Generation of Synthetic Wildfire Smoke Images with Generative Adversarial Networks

Dunja Božić-Štulić, Damir Krstinić, Darko Stipaničev, and Jakov Bejo

Original scientific article

***Abstract*—The scarcity of annotated images significantly hinders the development of robust deep learning models for early wildfire smoke detection. Traditional augmentation methods, such as rotation or mirroring, are often insufficient. This is particularly true for detecting subtle smoke formations at long distances, where smoke is often barely perceptible even to human observers. Existing smoke datasets predominantly feature developed smoke plumes or closer views, making them unsuitable for training models for this critical early-phase detection. To address this, we propose generating synthetic images using Generative Adversarial Networks. Unlike typical GAN applications that aim for high-fidelity object representation, our objective is different. We synthesize realistic, fuzzy images of subtle, distant smoke — blurred and blended with the background — yet retaining characteristic features essential for classifier training. We propose a GAN architecture based on a modified Super-Resolution GAN, specifically adapted without Residual blocks, in order to produce realistic images of smoke at long distances. Experimental evaluation demonstrates that augmenting datasets with GAN-generated smoke images significantly improves the performance of classifiers in detecting early-stage wildfire smoke, affirming the utility of GANs for data enhancement even when generating low-quality, realistic imagery. This method mitigates data scarcity, offering a viable solution for training effective early wildfire detection systems.**

***Index Terms*—generative adversarial networks, synthetic smoke images, deep learning.**

I. INTRODUCTION

The supervised training of deep convolutional neural networks has become a dominant approach to solving many computer vision tasks, such as image classification, object recognition and localization. In recent years, significant advances have been achieved in many problems where traditional computer vision algorithms have approached their limits. Provided with a large enough dataset of labeled samples, deep learning algorithms tend to produce very good results and usually significantly outperform traditional approaches. However, a basic prerequisite for the supervised training of Deep Neural Networks (DNNs), the availability of large datasets, can be a major obstacle in model development. DNNs may have up to hundreds of millions of parameters, which require huge sets of labeled images to generalize key features and distinguish different phenomena. Training a complex network model on a limited dataset will probably result in overfitting and poor performance on unseen images.

Collecting and annotating large datasets is time-consuming and usually an expensive process. Moreover, collecting a sufficiently large set of samples for some problems can be further hampered by other factors, such as regulations on personal data protection or data ownership. For some phenomena, their inherent feature could be a rare occurrence, making it even more difficult to gather a relevant set of examples. Building a prediction model based on an unbalanced dataset could result in models that tend to overlook the minority concepts [1]. Unfortunately, these underrepresented concepts or phenomena are usually of the highest interest and there is a high cost for misclassifications [2]. False negative detection could pose a major risk in applications such as diagnosing rare diseases [1] or smoke detection [3].

To overcome this limitation, researchers and developers often use augmentation strategies to increase the number of labeled images [4] and the diversity of information in input data. These techniques commonly include simple modifications to original images, such as translations, small angular rotations, flipping and scaling. However, this approach has limitations, especially for datasets where such modifications could result in unrealistic images. For instance, flipping an image horizontally, or applying angular rotation to an image of open space, could easily result in a very unrealistic image with the sky at the bottom of the scene, or an oblique horizon. This motivates the use of synthetic data samples, which can introduce more variability to the dataset to improve the training process.

Our research primarily focuses on the segmentation and classification of natural landscape images [5], [6], [7], particularly emphasizing smoke detection using wildfire surveillance cameras that monitor vast open areas [5], [8], [9]. This led to the initiation of the Intelligent Forest Fire Video Monitoring and Surveillance System [10], [11], [3]. The system is deployed in more than 116 locations in a number of Croatian forests and National Parks, as well as in several locations in Bosnia and Herzegovina, Montenegro and Albania. Surveillance cameras for wildfire detection are strategically mounted in remote locations, operating year-round, with detection distances varying from several hundred meters to over ten kilometers. The specific characteristics of the monitored territory, which often includes sea surfaces and areas of intense human activity, present unique challenges. A typical Mediterranean

Manuscript received November 14, 2025; revised January 22, 2026. Date of publication March 16, 2026. Date of current version March 16, 2026. The associate editor prof. Renata Lopes Rosa has been coordinating the review of this manuscript and approved it for publication.

Authors are with the University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia (e-mails: {dgotovac, dkrst, dstip, jbejo00}@fesb.hr).

Digital Object Identifier (DOI): 10.24138/jcomss-2025-0245

Fig. 1. Examples of early-stage wildfire smoke captured by operational surveillance cameras. Image quality is degraded by atmospheric conditions, sunlight effects, and dirt or insects on lenses, making smoke barely perceptible even to a human observer.

climate further compounds these difficulties with morning mists, strong daily westerly winds (leading to airborne dust and camera shake), pronounced light effects, and reflections of sunlight on sea surface and atmospheric particles. These environmental factors, coupled with potential obstructions like dirt and insects on camera lenses, collectively result in significant image quality degradation. As a consequence, the acquired images frequently lack crucial low- and mid-level features, such as corners, edges, and other distinctive shapes [5]. The unique characteristics of the Croatian coast — with small towns interspersed within dense Aleppo pine forests — require early fire detection and an efficient response, preferably within five minutes of ignition.

Emerging smoke is, fortunately, a relatively rare phenomenon to capture on camera. In its early stage, minutes after ignition, and at the required operational distance of the surveillance system, smoke can be fuzzy and transparent and blended with the background. In addition to this, it is represented within a very small area in high-resolution images, usually less than 0.1 ‰of the area of an image. Examples of smoke from real forest fires captured by surveillance cameras during the early ignition phase are shown in Figure 1. In both images, the smoke is barely visible due to fog and sunlight effects, making accurate image annotation challenging, even for human experts tasked with creating ground truth segmentation. It should be emphasized that the smoke in both images is less than 5 km from the camera, which is approximately half of the expected detection distance.

The aforementioned issues make it hard to collect a large enough dataset to train state-of-the-art neural network architectures. The application of common augmentation techniques is limited due to the possible loss of scene realism and low quality of recorded samples of early-phase of smoke at the required detection distance. The lack of a large enough set of images can be partially compensated by a transfer learning approach [5]. However, wildland surveillance images significantly differ from urban and close-range scenes, which also limits the possible benefits of retraining pretrained networks. In this work we propose a method of synthetic wildfire smoke image generation based on Generative Adversarial Networks (GAN) [12], [13]. We aim to create a dataset of synthetic images of the emerging smoke at a very early phase, which can be used to train state of the art deep neural networks. These samples must reflect the complexity of real surveillance camera images, including all interference and degradation. At the same time, they must retain the characteristic features necessary for classifier training and smoke detection. The images collected from operational wildfire surveillance cameras containing emerging smoke of real fire incidents in the early ignition state are used to train a generative model based on the modified Super-Resolution GAN network (SRGAN) [14]. To evaluate the effectiveness of the proposed approach, an experiment was conducted where the same Convolutional Neural Network (CNN) was trained on the original dataset collected from surveillance cameras and a dataset augmented with the synthetic smoke images generated by the proposed GAN architecture.

Theoretical considerations might suggest that GANs have limited utility for data augmentation, as they may merely replicate existing data. However, our experimental results strongly refute this in the context of early smoke detection. These findings robustly support our hypothesis that GAN-based augmentation significantly surpasses conventional, simpler data augmentation techniques. Specifically, these results suggest its potential for other domains where objects of interest are characterized by fuzziness, low image quality (e.g. lacking distinct low- and mid-level features like corners or edges), and an inherent rarity that makes real data collection exceptionally difficult.

Main contributions of this research are as follows:

- A GAN architecture is proposed which can intentionally produce ”fuzzy” and low-quality smoke images accurately mimicking the degraded nature of real surveillance footage;
- An experimental framework is established for evaluation of different data augmentation strategies;
- A dataset for early smoke detection is created and made publicly available.

The remainder of the paper is structured as follows: Section II provides an overview of related work in synthetic image generation for wildfire detection. Section III outlines the methodology of this research. Section IV details the proposed Generative Adversarial Network architecture, followed by Section V which presents the experimental setup used for evaluating the quality of synthetically generated data. Section VI concludes the paper, summarizing our findings and outlining directions for future research.

II. Related Work

Deep learning algorithms have demonstrated great performance in different computer vision tasks, such as image classification [15], [16], [17], [18], region proposal [19], [20], [21], segmentation [22], [23], [24] and many others. However, their efficacy depends on the availability of extensive and diverse datasets, which can be particularly challenging to acquire for rare or complex phenomena such as wildfire smoke [25]. Other applications limited by a lack of sufficiently large dataset include medical imaging [26], plant disease recognition [27] and many other fields. This reliance on data frequently leads to issues such as overfitting when training models on limited datasets, thereby hindering generalization to new, unseen data [28]. To address the limitations of small datasets various Data Augmentation (DA) techniques are employed. Traditional approaches include standard methods like color conversion, image blurring, cropping, adding noise, resizing, flipping, rotating and other transforms. In contrast, Deep Generative Adversarial Networks offer a more advanced data augmentation strategy by employing algorithmic architectures composed of two neural networks to create new synthetic data instances. GANs are composed of a generative model **G** that captures the data distribution, and a discriminative model **D** that estimates the probability that a sample came from the training data rather than **G**. This generative capability allows for the creation of new objects and varied semantic features. This is a significant advantage over traditional DA methods, where an object's shape often remains largely unchanged.

The first images that generator outputted in Goodfellow's 2014 paper [12] were looking good, with digits and faces, but were very fuzzy and blurred on the CIFAR-10 dataset. In order to fix this problem, authors in [29] created a GAN using Laplacian Pyramid. Their main contribution is a type of GAN that is capable to produce high-quality images that are easily mistaken for real images in 40% of all data when assessed by human experts. The main difference from Goodfellow's [12] GAN architecture is the number of image generators; instead of only one, the authors used a series of CNNs to slowly create the whole image (the resolution was increased in each step). The idea was to create a pyramid-style model in which every level uses its own CNN and is trained on two components. The authors in [30] went a step further, with a model called Super-Resolution GAN (SRGAN). The main focus was finding a way to recover all textures and details from images that were lost during the downsampling phase. To fix this problem, the authors proposed a new loss function that can be divided in two parts. The first part is the adversarial loss that encourages images to look more natural. The second part is the content loss that makes sure that the new resolution image has the same features as the original low-resolution image. In [31], the authors present a DCGAN (Deep Convolutional GAN) architecture model using CNN, as its name suggests. This paper introduces various techniques for successful learning by using correct parameters. In [32], the authors used variants of GANs, considering label conditioning. This approach resulted in 128×128 resolution image samples, showing global coherence.

In [33], the authors used a Progressive Growing of GANs (PGGAN) network to generate synthetic data. They chose PGGAN because of its training stability at large image sizes and its robustness to hyperparameter selection. The authors showed that the use of GANs when facing limited data is possible. Researchers at NVIDIA proposed a StyleGAN architecture [34]. Previous GAN models have already shown the ability to generate human faces, but one challenge is being able to control some features of the generated images, such as hair color or pose. StyleGAN attempts to tackle this challenge by incorporating and building on progressive training to separately modify each level of detail. In the literature GANs are usually used for data augmentation. The results reported in [13], [35] suggest that GANs can have significant benefits when used for data augmentation in some classification tasks. This research shows that GANs have great potential in image processing applications, and possibly represents a step towards achieving strong artificial intelligence. GANs were also used for the classification of smoke images and videos. In [36], the authors used a DCGAN architecture to classify smoke and non-smoke images. In [37] a GAN was used for image enhancement, and in [38], GANs were applied to predict future frames and smoke trend heatmap. In our paper, a GAN is used for synthetic image generation. In the literature, several attempts at synthetic smoke image data generation were reported. In [39], the authors used a CycleGAN model in order to generate new smoke data samples. The main idea was to reduce data imbalance in their dataset. In [40], the authors used a DCGAN network to produce new smoke images, which were as realistic as possible. The authors in [41] used a CycleGAN-based pixel-level domain adaptation method for image translation. In [41], the authors also employed synthetic images generated through a combination of 3D modeling enhanced by both pixel-level and feature-level domain adaptation techniques. 3D modeling based on Unreal Engine was used in [42] to generate a multi-faceted synthetic dataset for wildfire detection. A method for generating smoke images based on multi-scale expansion fusion GAN was proposed in [43]. A novel training-free and mask-guided diffusion framework for synthesizing realistic wildfire images was proposed in [44]. This method precisely controls the placement of wildfire elements and utilizes augmented masks and Perlin noise to enhance the diversity and realism of generated images. Two deepfake generation approaches for creating synthetic wildfire images to augment a dataset for training detection models were evaluated in [45].

Beyond GAN-based approaches, diffusion models have emerged as a powerful alternative for synthetic data generation [46], underpinning recent wildfire-specific methods such as Flame Diffuser [44] and showing promise for augmentation in limited-data settings [47].

Most of the mentioned methods were tuned for images of well-developed wildfire where fire and/or smoke was clearly visible. Our images were intentionally selected to serve as typical representations of smoke in its early phase. The automatic detection system was required to detect smoke in the first 5 minutes. In this phase, smoke may be scarcely visible even for a human observer.

III. Methodology

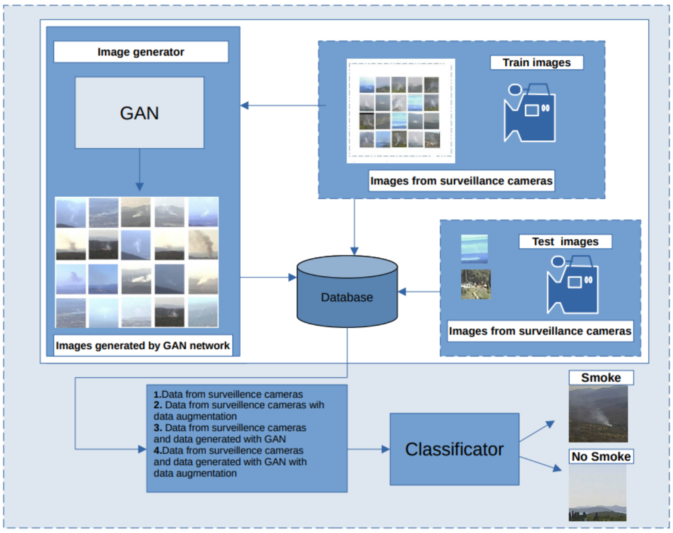

Fig. 2. Workflow of the proposed synthetic image generation method and evaluation framework

Research workflow is shown in Figure 2. The training dataset was collected from the operational wildfire surveillance system. Images in the dataset contain emerging smoke of real fire incidents in the early ignition phase. As the main requirement for the detection algorithm is very early smoke detection, only images of smoke from the first 5 minutes after ignition are used to create the dataset. All collected image samples are recordings of real forest fire incidents. The collected images are used to train the proposed GAN architecture.

To rigorously evaluate the capacity of the proposed Generative Adversarial Network generator to produce effective synthetic data for deep neural network training, a comparative experimental framework was established. This methodology comprised three distinct phases:

1) *Dataset Preparation:*
   - *Training Datasets:* Two distinct training datasets were constructed
     - *(a) Original Real Dataset:* This dataset consisted solely of real images of early-stage smoke, collected from operational surveillance cameras. This is the same set of images used to train the proposed GAN model.
     - *(b) Hybrid Augmented Dataset:* This dataset combined all images from the Original Real Dataset with synthetic images generated by the proposed GAN.
   - *Independent Test Dataset:* For objective evaluation, the test dataset was composed exclusively of real images. Crucially, these images were acquired from a geographically separate set of surveillance locations that were not utilized in the creation of either training dataset. This ensured complete data independence, preventing any overlap between the training and testing environments.
2) *Model Training:* The same CNN architecture was sequentially trained. First, it was trained on the (a) Original Real Dataset, and subsequently, on the (b) Hybrid Augmented Dataset.
3) *Performance Evaluation:* The performance of the CNN models, after training on each respective dataset, was then assessed using the Independent Test Dataset. The integrity and objectivity of the experiment were maintained by ensuring that neither the CNN classifier nor the GAN generator had any prior exposure to the images or surveillance locations represented in the test dataset. This rigorous separation demonstrates the models' true generalization capabilities.

The data used in this work were mostly taken from the OIV Fire Detect AI Forest Fire Monitoring and Surveillance System [48]. The company Odašiljači i veze d.o.o. reserves the right of ownership over images, video materials and other products of the OIV Fire Detect AI surveillance system. With the permission of the data owner, the data used in this paper are made public and available for download without special restrictions or fees at http://wildfire.fesb.hr/gan/data.

IV. GAN Architecture

In this section, we describe the GAN architecture used for the generation of synthetic images. Figure 3 illustrates the basic GAN concept, comprising a generator network and a discriminator network. Both the generator and the discriminator are neural networks. Generator network takes a random noise vector and produces an image. Discriminator works as a binary classifier; it needs to classify whether the data is real or synthetically generated (fake). The generator output is connected directly to the discriminator input. Using backpropagation, the discriminator's classification provides a signal that the generator uses to update its weights.

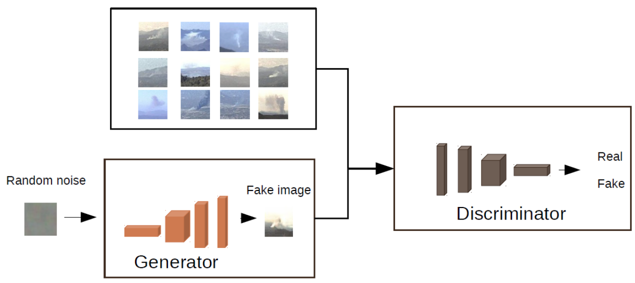

Fig. 3. Basic GAN concept illustrating the interaction between the generator and discriminator networks

The discriminator uses a cross-entropy loss, where p represents true distribution, and q the estimated distribution:

$$H(p, q) = -\sum_i p_i log(q_i) \tag{1}$$

The intention of the loss function H in equation 1 is to push the predictions of the real image towards 1 and the fake images to 0. We do so by the log probability term.

$$\min_G \max_D V(D,G) = \\ : \mathbb{E}_{x\sim p_{data}(x)}[log(D(x))]+ \\ : \mathbb{E}_{z\sim p_{data}(z)}[log(1-D(G(z)))] \tag{2}$$

If we observe the joint loss function, we maximize the discriminator term, which means $logD(x)$ should inch closer to zero, and $logD(G(z))$ should be closer to 1. Here, the generator is trying to make $D(G(z))$ inch closer to 1, while the discriminator is trying to do the opposite, as shown in Fig. 4.

The proposed architecture is inspired by the SRGAN network [14]. SRGAN is a generative adversarial network for the generation of super-resolution images. The network consists of two parts: an image generator and an image discriminator. The generator produces data based on the probability distribution and the discriminator tries to distinguish whether the data are real or produced from a generator network. The generator is trying to optimize the generated data, so that the produced data can fool the discriminator. There are two specific characteristics of the SRGAN network, B residual blocks and perceptual loss function. SRGAN consists of 16 B residual blocks, originated by ResNet. Within the residual block, two convolutional layers with small 3×3 kernels are used. The number of feature maps is 64, followed by batch-normalization layers and ParametricReLU as the activation function. Since the main goal of SRGAN is the generation of super-resolution, two sub-pixel convolution layers are added to increase the resolution. For the loss function network, a perceptual loss function was used. The perceptual loss function has two parts: an adversarial loss and content loss. The first part pushes the current generated image to the natural image manifold using a discriminator network. The second part is content loss, which is used for perceptual similarity instead of similarity in pixel space. There are three main differences in the proposed architecture:

- The input to the generator is a 256-dimensional latent vector, followed by upsample layers.
- B residual blocks are removed.
- The discriminator uses the LeakyReLU activation function.

Our intention was to use SRGAN to obtain important features from existing images. In SRGAN, B residual blocks preserve fine-grained detail for super-resolution — the opposite of our goal. Removing them deliberately reduces the generator's capacity to reconstruct sharp details, resulting in softer and more diffuse visual characteristics typical for early-stage smoke captured under adverse atmospheric conditions. Further, this modification decreases computational complexity and makes the model easier to train, particularly in scenarios where the available dataset is relatively limited. Figure 4 shows our architecture. The generator input is a 256-dimensional vector, followed by blocks of convolution, batch normalization and ReLU activation function. Upsample layers were added, because the input in the original SRGAN was an image. ReLU activation was used in all generator layers, except the final layer where Tanh activation function was used, similar to the approach in [31]. Two sub-pixel convolution layers are added to increase the resolution. For the discriminator we use LeakyReLU activation function in all layers, except the last one, where we use sigmoid. The ReLU activation function was originally used in the discriminator network. Our kernel size was 3, stride was 2, and the batch size was 64.

V. Experiment

The main goal of this research was to determine the feasibility of generating synthetic smoke images using the GAN architecture. There are two basic prerequisites for synthetic images:

1) Generated images must be as similar as possible to the images taken from surveillance cameras, where the smoke is in a very early stage, fuzzy and often blends into the background.
2) Although fuzzy and hard to see, even for a human observer, smoke should retain all of the features that are necessary to train the classifier.

Another concern was the theoretical observation that synthetically generated images can only contain the information that was already represented in the original dataset. As smoke in original surveillance camera images is often transparent and barely visible, without clear low- and mid-level visible features (e.g. edges or corners), there was an objective possibility that GAN will recreate variations of the existing backgrounds, without adding new information content that is relevant to smoke detection.

To evaluate the quality of the generated synthetic images and their applicability for training DNN architectures, an experimental scenario was set up in which the same CNN classifier was trained on different datasets: one containing only real data taken from surveillance cameras, and the other containing both real and synthetic images. Further, to ensure the objectivity of the experiment, a subset of surveillance locations was excluded as a source of the training data. These locations were used exclusively to collect the test dataset. No image from this subset of locations was seen by the GAN generator. In the experiment phase, images collected from these additional locations were only used to evaluate the trained classifier, and no image from these test locations was present in the classifier training dataset. Additionally, no images from locations used to collect training data were present in the test dataset. The performance of the same CNN architectures trained on different datasets was evaluated and results were compared.

A. Dataset

The dataset used in this study was created from the images taken from the Intelligent Forest Fire Video Monitoring and Surveillance System [3] installed on more than 116 locations in the coastal counties of the Republic of Croatia. The peculiarity of the Croatian coast is its large indentation with numerous settlements surrounded by a dense forest of Aleppo pine. This

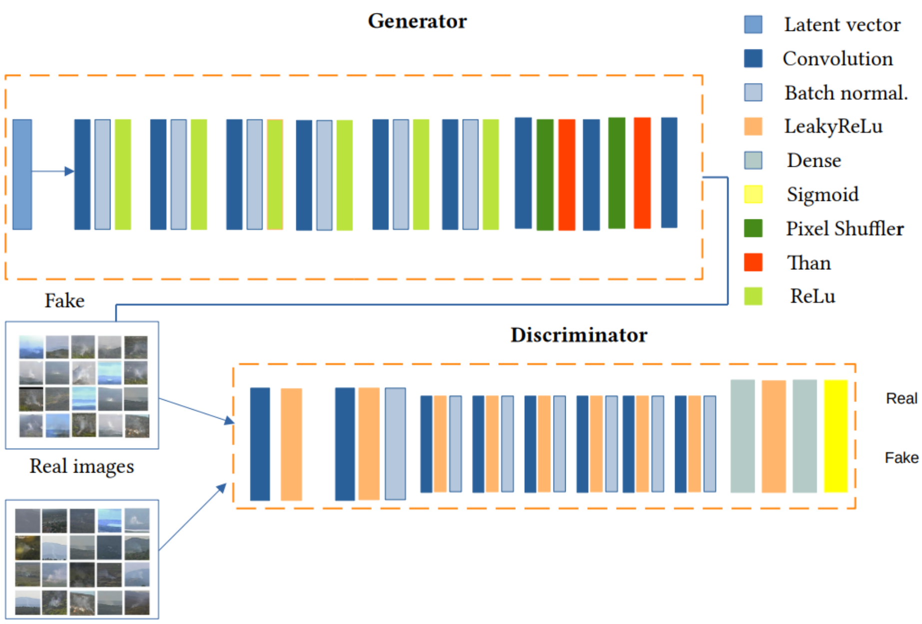

Fig. 4. Proposed GAN architecture for synthetic wildfire smoke image generation

Fig. 5. Original 1920×1080 resolution image taken from a surveillance camera and a positive 224×224 sample containing emerging smoke in its early phase. The smoke sample is shown magnified by a factor of 2.

means that practically the whole area is a Wildland Urban Interface [49], and the main requirement for the surveillance system is early forest fire detection, preferably in the first 5 minutes of ignition. Therefore, only images within the first 5 minutes of the first visible signs of smoke were taken into account when forming the dataset. To ensure maximum objectivity of the experiment, 16 randomly selected locations of the surveillance system were excluded as a source for the training data. No images from these locations were used to create the training dataset for either positive or negative samples. These 16 locations were used to collect images for the test dataset. This ensured that no sample in the test dataset was related in any way to any image used in the training process, for training a GAN or for training a classifier used for evaluation of the resulting GAN model.

From the image archive collected from the set of training locations a total of 4127 positive samples were collected, i.e., samples containing emerging smoke of the forest fire in early ignition phase. Samples were taken as 224x224 patches from 1920x1080 resolution images from surveillance cameras. Samples were taken without any processing, including scaling or improving image quality in any way. An example of a positive sample taken from surveillance camera images is shown in Figure 5. The same number of negative samples, i.e., samples not containing smoke, was collected from the same source. From the set of 16 surveillance locations set aside to collect test images, a balanced test dataset was created containing 681 positive and the same number of negative samples.

B. GAN Training

The proposed GAN architecture was trained on 4127 positive samples from the training dataset. The discriminator and generator were optimized using the Adam optimizer with learning rates of 0.0003 and 0.0001, respectively, with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The proposed GAN architecture hyperparameters were selected empirically based on preliminary experiments, to ensure stable training. The network was trained for 2000 iterations, with convergence assessed based on the stabilization of losses and visual inspection of the generated samples. Table I shows the parameters that were used for GAN training. For input in a generator network, we used a 256-dimensional latent space vector. The discriminator and generator use different learning rates to prevent the discriminator from becoming too dominant during training. Figure 6 shows different iterations of the training process, illustrating progressive improvement of the generator output with each iteration. The discriminator and generator loss curves during the training process are shown in Figure 7.

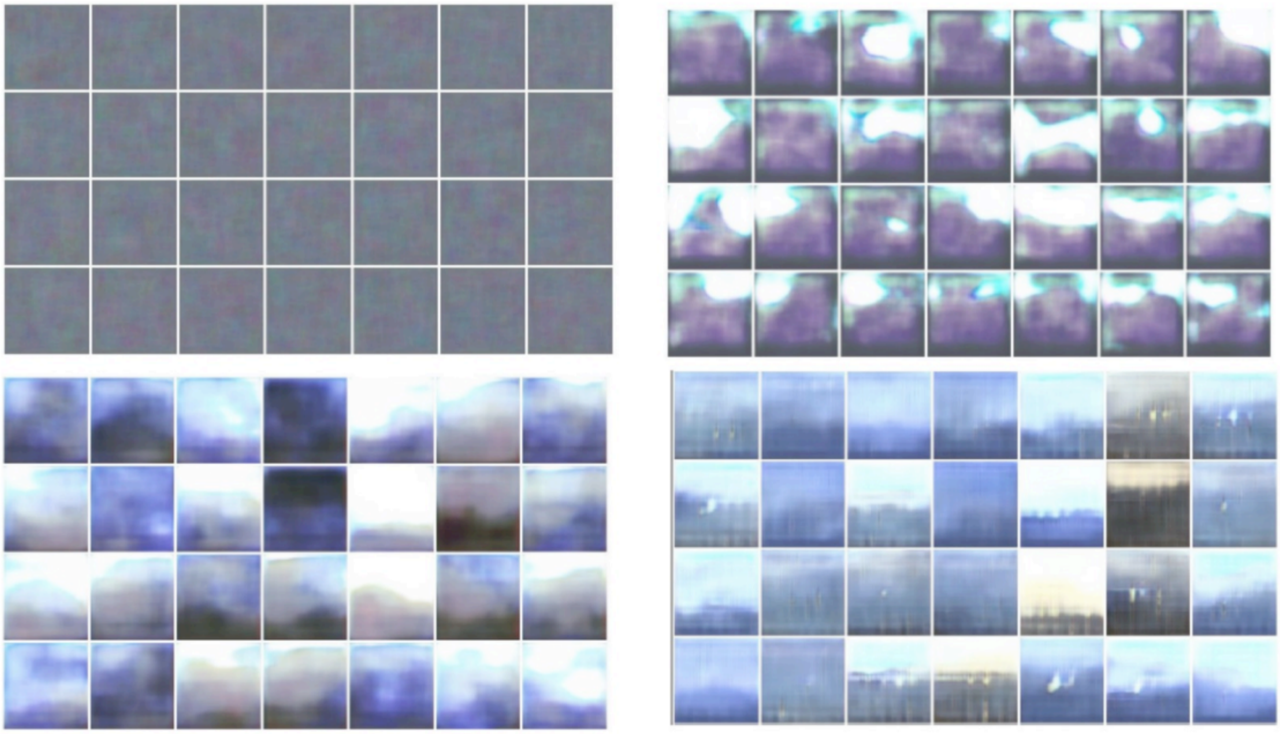

Fig. 6. GAN-generated output samples at the first four training iterations, illustrating progressive improvement of the generator

TABLE I
GAN TRAINING PARAMETERS

Size of the input image to discriminator	224x224x3
Size of discriminator output	real or fake
Size of latent vector to provide the generator	256
Size of discriminator output (generated image)	224x224x3
Optimizer	Adam optimizer
Discriminator learning rate	0.0003
Generator learning rate	0.0001
β_1	0.5
β_2	0.999
Number of iterations	2000

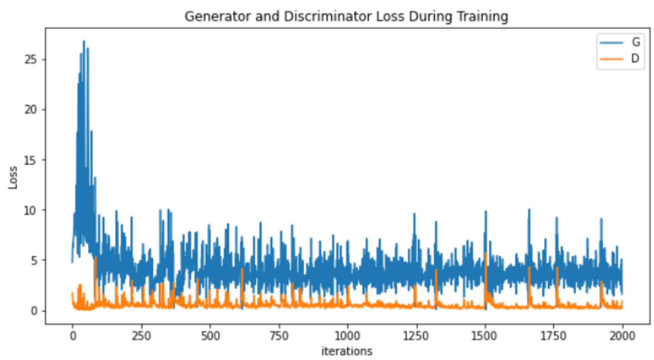

Fig. 7. Discriminator and generator loss curves during the training process

C. GAN Evaluation

We aim to examine the quality of the synthetic images produced by the proposed GAN model in order to evaluate the suitability of using the GAN in problems where the object or phenomenon of interest in an image is fuzzy, and a large enough set of annotated data is not available, specifically to generate smoke images in the early stages of a forest fire. To compare real and synthetic data and assess the contribution to the information diversity that can be obtained by adding synthetically generated images to a training dataset, an experiment was set up in which the same CNN architecture was trained using different datasets.

Standard metrics such as Fréchet Inception Distance (FID) and Inception Score (IS) measure perceptual realism for sharp, high-quality images. In specialized domains, these metrics may not reliably reflect the actual usefulness of the generated samples. As a result, improvements in FID or IS do not necessarily translate into better performance in domain-specific applications such as the one considered in this study. Therefore, we focused on evaluating the effectiveness of the generated images through their impact on classifier performance.

In the experiment, the VGG16 network architecture [50] was used. VGG16 was chosen for its generally good performance in classifying small objects in an image. Most of the parameters of the VGG16 model were left at their default values, as the focus of this study was the proposed GAN architecture and the classifier was only used as a tool for GAN evaluation. We used transfer learning with a pre-trained VGG16 network trained on Imagenet dataset and we trained only the last layer. The parameters that were used for training are shown in Table II. The same parameters were used in all experiments.

A trained GAN model was used to create a set of 2500 synthetic smoke images. Using these data, two distinctive datasets were created:

1) Real dataset, containing a total of 8254 images (4127

TABLE II
VGG16 TRAINING PARAMETERS

Optimizer	Adam
Learning rate	1e-5
Batch size	16
Class mode	binary
Number of iterations	500

positive and 4127 negative samples) taken from surveillance cameras without any preprocessing. The structure of the real dataset is shown in Table III.

2) Hybrid dataset, containing 4127 positive samples from cameras and 2500 synthetically generated images of smoke. To produce a balanced dataset, an additional 2500 negative samples were extracted from the surveillance system archive. These additional negative samples were taken from the same set of surveillance locations that were used to collect training data. The structure of the hybrid dataset is shown in Table IV.

TABLE III
REAL DATASET STRUCTURE

samples	*number of samples*	*source*
positive	4127	surveillance cameras
negative	4127	surveillance cameras

TABLE IV
HYBRID DATASET STRUCTURE

samples	*number of samples*	*source*
positive	4127	surveillance cameras
positive	2500	GAN-generated
negative	6627	surveillance cameras

Examples of original data samples taken from surveillance cameras are shown in Figure 8, and synthetically generated examples are shown in Figure 9.

In the first experiment, a VGG16 classifier was trained on both data sets: the real dataset (Table III) and the hybrid dataset (Table IV), with data augmentation intentionally disabled during training. To compare the GAN architecture's ability to increase the diversity of the training data with the effect of common data augmentation techniques, training was started again on both datasets, this time with data augmentation enabled. A total of four models were trained as follows:

1) VGG16 trained on real data with data augmentation disabled.
2) VGG16 trained on a hybrid dataset (real data combined with GAN-generated images) with data augmentation disabled.
3) VGG16 trained on real data with data augmentation enabled.
4) VGG16 trained on a hybrid dataset (real data combined with GAN-generated images) with data augmentation enabled.

An illustration of the four experiments is shown in Figure 10. All four trained models were evaluated on the same balanced test dataset, containing 681 positive and an equal number of negative samples. Test data contains only real images, collected from a separate set of surveillance locations that were not used to form the training dataset.

Evaluation results are shown in Table V. Classifier (a) trained only on real data without augmentation has achieved 73.1% accuracy and 75.1% precision. The same classifier, trained on hybrid dataset without augmentation (b) improves accuracy to 77.4% and precision to 79.4%. Training the classifier on real data with enabled augmentation (c) achieved 74.9% accuracy and 77.4% precision. As expected, augmentation improved classifier performance. However, the achieved improvement is lower when compared to scenario (b), where the training dataset was expanded by GAN-generated samples. Finally, the classifier (d) that was trained on hybrid data (real data and data generated with GAN) with augmentation enabled achieved 78.8% accuracy and 81.9% precision. Figure 11 presents examples of images from the test dataset that were correctly labeled as *smoke* by model (d), but were misclassified as *no smoke* by models (a), (b), and (c).

The obtained results shown in Table V clearly indicate that the use of synthetically generated smoke produced by the proposed GAN model outperforms standard augmentation techniques (flipping, random rotation, brightness, etc.). Moreover, these two techniques are not necessarily mutually exclusive, as the classifier trained on a hybrid dataset with augmentation enabled outperformed the other three models.

VI. DISCUSSION AND CONCLUSION

The scarcity of sufficiently large labeled datasets presents the primary challenge for applying state-of-the-art deep learning algorithms. To mitigate this problem and prevent overfitting, data augmentation techniques are commonly employed, ranging from simple operations like rotation or image mirroring to more advanced synthetic image generation algorithms. This research is focused on generating synthetic images of forest fire smoke in its earliest stages, only minutes from ignition, at large distances from the camera. Smoke, as the first visible sign of a forest fire, is poorly developed at this stage, often semi-transparent and blended with the background landscape. In addition, smoke at a distance measured in kilometers often occupies a very small part of the image. The quality of images captured by forest fire surveillance cameras is further reduced by interference such as atmospheric conditions, lighting effects, and dirt on the camera lens. This presents an additional obstacle to successful early-stage fire detection. Acquiring an appropriate training dataset for such conditions is challenging because only the earliest stages of fire are relevant, and forest fires are relatively rare occurrences.

In this paper, we investigated the feasibility of using generative adversarial networks to create synthetic images of smoke at large distances. Synthetically generated smoke images must be fuzzy and of low quality, to be as similar as possible to real images taken from surveillance cameras and suitable for training DNN-based classifiers. The fundamental problem was the lack of distinctive visible smoke features in a large number of images from the training set, which often presents a problem even for the human expert in charge of labeling the images.

Fig. 8. Original wildfire smoke samples collected from surveillance cameras

TABLE V
EXPERIMENT RESULTS

train data	*Aug*	*TP*	*TN*	*FP*	*FN*	*Acc*	*P*	*R*	*F1*
(a) real	NO	509	487	169	197	73.1%	75.1%	72.0%	73.5%
(b) hybrid	NO	539	511	139	173	77.4%	79.4%	75.7%	77.5%
(c) real	YES	524	498	153	187	74.9%	77.4%	73.7%	75.5%
(d) hybrid	YES	551	523	121	167	78.8%	81.9%	76.7%	79.2%

The main concern was that the synthetic image generator would output variations on the landscapes that existed in the training image set, while ignoring the smoke itself.

A GAN model is proposed based on the SRGAN architecture, primarily designed for generating super-resolution images from lower-quality images. The proposed model uses the SRGAN architecture's ability to extract important features from low-quality images, while removing the output blocks responsible for generating super-resolution images reduces the quality of the synthetically generated images. The input image originally used in the SRGAN architecture was replaced by a random noise vector to increase the diversity of the generated samples. Using this approach, synthetic images of smoke were obtained that are similar to real images with all the degradations that reduce the quality of images obtained from surveillance cameras.

The proposed GAN architecture's ability to generate images of the emerging forest fire smoke that can improve classifier performance was tested in an experiment in which the same CNN architecture was trained on two datasets: one including only real images and the other including a set of real images augmented by the synthetic images obtained by the GAN generator. The obtained results confirm that the proposed GAN architecture can create synthetic data samples that can significantly improve the efficiency of deep learning algorithms. Its demonstrated effectiveness encourages further development of the GAN architecture and its adaptation for generating emerging smoke images. In the future, focus will also include the adaptation and development of deep neural network architectures suitable for smoke detection on low-quality images. The proposed technique can be adapted for landscapes that differ morphologically and vegetatively from

Fig. 9. Examples of synthetic wildfire smoke images generated by the proposed GAN model

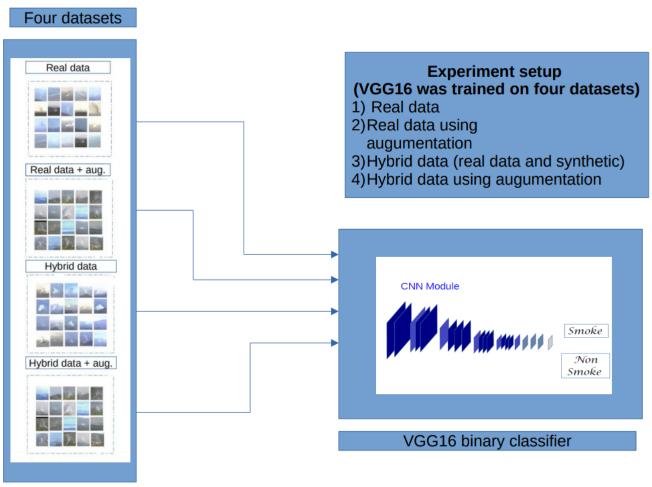

Fig. 10. Illustration of the four experimental configurations used to evaluate the proposed GAN-based data augmentation approach

the typical Mediterranean environment. It is also applicable to other domains where images are of lower quality, the object of interest is not clearly defined, or its rare occurrence makes it challenging to collect a sufficiently large training dataset.

All datasets used in this research are publicly available for download by the courtesy of the legal data owner, the company Odašiljači i veze d.o.o. We plan to expand the datasets with new images from surveillance cameras, and updated datasets will be published periodically.

REFERENCES

[1] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, “Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier],” *IEEE Computational Intelligence Magazine*, vol. 13, no. 4, pp. 59–76, 2018.

[2] V. López, A. Fernández, J. G. Moreno-Torres, and F. Herrera, “Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. open problems on intrinsic data characteristics,” *Expert Systems with Applications*, vol. 39, no. 7, pp. 6585–6608, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417411017143

[3] M. Štula, D. Krstinić, and L. Šeri ’c, “Intelligent forest fire monitoring system,” *Information Systems Frontiers*, vol. 14, no. 3, pp. 725–739, 2012.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1106–1114.

[5] D. Krstinić, M. Braović, and D. Božić-Štulić, “Convolutional neural networks and transfer learning based classification of natural landscape images,” *Journal of universal computer science*, vol. 26, no. 2, pp. 244–267, 2020.

Fig. 11. Examples of images correctly classified as *smoke* by the VGG16 model trained on the hybrid dataset (d), but misclassified as *no smoke* (False Negative) by models trained without GAN-generated data. Smoke locations are marked manually.

[6] M. Braović, D. Stipaničev, and D. Krstinić, "Cogent confabulation based expert system for segmentation and classification of natural landscape images," *Advances in Electrical and Computer Engineering*, vol. 17, pp. 85–94, 2017. [Online]. Available: http://dx.doi.org/10.4316/AECE.2017.02012

[7] M. Bugarić, M. Braović, and D. Stipaničev, "Augmented reality based segmentation of outdoor landscape images," in *8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2013, pp. 43–48.

[8] T. Jakovčević, D. Stipaničev, and D. Krstinić, "Visual spatial-context based wildfire smoke sensor," *Machine Vision and Applications*, vol. 24, no. 4, pp. 707–719, 2013.

[9] D. Krstinić, D. Stipaničev, and T. Jakovčević, "Histogram-based smoke segmentation in forest fire detection system," *Information Technology and Control*, vol. 38, no. 2, pp. 237–244, 2009.

[10] L. Šerić, A. Ivanda, M. Bugarić, and M. Braović, "Semantic conceptual framework for environmental monitoring and surveillance—a case study on forest fire video monitoring and surveillance," *Electronics*, vol. 11, no. 2, pp. 1–18, 2022.

[11] D. Stipaničev, "Holistic forest fire protection -- ipa adriatic project 2014 – 2016," 2015.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[13] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. N. Gunn, A. Hammers, D. A. Dickie, M. del C. Valdés Hernández, J. M. Wardlaw, and D. Rueckert, "GAN augmentation: Augmenting training data using generative adversarial networks," *CoRR*, vol. abs/1810.10863, 2018. [Online]. Available: http://arxiv.org/abs/1810.10863

[14] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016. [Online]. Available: http://arxiv.org/abs/1609.04802

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[16] D. Buscombe and A. Ritchie, "Landscape classification with deep neural networks," *Geoscience*, vol. 8, 2018.

[17] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, Dec 2015.

[18] Y. Li, M. Dixit, and N. Vasconcelos, "Deep scene image classification with the mfafvnet," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[19] R. Girshick, "Fast r-cnn," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. USA: IEEE Computer Society, 2015, pp. 1440–1448. [Online]. Available: https://doi.org/10.1109/ICCV.2015.169

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'15. Cambridge, MA, USA: MIT Press, 2015, pp. 91–99.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 779–788.

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *CoRR*, vol. abs/1511.00561, 2015. [Online]. Available: http://arxiv.org/abs/1511.00561

[23] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014. [Online]. Available: http://arxiv.org/abs/1411.4038

[24] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Cham: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[25] M. Park, D. Tran, J. Bak, and S. Park, "Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly supervised object localization," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, p. 103052, 11 2022.

[26] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "Gans for medical image analysis," *Artificial Intelligence in Medicine*, vol. 109, p. 101938, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719311510

[27] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[28] C. Pérez-Guerrero, J. F. Ciprián-Sánchez, A. Palacios, G. Ochoa-Ruiz, M. Gonzalez-Mendoza, V. Foroughi, E. Pastor, and G. Rodriguez-Hernandez, "Computer vision-based characterization of large-scale jet flames using a synthetic infrared image generation approach," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107275, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623014598

[29] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," *CoRR*, vol. abs/1506.05751, 2015. [Online]. Available: http://arxiv.org/abs/1506.05751

[30] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016. [Online]. Available: http://arxiv.org/abs/1609.04802

[31] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015,

cite arxiv:1511.06434Comment: Under review as a conference paper at ICLR 2016. [Online]. Available: http://arxiv.org/abs/1511.06434

[32] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. International Convention Centre, Sydney, Australia: PMLR, 06–11 Aug 2017, pp. 2642–2651. [Online]. Available: http://proceedings.mlr.press/v70/odena17a.html

[33] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *CoRR*, vol. abs/1710.10196, 2017. [Online]. Available: http://arxiv.org/abs/1710. 10196

[34] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *CoRR*, vol. abs/1812.04948, 2018. [Online]. Available: http://arxiv.org/abs/1812.04948

[35] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," *CoRR*, vol. abs/1801.02385, 2018. [Online]. Available: http://arxiv.org/abs/1801.02385

[36] S. Aslan, U. Güdükbay, B. U. Töreyin, and A. Enis Çetin, "Early wildfire smoke detection based on motion-based geometric image transformation and deep convolutional generative adversarial networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8315–8319.

[37] K. Yun, J. Bustos, and T. Lu, "Predicting rapid fire growth (flashover) using conditional generative adversarial networks," 2018.

[38] S. Cheng, J. Ma, and S. Zhang, "Smoke detection and trend prediction method based on deeplabv3+ and generative adversarial network," *Journal of Electronic Imaging*, vol. 28, no. 3, May 2019. [Online]. Available: https://doi.org/10.1117/1.JEI.28.3.033006

[39] M. Park, D. Q. Tran, D. Jung, and S. Park, "Wildfire-detection method using densenet and cyclegan data augmentation-based remote camera imagery," *Remote Sensing*, vol. 12, no. 22, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/22/3715

[40] H. Yin, Y. Wei, H. Liu, S. Liu, C. Liu, and Y. Gao, "Deep convolutional generative adversarial network and convolutional neural network for smoke detection," *Complexity*, vol. 2020, p. 6843869, Nov. 2020.

[41] J. Mao, C. Zheng, J. Yin, Y. Tian, and W. Cui, "Wildfire smoke classification based on synthetic images and pixel- and feature-level domain adaptation," *Sensors*, vol. 21, no. 23, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7785

[42] G. Wang, H. Li, P. Li, X. Lang, Y. Feng, Z. Ding, and S. Xie, "M4sfwd: A multi-faceted synthetic dataset for remote sensing forest wildfires detection," *Expert Systems with Applications*, vol. 248, p. 123489, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424003543

[43] Y. Huo, Q. Zhang, C. Wang, H. Wang, and Y. Zhang, "Enhancing wildfire detection: a novel algorithm for controllable generation of wildfire smoke images," *International Journal of Wildland Fire*, vol. 33, no. 11, p. WF24068, 11 2024. [Online]. Available: https://doi.org/10.1071/WF24068

[44] H. Wang, S. P. Haeri, X. Chen, A. Bastola, H. Li, W. Zhu, and A. Razi, "Flame diffuser: Wildfire image synthesis using mask guided diffusion," 12 2024, pp. 6171–6179.

[45] J. Bejo, D. Krstinić, A. Bartulović', and M. Braović', "A dual approach deepfake generation for synthetic wildfire images," in *33rd International Conference on Software, Telecommunications, and Computer Networks (SoftCOM2025), Split, Croatia*, 9 2025.

[46] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[47] B. Trabucco, K. Doherty, M. Gurinas, and R. Salakhutdinov, "Effective data augmentation with diffusion models," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024, arXiv:2302.07944.

[48] Odasiljaci i veze d.o.o, "Oiv fire detect ai, intelligent solution for early wildfire detection," 2018, last accessed 19 August 2022.

[49] S. I. Stewart, V. C. Radeloff, R. B. Hammer, and T. J. Hawbaker, "Defining the Wildland–Urban Interface," *Journal of Forestry*, vol. 105, no. 4, pp. 201–207, 06 2007. [Online]. Available: https: //doi.org/10.1093/jof/105.4.201

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

Dunja Božić-Štulić received her Master of Engineering degree in Electronics and Computer engineering in 2014, from Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Croatia. She received her Ph.D from the same university in 2020. She is currently working as assistant professor at Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split. Her research interests include machine learning, deep learning and artificial intelligence.

Damir Krstinić is a Full Professor at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture of the University of Split. His main areas of research are computer vision, image understanding and machine learning. He has co-authored more than 40 scientific papers. He is a member of the Program Council of the Center for Wildfire Research (Split, Croatia), and has been involved in several research and technological projects related to forest fires. He is the project leader of the Wildfire Early Detection and Monitoring System developed for Croatian Forests. The developed system has been in operation since 2018 and has been installed at more than 120 monitoring locations in Croatia, Bosnia and Hercegovina, Montenegro and Albania.

Darko Stipaničev is Professor Emeritus of Computer Science and Automatic Control at Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture University of Split. His research interests include complex systems modeling and control, intelligent systems analyses and design, digital image analyses, advanced Internet technologies and recently, the application of AI in environmental protection, particularly wildfire prevention and management, including vision based wildfire smoke detection, wildfire risk estimation, wildfire behaviour modelling and wildfire spread simulation.

Jakov Bejo is a PhD student at the Faculty of Electrical Engineering, Mechanical Engineering, and Naval Architecture of the University of Split. He works at Code Fire d.o.o., focusing on the early wildfire detection. His research covers real-time object detection, synthetic data generation, EEG-driven robotic control, 3D simulation soldier and military vehicle recognition, and digital twin systems development using Unreal and Unity engines. He has co-authored three scientific papers in the area of wildfire detection.