Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 9 (2023) 247–257
www.elsevier.com/locate/egyr

TMREES23-Fr, EURACA 06–08 February 2023, Metz-Grand Est, France

# CNN-based, contextualized, real-time fire detection in computational resource-constrained environments

Eleni Tsalera[a,*], Andreas Papadakis[b], Ioannis Voyiatzis[a], Maria Samarakou[a]

[a] *Department of Informatics and Computer Engineering, University of West Attica, Agiou Spyridonos, Egaleo, 12243, Greece*
[b] *Department of Electrical and Electronics Engineering Educators, School of Pedagogical and Technological Education, Athens, 14122, Greece*

Received 19 May 2023; accepted 29 May 2023
Available online xxxx

**Abstract**

The increasing occurrence of wildfires, amplified by the changing climate conditions and drought, poses threats to human lives, the environment and the geographically dispersed infrastructures. Such impact necessitates the prompt identification of wildfires so that appropriate countermeasures are taken. The availability of electronic equipment, such as Unmanned Aerial Vehicles, allows for images from dynamically changing, geographical areas, which must be directly processed for wildfire identification and contextualization. In this work, we identify the requirements and the constraints in terms of computational resources of this workflow, and investigate lightweight CNNs to be used. SqueezeNet, ShuffleNet, MobileNetv2 as well as ResNet50 are used for fire identification. To simulate the realistic conditions, we have investigated multiple datasets, selecting Forest-Fire and Fire-Flame datasets and images from 3rd party sources and performed cross-dataset identification evaluation. To rationalize the required computational resources and the operation cost, lightweight networks have been selected and compared with ResNet-50, which is more complex. The contextualization, i.e. the detection of elements related to energy infrastructures, has been based on image semantic segmentation, performed through ResNet-18. The identification results, expressed as classification accuracy has reached 96%, with satisfactory results in the cross dataset scenarios, while we have identified five classes from the CamVid dataset which can be used for the contextualization needs.


Peer-review under responsibility of the scientific committee of the TMREES23-Fr, EURACA, 2023.

*Keywords:* Fire detection; Semantic segmentation; Environmental protection; Identification and contextualization wildfire; Energy infrastructures; CNNs

## 1. Introduction

Wildfires pose serious threat to human lives, natural environment, and the infrastructures. The climate change amplifies the frequency and impact of wildfires. This increasing tendency is verified by statistics, as in 2021, fires were mapped in 22 of the European Union 27 Member States, burning 500,566 hectares (ha) in total, more than the approximate 340,000 ha of 2020 (with the exception of 2017 when over 10,000 km$^2$ had burnt) [1]. The situation

* Corresponding author.
*E-mail address:* etsalera@uniwa.gr (E. Tsalera).

https://doi.org/10.1016/j.egyr.2023.05.260

Peer-review under responsibility of the scientific committee of the TMREES23-Fr, EURACA, 2023.

is similar in the United States, where 100,000 wildfires occur each year with the National Interagency Fire Centre reporting in 2022 64,127 fires affecting 7,343,939 acres and in 2021 54,976 fires affecting 6,814,073 acres [2]. Especially for the infrastructure, including the electrical power and energy systems, wildfires can disrupt the business continuity of critical elements in power generation and distribution and cause impactful security incidents.

Considering the increasing difficulties in the confrontation of wildfires, effort is dedicated to the early detection. Towards this direction, new remote sensing approaches become available, including ground sensor -based systems, manned and unmanned aerial vehicle-based systems, and satellite-based systems. While a rich set of sensors (such as temperature, smoke, infrared, gas, optical spectroscopic carbon monoxide sensor) can be effectively used for detection in (geographically) restricted environments, the need for extensive coverage creates challenges in the cases of remote, open air and forest areas. To be able to make decisions in real time, the processing of the sensed data have to take place promptly so that fire is localized. In parallel context-based data related to infrastructure potentially affected by the wildfire combined with risk analysis can allow for intelligent decisions. However, fire and smoke detection is a challenging image-processing task, due to the highly dynamic nature of these phenomena, their irregular shapes and forms, as well as their asymmetrical distribution in space. This dynamic nature is amplified when images are taken irregularly in time from dynamically changing geographical settings. In addition, the quality of the images may also vary, due to the capturing technique, the changing field of view, the employed equipment, the distance, the physical barriers and even the compression applied.

In this work, we focus on the early detection and identification of fire in open air areas, endangering critical energy-related infrastructures. As the early fire detection belongs to the family of critical applications, involving real time decision making, a complete solution is designed including the identification and the contextualization (achieved through semantic segmentation). We assume the availability of distributed image retrieval mechanisms (such as UAV and/or satellite) which provide multiple, near real-time images, of medium or even low quality, in dynamically changing geographic areas. We also assume that the sensing systems are equipped with computational resources allowing the on-the-spot processing of the available visual material. From the technological perspective, we employ Convolutional Neural Networks (CNNs) for early fire detection and contextualization/ semantic segmentation. The particularities of the problem at hand involve a) the existence of a limited number of datasets (especially considering that the characteristics of the datasets can orient or even define the approach to solve a problem and influence the performance of the algorithms) and b) the need to restrict the required computational resources. For this we have considered a wide range of datasets, selecting two of them with the addition of images coming from 3rd party sources, to confront different settings in terms of resolution, noise and compression, as these may not be easily perceptible by the human visual system may affect the contained visual information used by the ML modules. To enhance robustness and flexibility, we employ *transfer learning, cross-dataset evaluation* and *noise inclusion.*

### 1.1. Similar work

Vision-based systems have used a variety of equipment, mechanisms and techniques such as background subtraction in video acquired by static cameras [3], the usage of synthetic images to train CNNs [4], linear color space conversion to detect fire pixels using Particle Swarm Optimization (PSO) for proper weights of the conversion matrix and K-medoids as a fitness metric [5], fuzzy logic systems using the Gaussian membership functions (GMFs) for the shape, size and motion variation of a fire from successive frames, along with the calculation of the distance between the camera and the fire region [6], and robotic systems with unmanned aerial vehicles (UAVs) with color decision rule to extract fire-colored pixels [7,8]. Image classification/ detection in fire detection can take place using traditional methods (where the selection of the most appropriate and representative features is an open and challenging issue [9]) and or neural networks. The former employs the features, such as color, texture, and shape of smoke and fire. The latter involve neural networks and specifically convolutional as a class of neural networks can process grid-based data and achieve high classification accuracy [10–12]. Zhang et al. [13] have presented a deep learning-based detection system for forest fires, using a full-image and fine-grained patch fire classifier. Mahmoud and Ren [14] have presented a forest fire detection system that applies a rule-based image processing technique and temporal variation. Muhammad et al. [15] have presented a deep CNN fire detection system applied upon video. Avazov et al. [16] developed a fire detector that accurately detects even small sparks and sounds an alarm within 8 s of a fire outbreak. A novel convolutional neural network was developed to detect fire regions using an enhanced You Only Look Once (YOLO) v4network. Hu et al. [17] propose multi-oriented detection based on a value conversion-attention mechanism module and Mixed-NMS (MVMNet). Xu et al. [18] present a lightweight fire-detection algorithm, Light-YOLOv5 (You Only Look Once) achieving detection speed of 91.1 fps., Zhang et al. [19] has designed a neural network architecture for forest fire detection and recognition based on Attention U-Net and SqueezeNet (ATT Squeeze U-Net), performing segmentation to extract the shape of forest fire, and classification to verify the detected fire area. Wang et al. [20] perform forest fire detection framework using superpixel-based suspicious fire region proposal and light-weight convolutional neural network and specifically the Expanded Squeeze-and-Excitation ShuffleNet (ESE-ShuffleNet). The usage of CNNs appears as a common denominator of recent solutions, with the computational cost rationalization coming into play. Aspects related to complex and dynamically changing backgrounds create challenges, especially due to the fact that the datasets typically focus on specific settings and backgrounds.

### *1.2. Structure of the paper*

The structure of the paper is as follows: Section 2 includes the methodological aspects, and specifically the selection of the CNNs and the application of transfer learning to cover the constraints applied. It also discusses the identified datasets, their characteristics, the consideration of noise and the mechanisms to extend them. Section 3 describes the implementation, including the preparation of the CNNs through transfer learning and the classification of images towards fire detection. Section 4 presents and discusses the results and, Section 5 includes the conclusion and the future work.

## 2. Methodology

### *2.1. Selection of CNNs and transfer learning*

CNN architecture is defined through the number of layers and their connections. The selection of CNNs is a challenging problem, depending on the case study, the requirements and constraints. In our case the key requirements the need for high classification accuracy, the generalization capabilities upon new data (due to the dynamically changing geographical context) and the constrained computational and telecommunications resources. In this view, we have selected four CNNs, namely a) SqueezeNet [21], as a simplified version of AlexNet, which achieves similar accuracy with 50x fewer parameters in classification tasks, b) ShuffleNet [22], which operates in constrained environments using techniques of pointwise group convolution and channel shuffle, c) MobileNet [23], which can be applied to mobile devices using depth-wise separable convolutions and d) ResNet-50 [24], as a representative of larger CNNs. Table 1 describes the characteristics of the selected CNNs.

**Table 1**. Characteristics of the selected CNNs.

| CNN | SqueezeNet | ShuffleNet | MobileNet_v2 | ResNet-50 |
|---|---|---|---|---|
| Depth | 18 | 50 | 53 | 50 |
| Parameters (millions) | 1.24 | 1.4 | 3.5 | 25.6 |

Considering limitations in the number and quality of labeled data, a typical bottleneck in supervised learning [25], we have applied transfer learning for fire detection and consequent semantic segmentation for object detection (contextualization). The selected CNNs SqueezeNet, ShuffleNet, MobileNet and ResNet-50 are trained on ImageNet[1] and they are retrained with the specific objective of fire identification. During retraining the training loss (the negative log likelihood, NLL) and its gradients per model parameter are calculated (backpropagation) and used to update the parameters with the optimizer. The retraining parameters is an open question, including the number of layers to be retrained, the optimizer, the number of epochs and the learning rate.

To make the framework adaptable to vertical application, such as the protection of energy infrastructure, we need to identify related objects such as buildings, roads, electricity transmission lines. Object detections cannot be continuously applied, rather in case fire identification is positive. Faster R-CNN, You Only Look Once (YOLO) v2-4, and Single Shot Detection (SSD) networks are focusing on semantic segmentation. We have used Deeplab v3+ [26] network pretrained on the CamVid database [27] and selected a set of interesting objects.

[1] http://www.image-net.org

### *2.2. Datasets*

Datasets differ regarding the type and number of items they contain, the origin of the images, the bias, the homogeneity (if for example the material has been created for this purpose or they have been queried from existing image databases), the technical characteristics, the curation and licensing aspects. Our research for reusable datasets related to wildfire identification has yielded relatively limited results, as in the existing datasets the visualization of large-scale fires which is straightforward to identify and additionally it is not appropriate for training in an early stage of the fire. For example, using FLAME[2] due to its specific background and clear fire visualization the accuracy can be maximized. In this view, the datasets considered include the following: (a) **Forest-Fire dataset** [28]: with 3-channeled images anonymized from existing databases with 1900 images and two classes (fire, no-fire), (b) **Fire-Flame**[3] dataset of 3 classes (fire, smoke, neutral) and 3000 images (for comparison reasons we excluded the "smoke" class), (c) **Unknown Images**: a set of images that have not been used in the networks' training procedure and have been collected from various publicly available sources, with 100 images in each class. Table 2 presents the characteristics of the selected datasets.

**Table 2**. The datasets.

| Dataset | Forest-Fire | Fire-Flame | Unknown Images |
|---|---|---|---|
| Classes | Fire – No Fire | Fire – No Fire | Fire – No Fire |
| Means | Terrestrial and aerial | Terrestrial and aerial | Terrestrial and aerial |
| Location | Forest | Everywhere | Everywhere |
| Resolution | 250 × 250 pixels | Not standard | Not standard |
| View | Front and top | Front | Front and top |
| Number of images | 1900 | 2000 | 200 |
| Distribution among classes | Uniform | Uniform | Uniform |

Sample images from each dataset are shown in Fig. 1.


**Fig. 1.** Sample images from the three datasets. (All images are publicly available).

[2] https://ieee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs
[3] https://github.com/DeepQuestAI/Fire-Smoke-Dataset

## 3. Implementation

Based upon the two selected datasets (i.e. Forest-Fire and Fire-Flame) and using the pre-trained neural networks mentioned in Section 2.1 (i.e. SqueezeNet, ShuffleNet, MobileNet_v2 and ResNet-50) we perform experiments based on transfer learning. Each set is divided into training set and testing set of 80% and 20% respectively. In the first experiment we investigate the retraining options evaluating the classification accuracy. After determining the optimal retraining parameters and the maximum of the classification accuracy, the images are subjected to Gaussian and Salt & Pepper noise at two different levels, i.e. medium and high. This is to determine which network is more robust against noise and which noise has the most destructive effects.

To approach realistic real-time scenarios, we also perform cross-dataset evaluations: the retrained networks are tested on a set of images that have not taken part in the training of the networks. The out-of-training-domain dataset (i.e. the "Unknown Images") is used as the test set and contains publicly available images from other datasets as described in Section 2.2. In the images of the last set, fire occupies a different part of the image (including also the more challenging early stage). Also, these images are examined after being subjected to noise. In parallel, the images are analyzed with semantic segmentation by associating each pixel of the image with an object category and in this way areas of interest (such as roads, buildings, people, vehicles, etc.) are detected in the images.

The workflow of the procedure is described in Fig. 2.

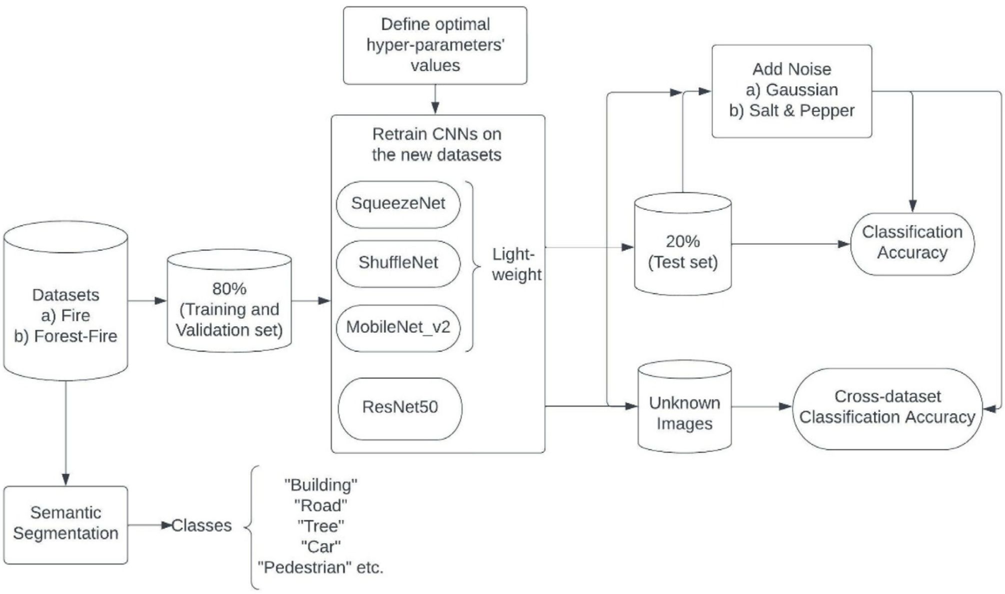


**Fig. 2.** Workflow scheme.

### 3.1. Training options

Determining the values of the transfer learning hyper-parameters is related to the adjustment of the weights using back-propagation to maximize classification accuracy and minimize the loss function. We have considered the Stochastic Gradient Descent with Momentum (SGDM) [29] and the Adaptive Moment Estimation (Adam) [30] with the former achieving superior results. Next, we investigate four different values of the mini-batch size (50, 100, 200 and 300) which is the divisor of the training set number of files, and the quotient that occurs is the number of iterations which are processed in each epoch in order to adjust the weights. The maximum number of epochs is the full passes of the training set, while validation patience is the number of times after which the loss stops decreasing. The learning rate refers to the size of corrective steps in the procedure of updating the weights. The combinations resulting from the two optimizers and the four values of mini-batch size (8 in number) were iteratively tested, while the maximum number of epochs was set to a flexible value which is combined with the value of validation patience.

The learning rate was also set to a low value so that small steps of updating the weights are made which delays the training but leads to more optimal weight values.

Specifically, the values of the hyper-parameters have been set as follows: Optimizer: SGDM, Mini-batch size: 100, Maximum Epochs: 10 (in all cases the training procedure stopped before the 10th epoch), Validation Patience: 2, Learning Rate: $2 \times 10^{-3}$. Furthermore, we shuffled the images in each epoch so that the update of the weights considers different set of images and augmentation operations (reflections, stretches and translations) to avoid overfitting.

### 3.2. Noise

Air and terrestrial monitoring images can be affected by noise, during capturing (e.g. due to aperture used and capturing parameters such as shutter speed), the (lossy) compression and transmission. This means that fire identification may have to take place in low quality or noisy images. Types of noise include Gaussian [31] and impulse [32] noise, missing image samples, packet loss in image transmission, and tampered images. Such noise can be easily ignored by the human optical system, but it can affect the performance of CNN during classification. Considering that when the models classify data of the same quality as the training data they achieve the highest accuracy, we decided to test the models on noisy images (while trained on clean ones) in order to estimate the lower boundary of the classification accuracy. The images have been subjected to Gaussian and Salt & Pepper noises by adjusting their parameters to obtain the same peak signal-to-noise ratio (PSNR). We consider two noise levels, specifically medium with PSNR set to 15 dB, and high with PSNR set to 10 dB. The effect of the two levels of noise on the images is shown in Fig. 3.

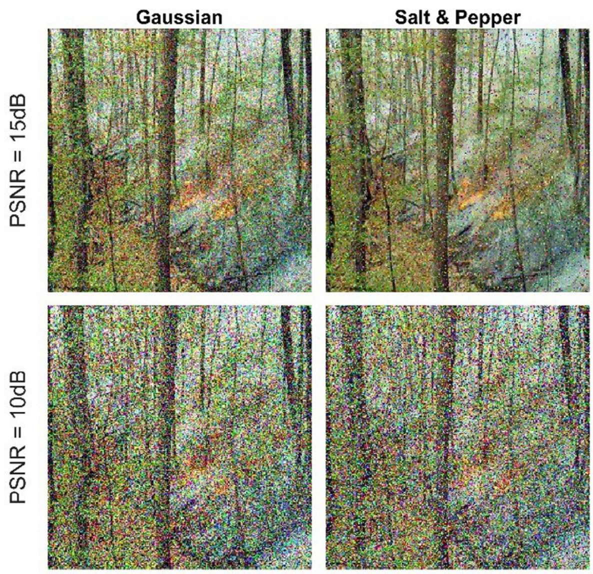


**Fig. 3.** Visualization of the effect of the two noise levels on image.

### 3.3. Semantic segmentation

The semantic segmentation is an additional process which, in combination with the detection of the fire, supports the determination of the level of risk depending on the area where it is located. The detection of interesting objects can give an indication on the type of the area, such as residential, forest, or affecting infrastructures (e.g. wind farm, photovoltaic or energy transmission lines). We have employed the ResNet-18 based Deeplab v3+ network (i.e. its initial weights are identical with those of ResNet-18) which is trained on the CamVid dataset. This dataset consists of street-level views of images with pixel-level labels of 32 classes including building, car, pedestrian, tree etc. According to the number of pixels that include the points of interest, the corresponding conclusions can be drawn.

Fig. 4 shows an example of semantic segmentation on a successfully classified image. Although this image does not depict a forest fire, it was chosen to highlight the function of semantic segmentation since it contains *objects* of interest.


**Fig. 4.** Semantic segmentation in an image with fire.

According to the color bar on the side we see that the pedestrian, the building, the electricity pylon, the rode and the plants are successfully detected. These classes can determine the level of immediacy of intervention and also the way of this (since there is a road).

## 4. Results and discussion

The classification accuracy and the training time for each dataset and CNN are shown in Table 3. The hyper-parameters are set to their optimal values and in each case, the training and the testing are performed on subsets of the same dataset. All four CNNs achieve high accuracy rates exceeding 95%. The training time of each network depends on its depth and, on the number of images used in the training process. The largest (deepest) network, MobileNet_v2, takes the longest training time but in exchange, it yields the highest classification accuracy for both datasets. Considering that the training procedure happens once, so far MobileNet_v2 is the dominant CNN. Regarding the two datasets, the classification results are quite similar with slightly better results being obtained for the Forest-Fire dataset.

**Table 3**. Classification accuracy and training time per dataset and CNN with optimal training options.

| Dataset | SqueezeNet | | ShuffleNet | | MobileNet_v2 | | ResNet-50 | |
|---|---|---|---|---|---|---|---|---|
| | Classification accuracy (%) | Training time (s) | Classification accuracy (%) | Training time (s) | Classification accuracy (%) | Training time (s) | Classification accuracy (%) | Training time (s) |
| Forest-Fire | 97.11 | 45 | 97.89 | 68 | 98.95 | 151 | 97.63 | 166 |
| Fire-Flame | 95.00 | 77 | 96.00 | 90 | 97.50 | 164 | 96.00 | 175 |

In the case of mildly noisy images, Table 4 shows the results of classification accuracy with Gaussian and Salt & Pepper noise (in the parentheses are the drops in classification accuracy compared to the values in Table 3). For comparison reasons, we have set both types of noise to have an average PSNR of approximately 15 dB.

**Table 4**. Classification accuracy per dataset and CNN for noised test images with PSNR = 15 dB.

| (PSNR = 15 dB) | SqueezeNet | | ShuffleNet | | MobileNet_v2 | | ResNet-50 | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper |
| Forest-Fire | 76.58 (−20.53) | 86.05 (−11.06) | 80.53 (−17.36) | 87.63 (−10.26) | 67.37 (−31.58) | 85.26 (−13.69) | 91.20 (−6.43) | 94.58 (−3.05) |
| Fire-Flame | 87.00 (−8.00) | 87.00 (−8.00) | 90.00 (−6.00) | 91.00 (−5.00) | 82.50 (−15.00) | 89.50 (−8.00) | 94.00 (−2.00) | 94.50 (−1.50) |

Regarding the two noises, Gaussian noise causes more destructive results in classification accuracy than Salt & Pepper noise. Concerning the two datasets, the Forest-Fire has an average decrease of classification accuracy around 19% with Gaussian noise and 9.7% with the Salt & Pepper. The Fire dataset appears to be less affected with an average decrease of classification accuracy about 7.8% and 5.4% with Gaussian and Salt & Pepper noise respectively. As for the CNNs, ResNet-50 is more the robust to noise with an average decrease in classification accuracy around 3.2%, followed by ShuffleNet with an average decrease of 9.7%, while the most vulnerable network to appear MobileNet_v2 with an average decrease of 17.1%. So, taking into account the comparable results of Table 3 but also the noise tolerance, ResNet-50 and ShuffleNet seem to be the best choices.

The results of classification accuracy for high level of noise, (PSNR of 10 dB) are shown in Table 5. For the light-weight CNNs the classification accuracy drops rapidly and fluctuates around 50%. ResNet-50 results are slightly better for the case of Forest-Fire dataset, while in the Fire-Flame dataset Gaussian noise has less impact upon the classification accuracy (80.8%) than Salt & Pepper (72%).

**Table 5**. Classification accuracy per dataset and CNN for noised test images with PSNR = 10 dB.

| (PSNR = 10 dB) | SqueezeNet | | ShuffleNet | | MobileNet_v2 | | ResNet-50 | |
|---|---|---|---|---|---|---|---|---|
| Dataset | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper | Gaussian | Salt & Pepper |
| Forest-Fire | 50.00 (−47.11) | 51.50 (−45.61) | 50.26 (−47.63) | 53.42 (−44.47) | 50.00 (−48.95) | 50.00 (−48.95) | 63.2 (−34.43) | 54.1 (−45.53) |
| Fire-Flame | 49.50 (−8.00) | 55.50 (−39.5) | 57.00 (−6.00) | 72.50 (−23.50) | 49.00 (−15.00) | 49.00 (−48.50) | 80.8 (−15.20) | 72.00 (−24.00) |

The retrained CNNs are evaluated on images out of the training domain (cross-dataset). The “Unknown Images” set is used first with clear images and then with highly noised (i.e. setting PSNR equal to 10 dB). In this way we examine a) whether ShuffleNet remains the dominant choice among the light-weight CNNs and b) whether either dataset is more suitable for training (i.e. contains more strongly differentiated images). The results are presented in Table 6.

**Table 6**. Cross-dataset classification results on “Unknown Images” for clear and noised images.

| | Training dataset | Forest-Fire | Fire-Flame |
|---|---|---|---|
| CNN | Test set | Unknown Images | Unknown Images |
| SqueezeNet | Clear Images | 85.50 | 85.50 |
| | Gaussian | 77.00 | 72.00 |
| | Salt & Pepper | 67.00 | 61.00 |
| ShuffleNet | Clear Images | 90.00 | 90.00 |
| | Gaussian | 71.50 | 71.00 |
| | Salt & Pepper | 65.50 | 65.00 |
| MobileNet_v2 | Clear Images | 86.50 | 86.00 |
| | Gaussian | 71.50 | 71.00 |
| | Salt & Pepper | 65.50 | 65.00 |
| ResNet-50 | Clear Images | 85.50 | 85.50 |
| | Gaussian | 82.50 | 79.00 |
| | Salt & Pepper | 73.00 | 71.00 |

According to the values of Table 6, ShuffleNet yields the highest classification performance for the cross-dataset test. It is also confirmed that ResNet-50 handles the effect of noises on images better. Regarding which set is more suitable for training, the results are comparable with slightly increased percentages in the case of networks trained on the Forest dataset.

## 5. Conclusions and future work

Wildfires represent a significant risk factor of environment debasement and they have impact upon human lives and activities. The main objective is the prevention and the early detection of fires, and to this end progress has been made in terms of a) electro-mechanical means able to retrieve images from geographically remote settings (such as UAVs) and b) of machine learning (ML) (especially CNN-based) means able to process figures and extract interesting information. These perspectives can be combined and offer near real time capabilities, with the constraint of limited computational resources.

To this end we have designed a CNN-based system able to process images in order to identify fires and extract context-based information through semantic segmentations. As the key constraint is the limitation of computational resources, we have selected three lightweight CNNs (ShuffleNet, SqueezeNet, and MobileNet_v2) and a larger one (ResNet-50) for comparison purposes. In principle the CNNs can achieve high accuracy even under normal conditions (i.e. using images of acceptable quality, without noise), with classification accuracy of more than 95%.

As a key particularity of the application of ML-based solution has been the limited availability of reusable datasets related to wildfires, we have performed a series of tests involving noisy images (at two levels of PSNR) as well as cross-dataset scenarios. For mild noise of up to 15 dB, light-weight CNNs (and especially the ShufleNet) appear preferable as they achieve high classification accuracy and do not require a lot of computing resources. If the noise is of a higher level and assuming that denoising mechanisms may improve the quality of the image for human perception but the noise still affects the ML algorithms, larger networks (in terms of number of parameters) have to be used. We have also observed that when the noise is at a moderate level the Gaussian causes affects more the classification accuracy, while when the noise level increases to high levels the Salt & Pepper is proved to be more destructive. In terms of the semantic segmentation, it has been verified that objects related to infrastructures (e.g power generating plants, electrical connectors, posts and transformers for the supply of electricity and energy), which are of interest in this work, exist in the main object detection sets (CamVid and COCO[4]) and the usage of CNNs trained for object detection (namely based upon ResNet-18) has yielded acceptable results.

As future work, technologically, we will continue the experimentation with and evaluation of additional CNNs in the area of fire identification. We will also work on semantic segmentation, with the aim of including further arbitrary objects, not necessarily included in pre-existing datasets. Furthermore, the proposed system can be integrated with a more *holistic* framework of fire prevention, detection, and extinction decisions. Such an example is presented in [33] who propose customized risk levels global danger of a fire and over a limited area of the region. Another interesting perspective is related to the prediction of the 'extreme' events [34]. Considering that the forest wildfires are such *extreme* events. i.e., they occur rarely and arise from seemingly normal conditions, an ambitious goal would be a prediction model for possible wildfires.

## CRediT authorship contribution statement

**Eleni Tsalera:** Conceptualization, Methodology, Algorithm development, Experiment, Results analysis, Writing – original draft. **Andreas Papadakis:** Conceptualization, Methodology, Supervision, Results analysis, Writing – original draft, Writing – review & editing. **Ioannis Voyiatzis:** Validation, Writing – review & editing, Formal analysis, Supervision. **Maria Samarakou:** Conceptualization, Supervision, Literature review, Writing – original draft, Writing – review & editing, Validation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

[4] https://cocodataset.org/

## Acknowledgments

The authors acknowledge financial support for the dissemination of this work from the Special Account for Research Grants, University of West Attica and the Special Account for Research of ASPETE, Greece through the funding program 'Strengthening ASPETE's research.

## References

[1] Jesus San-Miguel-Ayanz, Tracy Durrant, Roberto Boca, Pieralberto Maianti, Giorgio Liberta', Tomas Artes Vivancos, et al., Forest fires in Europe, Middle East and North Africa 2021, Publications Office of the European Union, Luxembourg, ISBN: 978-92-76-58616-6, 2022, 2022.
[2] Ahmad Alkhatib, A review on forest fire detection techniques, Int J Distrib Sens Netw 10 (3) (2014) 2014.
[3] Thierry Bouwmans, Javed Sajid, Maryam Sultana, S. Ki Jung, Deep neural network concepts for background subtraction: A systematic review and comparative evaluation, Neural Netw 117 (2019) (2019) 8–66.
[4] Qi-xing Zhang, Gao-hua Lin, Yong-ming Zhang, Gao Xu, Jin-jun Wang, Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images, Procedia Eng 211 (2018) (2018) 441–446.
[5] Amin Khatami, Saeed Mirghasemi, Abbas Khosravi, C. Peng Lim, Saeid Nahavandi, A new PSO-based approach to fire flame detection using K-medoids clustering, Expert Syst Appl 68 (2017) (2017) 69–80.
[6] ByoungChul Ko, Jin-Hun Jung, Nam Jae-Yeal, Fire detection and 3D surface reconstruction based on stereoscopic pictures and probabilistic fuzzy logic, Fire Saf J 68 (2014) (2014) 61–70.
[7] Chi Yuan, Zhixiang Liu, Zhang Youmin, Vision-based forest fire detection in aerial images for firefighting using UAVs, in: Proceedings of international conference on unmanned aircraft systems, 2016, pp. 1200–1205, 2016.
[8] Yi Zhao, Jiale Ma, Xiaohui Li, Jie Zhang, Saliency detection and deep learning-based wildfire identification in UAV imagery, Sensors 18 (3) (2018) 712, 2018.
[9] Eleni Tsalera, Andreas Papadakis, Maria Samarakou, Novel principal component analysis-based feature selection mechanism for classroom sound classification, Comput Intell 37 (4) (2021) 1827–1843.
[10] Anabel Gómez-Ríos, Siham Tabik, Julian Luengo, A.S.M. Shihavuddin, Bartosz Krawczyk, Francisco Herrera, Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation, Expert Syst Appl 118 (2019) 315–328, 2019.
[11] Eleni Tsalera, Andreas Papadakis, Maria Samarakou, Ioannis Voyiatzis, Feature extraction with handcrafted methods and convolutional neural networks for facial emotion recognition, Appl Sci 12 (17) (2022) 8455.
[12] Fengju Bu, M. Samadi Gharajeh, Intelligent and vision-based fire detection systems: A survey, Image Vis Comput J 91 (2019) 2019.
[13] Qingjie Zhang, Jiaolong Xu, Liang Xu, Haifeng Guo, Deep convolutional neural networks for forest fire detection, in: Proceedings of the international forum on management, education and information technology application, 2016, pp. 568–575, 2016.
[14] A.I. Mubarak Mahmoud, Honge Ren, Forest fire detection using a rule-based image processing algorithm and temporal variation, Math Probl Eng (2018).
[15] Khan Muhammad, Jamil Ahmad, Zhihan Lv, Paolo Bellavista, Po Yang, Sung Wook Baik, Efficient deep CNN-based fire detection and localization in video surveillance applications, IEEE Trans Syst, Man, Cybern: Syst 49 (7) (2018) 1419–1434, 2018.
[16] Kuldoshbay Avazov, Mukhriddin Mukhiddinov, Fazliddin Makhmudov, Young Im Cho, Fire detection method in smart city environments using a deep-learning-based approach, Electronics 11 (1) (2022) 73.
[17] Yaowen Hu, Jialei Zhan, Guoxiong Zhou, Aibin Chen, Weiwei Cai, Kun Guo, et al., Fast forest fire smoke detection using MVMNet, Knowl-Based Syst 241 (2022) 2022.
[18] Hao Xu, Bo Li, Fei Zhong, Light-YOLOv5: A lightweight algorithm for improved YOLOv5 in complex fire scenarios, Appl Sci 12 (23) (2022) 2022.
[19] Jianmei Zhang, Hongqing Zhu, Pengyu Wang, Xiaofeng Ling, ATT squeeze U-net: A lightweight network for forest fire detection and recognition, IEEE Access 9 (2021) (2021) 10858–10870.
[20] Pengyu Wang, Jianmei Zhang, Hongqing Zhu, Fire detection in video surveillance using superpixel-based region proposal and ESE-ShuffleNet, Multimedia Tools Appl (2021) (2021) 1–28.
[21] N. Forrest Iandola, Song Han, W. Moskewicz Moskewicz, Khalid Ashraf, J. William Dally, Kurt Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size, 2016, 2016.
[22] Ninging Ma, Xiangyu Zhang, Hai-Tao Zheng, Jian Sun, ShuffleNet V2: Practical guidelines for efficient CNN architecture design, Proc Eur Conf Comput Vis (2018) (2018) 116–131.
[23] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, MobileNetV2: Inverted residuals and linear bottlenecks, in: Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition, 2018, pp. 4510–4520, 2018.
[24] Kaiming He, Zhang Xizngyu, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778, 2016.
[25] Alexander Kensert, J. Philip Harrison, Ola Spjuth, Transfer learning with deep convolutional neural networks for classifying cellular morphological changes, SLAS Discov: Adv Life Sci R & D 24 (4) (2019) 466–475.
[26] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Adam Hartwig, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision, 2018, pp. 801–818, 2018.

[27] J. Gabriel Brostow, Julien Fauqueur, Roberto Cipolla, Semantic object classes in video: A high-definition ground truth database, Pattern Recognit Lett 30 (2) (2009) 88–97.
[28] Ali Khan, Bilal Hassan, Somaiya Khan, Ramsha Ahmed, Abuassba Adnan, DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection, Mob Inf Syst (2022).
[29] Ali Ramezani-Kebrya, Ashish Khisti, Ben Liang, On the generalization of stochastic gradient descent with momentum, 2021, arXiv preprint arXiv:1809.04564, 2018.
[30] S. Nitish Keskar, Richard Socher, Improving generalization performance by switching from adam to sgd, 2017, arXiv:1712.07628, 2017.
[31] C. Sandeep Kumain, Maheep Singh, Navjot Singh, Krishan Kumar, An efficient Gaussian noise reduction technique for noisy images using optimized filter approach, in: Proceedings of the first international conference on secure cyber computing and communication, IEEE, 2018, pp. 243–248, 2018.
[32] Bo Fu, Xizoyang Zhao, Chuanming Song, Ximing Li, Xianghai Wang, A salt and pepper noise image denoising method based on the generative classification, Multimedia Tools Appl 78 (2019) (2019) 12043–12053.
[33] Manuel Casal-Guisande, Jose-Benito Bouza-Rodríguez, Jorge Cerqueiro-Pequeño, Alberto Comesaña-Campos, Design and conceptual development of a novel hybrid intelligent decision support system applied towards the prevention and early detection of forest fires, Forests 14 (172) (2023) 2023.
[34] Ethan Pickering, Stephen Guth, E. George Karniadakis, P. Themistoklis Sapsis, Discovering and forecasting extreme events via active learning in neural operators, Nature Comput Sci 2 (2022) (2022) 823–833.