it using the radio-frequency (RF) protocol. Meanwhile, the weather data is continuously displayed on the converted screen, and the time-wise (red: 1 h, orange: 2 h, and yellow: 3 h) spreading range is visualized as a cone on the map based on to the weather data (wind direction and speed). This visualization is based on the table for prediction of wildfire spread (Kim et al., 2005) in Korea, which is used for quick decision-making in an urgent situation. Further, it is used as a reference for determining the spread of the fire. Additionally, evacuation sites and important facilities are displayed on the map to help decision makers receive support information for disaster response.

## 5. Conclusion

This study focused on video-surveillance to detect early fires to minimize environmental destruction and pollution from wildfire disasters. There are challenges in developing a forest fire detection deep learning model, such as mis-detection and non-detection of small smoke caused by lack of data.

In order to solve these fundamental problems, our research has contributed as follows:

- We developed a virtual wildfire image through GAN and performed object localization with weakly supervised learning for annotating. GAN DA helps to improve false positives and false negative in wildfire detectors.
- Although auto-annotated (also known as pseudo-labeled) data from WSOL can occur error of bounding box setting, they eventually help to increase accuracy of the wildfire detection model. Additionally, threshold modification aids in reducing inaccuracy from the auto-annotation.
- We proposed a method to improve the existing YOLOv5 model so that the main feature information about smoke and flame can be learned more effectively in the learning stage of the model.
- We established a drone and CCTV-based monitoring system based on a CNN-based artificial intelligence model that can automatically detect wildfires via visual alerts.

Based on the revealed insights, this study provides a direction for future research on what should be considered as follows:

- To build a learning dataset, data collection and processing can be supplemented via DA and WSOL to reduce human and material costs. It is very difficult to reproduce and collect real data in the disaster field. Given that the boundary of the object is ambiguous, subjective intervention from the workers is necessary in the annotation process. When several workers work at the same time, some irregularity is possible. GAN reduces the burden of data collection, and the CAM obtained from a good-performance classifier allows metadata to be added via computational formulas to multiple images.
- False positives occur mainly in objects similar to detection targets such as fog, clouds, and streetlights. Furthermore, forests have an environment that is frequently exposed to the similar objects. Hence, it is necessary to accurately distinguish fine-grained objects to minimize false alarms.
- Based on the common and frequently occurring errors in this experiment, a method of constructing DA of training datasets, such as image shooting angles, was evoked.
- The proposed model can be extended to include multiple classes such as rescuers and vehicles, and important structural facilities such as gas stations or cultural properties; they can be primarily defended in the case of fire suppression.

The proposed model has been installed in the wildfire response decision support system and the system prototype has been tested. We intend to optimize the model and improve accuracy via future research. This is related to the aforementioned extensibility of the model and is annotated in the same manner for various instances to provide information such as "the presence or absence of survivors" on disaster sites. Additionally, it aims to minimize damage by establishing a system to support decision-makers' responses by providing various Big data analysis information required for wildfire response using public data provided by the government.

## CRediT authorship contribution statement

**Minsoo Park:** Conceptualization, Methodology, Software, Project administration, Writing – original draft. **Dai Quoc Tran:** Resources, Data curation. **Jinyeong Bak:** Software, Validation, Writing – review & editing. **Seunghee Park:** Supervision, Funding acquisition.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Seunghee Park reports financial support was provided by Korea Ministry of the Interior and Safety. Seunghee Park reports financial support was provided by Korea Ministry of Science and ICT.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

This research was supported by a grant [2022-MOIS38-002] from the Ministry of Interior and Safety (MOIS)'s project for the development of accident prevention technology for vulnerable groups. In addition, this research was supported by a grant from the Ministry of Science and ICT(MSIT) to the National Research Foundation of Korea (NRF) [NRF-2021R1A4A3033128].

## References

Aguilera, R., Corringham, T., Gershunov, A., Benmarhnia, T., 2021. Wildfire smoke impacts respiratory health more than fine particles from other sources: Observational evidence from Southern California. Nature Commun. 12 (1), 1–8.

Ahn, J., Cho, S., Kwak, S., 2019. Weakly supervised learning of instance segmentation with inter-pixel relations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2209–2218.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D.A., Hernández, M.V., Wardlaw, J., Rueckert, D., 2018. Gan augmentation: Augmenting training data using generative adversarial networks. arXiv preprint arXiv:1810.10863.

Center for wildfre research University of Split Faculty of Electrical Engineering, M. E., Architecture, N., 2014. FESB MLID dataset. http://wildfire.fesb.hr.

Chen, R., Huang, W., Huang, B., Sun, F., Fang, B., 2020. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8168–8177.

Chenebert, A., Breckon, T.P., Gaszczak, A., 2011. A non-temporal texture driven approach to real-time fire detection. In: 2011 18th IEEE International Conference on Image Processing. IEEE, pp. 1741–1744.

Chi, R., Lu, Z.-M., Ji, Q.-G., 2017. Real-time multi-feature based fire flame detection in video. IET Image Process. 11 (1), 31–37.

Chu, M., Thuerey, N., 2017. Data-driven synthesis of smoke flows with CNN-based feature descriptors. ACM Trans. Graph. 36 (4), 1–14.

Dozat, T., 2016. Incorporating nesterov momentum into adam.

Elfwing, S., Uchibe, E., Doya, K., 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Netw. 107, 3–11.

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., 2010. Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. 32 (9), 1627–1645.

Foggia, P., Saggese, A., Vento, M., 2015a. Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. IEEE Trans. Circuits Syst. Video Technol. 25 (9), 1545–1556.

Foggia, P., Saggese, A., Vento, M., 2015b. Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. IEEE Trans. Circuits Syst. Video Technol. (ISSN: 1051-8215) 25 (9), 1545–1556.

Guo, K., He, C., Yang, M., Wang, S., 2022. A pavement distresses identification method optimized for YOLOv5s. Sci. Rep. 12 (1), 1–15.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.

Hüttner, V., Steffens, C.R., da Costa Botelho, S.S., 2017. First response fire combat: Deep leaning based visible fire detection. In: 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics. SBR, IEEE, pp. 1–6.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Jager, H.I., Coutant, C.C., 2020. Knitting while Australia burns. Nature Clim. Change 10 (3), 170.

Jang, H., 2020. Fire occurs prediction video dataset. URL https://aihub.or.kr.

Jeong, C., Jang, S.-E., Na, S., Kim, J., 2019. Korean tourist spot multi-modal dataset for deep learning applications. Data 4 (4), 139.

Jung, D., Tran Tuan, V., Quoc Tran, D., Park, M., Park, S., 2020. Conceptual framework of an intelligent decision support system for smart city disaster management. Appl. Sci. 10 (2), 666.

Karbhari, Y., Basu, A., Geem, Z.W., Han, G.-T., Sarkar, R., 2021. Generation of synthetic chest X-ray images and detection of COVID-19: A deep learning based approach. Diagnostics 11 (5), 895.

Kato, S., Miyamoto, H., Amici, S., Oda, A., Matsushita, H., Nakamura, R., 2021. Automated classification of heat sources detected using SWIR remote sensing. Int. J. Appl. Earth Obs. Geoinf. 103, 102491.

Khan, A., Hassan, B., 2020. Dataset for forest fire detection. URL https://data.mendeley.com/datasets/gjmr63rz2r/1.

Kim, D.-H., Won, M.-S., Lee, M.-B., 2005. A case study of forest fire spread in yangyang. In: Proceedings of the Korean Society of Agricultural and Forest Meteorology Conference. Korean Society of Agricultural and Forest Meteorology, pp. 109–113.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kortylewski, A., He, J., Liu, Q., Yuille, A.L., 2020. Compositional convolutional neural networks: A deep architecture with innate robustness to partial occlusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8940–8949.

Langnickel, L., Fluck, J., 2021. We are not ready yet: limitations of transfer learning for disease named entity recognition. BioRxiv.

Lee, J., Kim, E., Lee, S., Lee, J., Yoon, S., 2019. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5267–5276.

Lee, H.-Y., Tseng, H.-Y., Huang, J.-B., Singh, M., Yang, M.-H., 2018. Diverse image-to-image translation via disentangled representations. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 35–51.

Li, Y., Li, S., Du, H., Chen, L., Zhang, D., Li, Y., 2020. YOLO-ACN: Focusing on small target and occluded object detection. IEEE Access 8, 227288–227303.

Li, H., Wu, Z., Shrivastava, A., Davis, L.S., 2022. Rethinking pseudo labels for semi-supervised object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, (2), pp. 1314–1322.

Li, T., Zhao, E., Zhang, J., Hu, C., 2019. Detection of wildfire smoke images based on a densely dilated convolutional network. Electronics 8 (10), 1131.

Mikołajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW). IEEE, pp. 117–122.

Moreno, S., 2021. Risk mitigation: Human factors II. Task Force 74.

Muhammad, K., Ahmad, J., Mehmood, I., Rho, S., Baik, S.W., 2018. Convolutional neural networks based fire detection in surveillance videos. IEEE Access 6, 18174–18183.

Namozov, A., Im Cho, Y., 2018. An efficient deep learning algorithm for fire and smoke detection with limited data. Adv. Electr. Comput. Eng. 18 (4), 121–128.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vis. 42 (3), 145–175.

Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2015. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 685–694.

Park, M., Tran, D.Q., Jung, D., Park, S., 2020. Wildfire-detection method using DenseNet and CycleGAN data augmentation-based remote camera imagery. Remote Sens. 12 (22), 3715.

Park, M., Tran, D.Q., Lee, S., Park, S., 2021. Multilabel image classification with deep transfer learning for decision support on wildfire response. Remote Sens. 13 (19), 3985.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 201.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 658–666.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.-L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Adv. Neural Inf. Process. Syst. 33, 596–608.

Sousa, M.J., Moutinho, A., Almeida, M., 2020. Wildfire detection using transfer learning on augmented datasets. Expert Syst. Appl. 142, 112975.

Steffens, C.R., Rodrigues, R.N., da Costa Botelho, S.S., 2015. An unconstrained dataset for non-stationary video based fire detection. In: 2015 12th Latin American Robotics Symposium and 2015 3rd Brazilian Symposium on Robotics (LARS-SBR). IEEE, pp. 25–30.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.

Tang, Z., Liu, X., Chen, H., Hupy, J., Yang, B., 2020. Deep learning based wildfire event object detection from 4K aerial images acquired by UAS. AI 1 (2), 166–179.

Toulouse, T., Rossi, L., Akhloufi, M., Celik, T., Maldague, X., 2015. Benchmarking of wildland fire colour segmentation algorithms. IET Image Process. 9 (12), 1064–1072.

Toulouse, T., Rossi, L., Campana, A., Celik, T., Akhloufi, M.A., 2017. Computer vision for wildfire research: An evolving image dataset for processing and analysis. Fire Saf. J. 92, 188–194.

Tran, D.Q., Park, M., Jeon, Y., Bak, J., Park, S., et al., 2022. Forest-fire response system using deep-learning-based approaches with CCTV images and weather data. IEEE Access 10, 66061–66071.

Tran, D.Q., Park, M., Jung, D., Park, S., 2020. Damage-map estimation using UAV images and deep learning algorithms for disaster management system. Remote Sens. 12 (24), 4169.

University of California San Diego, California, America, 2021. The high performance wireless research and education network: An overview.

Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., Pinheiro, P.R., 2020. Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. Ieee Access 8, 91916–91923.

Wang, J., Chen, Y., Gao, M., Dong, Z., 2021a. Improved YOLOv5 network for real-time multi-scale traffic sign detection. arXiv preprint arXiv:2112.08782.

Wang, C.-Y., Yeh, I.-H., Liao, H.-Y.M., 2021b. You only learn one representation: Unified network for multiple tasks. arXiv preprint arXiv:2105.04206.

Xu, G., Zhang, Y., Zhang, Q., Lin, G., Wang, J., 2017. Deep domain adaptation based video smoke detection using synthetic smoke images. Fire Saf. J. 93, 53–59.

Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., Ye, Q., 2019. Danet: Divergent activation for weakly supervised object localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6589–6598.

Zhan, W., Sun, C., Wang, M., She, J., Zhang, Y., Zhang, Z., Sun, Y., 2022. An improved Yolov5 real-time detection method for small objects captured by UAV. Soft Comput. 26 (1), 361–373.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018a. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595.

Zhang, Q.-x., Lin, G.-h., Zhang, Y.-m., Xu, G., Wang, J.-j., 2018b. Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images. Procedia Eng. 211, 441–446.

Zhang, X., Wei, Y., Feng, J., Yang, Y., Huang, T.S., 2018c. Adversarial complementary learning for weakly supervised object localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1325–1334.

Zhang, X., Wei, Y., Feng, J., Yang, Y., Huang, T.S., 2018d. Adversarial complementary learning for weakly supervised object localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1325–1334.

Zhang, X., Wei, Y., Kang, G., Yang, Y., Huang, T., 2018e. Self-produced guidance for weakly-supervised object localization. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 597–613.

Zhao, Y., Ma, J., Li, X., Zhang, J., 2018. Saliency detection and deep learning-based wildfire identification in UAV imagery. Sensors 18 (3), 712.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016a. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2921–2929.

Zhou, H., Sattler, T., Jacobs, D.W., 2016b. Evaluating local features for day-night matching. In: European Conference on Computer Vision. Springer, pp. 724–736.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232.

**Minsoo Park** received a BS in Civil, Architectural Engineering, and Landscape Architecture from Sungkyunkwan University, Suwon, South Korea, in 2017. He is currently a Ph.D. candidate in Civil Engineering with the School of Civil, Architecture, and Environmental Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include Bigdata analysis and AI applications for natural disaster response and safety management.

**Dai Quoc Tran** received a BS in transportation engineering from the Mientrung Un iversity of Civil Engineering, Vietnam, in 2018. He is currently working toward a Ph.D. in civil engineering with the School of Civil, Architecture, and Environmental Engineering, Sungkyunkwan University.

**Jinyeong Bak** has been an assistant professor at Sungkyunkwan University since September 2020. He received his Ph.D. in computing from Korea Advanced Institute of Science and Technology (KAIST) in August 2020. Previously, he had received a BS in computer engineering from Sungkyunkwan University, South Korea, in 2011. His research interests include machine learning and natural language processing.

**Seunghee Park** has been an associate professor at Sungkyunkwan University since March 2009. He received his Ph.D. from the Korea Advanced Institute of Science and Technology (KAIST) in 2008. He worked as a post-doctoral researcher at KAIST from February to September, 2008. He also worked as a post-doctoral research fellow at the Center of Intelligent Material Systems and Structures, Mechanical Engineering, Virginia Tech., Virginia, USA from October 2008 to February 2009. He was a visiting professor at the Center of Intelligent Material Systems and Structures, Mechanical Engineering, Virginia Tech., from July to August 2009.