Merve TAŞ

A Ph.D. Thesis

AGU 2022

EARLY DETECTION OF FOREST FIRE FROM VIDEO UTILIZING TEMPORAL INFORMATION

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Merve Taş
October 2022

EARLY DETECTION OF FOREST FIRE FROM VIDEO UTILIZING TEMPORAL INFORMATION

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Merve Taş
October 2022

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Merve Taş
Signature :

REGULATORY COMPLIANCE

Ph.D. thesis titled “Early Detection of Forest Fire From Video Utilizing Temporal Information” has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor	Co-Advisor
Merve TAŞ	Assist. Prof. Kasım TAŞDEMİR	Assoc. Prof. Zafer AYDIN

Head of the Electrical and Computer Engineering Program

Assoc. Prof. Zafer AYDIN

ACCEPTANCE AND APPROVAL

Ph.D. thesis titled "Early Detection of Forest Fire From Video Utilizing Temporal Information" and prepared by Merve TAŞ has been accepted by the jury in the Electrical and Computer Engineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

/ / 2022

JURY:

Advisor : Assist. Prof. Kasım TAŞDEMİR

Member : Prof. Behçet Uğur TÖREYİN

Member : Assist. Prof. Mustafa COŞKUN

Member : Assist. Prof. Gülay YALÇIN ALKAN

Member : Assist. Prof. Nuh AZGINOĞLU

APPROVAL:

The acceptance of this Ph.D. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated ….. /….. / 2022 and numbered

……….. /……….. / ………..

Graduate School Dean
Prof. Dr. İrfan ALAN

ABSTRACT

EARLY DETECTION OF FOREST FIRE FROM VIDEO UTILIZING TEMPORAL INFORMATION

Merve TAŞ
Ph.D. in Electrical and Computer Engineering
Advisor: Assistant Professor Kasım TAŞDEMİR
Co-Advisor: Assoc. Prof. Zafer AYDIN

October 2022

Forest fires are considered as the major threats to lives, properties and to the integrity of the ecosystem around the world. In most cases, the fire damage can be reduced, when the initial signs of the fire are detected in a timely manner. Since smoke is considered as the first visual sign of fire, detection of smoke is vital. Hence, a successfully designed smoke detection system is essentially critical in the early detection of smoke for outdoor environments. The existing smoke detection methods suffer from high false alarm rates and cannot accurately detect smoke in hazy environments.

To address these problems, this thesis is focused on smoke detection model at an early stage that utilizes deep learning (DL) based techniques for outdoor locations. This work contributes mainly to four aspects of smoke detection: (1) new datasets preparation for three smoke detection tasks classification, detection-segmentation, and video classification, (2) utilizing transfer learning to detect the smoke on the relatively small dataset, (3) image dehazing process that includes removing the haze from the dataset images to enhance the system performance, (4) designing a novel hybrid video classification model by combining the two DL based video classification structures.

This work will be a resourceful reference for researchers working in the fields of forest fire or smoke detection studies at an early stage. The experiments, research findings, and enhanced performance of the smoke detection system provide a source of information about smoke detection. Current studies can be utilized to further improve the design of efficient and reliable fire safety models.

Keywords: Deep Learning, Spatio-Temporal Information, Forest Fire Early Detection, Smoke Detection, Image Dehazing.

ÖZET

ZAMANSAL BİLGİDEN FAYDALANARAK VİDEODAN ORMAN YANGINLARININ ERKEN TESPİTİ

Merve TAŞ
Elektrik ve Bilgisayar Mühendisliği Anabilim Dalı Doktora
Tez Yöneticisi: Dr. Öğr. Üyesi Kasım TAŞDEMİR
İkinci Tez Yöneticisi: Doç. Dr. Zafer AYDIN

Ekim 2022

Orman yangınları, tüm dünyada yaşamlara, mülklere ve ekosistem bütünlüğüne en büyük tehdit olarak kabul edilmektedir. Orman yangınlarının erken tespiti ile yangının yol açacağı hasarlar azaltılabilir. Duman, yangınların ilk görsel işareti olduğundan, dumanın tespiti oldukça önemlidir. Başarılı şekilde tasarlanmış bir duman algılama sistemi, dış ortamlarda dumanın erken tespitinde kritik öneme sahiptir. Mevcut duman algılama yöntemleri yüksek yanlış alarm problemi ile karşılaşmaktadır ve puslu ortamlarda duman tespiti konusunda tam olarak başarılı değildir.

Bu tez, orman yangınlarının erken aşamada tespitindeki problemleri çözmek için derin öğrenme tabanlı yöntemlerin kullanılmasını önermektedir. Bu çalışma, dumanın tespiti için dört farklı öneri sunmaktadır. (1) Dumanın görüntüler üzerinde tespit edilebilmesinde kullanılan duman sınıflandırması, tam olarak yerinin belirlenmesi ve videodan dumanın tespiti gibi yöntemlerde kullanılmak üzere üç farklı veri setinin hazırlanması. (2) Nispeten daha küçük veri setleri için öğrenme aktarımı yönteminin kullanılması. (3) Sistem performansını artırmak için veri seti görüntülerinden bulanıklığın kaldırılması. (4) Derin öğrenme tabanlı iki farklı yapının kullanılarak hibrit bir video sınıflandırma modelinin tasarlanması.

Bu çalışma, erken aşamada orman yangını veya duman algılama çalışmaları alanlarında çalışan araştırmacılar için kaynak niteliğinde olacaktır.

Anahtar kelimeler: Derin Öğrenme, Mekan-Zamansal Bilgi, Erken Aşama Orman Yangınları Tespiti, Duman Tespiti, Görüntü Bulanıklığı Giderme.

Acknowledgements

I would like to thank to my respectable adviser, Kasım Taşdemir, for his support, valuable advice and guidance during my PhD journey. I am grateful that he has provided me constant motivation thought my studies.

I would like to special thanks to my second adviser, Zafer Aydın, for his patience, great support, valuable ideas and suggestions. I grateful to him for all his help whenever I needed.

I would like to extend my gratitude to Oğuzhan Balki, for his continuous assistance and patience. He has given substantial effort and extended his advice to all my work during my PhD. I also don't forget my uncle, Yusuf Balki; he was there whenever I needed his assistance and care, and if he was alive, he would be really proud of me as I am finishing this part of my career.

I also would like to thank to my collages, Zeynep Şenel, Hilal Hacılar, Sena Yengeç Taşdemir and Sema Keleş Çetin for their motivations and supports.

My dear parents, Balki family. I can only express my warmest feelings to my father, Dr. Nihat Balki, and my mother, Sevgi Balki, for their encouragement, unconditional support, love, and sacrifice. My father has been a truly inspiration for my life and eventually I have followed his path for my career.

My final whole gratefulness goes to my spouse, Dr. Yusuf Taş, and to my son, Hamza Eren, for their endless patience and support. The completion of this work would not be possible without their inspiration, encouragement, and understanding. Finally, special thanks to my little baby girl on the way, for being a good listener, joy, and motivation for writing this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
AVHRR	Advanced Very High Resolution Radiometer
BLSTM	Bi-Long Short-Term Memory
BPNN	Propagation neural network
CNN	Convolutional Neural Network
COCO	Common Objects in Context
DCGAN	Deep Convolutional Generative Adversial Network
DCLRN	Deep Convolutional Long-Recurrent Networks
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DNCNN	Deep Normalization and Convolutional Neural Network
FAR	False Alarm Rate
FBM	Fractional Brownian Motion
FC	Fully Connected
FCN	Fully Convolutional Network
FHE	Fully Homomorphic Encryption
FN	False Negative
FoHIS	Foggy and Hazy Image Simulator
FP	False Positive
FPN	Feature Pyramid Network
FPS	Frames Per Second
GAN	Generative Adversial Networks
GDF	General Directorate of Forestry
GMM	Gaussian Mixture Model
GOES	Geostationary Operational Environmental Satellite
GRU	Gated Recurrent Unit

IOT	Internet of Things
IOU	Intersection over Union
IP	Image Processing
LBP	Local Binary Pattern
LK-M	Lucas Kanade Method
LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
MOG	Mixture of Gaussian
ML	Machine Learning
MLP	Multi-Layer Perception
NMS	Non-Maximum Suppression
NICC	National Interagency Coordination Center
PKC	Public Key Cryptography
PKI	Public Key Infrastructure
PQC	Post-Quantum Cryptography
PRELU	Parametric Rectified Liner Unit
PSNR	Peak Signal-To-Noise Ratio
PSO	Practical Swarm Optimization
RCNN	Region based Convolutional Neural Network
RELU	Rectified Liner Unit
RESNET	Residual Neural Network
RGB	Red, Green Blue
RNN	Recurrent Neural Networks
ROI	Regions of Interest
SG	Smart Grid
SMC	Secure Multiparty Computation
SLIC	Simple Linear Iterative Clustering
SRoFS	Suspected Regions of Fire
SSD	Single Shot Detector
SSIM	Structural Index Similarity
SWIN	Shifted Window Transformer
SW-MSA	Shifted Window Multi Head Self-Attention Module
TN	True Negative

TP	True Positive
UAV	Unmanned Ariel Vehicle
VGG	Visual Geometry Group
ViT	Vision Transform
W-MSA	Window Multi Head Self-Attention Module
YOLO	You Look Only Once

To my family

Chapter 1

Introduction

Severe wildland fires are considered to be one of the major threats on human life and the environment. Industrial growth without taking any precautions for the environment give rise to increase the Earth’s temperature. Climate changes or accident by people lead to forest fire all over the world, which not only cause the economic losses and destroy the ecological balance, but also has the hazardous effects on the safety of human life. Forest fires can cause smoke pollution, release greenhouse gases and degrade the ecosystem.

According to the National Interagency Coordination Center (NICC) of USA, annually there were an average of 61,289 wildfires and consequently average of 7,4 million acres impacted between from 2012 to 2021 years in the USA [1]. Figure 1.1 illustrates annual wildfires and acres burned in USA for 30 years.

In addition, during the forest fires 6,000 structures such as residential, buildings destroyed in Canada in 2021.

The Australian 2019-2020 bushfire season was one of the most disaster effect in recent years all over the world. During the 2020 summer season, millions of hectares burned, thousands of properties damaged and countless numbers of wildlife exposed in Australian wildfires [2].

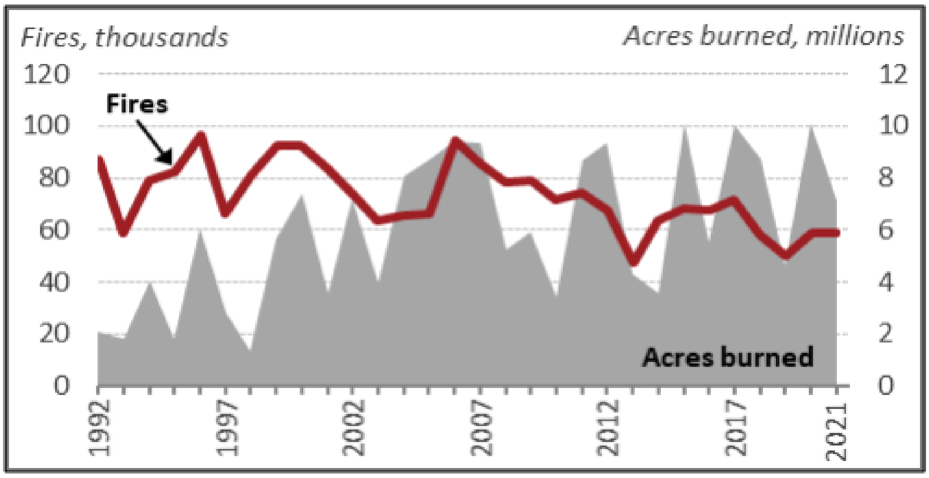
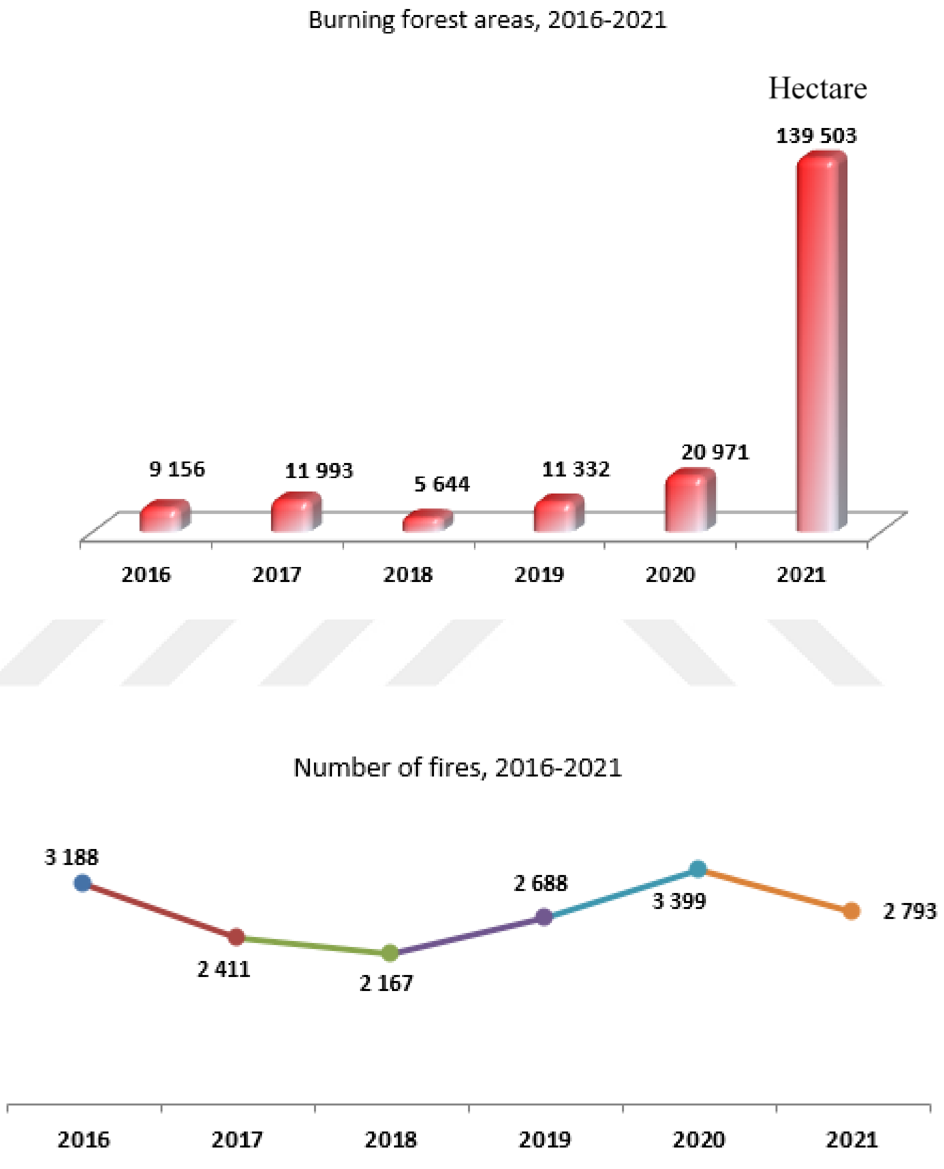


Figure 1.1. Annual wildfires and Acres Burned in USA for 30 years [1].

According to the General Directorate of Forestry (GDF) statistics, forest fire activity between the years of 2016 and 2021 in Turkey is shown in Figure 1.2. In these years, a total of 16,676 wildfires occurred, resulting in damage to 198,599 hectares [3].



Burning forest areas, 2016-2021

Number of fires, 2016-2021

Figure 1.2. Forest fire activity between the years of 2016 and 2021 in Turkey (GDF) [3].

Furthermore, there are many reasons for wildfire accidents such as negligence, intentional by the people, natural from lighting and unknown reasons. Statistics on the cause of wildfires in 2021 are shown in Figure 1.3.

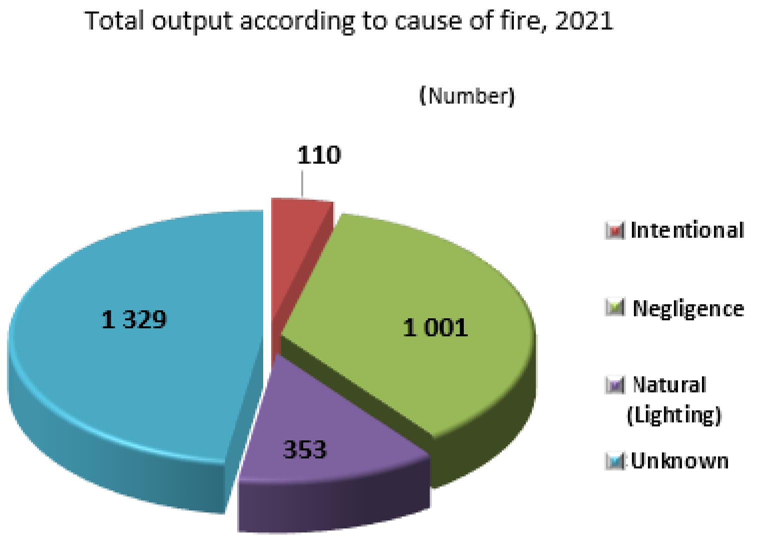


Figure 1.3. Causes of wildfires in Turkey, 2021 [3].

Ecological, biological, technical, economic, administrative, and social factors all play a role in forest fires. Forest fires are natural event which have destructive effects on the economic, social and environment. In most cases, the fire damage could have been reduced, if not fully prevented, if the initial signs of the fire were detected in a timely manner. Thus, early detection of forest fires are very crucial.

1.1 Early Detection Approaches of Forest Fires

Several traditional and sensor-based computer vision techniques are developed for forest fire prevention and minimization of its devastating effects over the years. Computer vision based systems use image and video analysis that obtained from surveillance cameras by using image processing (IP) and machine learning (ML) algorithms for early detection of forest fires.

Traditional smoke detection sensors generally use indoor such as residential buildings, commercial complexes, industrial locations, critical care areas. These systems detect the smoke and fire by sensing the smoke/fire particles or rising in temperature. These sensors are also cheap and simple to use [4]. For this system, location and distance of sensor is very critical to determine the smoke/fire areas. In outdoor areas, background scene changes with weather conditions and time. Moreover, the detection from video is more challenging task in outdoor environment when the presence of low illumination conditions such as fog and haze. However, traditional sensor based smoke detectors that have limited range are not applicable for large covered outdoor areas. Wildfire detection in early stage is very crucial task that generally uses cameras that located at surveillance towers at a certain height, through aerial surveys or satellite images. However, environmental conditions limit the detection performance of this vision-based systems and it leads to false alarms. Nonetheless, several algorithms have been proposed for vision-based smoke/fire detection that are feasible to solve the problems related to traditional detectors. These solutions generally use IP and ML techniques to analyze images and videos captured by monitoring systems and detect the smoke/fire in early stage [5]. The most of existing vision-based systems focus on the color, shape, motion and texture features for the smoke/fire detection techniques. However, effective feature representation is very challenging task due to smoke/fire has chaotic behavior, shapes, movement, color, texture and density in surveillance videos. Thus, these systems have high false alarm rates.

In computer vision-based smoke and fire detection systems, features are crucial. The classifiers utilize the features with the hand-crafted extraction process for the training process. Hand-crafted feature extraction methods are both expensive and time consumer. Although, deep learning (DL) methods automatically learn the features directly from input data instead of manual feature extraction methods. In last decades, deep learning methods have incredible performance in many object detection and recognition tasks [6]. In literature, many studies have been published related to fire and smoke detection method using IP, ML and DL techniques. In outdoor environments, detection of smoke that first sign of forest fires is very important task to minimize the effects of fire.

In general, image and video based smoke detection methods consist of three tasks. First of all is classification task that images are classified as smoke and non-smoke. Second is detection task which are based on localization and detection of smoke in an image or video frames. The final is segmentation task in the pixel level. All these tasks are used IP, ML and DL algorithms shown in Figure 1.4.

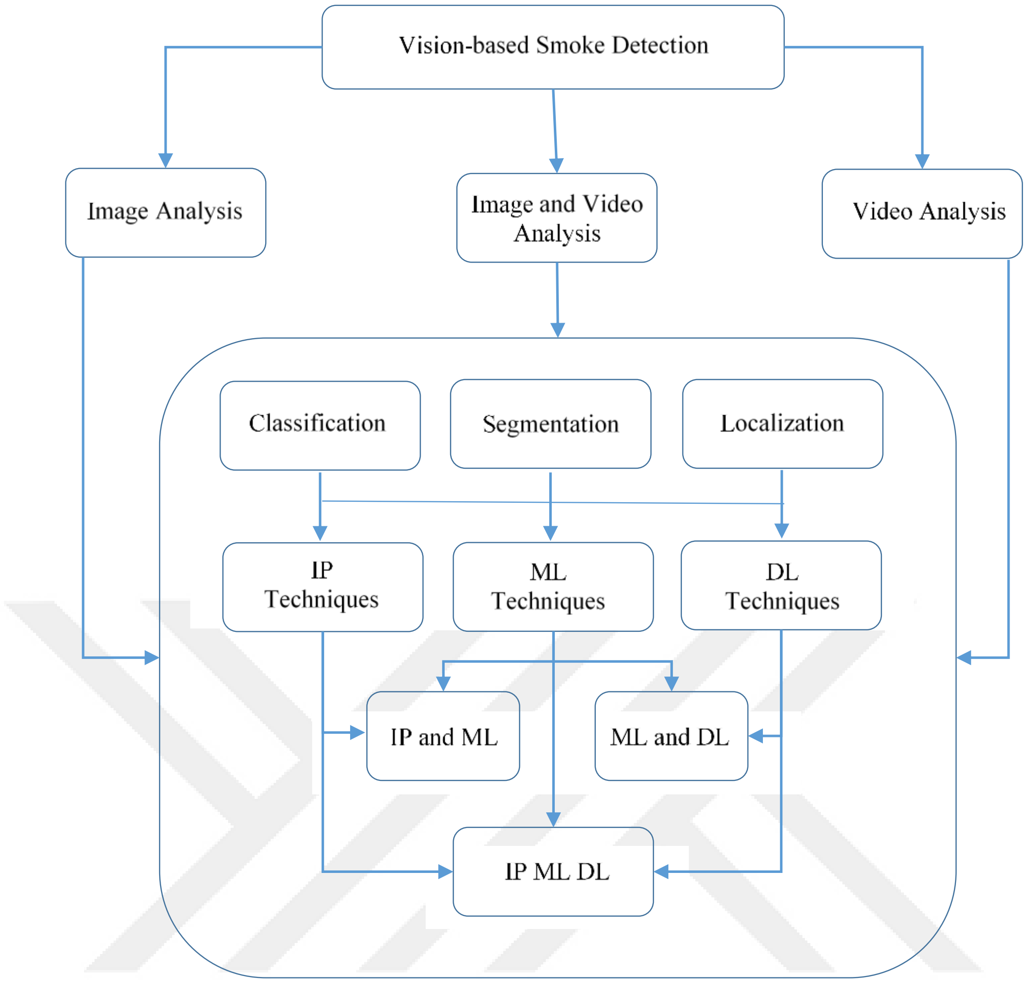


Figure 1.4. Using techniques in vision-based smoke detection [7].

Smoke classification methods mostly focus on the prediction smoke and non-smoke scenes in an image or video frames and also can be predicted fire and non-fire scenes for some cases. Smoke detection methods utilize both location of smoke and fire by using bounding boxes. Smoke segmentation is mainly focus on the images and videos from forests, city scenes [7].

1.1.1 Research Problems and Research Questions of the Thesis

Every year, many forest fires break out and threaten human lives and properties around the world. In most cases, the fire damage can be prevented, or at least reduced, if the fires are detected earlier. Therefore, developing an early fire detection system is essentially vital. Since the first visual sign of a forest fire is smoke, detecting the smoke is vital for preventing forest fire events. Therefore the detection of smoke is always the first step in fire-alarm systems.

The broad question of the thesis is, in the knowledge of the recent advancements in machine learning, can a system with higher detection accuracy be developed? However, this question begs more specific questions such as which recent methods are promising in smoke detection problem, and, which disadvantages of the current methods can be alleviated?

Based on the literature review, research consists of three research problems:

1. Even if there are some attempts to use temporal information, the literature overlooks the temporal changes in a smoke object, so to speak, behavior of a smoke object over time. Would using temporal data along with the spatial one yield any significant improvement in the accuracy?
   a. How to incorporate the temporal information?
   b. In order to use temporal information, can we employ popular deep learning methods in smoke detection problem?

   CNN based smoke detection algorithms focus on 2D images that have only spatial information. However, video sequences consist of both spatial and temporal information. There are many action recognition from videos methods to incorporate the temporal information. The main aim is to determine the convenient smoke detection from video method in terms of high accuracy and low false alarm rate.

2. How well the most recent and often mentioned Deep learning methods for segmentation tasks such as Mask RCNN, perform on smoke detection problem?

   This question addresses utilizing of the smoke segmentation architecture for detection of smoke in early stage. The aim of this study is selection of the best accurate smoke detection system.

3. How to improve the system performance the existing deep learning based smoke detection models?

   For the improving of the system performance, both fine-tuning to models and pre-processing method such as image dehazing are utilized on the dataset images.

1.1.2 Objectives

The overall aim of this thesis is to develop a smoke detection system to detect the forest fires in early stage for complex outdoor environment. For this purpose, the thesis includes the following objectives:

- Curating and labeling dataset and extraction of meaningful features for detection of smoke in early stage.
- Applying pre-processing methods to the dataset to increase the quality of the dataset images.
- Utilizing CNN based classification, detection and segmentation methods. Enhancing the DL based state-of-the-art methods by using fine-tuning methods and hyperparameter optimization and comparison of performance these methods.
- Design a video-based smoke/non-smoke classification model by using the popular DL based action recognition techniques.

1.1.3 Contributions

The methods proposed in this thesis are aimed at increasing system performance which depends on high detection accuracy and low false alarm rate. The main contributions of this study are as follows:

1. Dataset preparation for three computer vision tasks such as image classification, detection and segmentation and video classification.
2. This is the first study which attempts to eliminate fog in the images before any further smoke analysis.
   a. Empirically it is shown that this significantly reduces the error rate for both forest smoke segmentation and classification tasks.

    b. The method generalizes well to current or future smoke related computer vision tasks specifically performed in hazy environment.

3. A novel hybrid model is then proposed by combining two methods that have used for video-based classification.

1.1.4 Scope

This thesis mainly focuses on the smoke detection task by using surveillance videos and its extracted frames. Feature extraction from input data by using different DL based models is challenging task due to chaotic behavior of smoke.

The study considers investigating of fine-tuned CNN based classification methods. The aim of fine-tuning process to use pre-trained weights to speed up the training process and enhancing the accuracy on the small dataset. Also pre-processing which removes the haze from images both improves the system performance and makes it easier to detect of smoke.

Moreover, it includes dataset preparation for smoke detection and segmentation in an image. Popular DL based smoke detection and segmentation methods can be applied on the created dataset. One aim of this study is to prove usability of the dataset for the recent deep learning based models on different tasks.

In addition, this thesis aims to design a video smoke/non-smoke classification model by combining the two different video-based method on the new created video databases. The main concept is to improve the existing DL based techniques for special task.

In this thesis, overall proposed models aim to develop a fire safety model in early stage to diminish the effect of forest fires. As a result, utilizing deep learning based methods can be effective techniques for smoke classification, detection, segmentation and video classification tasks.

1.1.5 Thesis Outline

The remaining chapters of this thesis are outlined as follows.

Chapter 2 contains the literature review of this thesis. Related works for smoke and fire detection are presented in this chapter which mainly focus on machine learning and deep learning based techniques.

Chapter 3 compares the CNN based and Transformer based that uses for natural language processing smoke/non-smoke classification methods. First, the dataset preparation section is described. Then, using all models are fine-tuned by using special hyperparameters. Also, pre-processing method is applied the dataset to obtain haze-free clear images. Then, again all results are compared in terms of accuracy in this chapter.

Chapter 4 utilizes smoke detection and segmentation frameworks. First, the dataset which are selected from smoke/non-smoke dataset in Chapter 2 is created in this chapter. State-of-the art object detection methods are implemented to created dataset. Then, DL based popular image segmentation method is applied to the same dataset. Then the smoke localization results are compared in terms of accuracy. In addition, pre-processing method is employed to smoke segmentation task to investigate the effects of image dehazing.

Chapter 5 develops a hybrid video smoke classification method which are classified videos as smoke and non-smoke. First, new video classification dataset is created in this section. Dataset is arranged as sub-videos of dataset videos such as existing video databases in literature. Then created dataset is utilized with several video-based classification models. After that, the hybrid video classification method that includes combining of Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) structures is proposed for our special task. Then all results are compared to each other by using evaluation metrics.

Chapter 6 includes conclusion, societal impact of this thesis and proposals for future work.

Chapter 2

Literature Review

2.1 Traditional Smoke Detection Methods

Smoke detection fall into categories such as smoke classification, detection and segmentation and also video classification by using surveillance cameras.

The smoke classification task primarily involves labeling the images or video frames as ‘smoke’ and ‘non-smoke’. Image labels may also include other classes such as smoke with fog or smoke without fog, depending on the environmental conditions at the first time of image capturing process. The labels may contain other classes such as smoke with fog, smoke without fog, non-smoke with fog and non-smoke without fog etc.

Another category is smoke detection which is related to bounding box estimation to detect the localization of the smoke in an image. This method predicts a region of interest by using the high, center and width of smoke containing bounding boxes.

Segmentation is a method of split the pixels of an image into multiple regions to obtain useful information from the image. Pixel-by-pixel segmentation provides information about the object of interest. Segmentation can be divided into two types: semantic segmentation and instance segmentation. In the semantic segmentation, all entities in an images are counted as single entity, whereas instance segmentation regards each entities as different entity.

2.1.1 Traditional Smoke Classification Methods

Early techniques are used traditional IP algorithms such as local binary pattern (LBP) [8], Lucas Kanade method for optical flow estimation (LK-M) [9], Gaussian Mixture Model (GMM) [10], Hue, saturation, value (HSV) color model [11], image enhancement [4], image matting [12] and background subtraction methods. ML techniques also widely preferred for the smoke classification tasks. Support vector machine (SVM), k-nearest neighbors (k-NN) [13] practical swarm optimization (PSO) [14] are most commonly used in ML. Cui at al. [15] proposed a neural network classifier for smoke and non-smoke texture analysis. Texture feature extraction was carried out using wavelet packets and grey level co-occurrence matrices (GLCM), and smoke as well as non-smoke textures were distinguished using a two-layer backpropagation neural network classifier. Their analysis based on the combination of GLCM and original textures features obtained better accuracy for the classification task.

Several studies have found that LBP can effectively extract smoke textures. Yuan at al. [16] suggested a method based on LBP and LBPV histogram sequences for classification. Classification was carried out using a neural network classifier. LBP was used to extract local information, while a 3-level image pyramid was used to extract global texture information. Their approach performed well on hundreds of test images, but it suffered when tested on video frames containing unseen objects during the training process. These videos included objects such as diesel fuels combustion smoke, and non-smoke video frames of traffic on streets. Vidal-Calleja and Agammenoni et al. [17] proposed a smoke detection approach based on the bag-of-words feature. The method could detect the amount of smoke in a given image. The method achieved good performance on the test set. Ho at al. [18] introduced a SVM classifier based smoke detection system based on laser light that was tested at night-time. In the complicated case, the algorithm showed good performance. Yuan et al. [4] suggested another LBP code-based approach for texture categorization. The experiment was carried out on four datasets comprising smoke images in order to detect smoke. The approach showed highly effective performance in texture classification, with a 95.3% correct texture classification rate on the dataset provided by Brodatz et al. [19].

Smoke classification on satellite images is another field of research, and a number of studies has been done in this area. One of the difficulties in detecting smoke with satellite images is its visual similarity to clouds, haze, fog, and so on. To address this issue, many researchers have offered methods based on classical ML techniques. Li et al. [20] suggested a method to discrete smoke plumes from the background in satellite images of forest fires. In the model, first object pixels are extracted by using threshold. Then the selection of feature vectors is performed to pass into the back propagation neural network (BPNN) as input. They achieved 97.63% classification accuracy rate in their model. Li et al. [21] suggested a neural network and threshold-based strategy to classifying images from the Advanced Very High Resolution Radiometer (AVHRR) dataset, with their classification model obtain 99.6% accuracy. Tian et al. [12] suggested a smoke classification approach for images involving problematic objects like fog that have an appearance similar to smoke. The technique depended on separating smoke from the background in a single video frame. To identify sparse representations for the smoke and background components, first the authors formulated the issue of separating a frame into quasi-smoke and quasi-background components as a convex optimization problem. Then, they created a feature using the sparse coefficients for detecting smoke in video frames. In the experiments, the authors used five datasets. The datasets included smoke images with various forms of smoke, such as heavy and light smoke, as well as non-smoke images with items that looked like smoke, such as garments, clouds, fog/ haze, glass, shadow, sky, steam, vehicle body, wall, water, and so on. The method achieved 94.9% accuracy on the dataset. The approach could perform well even for images with complicated backgrounds such as cloud and fog/haze that were visually comparable to smoke. Hossain et al. [22] suggested a method for classifying smoke and flame in images. In their algorithm, an input image is first separated into blocks, and then an artificial neural network (ANN) is trained to categorize these blocks for smoke and flame. They have used color and texture combination to classify these blocks as fire/smoke/neutral. The model is built with 20 train and 5 test images gathered from the internet. Their model achieved 84.8% for average block categorization accuracy. Although their model performed texture classification, they used less images.

According to a review of these methods, neural networks and SVM have proven quite popular in smoke categorization. Various techniques for identifying smoke texture have also been developed in order to appropriately classify smoke images. Despite the fact that conventional IP and ML based solutions perform well to address the many issues, real-time smoke detection and position estimation remains a challenge. This area of research has been kept alive by a variety of difficult background conditions, clouds, fog, and other smoke-like objects, as well as the additional challenges of low illumination. With the improvements of DL, researchers have proposed various new strategies for smoke detection that take advantage of CNNs feature extraction capabilities. In the following section, we will go through some of the more traditional methods for smoke detection and segmentation.

2.1.2 Traditional Smoke Detection and Segmentation Methods

For object localization task, bounding boxes provide to locate the certain object in an image. Therefore, identifying the smoke at a certain location serve as the prevention of the disaster. Previously, conventional image processing and machine learning approaches were widely used to locate smoke in an image or video. For video smoke analysis, IP/ML techniques based on bounding boxes for smoke localization were generally proposed. For localizing of smoke in images and videos. Gagliardi et al. [23] used both IP and DL approaches. The method utilizes a Kalman filter for motion detection, color segmentation, the extraction of bounding boxes around a moving gray object, and prediction via a CNN. For this purpose training the CNN dataset is used that comprise smoke and non-smoke images. The authors used two datasets to assess the performance of their method: the Firesense dataset Dimitropoulos et al. [24] which contains videos of flame and smoke, and the Gagliardi and Saponara et al. [25] dataset, which contains 42 videos. Even in settings with moving objects like clouds, the approach looked excellent. Because a lightweight CNN is utilized, the method can also be used to detect smoke in IoT devices. Hossain et al. [26] proposed a method for localizing forest fires and smoke using aerial images based on LBP and ANN. Different environmental conditions, their technique outperformed SVM and You Only Look Once (YOLOv3). Their approach processed 19 frames per second efficiently.

Despite the dominance of conventional IP and ML approaches in the 1990s with considerable improvements in the performance of DL methods for general object identification, researchers focused on studies CNN-based bounding box detection algorithms.

Smoke segmentation is a type of detection problem. Segmentation is difficult task due to multiple problems such as identical backgrounds, hazy images, similar color and shape etc. In smoke segmentation, color and texture features plays vital role. Several researchers have utilized color enhancement and color channel analysis to segment the smoke regions. Some have concentrated on smoke motion features. Background subtraction, morphological procedures, region growth, color enhancement, and other vision-based approaches are commonly utilized by researchers. Long et al. [27] suggested a smoke segmentation approach that uses the dark channel before to estimate smoke transmission. The experimental evaluation was carried out on real-life images with heavy and light smoke. The approach could only detect gray and white smoke. The model's performance may have improved with more color range and motion feature analysis. Wang et al. [28] segmenting smoke in forest fire images utilizing region growth and Fractional Brownian Motion (FBM). FBM was used to evaluate the images, and the Hurst exponent was analyzed. Segmentation was done using the Hurst exponent threshold values, and a threshold values between 0.75 and 0.875 was found to be better; a threshold of 0.81125 was found to be the best. Their strategy outperformed edge detector operators. Li et al. [29] introduced a clustering-based statistical segmentation technique that uses both spectral and spatial information in a model and outperforms histogram-based algorithms such as k-means algorithm. The method's performance was evaluated using two images. One image shows smoke rising from a fire near Cuiaba, Brazil, while another shows smoke rising from a fire in California, USA. They used extremely few images to assess the efficacy of their method, therefore it is unclear how well the system would operate on unseen images. Xing et al. [30] also developed a color-based technique that makes use of HSV and LAB space. Locate the smoke region, they used pixel clustering with k-means. The model is constructed utilizing smoke images in an outside area, with three main image categories: human, smoke, and weed. However, their model fails to address the issue of over/under-segmentation. Xiong and Yan et al. [31] introduced a segmentation approach called Simple Linear Iterative Clustering (SLIC). Reduce the false detection rate, the method first groups the pixels based on their similarity in location and color, and then detects the boundary line between the sky and the ground, which eliminates the interfering object cloud. SVM is used to classify the super pixels. In forest scenes, the approach achieved 77% accuracy. However, the approach is biased by the daylight lighting circumstances.

2.1.3 Traditional Video Based Smoke Classification Methods

Video smoke analysis plays a critical role in smoke and fire detection as most of the surveillance systems. The motion characteristics and dynamics of smoke are essential in early fire detection. A considerable amount of work has been done in the area of video smoke analysis, and a large number of research articles reviewed in this study fall under this category.

Apart from spatial image characteristics, traditional IP approaches heavily used transform domain analysis. The wavelet transform was the most used transform domain technique. Toreyin et al. [32] suggested an approach based on the idea that as time passed, the wavelet energy of edge areas decreased. Their approach could identify smoke in real time within 10 ms of processing time per frame. Chen et al. [33] discussed two essential decision rules for smoke pixel evaluation, chromaticity and diffusion based decision procedures. Xu and Xu et al. [34] developed another method for training an ANN that uses both static and dynamic properties of smoke. Their method was effective in detecting a moving target by extracting its contour. The neural network was then taught to detect smoke. They tested 50 videos and got a hit rate of 0.94 out of them. Chunyu et al. [35] utilized a neural network classifier to identify fire smoke in real time. They used a texture analysis method based on image block processing to differentiate smoke texture features from non-smoke texture. The system was tested using two smoke videos and two non-smoke videos from the VisiFire Dataset [36]. It could generate more than 50% of the alarm frames in smoke videos, but less than 5% in non-smoke videos. Toreyin and Cetin [37] suggested a four-step system for wildfire detection that took into account slow-motion of smoke objects, grey-colored region distinction, upward movement of smoke, and shadow regions. The method's efficacy was proved by testing it on 5 video clips collected from 6 hours of woodland shooting. Chunyu et al. [9] suggested a color and motion-based technique. The Lucas Kanade technique was used to calculate the optical flow of potential locations. Testing was carried out on eight smoke and seven non-smoke videos from the VisiFire Dataset. Because of the utilization of motion and color information, objects with similar color distributions, such as a reflection from a car headlight, were easily identified. Tung and Kim et al. [38] suggested a classification approach for video sequences based on motion, color, and area difference in the probable smoke regions of consecutive frames as feature vectors in the SVM for identifying smoke. Indoor, outdoor, non-smoke, and smoke-colored objects in motion were shown in training videos. Their approach was evaluated on nine indoor-outdoor smoke video sequences, comprising six positive (smoke) and three negative (non-smoke) videos. SVM demonstrate robust performance when the smoke in the previous ten frames, only the alarm was considered valid. Therefore the changes of generating false alarm rate reduced. On real and simulated frame sequences, Labati et al. [39] suggested a computational intelligence technique for recognizing and segmenting wildfire smoke. To improve the algorithm's robustness, several adverse situations were introduced into the data. These modifications included the adding of noise, changes in lighting, and fog effects. Their technique was evaluated on datasets with low and medium quality video frames, actual and synthetic frames, and smoke/non-smoke frames in a variety of environments. The authors Labati et al. [39] tested their method using the VisiFire dataset and the dataset provided in Yuan et al. [40]. Their method detected smoke in long-range movies in real-time applications and in a variety of environmental conditions. Torabnezhad et al. [13] proposed another work for short-range smoke detection. It was a two-phase method: first, a candidate smoke mask was produced, and then, using energy calculation, a more accurate smoke mask was identified. Six video sequences from the VisiFire dataset were used to test the efficiency of method. In the second phase of the algorithm, the average smoke detection rate on 5 of the 6 videos was 90.09%, and the false smoke detection rate was 7.8% in the sixth video. Barmpoutis et al. [11] used various smoke properties such as texture, spatio-temporal dynamics, and other motion information to determine the potential smoke patch in an image. Twenty video sequences of smoke and non-smoke indoor and outdoor situations were evaluated experimentally. The videos were taken from the VisiFire dataset and the VISOR [41]. An average detection rate of 93.37% was obtained.

Image enhancement and edge detection techniques have also been utilized by researchers to increase the efficacy of IP-based smoke detection systems. Yuanbin et al. [8] suggested an IP technique that uses image enhancement to extract the smoke region with the help of GMM. For recognition, static and dynamic features are taken and given into the SVM classifier. LBP is another popular feature extraction method utilized by researchers in smoke detection challenges. Alamgir et al. [42] suggested an LBP-based feature extraction method that takes into account both local and global information on the texture and color features of smoke. SVM is used for training and classification once features are extracted from candidate smoke regions. The studies were carried out using public datasets Çetin et al. [5], Ko et al. [43], Töreyin et al.[44]. The videos include outdoor views such as mountains with clouds, sparkling lights, walking person, smoke from dustbins, and so on. On publicly available datasets, the technique performs excellent results, with a True Positive Rate (TPR) of 92.02% on average. Islam et al. [10] suggest another smoke classification and segmentation approach based on GMM and HSV color models. The approach performs well in smoke segmentation in images with complicated backgrounds. The target area is first established using smoke growth analysis, then features for classification are extracted using SVM. Eight videos from the VisiFire and VISOR datasets were utilized to develop the SVM-based ML model. A classification accuracy of 97.34% was achieved on average. However, their model has not been trained on difficult environmental conditions such as fog. Wu et al. [45] recently presented a forest fire smoke detection technique that utilizes a pixel block rule, wavelet transform, and dictionary learning. The approach demonstrates good results in detecting smoke in difficult settings such as hazy conditions, shadows, and shaking trees. However, it also has a high FAR.

2.2 Deep Learning Based Smoke Detection Methods

The CNNs have excellent performance on smoke detection task. Over the decades a number of deep CNN based classification structure have been developed to improve the smoke classification performance. Some popular CNN based classification methods such as AlexNet [46], VGGNet [47], MobileNetv2 [48], GoogleNet [49], ResNet [50], DenseNet [51] considered as state-of-the-art deep learning methods which have outstanding performance on the benchmark datasets. Several researchers have been utilized deep learning methods to address the smoke classification problem with the inspiration from these CNN based structures. Deep learning (DL) has gained popularity for fire and smoke detection.

2.2.1 Deep Learning Based Smoke Classification Methods

Yin et al. [52] proposed a deep normalization and convolutional neural network (DNCNN) system. They used normalized layers as a replacement of convolutional layers to boost the smoke detection system performance and they achieved lower than 60% false alarm rate (FAR) and higher than 96.37% accuracy rate. The experiment was conducted on VSD dataset. Filonenko et al. [53] performed a comparative study of several most popular CNN architectures for smoke detection task on the different combinations of VSD dataset. Tao et al. [54] also proposed a CNN based smoke classification model which was tested on VSD dataset. They achieved a low FAR of 0.44% and a high detection rate of 99.4%.

In some cases, smoke classification algorithms do not provide the expected performance despite the using of transfer learning methods and they generally need to large datasets. For the solution of this problem, some researchers proposed data augmentation method by using Generative Adversial Networks (GAN) to generate the more data containing smoke. Namazov and Im Cho [55] proposed a CNN based classification model that uses GAN to tackle of the overfitting on limited dataset. The model uses the adaptive piecewise linear unit as an activation function. In another work by Yin et al. [56] the authors have proposed a model that combines a deep convolutional generative adversarial network (GAN) for data augmentation and a convolutional neural network to extract more descriptive smoke features. They used the vibe algorithm to generate smoke and non-smoke images in a dynamic scene. Then, they classified the images as smoke or non-smoke. Gu et al. [57] proposed a deep dual-channel neural network (DCNN) to detect the smoke. The DCNN is end-to-end network which consists of dual channels of deep subnetworks. The first subnetwork extracts the texture information of smoke in detail and the second captures the base contours information of smoke. The combined subnetworks achieved high accuracy rate more than 99.5% on the VSD dataset. On the other hand, proposed method has weakness for detection of the smoke that have poor texture information. Liu et al. [58] have also proposed a dual-channel smoke detection model. First channel is residual network based on AlexNet which extracts the meaningful features and the second is CNN network for extracting features from dark channel images in detail. The network was trained with VSD dataset with some different scene images similar to smoke such as cloud, wall, water surface, etc. the system reached the 98.56% accuracy rate. Zhang et al. [59] proposed a dual channel CNN model to improve classification performance. In their model, they used transfer learning with AlexNet as the first channel to extract generalized features and the second channel as a small CNN for the extraction of detailed features. The proposed method achieved good classification performance with 99.33% accuracy rate on the public VSD dataset.

The existing methods discussed above suffer from a high false alarm rate (FAR) and limited accuracy in severe hazy environments. To address these problems, Khan et al. [60] proposed a VGG-16 based energy-efficient deep CNN smoke detection framework which is trained on foggy data for early detection of smoke in both normal and foggy Internet of Things (IoT) environments. They increased the smoke detection accuracy and reduced the FAR using their method on their benchmark smoke detection dataset. Muhammad et al. [61] have also proposed deep CNN based fine-tuned with MobileNetv2 [48] smoke detection model in foggy surveillance environments. They compared their method with state-of-the-arts structures. Both Khan et al. and Muhammad et al. measured model performance on the VisiFire dataset and VSD dataset. They reported 94.76% accuracy rate and 2.06% FAR on the datasets. In recent times, He et al. [62] proposed an attention based deep fusion CNN based classification model using VGG architecture in a foggy environment on a self-created dataset. They used an attention mechanism a feature-level, and a decision-level fusion model in VGG together. For the satellite images, Ba et al. [63] designed a SmokeNet based on CNN with spatial and channel-wise attention for the six classes (dust, haze, land, seaside, smoke, cloud) classification task. The proposed model achieved 92.75% accuracy rate on the SmokeRS dataset [63].

2.2.1 Deep Learning Based Smoke Detection and Segmentation Methods

CNNs are preferred for smoke localization problem due to they learn the features from the input data automatically [64-66] and they have shown promising results in localization task [67]. Therefore, the majority of researchers in literature are widely used DL based smoke and fire localization methods. For the smoke localization task, Zeng et al. [68] compared the performances of different object detection structures and their different backbones such as Faster RCNN, SSD, and ResNetv2, MobileNet, InceptionNet, Inception ResNetv2, respectively. Faster RCNN with Inception ResNetv2 backbone achieved 56.04% mAP rate. In a similar study, Wu and Zhang [69] applied Faster RCNN, YOLO, and SSD structures to identify the localization of smoke. They claimed that real-time SSD performance was satisfactory despite the real-time YOLO performance sufficiently not good on small size and dark objects of smoke and fire. Xu et al [70] proposed a CNN and wavelet based model which performs two tasks such as color segmentation to find the candidate regions for flame and wavelet for generating smoke regions. System performance was reported as good and near-real-time processing time for localization of flame. Real time YOLOv3 [67] based smoke localization method was proposed by Jio et al. [71] for Unmanned Ariel Vehicle (UAV) application. The model accuracy and speed was good for UAV application in forest fire detection. The proposed method was not successful enough for localization of small fire regions due to the limitation of YOLOv3 structure. Shi et al. [72] also proposed a YOLOv3 based smoke and fire detection system. They reached the 83.7% average precision on the three open public forest fire dataset. In recent, Li and Zhao [73] also performed Faster RCNN, SSD and YOLOv3 based smoke and fire localization models. They compared the performances of the structures each other on the three open public forest fire databases.
For real time smoke localization, Saponara et al. [74] was proposed YOLOv2 based model. They evaluated the system performance on the three different dataset.
In literature, existing studies demonstrated that CNN based smoke and localization algorithms have performed outstanding performance.

In the field of smoke detection and segmentation have been tremendous growth with enhance the DL techniques. Over the last decades, several DL based smoke segmentation models have been proposed. Yuan et al. [40] proposed a two-path encoder-decoder fully convolutional network (FCN) with skip connections architecture that to extract the global information of smoke and to keep fine spatial details of smoke from local information. The network was trained on dataset which contains real and synthetic smoke images. The system performance was better than some the state-of-the-arts segmentation structures such as FCN, SegNet and Deeplab. In another work by Frizzi et al. [75] proposed a new VGG based smoke and fire segmentation model in RGB images. That network is a combination of coding and decoding phases to achieve significantly higher segmentation accuracy. They also showed that the quality of images and image set's diversity improve the segmentation performance. Khan et al. [76] was proposed a smoke detection and segmentation framework by using EfficientNet for classification and DeepLabv3+ for semantic segmentation for clear and hazy environments. They arranged the dataset into four different classes (smoke, non-smoke, smoke with fog, and non-smoke with fog) to improve the smoke detection accuracy in hazy environments. In their study, they used their created dataset. Wen and Burke [21] proposed a deep learning model using U-Net for segmentation of the smoke. In their model, they replaced Rectified Liner Unit (ReLU) activation function with Parametric Rectified Liner Unit (PReLU). They used Geostationary Operational Environmental Satellite (GOES) data. The model was analyzed by using binary cross entropy and mean absolute error (MAE). They also performed data augmentation effectively to improve their system performance. Another CNN based decoder-encoder segmentation architecture was proposed by Larsen et al. [77] for smoke segmentation. The model achieved 57.6% mean IoU and has the good performance in satellite imagery.

DL based smoke segmentation methods have attracted for the many researchers due to the outstanding performance of CNN based models in recent years.

2.2.3 Deep Learning Based Video Classification Methods

Prediction of smoke movement from video frames is more challenging task than 2D image based smoke detection tasks. Hu and Lu [78] was introduced a model that catches the movement information from ordered frames by using spatio-temporal CNN. He trained the model on their dataset. For the estimation of movement, they proposed two stream ConvNet which spatio-temporal learning based two streams are trained separately and then combined by SVM fusion. They reported that the system reached 97.0% detection rate and 3.5% FAR. In the study of Nguyen et al. [79], CNN based motion detection method was performed to classify smoke and non-smoke scenes from video stream. The model composed of three step algorithms. First, determining the candidate regions by using Mixture of Gaussian (MOG) Background Modeling for detection background pixel changes. Then images were classified as smoke and non-smoke by using cascade model which compounds of multiple smoke classifiers. Final step is temporal analyses of video streams where history of image frames is examined over time for video based smoke classification. Their method achieved to tackle 40 frames per second (fps) and then smoke is detected between 3 to 10 s. For the capturing spatio-temporal features in a video, two-stage Deep Convolutional Generative Adversial Network (DCGAN) model by using motion-based transformation as a pre-processing was proposed by Aslan et al. [80]. True negative rate and true positive rate of the system were obtained as 99.45% and 86.23%, respectively. Yang and Sun [81] performed a DenseNet [51] based method. After extraction of features by using GMM related to motion and color of smoke, the model was trained helping with DenseNet The method achieved 99.27% accuracy rate on the public dataset. Shi et al. [82] proposed two-module method for smoke detection in a video. The first module uses an optical flow estimator and the LBP to extract features, which are then sent to the MobilNetv2 network. They have also conducted case studies by using three dataset. In recently, Pan et al. [83] and Pan et al. [84] proposed a transfer learning based method with MobileNetv2 backbone by using image blocks which down sampling the high resolution images to prevent the information loss depends on the small size of images. They reported as the model has sufficiently good results on the day and night forest fire videos.

RNN can be utilize to capture the dynamic behavior of smoke. LSTM is a type of RNN model that overcomes the vanishing gradient problem of RNN. Lin et al. [16], developed a joint detection framework that uses region based CNN (RCNN) and 3D CNN for smoke detection on video sequences. Qiang et al. [17] proposed a new feature extraction method based on VGG network which possess spatial (static) and time (dynamic) stream. VGG and Bi-Long Short-Term Memory (BLSTM) are used to extract static and dynamic features, respectively. Finally the two sets of features are fused to achieve higher forest fire smoke detection performance. Hu et al. [78] proposed a method that combining of Deep Convolutional Long-Recurrent Networks (DCLRN) and optical flow for real time smoke and flame detection from video in outdoor scene. The method was used spatial and sequence learning by using CNN and deep LSTM model together where the CNN extracts the features from optical flows of consecutive frames, and temporally accumulated in an LSTM network. They achieved good accuracy and dependability in the detection and identification of fire monitoring videos. Kim and Lee [85] proposed a DL based method that uses Faster RCNN to extract the spatial features from video sequences of the suspected regions of fire (SRoFs) and non-fire. The extracted features were accumulated by LSTM for the temporal information and the final video classification was performed as smoke and fire. They achieve high detection accuracy and low FAR. Yu Zhao et al. [86] proposed a Deep Gated Recurrent Unit (GRU) based model is proposed to detect the forest fires at early stage by using GOES-R satellite time series data. The model was implemented as a 6-level architecture, which consists of 5 GRU layers with many-to-many architecture and one level of Dense network to generate the output. They obtained good detection accuracy and lower FAR.

Chapter 3

Deep Learning Models for Smoke Classification

In this section, I investigated state-of-the-art deep learning CNN architectures and effect of image dehazing for smoke detection on our created dataset. For this purpose, we used pre-trained fine-tuned CNN structures. Most popular of current classification based smoke detection systems focus on distinguish between smokes and other moving objects. Moreover, environmental conditions determines the shape and motions of the smoke. Thus, classification and detection of smoke is very hard and crucial task.

3.1 Transfer Learning for Smoke Classification

Transfer learning, the method that we preferred does not require extremely large training dataset and computational power. In transfer learning, a pre-trained Convolutional Neural Network is used for feature extractor. Fine-tuning is a type of transfer learning. Applying to fine-tuning to deep CNN models that have already been trained on ImageNet database. Applying fine-tuning to the model causes the building of a new fully-connected head which places top of the original architecture.

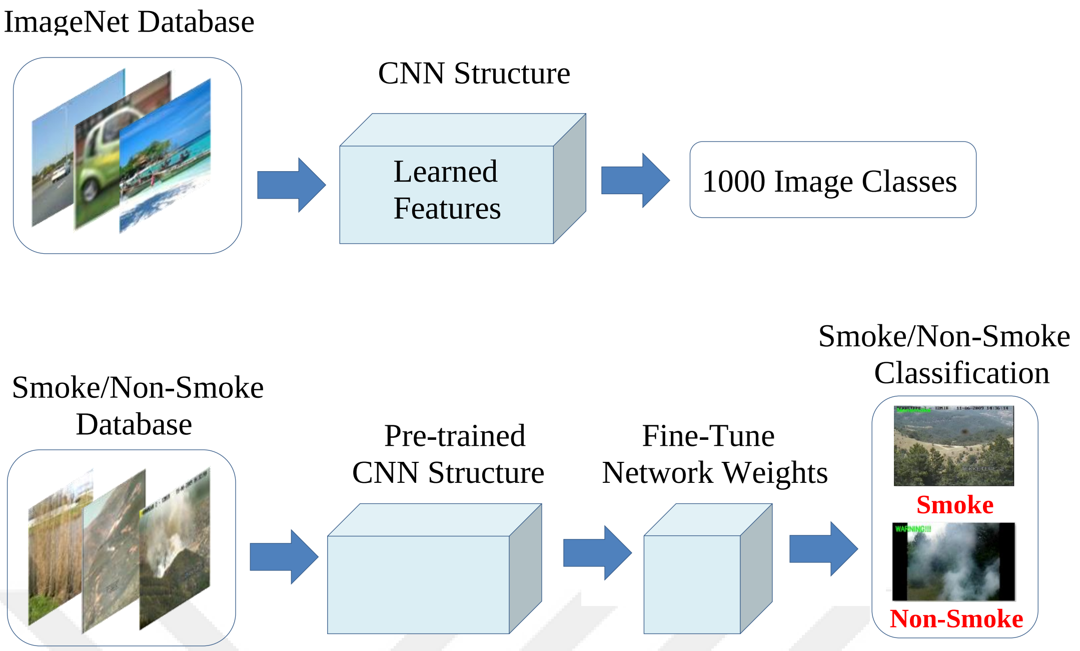


Figure 3.5. Transfer Learning Structure for adapting Smoke/Non-Smoke classification.

3.1.1 Smoke Classification Dataset Preparation

For the experiments, VisiFire [87] dataset was used in the training and test process. A total of 50 Smoke/Non-Smoke videos were selected from the dataset; 33 videos for training, 5 for validation and 12 for testing. We extracted the images from these videos and the Smoke/Non-Smoke dataset was created as a balanced dataset. Extracted images were chose as clearly indicated that smoke and Non-Smoke images in video frames. These images were selected manually for categorization of classes. Totally we used 4,631 images for training, 518 for validation and 802 for the test. These datasets is made publicly available for research community (https://github.com/eem-merve/Smoke-Segmentation-Dataset). Sample frames from this dataset are shown in Figure 3.2 and Figure 3.3.

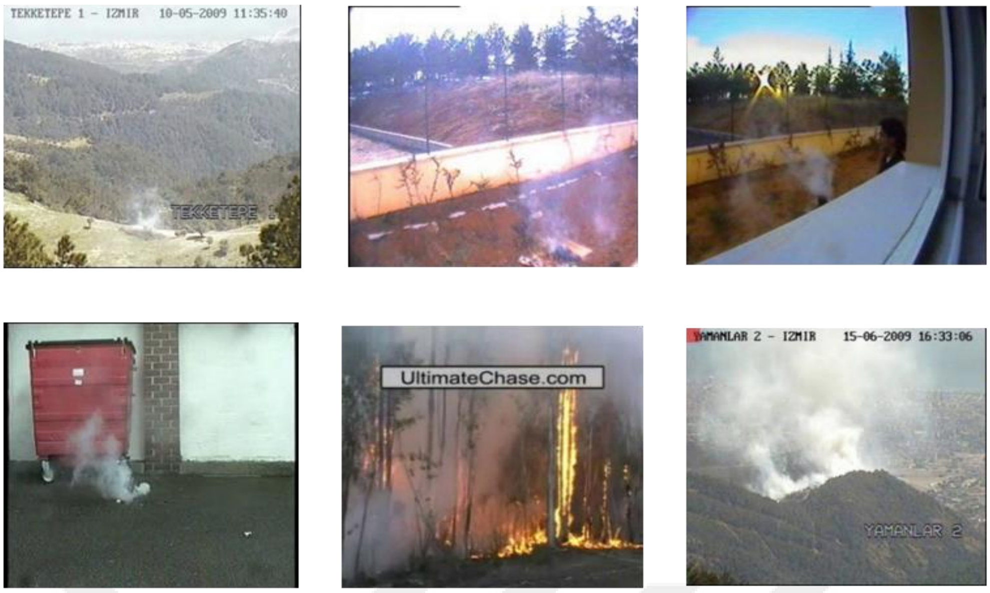

TEKKETEPE 1 - IZMIR 10-05-2009 11:35:40
TEKKETEPE 1
UltimateChase.com
YAMANLAR 2 - IZMIR 15-06-2009 16:33:06
YAMANLAR 2

Figure 3.6. Some sample images of our smoke dataset obtained by extracting images from the video.

YAMANLAR 1 - IZMIR 08-06-2009 15:01:41
PRESET 1
TEKKETEPE 1 - IZMIR 10-05-2009 11:25:50
TEKKETEPE
Marmaris Tansas 28-11-2008 14:25:59

Figure 3.7. Some sample images of our non-smoke dataset obtained by extracting images from the video.

3.1.2 Fine Tune Network

In this thesis, we used AlexNet [46], VGG16 [47], ResNet50 [50], EfficientNetB0 and EfficientNetB1 for the smoke detection in images. ResNet50, VGG16, EfficientNetB0 image classification networks require 224x224x3 size images while AlexNet and EfficientNet [88] network requires 227x227x3 and 240x240x3 image size for network training, respectively. However the extracted images from the videos have different image sizes. Thus, we cropped the dataset images as 224x224x3, 227x227x3 and 240x240x3 image size according to using classification models. For fine tuning process, state-of-the-arts deep learning classification networks are modified and then networks are re-trained. When removing the original FC layers end of the network and then place it with new fully connected head. These new FC layers can be fine-tuned to our classification dataset. After that network is trained for two classes as smoke/non-smoke classification. All classification steps that fine-tuned CNN networks on our prepared dataset are shown in Figure 3.4.

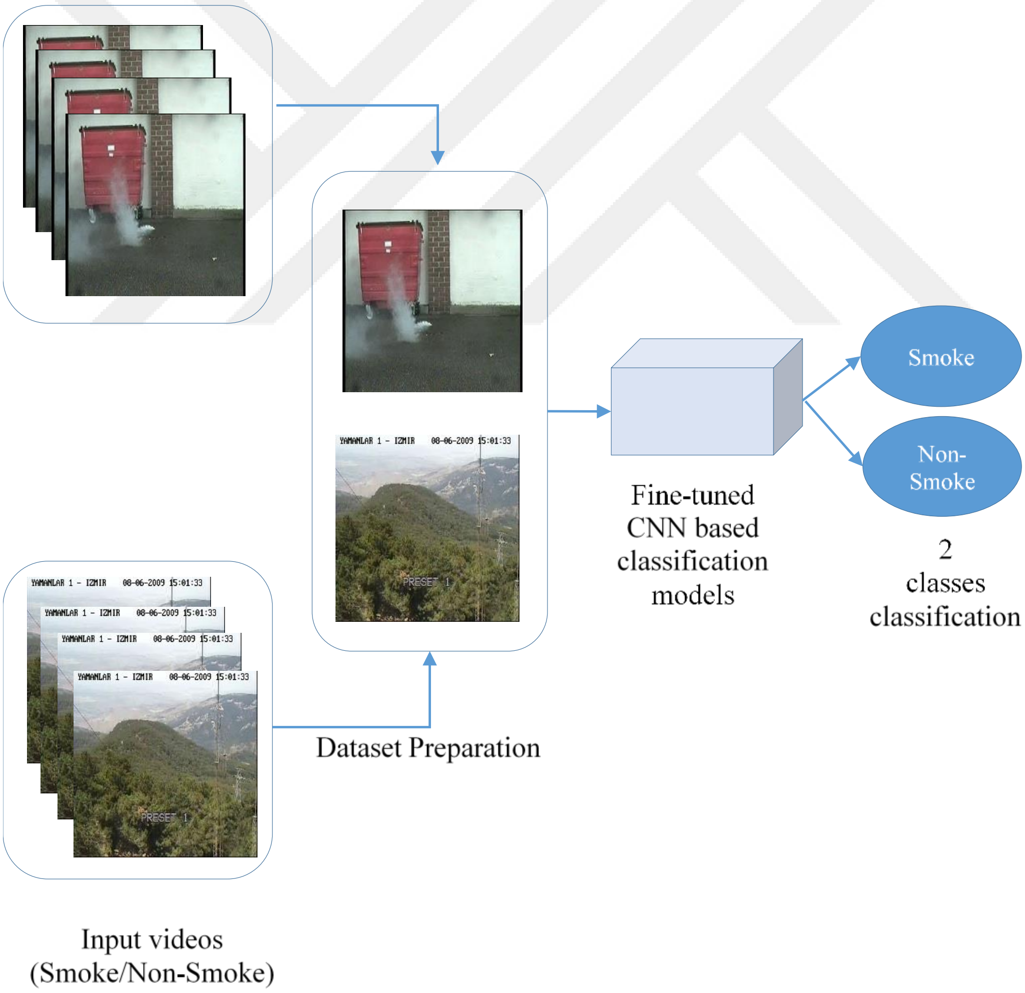

Smoke
Non-Smoke
Fine-tuned CNN based classification models
2 classes classification
Dataset Preparation
Input videos (Smoke/Non-Smoke)

Figure 3.8. Overall network architecture for classification task. First step is dataset preparation from the videos. Second step is classification process via fine-tuned CNN network.

There are EfficientNet variants called B0 to B7. We also used these variants for the classification task on our dataset. We obtained good results in terms of classification accuracies with the B0 and B1. These models need to small size of input resolution during network training. B0 and B1 require input resolutions of 224 by 224 and 240 by 240, respectively. Other variants did not give successful results in terms of performance metrics in comparison to alternative classification models.

3.1.3 Performance Metrics for Smoke Classification

Precision, F1 score, and recall were used as performance metrics for the evaluation of smoke classification system. When a smoke formation is detected according to ground truth by the smoke classification system; it was considered as a true positive (TP) output. In the case that the smoke classification system identifies the smoke that does not match the ground truth, it is considered as a false positive (FP) output. Detection of the non-smoke in the video frame is regarded as a true negative (TN), whereas no detection of non-smoke in the video is regarded as a false negative (FN). Precision, recall and F1 score are formulated as:

$$Precision\ (P) = \frac{TP}{TP+FP} \tag{3.1}$$

$$Recall\ (R) \quad = \frac{TP}{TP+FN} \tag{3.2}$$

$$F1\ score \quad = \frac{2*P*R}{P+R} \tag{3.3}$$

Another evaluation metric is False Alarm Rate (FAR) to demonstrate the efficiency of pre-processing method. FAR is calculated as the ratio of total number of FP images to the total number of non-smoke images (NS) used for smoke/non-smoke images classification task. FAR is calculated by using Eq. 3.4.

$$FAR = \frac{FP}{NS}\ x\ 100 \tag{3.4}$$

Table 3.1 is obtained when using performance metrics such as P, R, F1 score. All networks are based on fine-tuned CNN network. These results are obtained on 802 test image samples in our Smoke/Non-Smoke dataset.

Table 3.1 Comparison of different classification models performance metrics on our dataset.

Methods	Precision	Recall	F1 score
AlexNet	0.79	0.75	0.75
VGG-16	0.88	0.86	0.86
ResNet50	0.92	0.90	0.90
EfficientNetB0	0.94	0.93	0.93
EfficientNetB1	0.98	0.97	0.97

Table 3.1 shows precision, recall and F1 score measures of the smoke classification models evaluated on our dataset. Based on this table, the best performing model is obtained as the EfficientNet. EfficientNet is achieved 97% accuracy when fine-tuned network for 2 classes classification task.

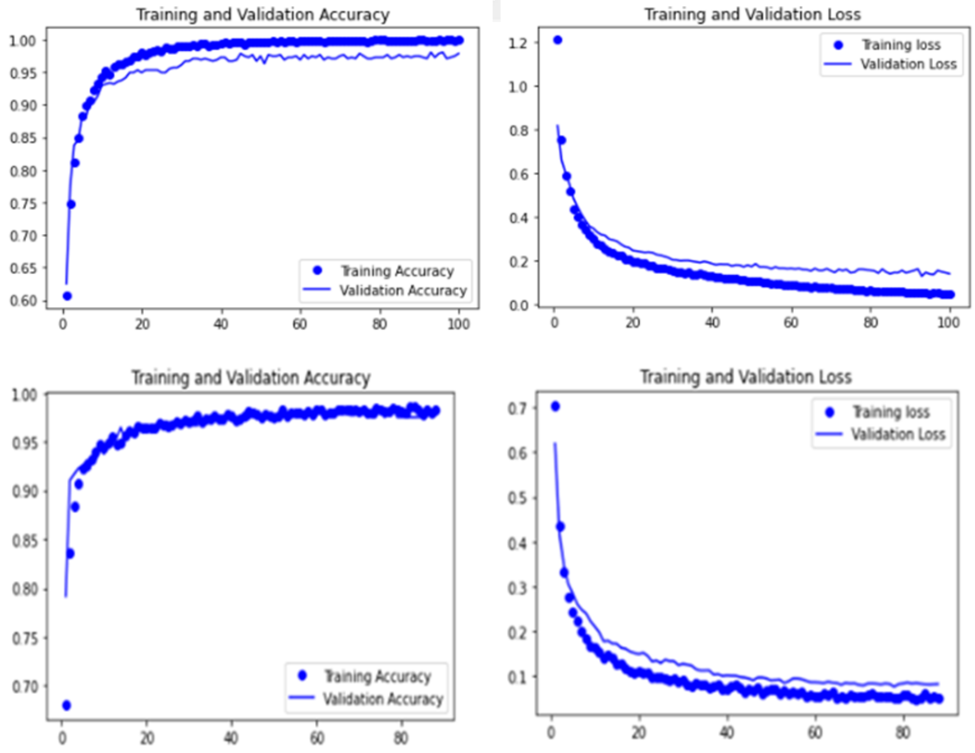


Figure 3.68. Classification models training accuracy and loss plots.

During the training process, accuracy and loss plots from the classification model are shown in Figure 3.5. These plots show that the training is performed successfully, where both training and validation loss curves saturate to their minima and the validation and training accuracies are close to each other, which indicates that there is no significant overfitting.

Table 3.2 shows the hyperparameters for the fine-tuned CNN classification structures. According to Table 3.2, we added GlobalMax Pooling, dropout (rate=0.2), batch normalization, fully connected layer to classify the smoke and non-smoke images when we used EfficientNet and also we used binary cross entropy as a loss function and Adam optimizer with 0.0001 learning rate in the training. Input image size is 224x224 and 240x240 to train the fine-tuned EfficientNetB0 and EfficientNetB1 CNN architectures, respectively. For AlexNet, we added fully-connected layer with L2 kernel regularization (L2=2e-4), batch normalization, dropout (rate=0.5), fully-connected, batch normalization, dropout (rate=0.5) and the final fully-connected layer to classify the smoke/non-smoke images. Input size is resized as 227x227 to train AlexNet. For the fine-tuned VGG16 model, we added fully-connected, dropout (rate=0.2) and final fully-connected layer. In the fine-tuned ResNet50 model, we added Average Pooling, fully-connected, dropout (rate=0.5) and final fully-connected for smoke/non-smoke classification. Moreover, we applied the early stopping (patience=10) regularization techniques for all models to prevent overfitting.

Table 3.2 Hyperparameters of the classification methods.

Model	Input Image Size	Optimizer	Learning Rate	Batch Size	Regularization Techniques
AlexNet	227x227	Adam	1e-3	32	Droput, Batch Normalization, L2 Kernel Regularizer, Early Stopping
VGG-16	224x224	Adam	1e-4	32	Dropout, Early Stopping
ResNet-50	224x224	Adam	1e-4	32	Dropout, Early Stopping
EfficientNetB0	224x224	Adam	1e-4	32	Dropout, Batch Normalization, Early Stopping
EfficientNetB1	240x240	Adam	1e-4	32	Dropout, Batch Normalization, Early Stopping

3.2 Transformer Based Image Classification Models

Transformers have great success for Natural Language Processing tasks. After improving transformer, transformers are used for image classification. The advantages of using Transformer based classification are that extraction of more powerful features by using attention mechanism. Such architectures are called as Vision Transformers (ViT) [89] . The ViT model works with self-attention mechanism instead of convolutional layers.

3.2.1 Vision Transformer Based Smoke Classification

Vision transformer (ViT) divides the images into visual tokens and splits an images into the fixed size of patches. These patches are linearly embedded and position embedding are added for input of Transformer encoder. ViT regards image patches as word, then it reaches embeddings of the patches to the transformer.

ViT achieves competitive performance such as ImageNet and CIFAR100 compare to CNN [89]. Results are even further improved when applied to larger datasets, where ViT was able to achieve similar results or beat CNNs in some benchmarks.

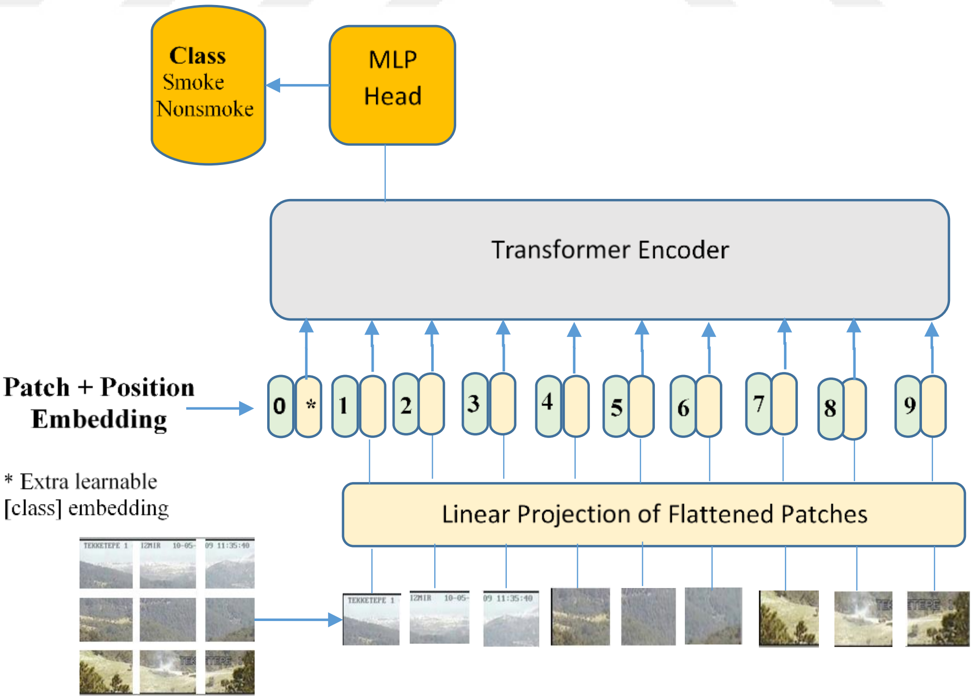

Class
Smoke
Nonsmoke
MLP Head
Transformer Encoder
Patch + Position Embedding
0 * 1 2 3 4 5 6 7 8 9
* Extra learnable [class] embedding
Linear Projection of Flattened Patches

Figure 3.9. Vision Transformer Architecture [89].

The 2D images divided into N pathes of size PxP. In the embedding stage of the VİT, each patch is flattened and linearly transformer into D dimention vector. Since a 2D position embedding based on x,y coordinate wasn't useful to the model the position is generated as a single value. The process over convert's image patches into tokens. The token input process is identical to standard NLP tasks. Thus, there is no modification encoder transformer model. The research suggests a hybrid approach that feeds CNN-generated feature maps rather than the original raw image. In this study VİT was used to detect of two classes as smoke and non-smoke. First images were resized 72x72 pixels size and patch size was selected during to training process Adam W optimizer with 1e-3 learning rate binary cross entropy was preferred as a loss function for two classes.

3.2.2 Shifted Windows Transformer Based Smoke Classification

Shifted windows (Swin) transformers [90] are hierarchical new vision transformer. Swin transformers work on two concepts that are hierarchical feature maps and shifted window attention. Hierarchically extraction of the feature maps and overall structure of Swin transformers are shown in Figure 3.7 and Figure 3.8, respectively.

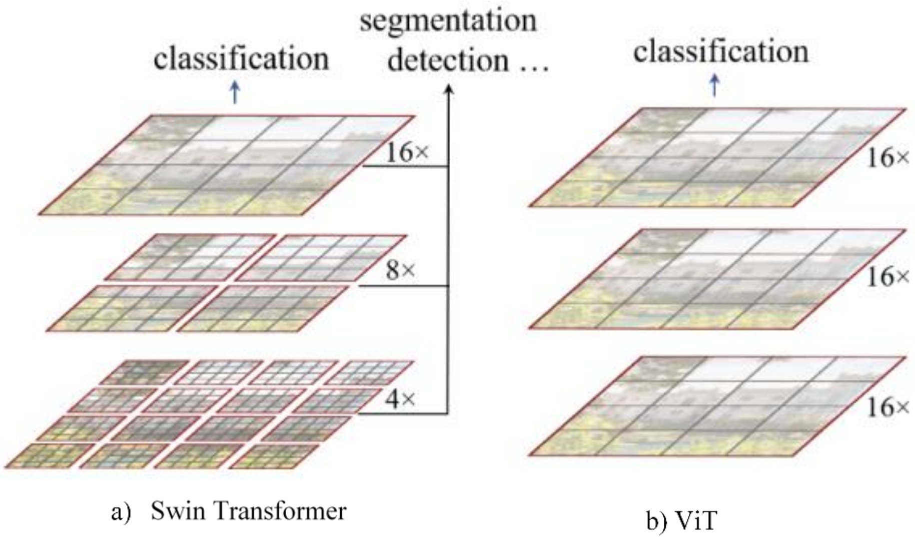


Figure 3.10. Comparison of Swin Transformer and ViT in terms of working principle.

Hierarchical structure are obtained with downsampling of the feature maps from one layer to another while ViT utilizes single feature maps in its architecture. Swin Transformer can also be used in segmentation tasks due to the hierarchical structure.

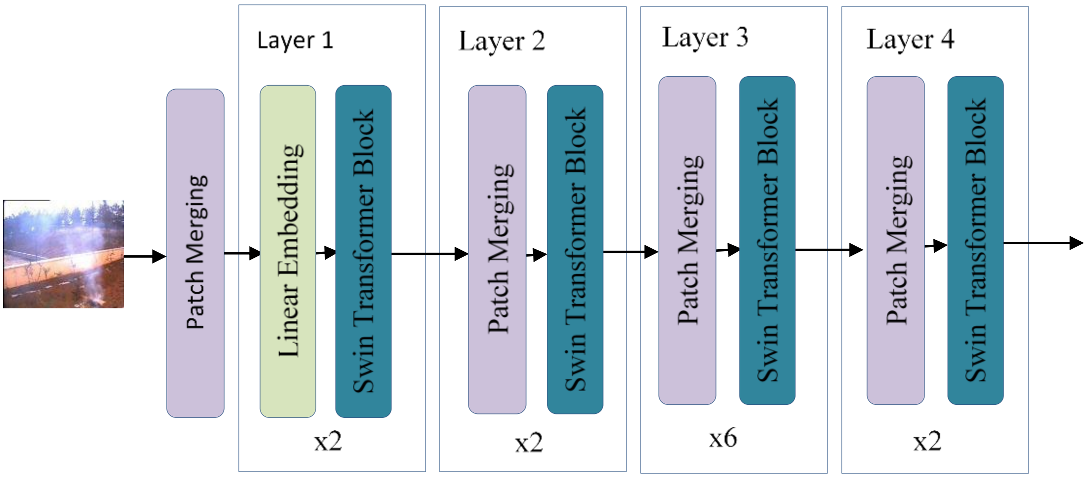


Figure 3.11. Swin Transformer Architecture [90].

Swin transformer divided into RGB images with non-overlapping patches. Each patch is considered a “token” with its features set to be concatenation of the RGB values of the individual pixels.

In Swin transformer, downsampling of feature maps are performed by using patch merging while CNN structure uses convolution operation for the downsampling. In the patch merging process, first input images are divided into the groups. Patches are stacked deep-wisely in each group and then, stacked groups are combined.

Swin transformer blocks used a window mutihead self-attention module (W-MSA) and shifted window MSA (SW-MSA) instead of standard mutihead self-attention module in ViT. In standard MSA has some problem on the high resolution images. This issue is addressed with Swin Transformer by using W-MSA. Yet, restricting self-attention to each window leads to limit modeling power of the network. To solve this issue, SW-MSA is used after the W-MSA.

In recently, Swin transformers broadly is utilized for classification and detection tasks due to the structures that have hierarchical feature maps and shifted window MSA.

Table 3.3 demonstrated that comparison of Transformer based classification models performance. According to the Table, Swin Transformer based smoke/non-smoke classifier achieved better performances than Vision Transformer.

Table 3.3 Comparison of Transformer based classification models performance metrics on our dataset.

Methods	Precision	Recall	F1 score
Vision Transformer	0.72	0.73	0.72
Swin Transformer	0.76	0.75	0.76

3.3 Image Dehazing Method

There are several hazy conditions involved when video data is used to detect fire formation such as fog and smoke particles in the atmosphere that absorb and scatter the light [91]. These effects obscure the view of the camera and cause significant degradations in the image quality. Based on the atmospheric scattering model, an image in hazy scene mainly consist of two parts including attenuation and scattering process shown in Figure 3.9. The first part is the attenuation process which is reflected light from the object surface to the camera. The second part is the scattering of air-light reaching the camera. These two parts establish the theoretical basis of hazy images.

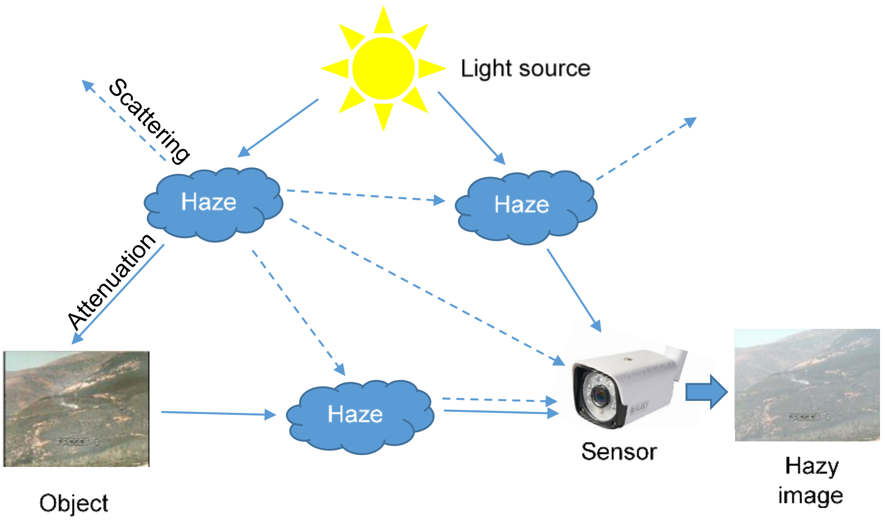


Figure 3.12. Hazy image formation process [92].

In the field of computer vision, a hazy image can be expressed in Eq. (3.5), which is derived based on the scattering model.

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{3.5}$$

where x is the distance coordinate, $I(x)$ is the hazy image, $J(x)$ is the hazy free image, A is the atmospheric light and $t(x)$ is the science transmittance i.e. the portion of sunshine that does not scatter and reaches directly to the camera. Moreover, the terms of $J(x)t(x)$ and $A(1 - t(x))$ indicate that direct attenuation and air-light, respectively. The science transmission is defined in Eq. (3.6).

$$t(x) = e^{-\beta d(x)} \tag{3.6}$$

where β is the scattering factor and $d(x)$ is the depth of pixel x. Eq. (3.5) and Eq. (3.6) are used to obtain direct transmission and airlight, respectively. The purpose of image dehazing methods is to recover $J(x)$ from $I(x)$ [93, 94].

3.3.1 Image Dehazing Model Selection

There are several image dehazing studies in the literature [93, 95-100]. We compared three dehazing methods proposed in Meng et al. [95], Li et al. [99], and Mondal et al. [93] in terms of peak signal-to-noise ratio (PSNR) and structural index similarity (SSIM), which are widely used for image quality [101]. We first obtained hazy samples by synthetically adding haze to a set of clear samples selected from our dataset. To achieve this, we used Foggy and Hazy Image Simulator (FoHIS) [102] to obtain the hazy images starting from clear images, which serve as a ground truth for their hazy counterparts.

We then used these methods as proposed in [95], [99], and [93] to remove the haze from the images. Figure 3.10 shows average PSNR and SSIM results of these methods, which are computed by comparing dehazed images to their clear versions.

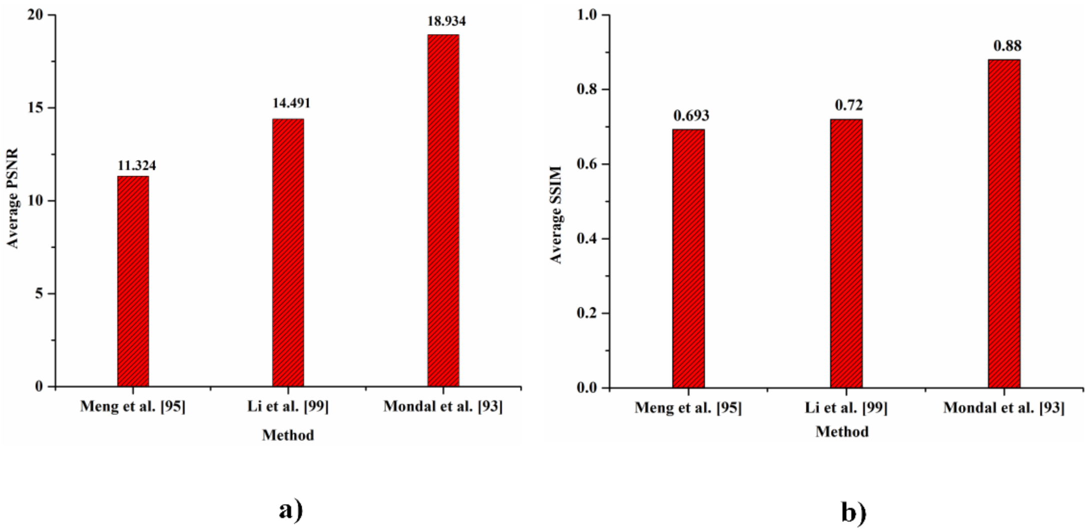


Figure 3.13. PSNR and SSIM comparison of three image dehazing methods. a) Average PSNR and b) Average SSIM are obtained using image samples from our dataset.

Based on this figure, the FCN based image dehazing method proposed in Mondal et al. [93] performs better than the other two methods for image dehazing task since higher PSNR and SSIM values mean that the dehazed images are on average closer to the reference haze-free images. According to test results, we utilized a FCN based image dehazing method [93].

3.3.2 Effects of Image Dehazing Method on Dataset Images

Image dehazing is a fog removal strategy as a pre-processing method to eliminate fog from the input images. Figure 3.11 shows the sample dataset images in the smoke/non-smoke database and the output images of the image dehazing model with improved quality. First row presents the original dataset images and the second row presents the dehazed counterparts. In the training process, these images were used as the input for smoke detection system.

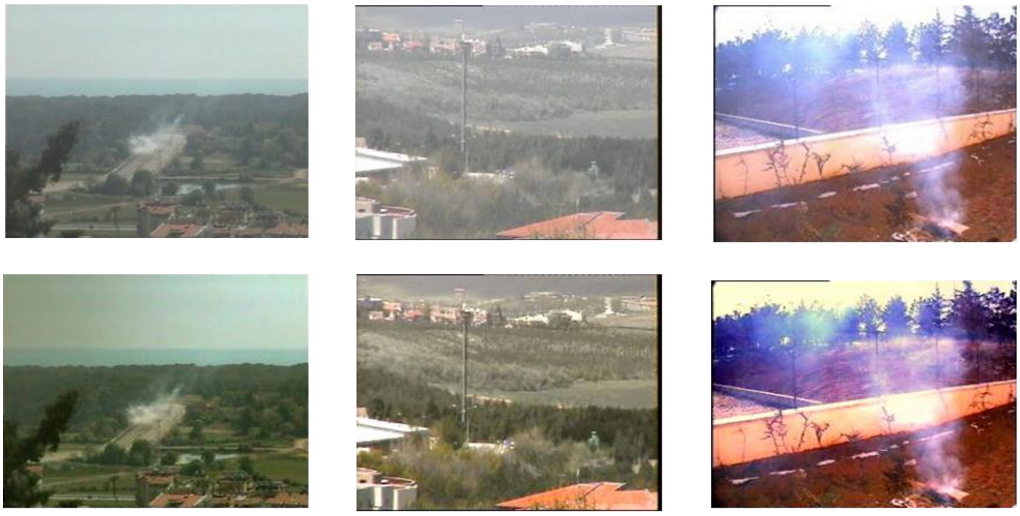


Figure 3.14. First row demonstrates three original dataset images in the smoke/non-smoke database, and the second row indicates their dehazed counterparts.

Figure 3.12 presents our architectures that combine image dehazing with smoke classification. Classified smoke/non-smoke images are the output images of the network as presented in Figure 3.12.

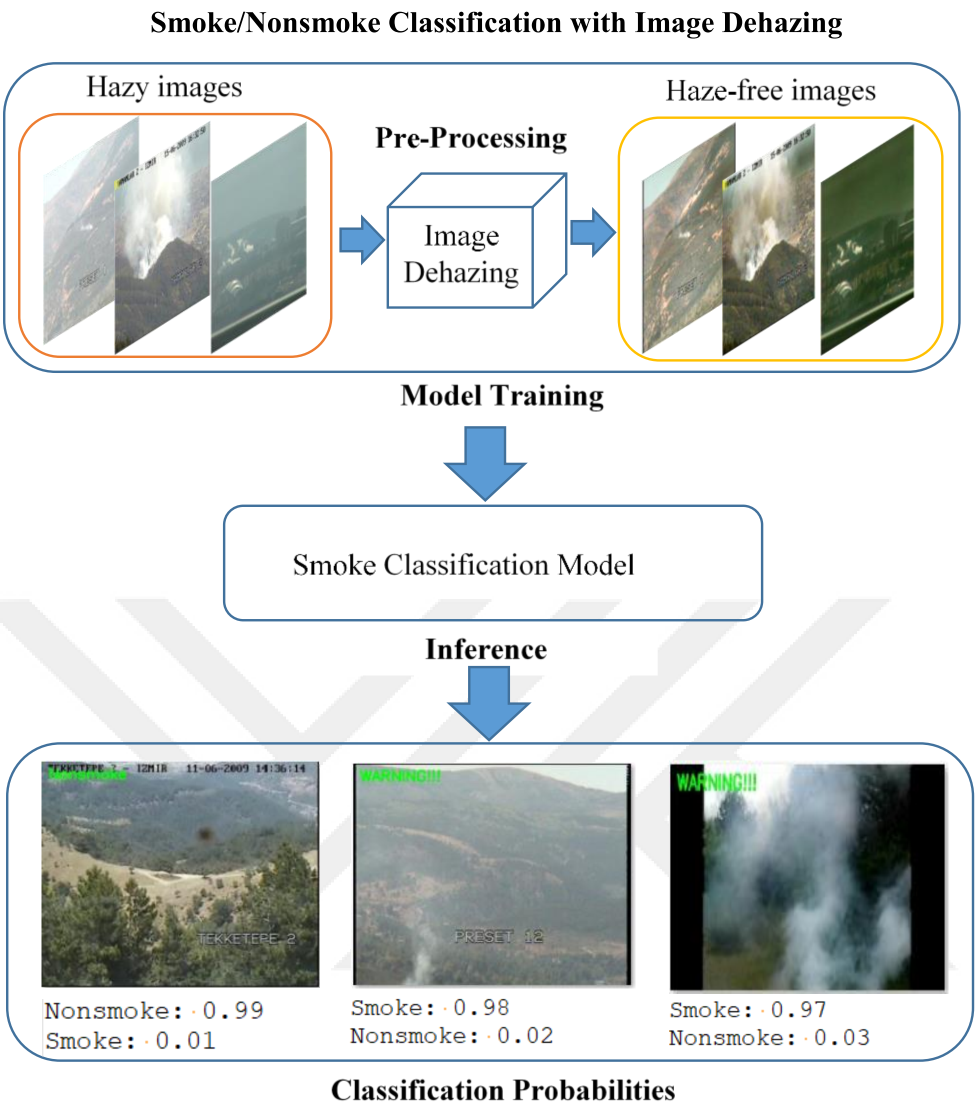


Figure 3.15. Steps of our classification method. Haze-free images are used as input for classification models to obtain smoke classification outputs.

Table 3.4 denotes the smoke classification performances of different classification methods with and without dehazing applied. Based on these results, the smoke classification performance is improved for all classifiers and for all metrics when dehazing is performed as a pre-processing step.

For the smoke classification task, the best results are achieved by the EfficientNet model when input images are dehazed. The amount of the improvements are ~4% in Precision, 5% in Recall, 5% in F1 score.

Table 3.4 The effect of image dehazing on smoke classification. Image dehazing improves the smoke classification metrics for all architectures.

Methods	**Precision**	**Recall**	**F1 score**	**FAR**
AlexNet	0.79	0.75	0.75	0.21
AlexNet w/dehazing	0.83	0.81	**0.81**	**0.17**
VGG-16	0.88	0.86	0.86	0.12
VGG-16 w/dehazing	0.93	0.92	**0.92**	**0.07**
ResNet50	0.92	0.90	0.90	0.08
ResNet50 w/dehazing	0.94	0.93	**0.93**	**0.06**
EfficientNetB0	0.94	0.93	0.93	0.06
EfficientNetB0 w/dehazing	0.98	0.97	**0.97**	**0.02**
EfficientNetB1	0.98	0.97	0.98	0.02
EfficientNetB1 w/dehazing	0.99	0.98	**0.99**	**0.01**

Chapter 4

Deep Learning Models for Smoke Detection and Segmentation

After smoke classification, smoke detection and segmentation tasks were performed on our created dataset, respectively. Single Shot Detector (SSD) [103], You Only Look Once (YOLO) [64] and Faster Regional CNN [66] object detectors were used for smoke detection task which are the most commonly used architecture for object detection and also Mask RCNN was used for segmentation of smoke in an image.

4.1 Smoke Detection

SSD, YOLO and Faster RCNN were successfully applied for object detection task in literature. We employed these three object detection structures to detect the smoke in an image. SSD with MobileNetv2 backbone, YOLOv5 and Faster RCNN with ResNet50 and ResNet101 backbones were implemented on our dataset.

4.1.1 Smoke Segmentation Dataset Preparation

VisiFire [87] dataset was used in the training and test process. A total of 38 smoke videos were selected from the dataset; 24 videos for training, 4 for validation and 10 for testing. We created a new smoke detection and segmentation dataset using a subset of our classification dataset. We have chosen 429, 50 and then 50 smoke images for training and for test and validation set, respectively. Figure 4.1 demonstrates that some smoke video samples to create the dataset for object detection task. The created dataset images were extracted from these videos.

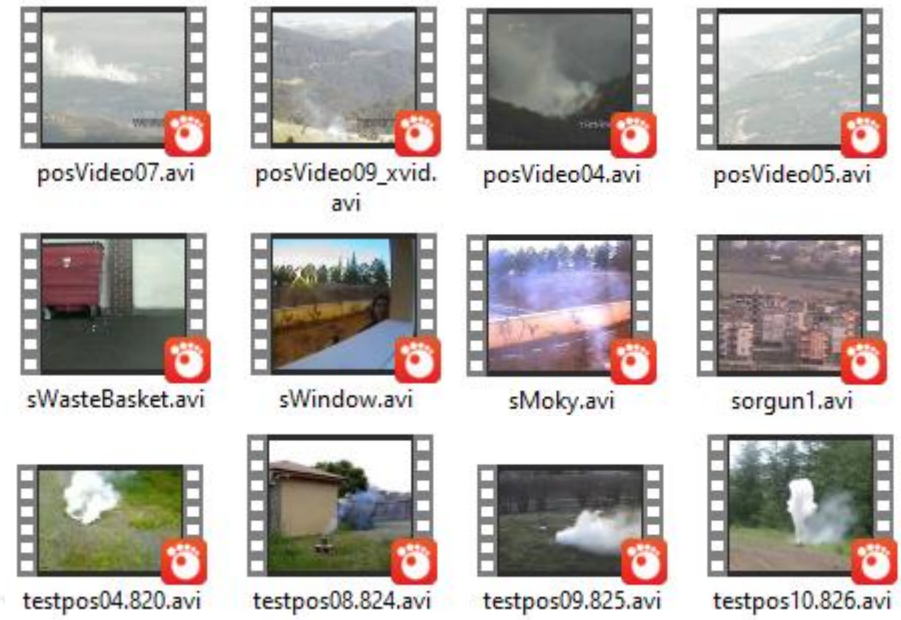


Figure 4.16. Some sample videos for smoke detection task.

We applied image labelling process as a rectangular shape for object detection. After the labeling, we export bounding box coordinates of smoke parts in image. These coordinates saved as "COCO JSON" format to use in training phase. We used the Detectron 2, a software powered by the Pytorch framework, containing many backbone structures and faster training process. Also we used ResNet50-FPN and ResNet101-FPN backbones with Faster RCNN, SSD with MobileNetv2 and YOLOv5 to detect the smoke in image. Image labelling process is shown in Figure 4.2.

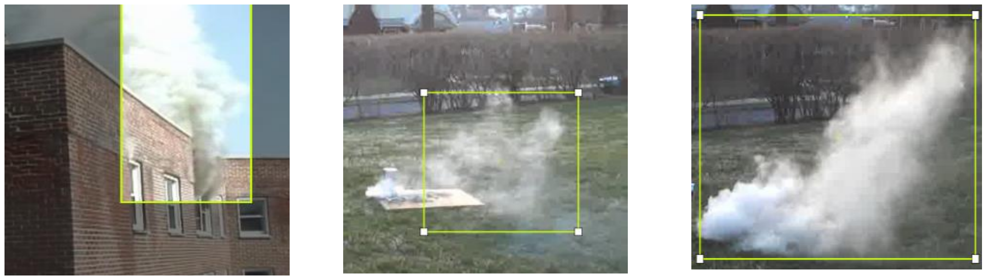


Figure 4.17. Image labelling process for object detection task.

4.1.2 Smoke Detection Models

Smoke detection is more challenging task compare to smoke classification. Smoke detection task includes smoke classification. In image classification task, class is predicted for an object in an image. In object localization, localization of object is determined and this is indicated its localization with bounding boxes in an image. The presence of object is located with bounding boxes and also it shows relevant class in an image when the use of object detection method. In the another computer vision task which is called object segmentation, instance of recognized objects are indicated by highlighting specific pixels of the object instead of bounding boxes. Figure 4.3 demonstrates that steps of smoke detection and segmentation tasks.

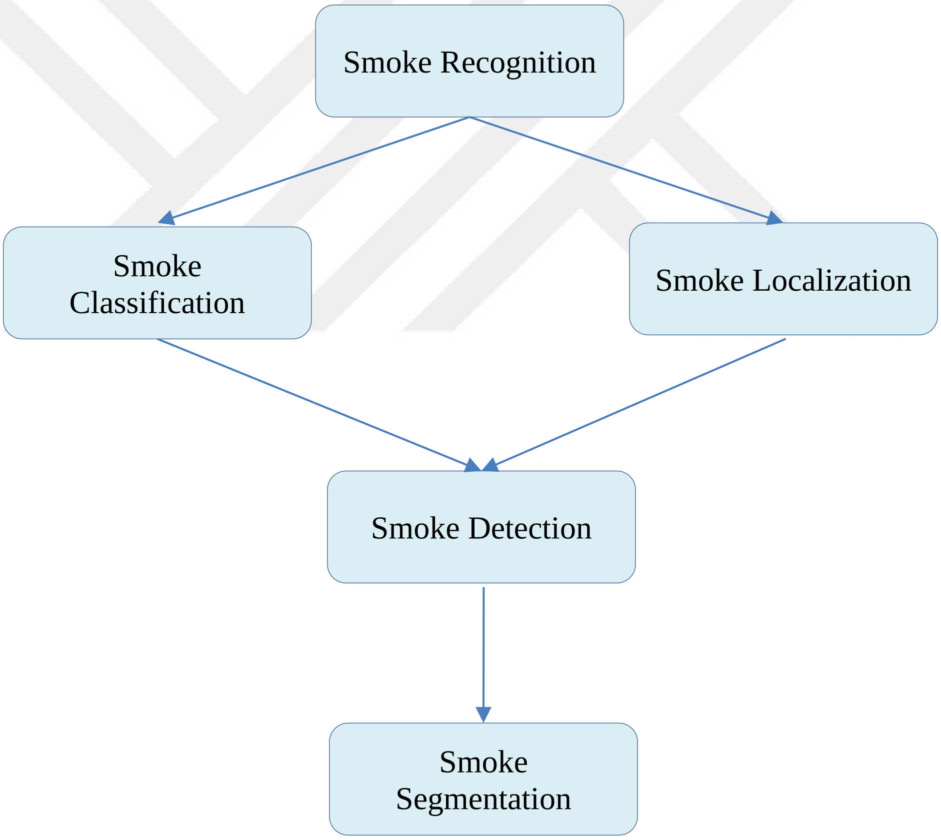


Figure 4.18. Overview of smoke recognition task.

In this part; SSD with MobileNetv2 backbone, Faster RCNN with ResNet50 and ResNet101 backbones and YOLOv5 object detectors are used to detect the smoke in an image. Figure 4.4 demonstrates that the stages of smoke detection task when using most popular object detectors. After the training process, smokes were detected with their detection probabilities on our test images.

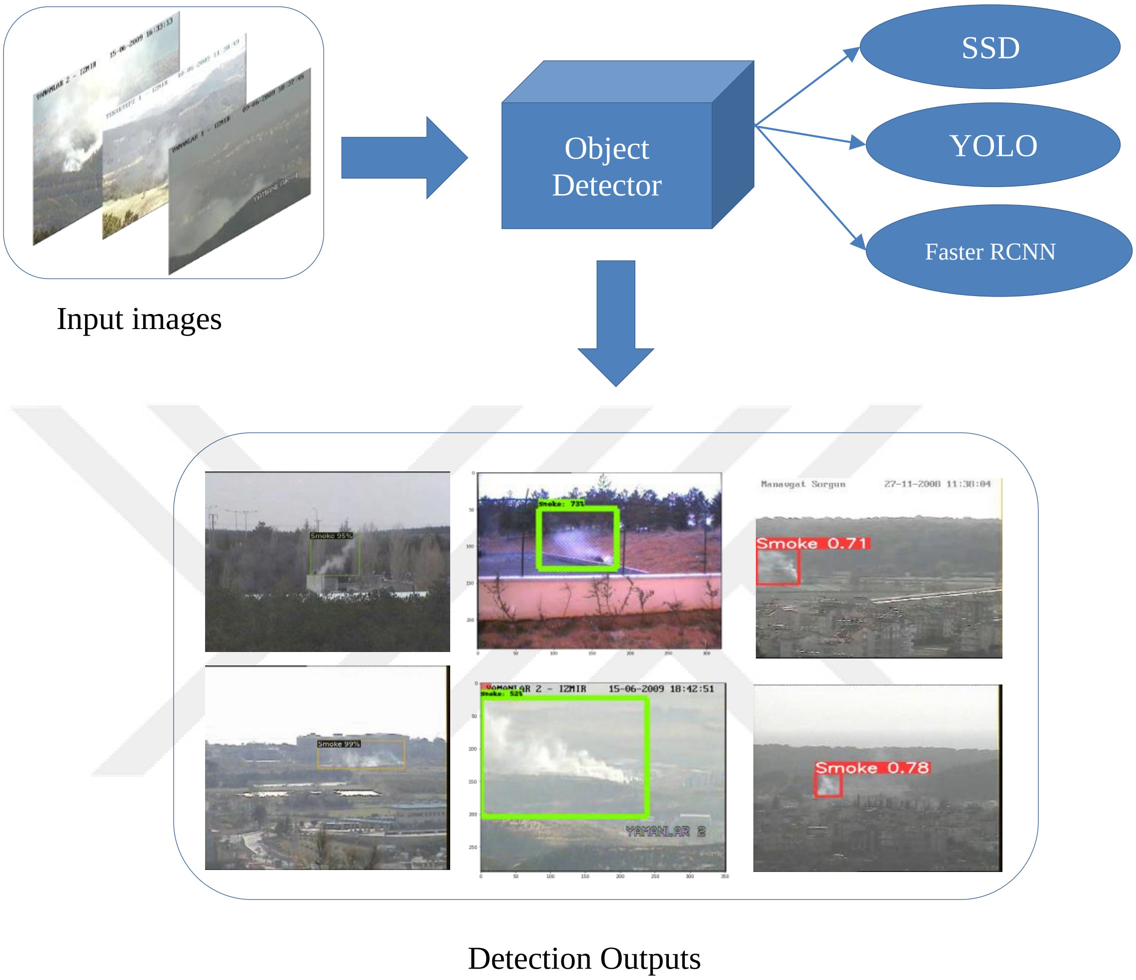


Figure 4.19. Smoke detection task when using different object detector structures on our dataset.

4.1.2.1 Single Shot Detector (SSD) Based Smoke Detection

Single Shot Detector (SSD) is designed for object detection task. SSD network works as 2 stages during the detection of objects. First stage is extraction of feature maps after that second stage is applying convolutional filters to detect the objects. SSD uses VGG16 network for extraction of feature maps. After the extraction of feature maps, SSD applies 3x3 convolution filters for each cell to compute both location and class scores. In the following step, SSD uses Non-Maximum Suppression (NMS) to eliminate the redundant predictions pointing to the same object. Figure 4.4 shows that SSD network architecture with smoke detection stages. SSD architecture frequently uses the single feed-forward convolutional network. This network generates a fixed-size collection of bounding boxes and scores for the presence of object class instances in those boxes to precisely estimate classes and region box (anchor) offset without second step per proposal classification operation [104].

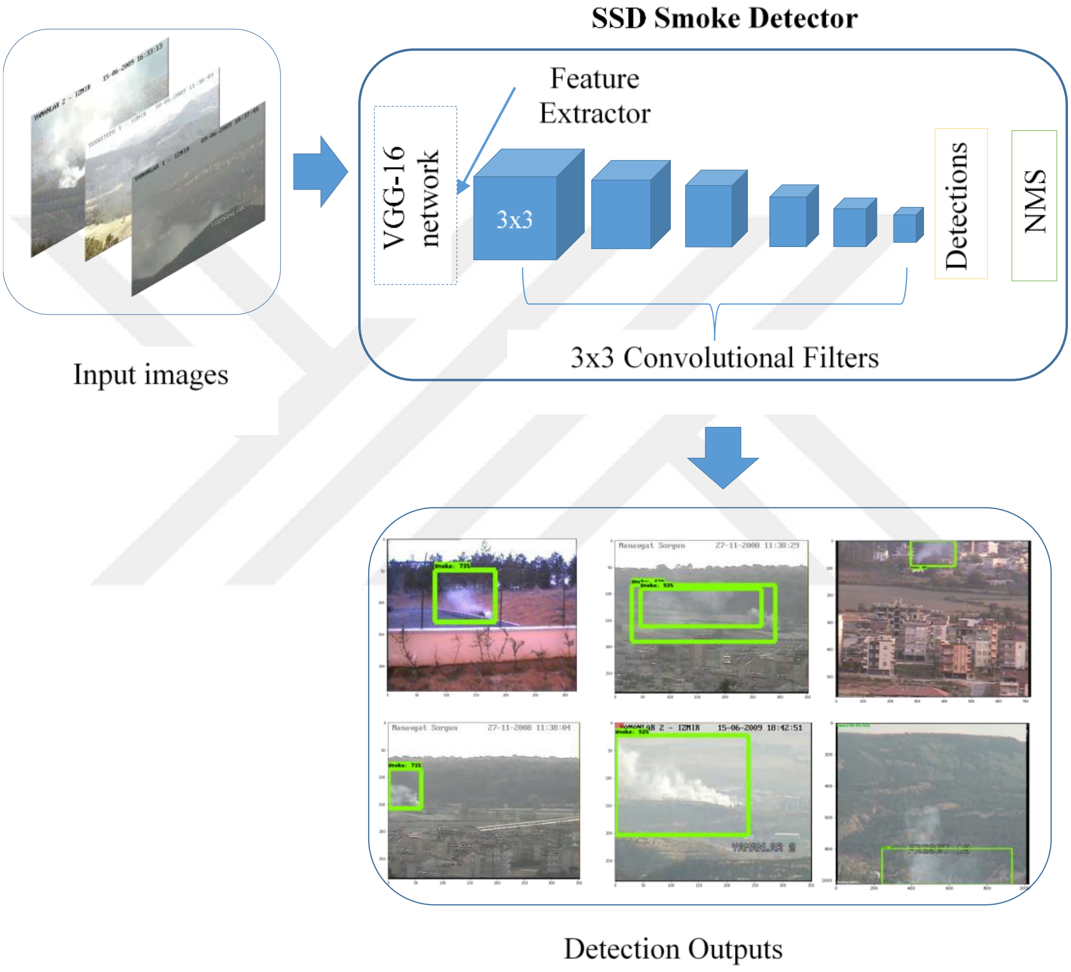


Figure 4.20. SSD architecture for smoke detection.

In this thesis, SSD object detection scheme with MobileNetv2 backbone was implemented to detect smoke in an image. Number of epoch was selected 30.000 iterations, batch normalization and l2 regularize were used for regularization techniques and sigmoid function were also used for the final scores due to our segmentation task is detection of only 1 class.

4.1.2.2 You Only Look Once (YOLO) Based Smoke Detection

One of the another object detection method is YOLO which initially proposed by Redmon et al. [64]. The latest form YOLOv5 was developed by Ultralytics is fast, easy to train and has high accuracy compare to other YOLO models. Yolov5 is single-stage object detector and consists of 3 parts. These parts are model backbone, neck and head. Model backbone is used for feature extraction of input image. Model neck is comprised feature pyramid network (FPN) [105] which uses to detect same object with different scales and sizes. The last part is model head that employs anchor boxes on features. Consequently, final output is obtained by using bounding boxes with a class score.

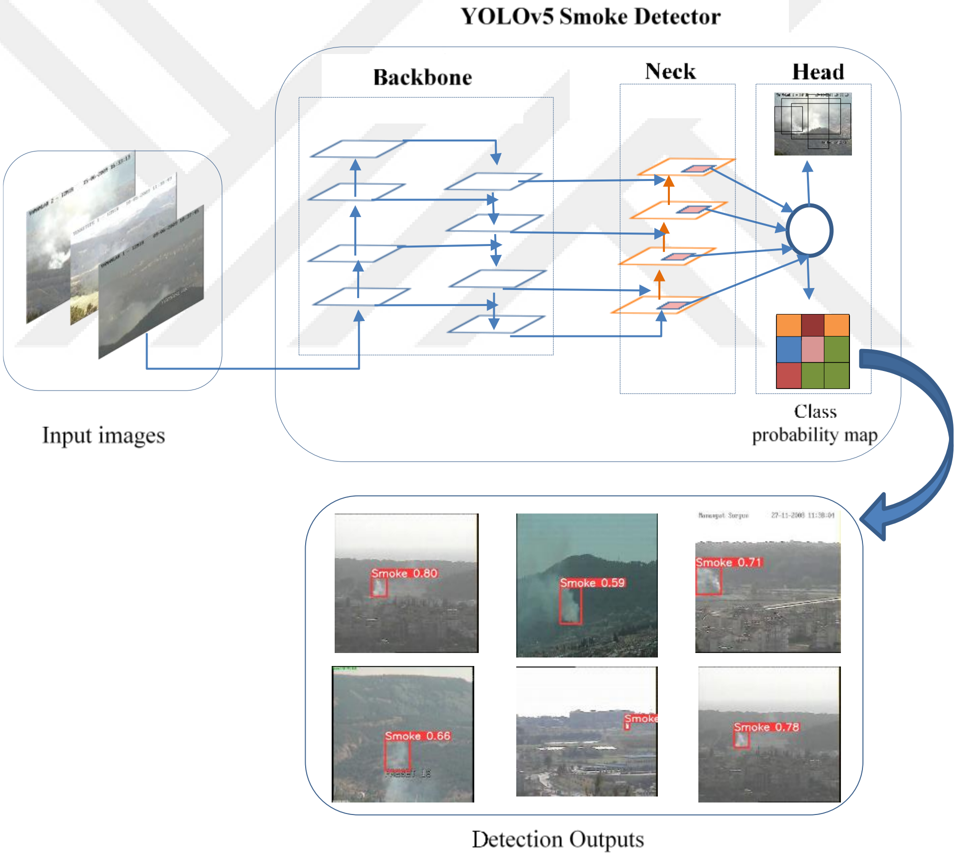


Figure 4.21. YOLOv5 architecture for smoke detection.

YOLO and its several versions are famous object detection structure. YOLO has many advantages such as easy to implement and can train the entire image directly. Thus, YOLO has grown steadily [106]. YOLO performs better performance in terms of processing time because it does not use a separate network to extract candidate regions.

4.1.2.3 Faster Regional CNN Based Smoke Detection

Region based CNN detection methods are one of the first large and successful application of CNNs for object localization, detection, and segmentation. RCNN structures mainly consist of three components. These are region proposal network (RPN), feature extractor and classifier. In the region proposal module, region proposals are generated and extracted. In feature extraction module, features that each candidate region comes from region proposal module are extracted by using CNN. Final step is classification of features by using preferred classifier model. Faster RCNN has achieved good detection performance on Microsoft COCO [107] and Pascal VOC dataset.

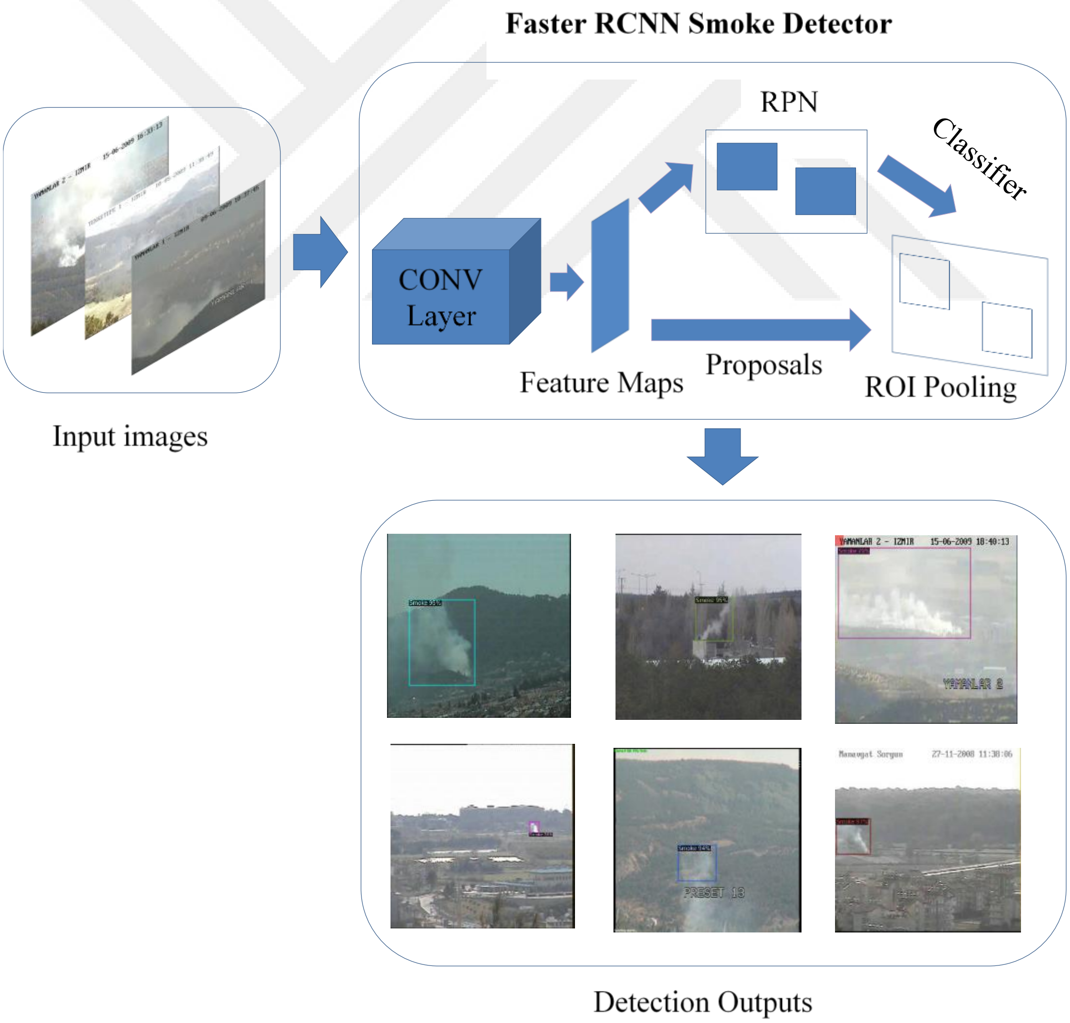


Figure 4.22. Faster RCNN architecture for smoke detection.

In faster RCNN, the input image passes through the convolution layer and feature maps are extracted. Then, a sliding window is used in RPN for each location over the feature map. Anchor boxes (default bounding boxes) are used to generate region proposals for each location. The output of RPN is a set of rectangular object proposals that have a probability of containing the objects of interest. Bounding box labels that are assigned to the boxes and their probabilities (objectiveness score) for each label and box are obtained. After RPN, different size proposed regions are found. Thus, ROI pooling solves the problem by scaling down the feature maps into the same size. Classifier layer determines the output of the system regardless of the presence of an object and the regression layer outputs for the box coordinates (box center coordinates, width and height). In this layer, regression is calculated while comparing the estimated bounding and the ground truth boxes [104] in Figure 4.7.

In the literature, Faster RCNN smoke detector is mostly used for detection of smoke/fire in an image. Faster RCNN has good detection performance to detect the smoke/fire. Thus, Faster RCNN based smoke detectors can be used for early detection of smoke in applications because of system performance and high accuracy rate.

In my experiment based on Faster RCNN with ResNet50 and ResNet101 backbones on Detectron2 [108] platform is performed to detect smokes. These backbones are used as a top-down structure called Feature Pyramid Network (FPN) [105]. FPN improves the standard feature extraction pyramid by adding a second pyramid that takes the high level features from the first pyramid and passes them down to lower layers. This approach allows features at every level to have access to both, lower and higher levels.

4.1.3 Performance Metrics for Smoke Detection Models

In the following equations, the TP, FP, and FN here are defined for the object detection task (i.e. smoke detection task) and are different from the TP, FP and FN defined for smoke classification task. For detection task, smoke detection system correctly identified the smoke according to the ground truth, it was regarded as a true positive (TP). In the false positive (FP) case, the system detected the smoke that did not match the ground truth. When the system did not detect the smoke in the video frame, it was regarded as a false negative (FN). There were no true negatives (TN), since there were no frames that did not include any smoke.

$$Precision\ (P) = \frac{TP}{TP+FP} \quad \textbf{(4.1)}$$

$$Recall\ (R) \quad = \frac{TP}{TP+FN} \quad \textbf{(4.2)}$$

$$F1\ score \quad = \frac{2*P*R}{P+R} \quad \textbf{(4.3)}$$

In Table 4.1, performance of smoke detectors were compared on our created dataset. Results show that Faster RCNN based smoke detector was more successful in terms of F1 score according to other detectors. YOLOv5 and Faster RCNN ResNet50-FPN have high Precision value that means all predictions are true. However, these detectors have weakness on prediction of some smoke images that means FN rate is high. Thus, Recall value is lower than Faster RCNN ResNet101-FPN.

Table 4.5 Comparison of smoke detection results on our dataset.

Detection Models	P	R	F1
SSD MobileNetV2	0.760	0.527	0.623
YOLOv5	1.000	0.667	0.800
Faster RCNN ResNet50-FPN	1.000	0.833	0.909
Faster RCNN ResNet101-FPN	0.892	0.983	0.935

4.2 Smoke Segmentation

Mask RCNN [109] is one of the best instance segmentation model. It detects the target in an image and also gives the predicted mask for each detected target. Mask RCNN is extended on the basis Faster RCNN and adds a new branch to obtain segmentation masks. The detailed framework of Mask RCNN is demonstrated in Figure 4.8.

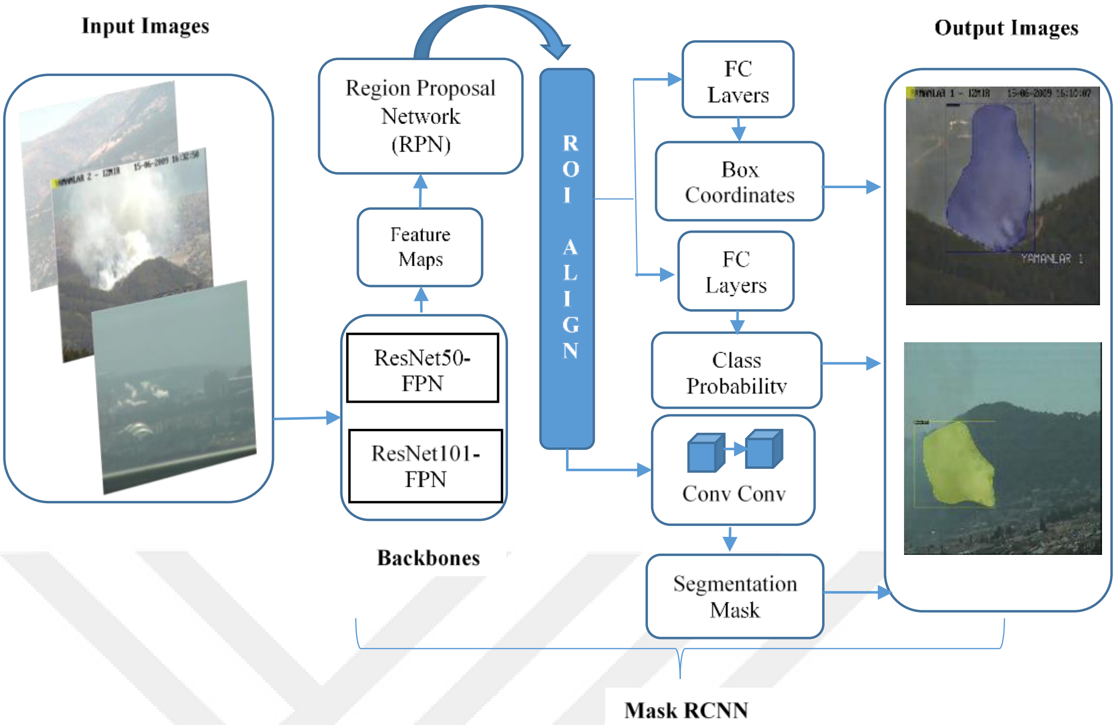


Figure 4.23. Mask RCNN architecture for smoke segmentation.

For smoke segmentation task, Mask RCNN with ResNet50-FPN/ResNet101-FPN backbones were used. Mask RCNN instance segmentation model with ResNet50 and ResNet101 backbones were implemented on Detectron2 platform to detect smokes. Using dataset was the same dataset in object detection task and also images in the dataset were resized as 800x1024 pixel size.

Image labelling process is applied as a polygon shape for instance segmentation. Labelled data was converted to the Creating Common Objects in Context (COCO) annotation format that is widely used by the instance segmentation and object detection community. A total of 1,511 instances were used during the training and testing process. Image labelling process is shown in Figure 4.9.

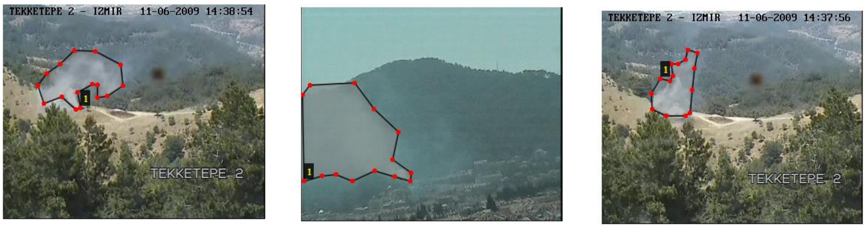

TEKKETEPE 2 - IZMIR 11-06-2009 14:38:54
TEKKETEPE 2
TEKKETEPE 2 - IZMIR 11-06-2009 14:37:56
TEKKETEPE 2

Figure 4.24. Image labelling process for instance segmentation

4.2.1 Performance Metrics for Smoke Segmentation

Intersection over Union (IoU) and Average Precision (AP) were used to evaluate the smoke segmentation system. Supporting metric called as Intersection over Union (IoU) **is** also required in order to determine the validity of a detection (predicted mask). In object detection systems, IoU is calculated as the area of intersection over union between the ground-truth mask and the predicted mask as shown in Figure 4.10.

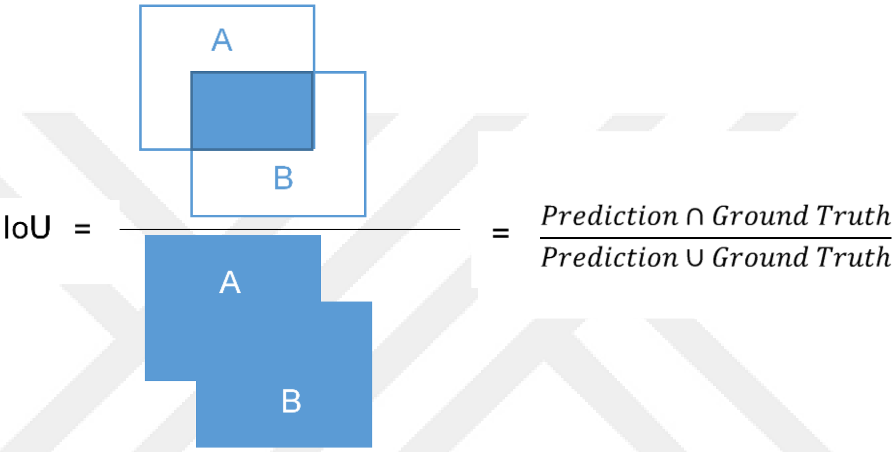


Figure 4.25. Intersection over Union (IoU) visual representation.

Intersection over Union (IoU), which is also called as the Jaccard index is defined as the size (i.e. the number of pixels) of the intersection divided by the size of the union between predicted box (region A) and its corresponding ground truth (region B) in Eq. 4.4 and Eq. 4.5. When the IoU value is bigger than the threshold, a prediction is regarded as True Positive (TP), otherwise it is regarded as False Positive (FP).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad \textbf{(4.4)}$$

$$J(A,B) = \frac{|A \cap B|}{|A|+|B|-|A \cap B|} \qquad \textbf{(4.5)}$$

In addition, IoU can also be defined in terms of TP, FP and FN in Eq. (4.6).

$$IoU = \frac{TP}{TP+FP+FN} \qquad \textbf{(4.6)}$$

I used ResNet50-FPN and ResNet101-FPN as the feature extractor backbone. Standard COCO metrics including AP (Average Precision), AP$_{50}$ and AP$_{75}$ are used to evaluate the model. The performance of the object detection and localization algorithm is evaluated by a metric called Average Precision. AP, which is calculated with the help of several other metrics such as IoU, confusion matrix (TP, FP, FN), precision and recall. When the AP is computed at a single IoU of 0.50 and a single IoU of 0.75, these correspond to the metrics AP50, and AP75, respectively.

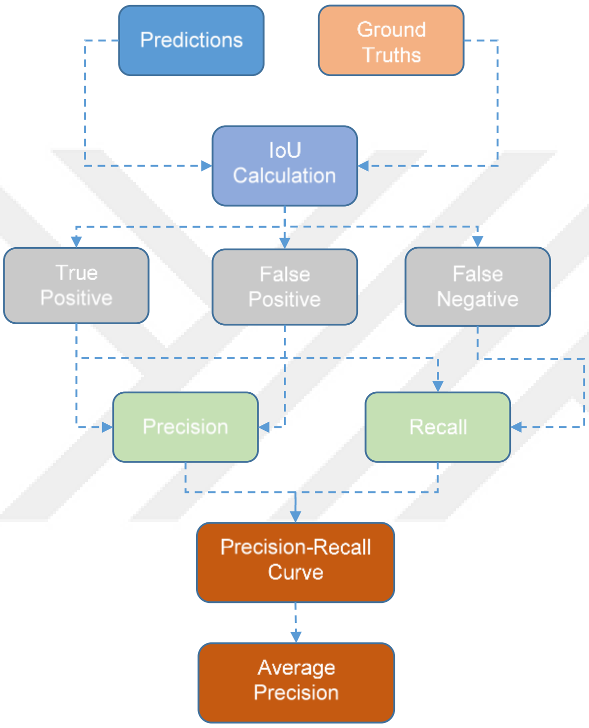


Figure 4.26. Calculation of Average Precision for one class.

AP is calculated from precision-recall curve. The Pascal VOC Challenge uses an 11-point interpolation. The recalls are divided into 11 points (0 to 1 with a step size of 0.1) and the value of precision at each recall point is calculated. AP is calculated as the average at these points in Eq. (4.7).

$$Average\ Precision\ (AP) = \frac{1}{11}\sum_{R\in(0,0.1,0.2,\ldots\ldots,1)} P_{interp}(R) \tag{4.7}$$

where R i.e. recall takes on 11 values, and $P_{interp}(R)$ is the interpolated precision at R values. In the Figure 4.11, we demonstrate that calculation of average precision as a graphical version for smoke segmentation task.

Table 4.2 Smoke segmentation results for different backbones on the Mask RCNN structure.

Backbones	Type	AP %	AP50 %	AP75 %
ResNet50-FPN	Box Mask	52.13 47.14	86.62 83.29	58.81 49.49
ResNet101-FPN	Box Mask	54.56 49.34	92.24 90.45	59.09 50.61

Table 4.2 shows smoke detection performances of ResNet50-FPN and ResNet101-FPN on our created dataset.

4.2.2 Effects of Image Dehazing on Smoke Segmentation

Image dehazing method obtains clear image which is important technique to improve the system performance in computer vision community. Thus, effects of image dehazing were investigated on smoke segmentation method as well. For this purpose, smoke segmentation system was trained by using dataset images and their dehazed counterparts together as system input. After that results were compared to the same test set. Figure 4.11 indicates that Mask RCNN implementation steps to demonstrate the effects of image dehazing. Two approaches are implemented for smoke segmentation depending on what is used as input images. Figure 4.11 (a) shows that smoke detection is performed with the Mask RCNN using extracted images from videos in the dataset. Figure 4.11 (b) shows that both the images and their dehazed counterparts were used together for system input during the training process.

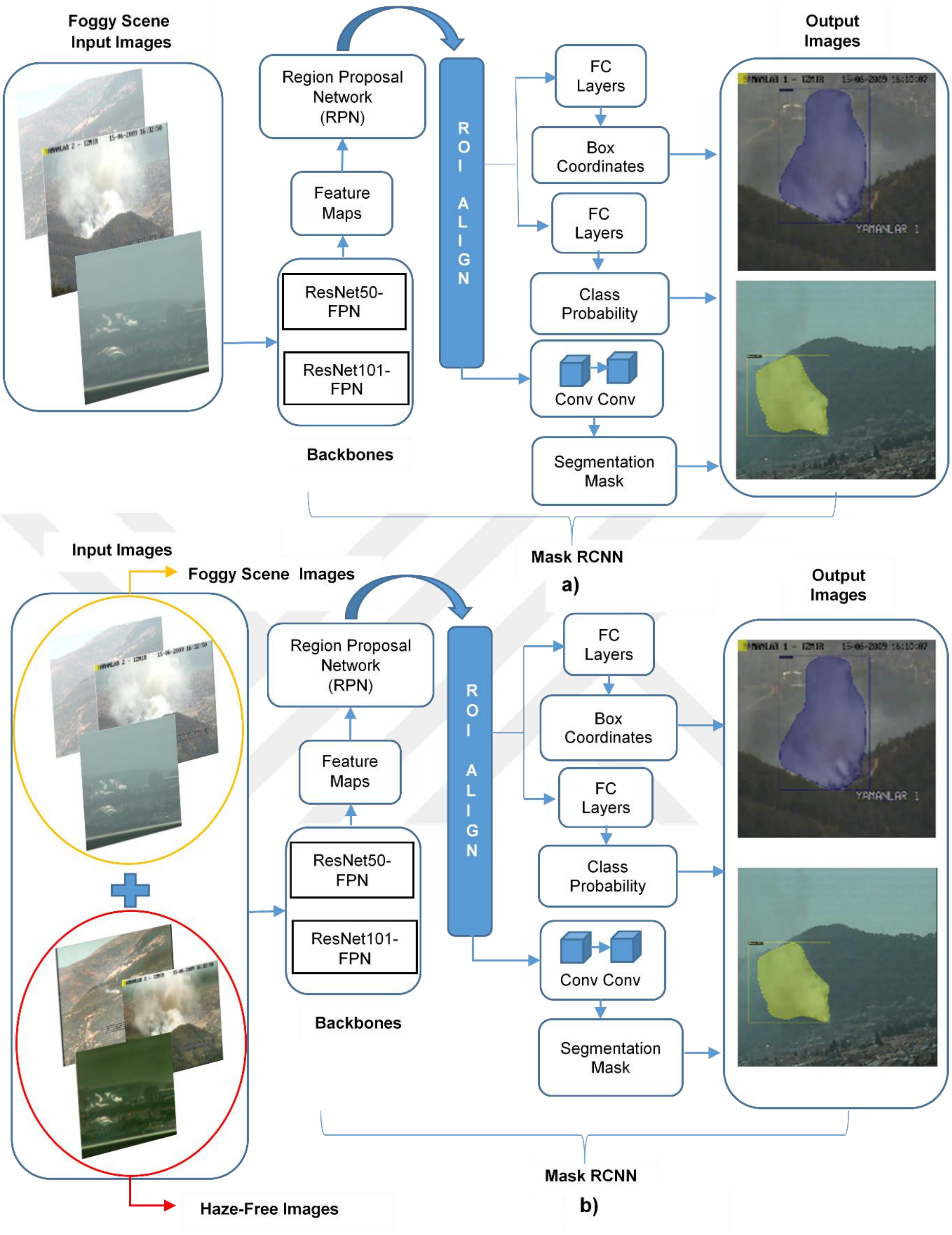


Figure 4.27. The Mask RCNN architecture is implemented in two ways. a) Foggy images are used as input for the system without the pre-processing method. b) Foggy images and their dehazed counterparts (haze-free images) are used as input for the system.

Table 4.3 shows smoke detection performances of ResNet50-FPN and ResNet101-FPN with and without dehazing applied as the pre-processing method. In this table, ResNet50-FPN and ResNet101-FPN represent segmentation methods without pre-processing applied and the ResNet50-FPN w/dehazing and ResNet101-FPN w/dehazing represent method with pre-processing applied.

Table 4.3 Smoke segmentation results with and without dehazing approach for different backbones on the Mask RCNN structure.

Backbones	Type	AP %	AP50 %	AP75 %
ResNet50-FPN	Box	52.13	86.62	58.81
	Mask	47.14	83.29	49.49
ResNet50-FPN w/dehazing	Box	54.67	90.82	67.95
	Mask	48.30	85.22	56.98
ResNet101-FPN	Box	54.56	92.24	59.09
	Mask	49.34	90.45	50.61
ResNet101-FPN w/dehazing	Box	**60.22**	**93.87**	**76.11**
	Mask	50.15	93.07	57.47

Table 4.3 indicates that the smoke detection performances in terms of AP, AP50 and AP75 with pre-processing yields, better results compared to case without pre-processing using the Mask RCNN smoke localization architectures with ResNet50-FPN and ResNet101-FPN backbones. For smoke localization task, we obtained ~6% and 1% improvements in terms of box AP and mask AP, respectively. In addition, each accuracy metric best performing results are highlighted in Table 4.3.

We investigated the influence of image dehazing method on wildfire smoke detection. We have implemented fully convolutional neural network based image dehazing method as a pre-processing approach from surveillance images extracted from video inputs. Our experimental results indicate that our approach gives significantly better performance increasing the accuracy for both smoke classification (in Chapter 3) and detection tasks. We compared our results based on pre-processing approach for the common classification and object detection methods and also Mask RCNN segmentation with two different backbones (ResNet50 and ResNet101). Gathered data indicates that the removal of haze from images before the training process yields a better outcome in terms of smoke classification and localization accuracy.

Chapter 5

Early Detection of Forest Fires from Video

CNN based detection approaches have excellent performances on generic visual detection tasks. Moreover, CNN based approaches mainly focus on images for detection of smoke or fire. However, video based detection methods should determine both spatial and temporal information. A video comprise of ordered sequence of frames. Each frame and ordered sequence contains spatial and temporal information, respectively. Recurrent Neural Networks (RNN) is utilized with sequence data and their output depends on previous steps output. RNN has a memory in which all information about the computations in the system is recalled. In the CNN networks, inputs and output independent from each other while RNN structure uses input information to obtain output.

In this thesis, GRU and LSTM [110] action recognition from video techniques with CNN structures which are commonly used in literature were performed together to detect the smoke in a video. The system consist of two parts. In the first step, CNN feature extraction process is used for spatial information while RNN utilizes for the temporal information in the second step.

5.1 Dataset Preparation for Video Smoke Detection

VisiFire [87] dataset was used in the training and test process. A total of 64 smoke and non-smoke videos were selected from the dataset; 38 videos for training, 11 for validation and 15 for testing. A total of 1000 sub-videos were obtained from dataset videos. First of all, the main videos were divided into sub-videos that each video includes between 40-50 frames. This prosess was performed in order to remove the parts that did not contain smoke from the smoke videos. Dataset was arranged like UCF101 [111] action recognition dataset. After that the dataset was arranged as two classes which are smoke and non-smoke.

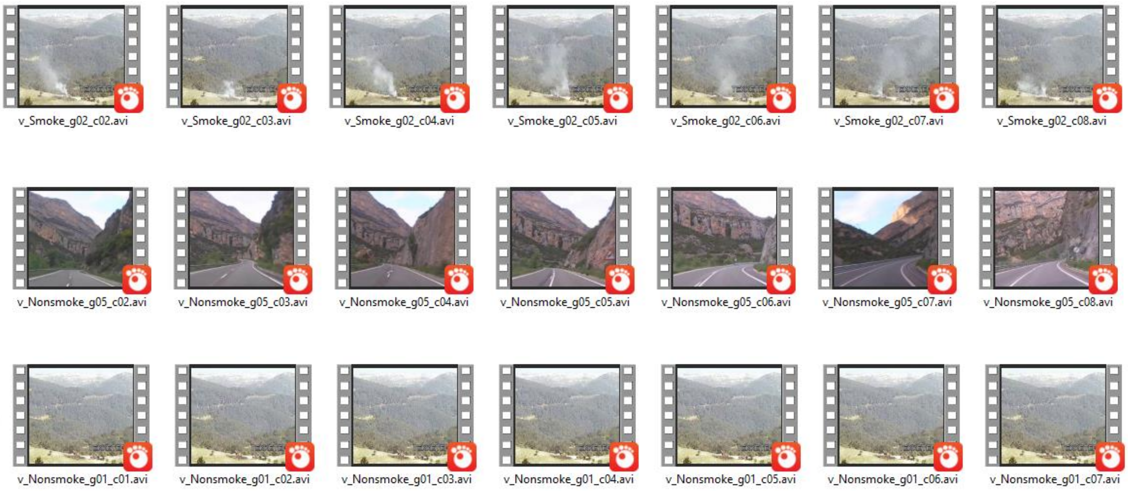


Figure 5.28. Some samples from our created video smoke/non-smoke dataset.

Dataset made up with 405 videos and 413 videos for non-smoke and smoke class, respectively. The dataset was arranged 70% training set, %10 validation and %20 test set in the final step. Figure 5.1 shows that some samples of our created dataset.

5.2 Recurrent Neural Network Based Video Smoke Detection Methods

5.2.1 CNN Feature Extraction

A CNN architecture is utilized to extract spatial features from video sequences. GoogLeNet [49] Inception v3 is a widely-used image recognition model on the ImageNet [112] dataset. Thus, Inception v3 pre-trained on ImageNet was preferred as a transfer learning. Inception v3 network stacks 11 inception modules where each module consists of pooling layers and convolutional filters with rectified linear units as activation function. The final pooling layer treats as a feature extractor when the fully connected (FC) layers remove from the end of Inception v3 structure.

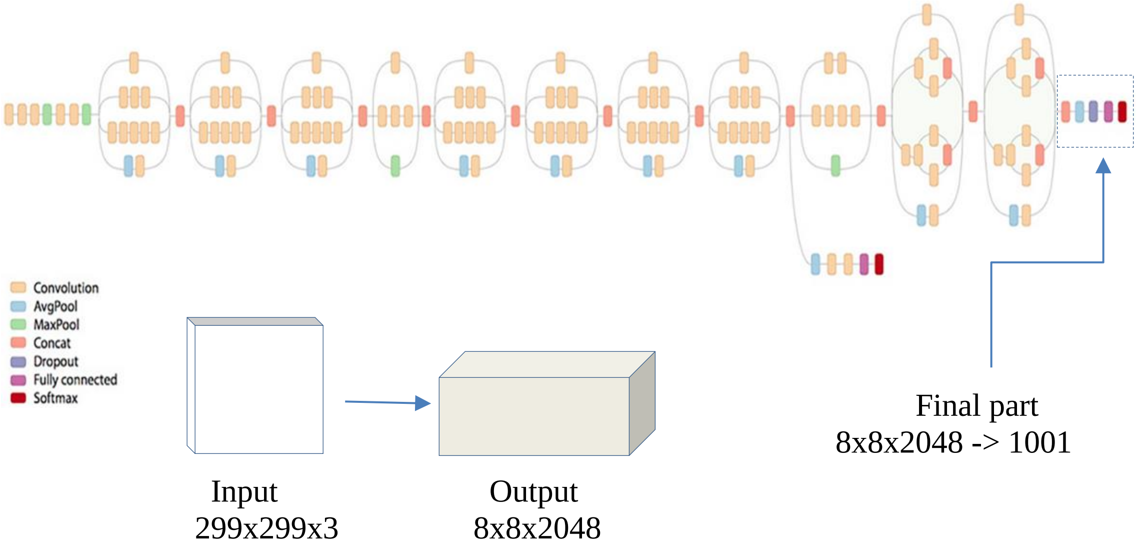


Figure 5.29. Inception v3 Architecture.

Inception network was used to extract features from videos. After the extracted features pass to a separate RNN. First, every frame from every video is run through Inception, saving the output from the final pool layer of the network. Thus, the top classification part of the network is cut off successfully so that 2,048 dimension vector of features can pass to RNN.

5.2.2 Temporal Feature Learning

LSTM, also known as the Long Short Term Memory is an RNN architecture with feedback connections, which enables it to perform or compute anything that a Turing machine can. A single LSTM unit is composed of a cell, an input gate, an output gate and a forget gate, which enables the cell to remember values for an arbitrary amount of time. The gates control the flow of information in and out the LSTM cell.

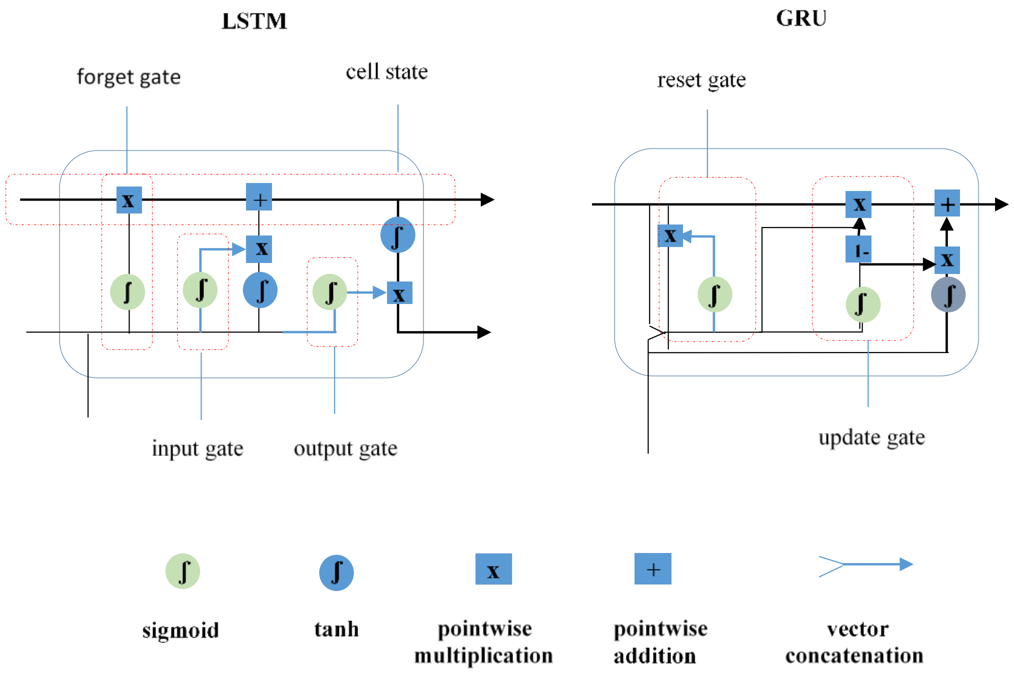


Figure 5.30. LSTM and GRU Structure.

The GRU, known as the Gated Recurrent Unit is an RNN architecture, which is similar to LSTM units. The GRU comprises of the reset gate and the update gate instead of the input, output and forget gate of the LSTM. The reset gate determines how to combine the new input with the previous memory, and the update gate defines how much of the previous memory to keep around. Gated recurrent unit (GRU) is improved to resolve the vanishing gradient problem which come from with standard recurrent neural networks (RNNs).

5.2.3 LSTM Based Smoke Detection Architecture

LSTM structure is used for temporal training in video based detection system. I also used LSTM with CNN structure for smoke detection from video task. For this task, extracted features by using Inception v3 served as input to LSTM blocks. Figure 5.4 demonstrates that our overall network architectural design for the video based smoke/non-smoke classification task.

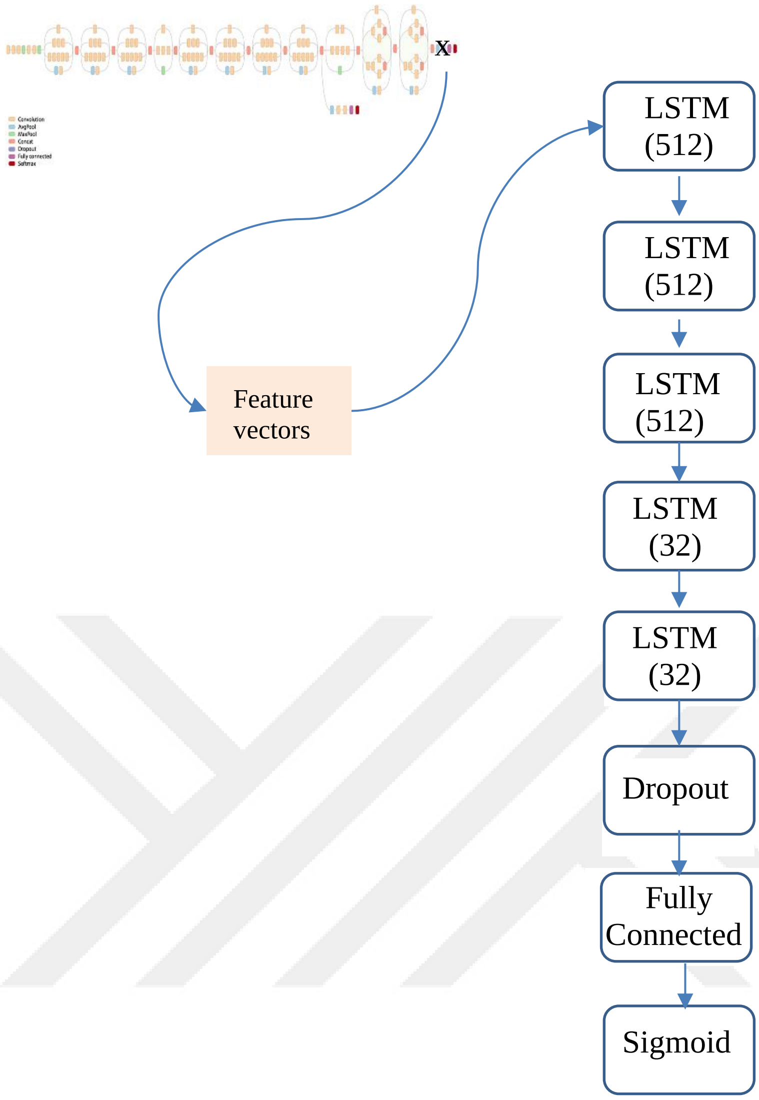


Figure 5.31. Overall video smoke detection network design based on CNN-LSTM Structure.

In this work, five stacked LSTM module, dropout, final dense layer and classification layer were used. In the training process, Adam optimizer with 0.001 initial learning rate, dropout regularization techniques with 40% dropout rate were used. Using LSTM hyperparameters is shown in Table 5.1.

Table 5.6 Hyperparameters of the LSTM based video classification methods.

Initial Learning Rate	Batch Size	Optimizer	Epoch	Regularization Techniques
0.001	16	Adam	100	Early stopping, Dropout

5.2.4 GRU Based Smoke Detection Architecture

GRU structure is also used for temporal training in video based detection system. GRU based system is utilized same way to LSTM. Extracted meaningful features which come from CNN network are used as input of GRU system. GRU network more speedier than LSTM network in the training. As shown in Figure 5.5, The CNN-GRU network was used to classify the smoke/non-smoke from video data.

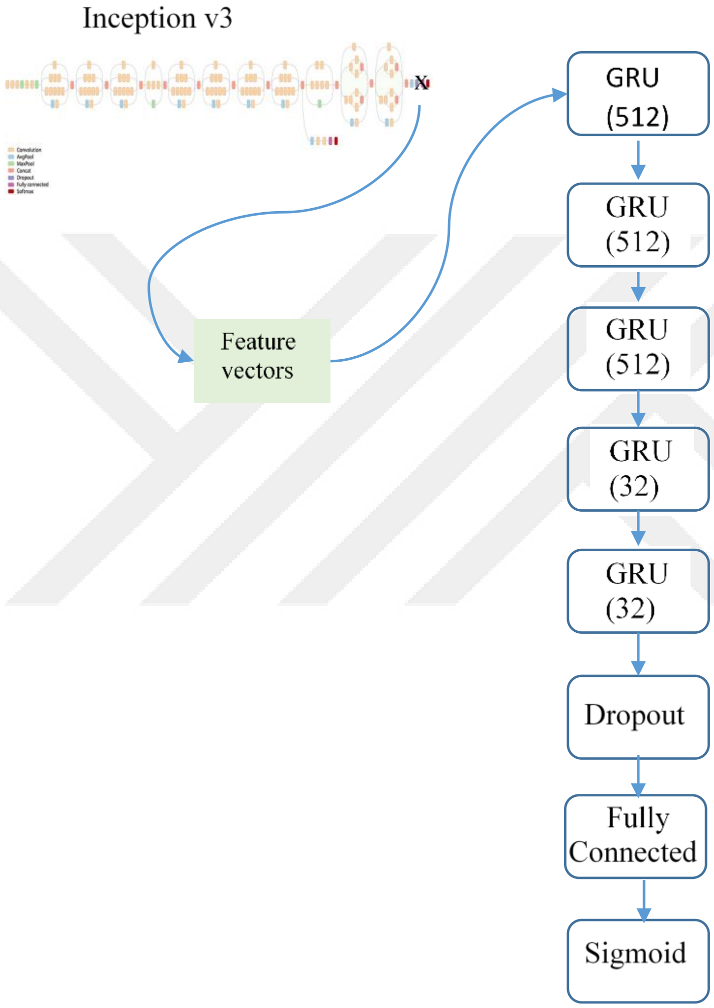


Figure 5.32. Overall video smoke detection network design based on CNN-GRU Structure.

In my work, I proposed five stacked GRU layers, then dropout and dense layer (fully connected) was added to network. Final was classification layer and Sigmoid function was used to predict the two classes.

Table 5.2 Hyperparameters of the GRU based video classification methods.

Initial Learning Rate	Batch Size	Optimizer	Epoch	Regularization Techniques
0.001	16	Adam	100	Early stopping, Dropout

During the training process, using the hyperparameters are shown in Table 5.2. Early stopping and dropout (dropout rate : 40%) were selected to prevent overfitting.

5.2.5 Hybrid Stacked GRU-LSTM Based Smoke Detection Architecture

In literature, GRU and LSTM architectures are mainly used to obtain spatiotemporal information from video. There is not clear winner between these architectures. These structures prefer according to working on the special task. Thus, I proposed hybrid structure which includes GRU and LSTM together for the video smoke detection. Proposed structure includes both GRU and LSTM network with CNN feature extractor. Similarly, Inception v3 was used to extract the meaningful features from the video frames. The extracted features were 2048x10 size for each video sequence due to each sequence comprising of 10 frames. Then the sequence of extracted features were passed through as input of the proposed hybrid structure. Figure 5.6 demonstrates the proposed hybrid smoke detection from video structure. In this structure five GRU and three LSTM blocks with two dropout layers were used to classify the smoke/non-smoke from video.

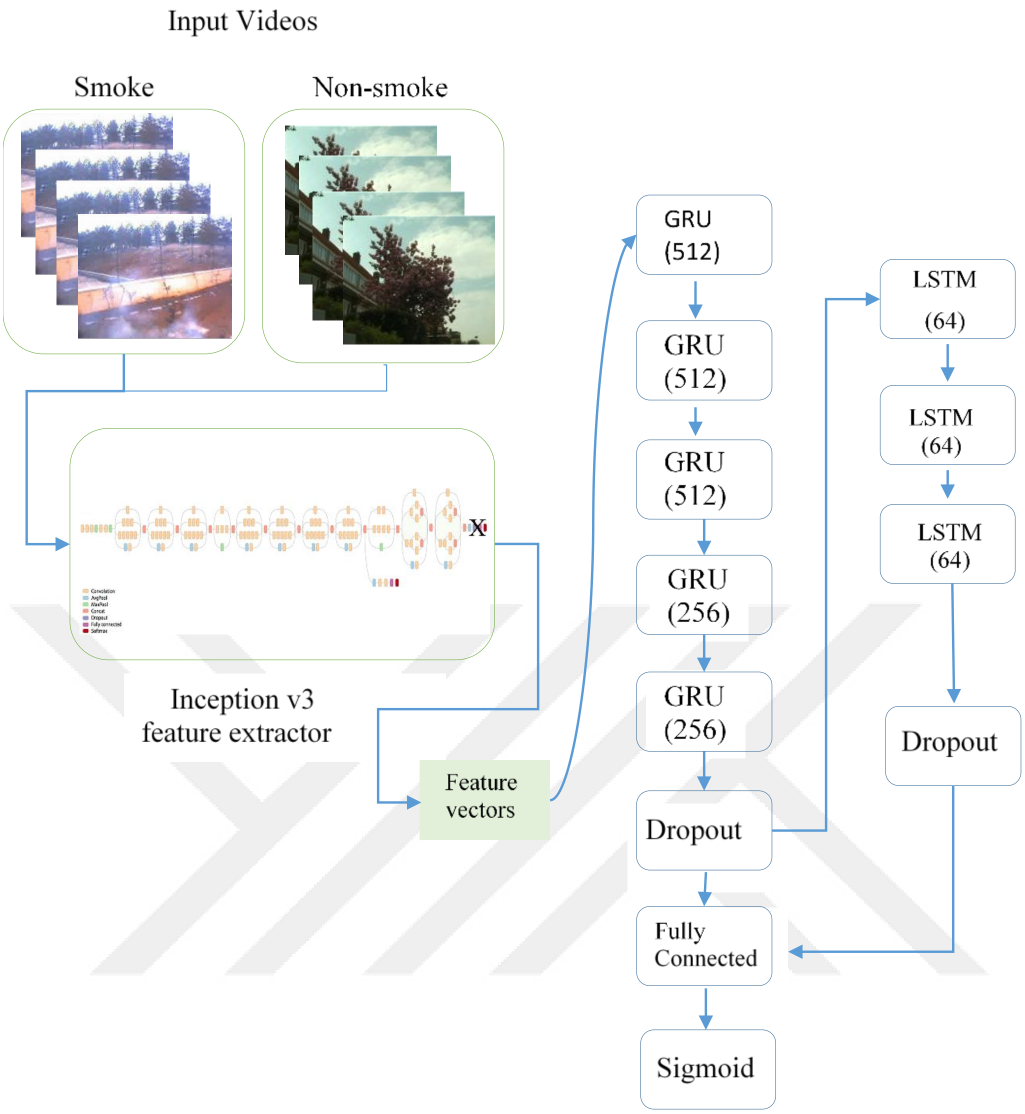


Figure 5.33. Proposed hybrid video smoke detection network design.

5.2.6 Performance Metrics for Evaluation of Video Smoke Detection System

When the smoke classification system correctly identified a smoke according to the ground truth, it was regarded as a true positive. In the false positive case, the system detected the smoke that did not match the ground truth. When the system did not detect the non-smoke in the video frame, it was regarded as a false negative. When the system detect the non-smoke in the video frame, it was regarded as a true negative. Related formulas were given in Chapter 3. The overall accuracy was calculated by using in Equation 5.1.

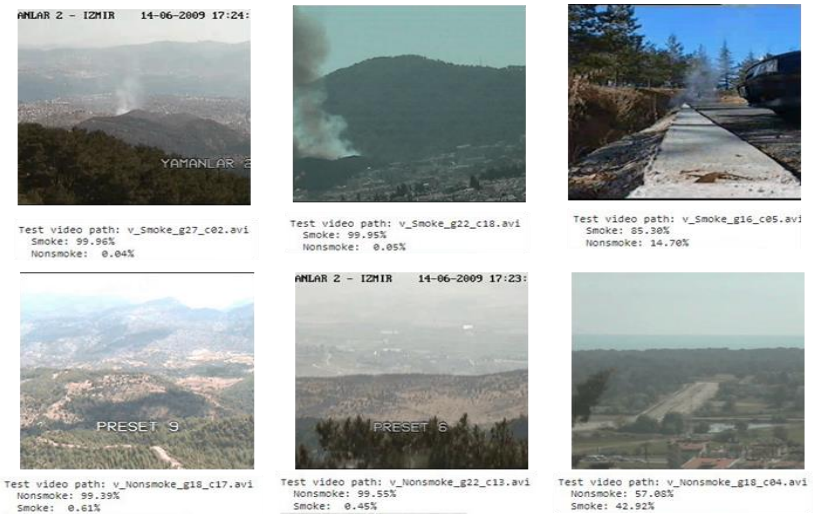
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5.1}$$

According to the results in Table 5.4, the hybrid structure overall accuracy were achieved 90.02%. The proposed hybrid structure was improved system performance with decreasing in FP rate when the LSTM and GRU structures used together.

Table 5.3 Confusion Matrix belong to hybrid structure results.

Confusion Matrix	**Negative (predicted)**	**Positive (predicted)**
Negative (actual)	TN: 67	FP: 3
Positive (actual)	FN: 15	TP: 97

Figure 5.7 demonstrates the results for video smoke detection. First row and second row related to smoke and non-smoke test videos, respectively. Figure shows that correctly prediction of smoke and non-smoke classes. All test videos comprising of between 40-50 frames. Predictions were obtained as an average predictions through every 10 frames.


ANLAR 2 - IZMIR 14-06-2009 17:24:
YAMANLAR 2
ANLAR 2 - IZMIR 14-06-2009 17:23:
PRESET 9
PRESET 6

```
Test video path: v_Smoke_g27_c02.avi
  Smoke: 99.96%
  Nonsmoke:  0.04%
```

```
Test video path: v_Smoke_g22_c18.avi
  Smoke: 99.95%
  Nonsmoke:  0.05%
```

```
Test video path: v_Smoke_g16_c05.avi
  Smoke: 85.30%
  Nonsmoke: 14.70%
```

```
Test video path: v_Nonsmoke_g18_c17.avi
  Nonsmoke: 99.39%
  Smoke:  0.61%
```

```
Test video path: v_Nonsmoke_g22_c13.avi
  Nonsmoke: 99.55%
  Smoke:  0.45%
```

```
Test video path: v_Nonsmoke_g18_c04.avi
  Nonsmoke: 57.08%
  Smoke: 42.92%
```

Figure 5.34. Video smoke detection system results.

Table 5.7 Comparison of all video based structures.

MODEL	CNN-LSTM	CNN-GRU	CNN-LSTM-GRU
Accuracy (%)	0.860	0.879	0.900

Table 5.4 shows that comparison of the three methods for detection of smoke in a video. According to results, hybrid model has better performance in terms of accuracy than the other two methods. Thus, proposed hybrid structure is more convenient for our task on created dataset.

Chapter 6

Conclusions and Future Prospects

Thesis outcomes are evaluated in terms of system performance which based on accuracy and false alarm rate.

6.1 Conclusions

This thesis has introduced many contributions for outdoor smoke detection tasks such as smoke classification, detection-segmentation and smoke classification in video sequences. We have introduced some methodologies such as dataset creating, fine-tuning, pre-processing, and network design.

Three different datasets were created to utilize in the smoke detection methods and the dataset was shared on internet so that the researchers work on the forest fire smoke detection can use it. Image dehazing based pre-processing method was applied to the dataset images to obtain clear images. These haze-free images were used in fine-tuned state-of-the-art CNN classification structures as input. Pre-processing method was improved system performance both smoke classification and detection/segmentation. Gathered data indicates that the removal of haze from images before the training process yielded a better outcome in terms of smoke classification and localization accuracy. Removal of the fog from images provided higher accuracy for both the classification and the segmentation tasks even when a small dataset is used.

Smoke detection from surveillance videos is a more challenging task than the image based detection task due to the video streams that need to be analyzed for the spatio-temporal information. We proposed a model that classifies the surveillance videos as smoke and non-smoke. We utilized a hybrid structure that is a combination of CNN and RNN to obtain of spatio-temporal information. The performance of designed hybrid structure was determined to be more accurate as compared to the individual LSTM and GRU models.

The experimental results clearly indicate that the proposed methods achieve a high smoke detection performance on forest fire image datasets. Proposed structures that are conducted to our datasets achieved high accuracy rate and low false alarm rate.

Our approach is feasible for current and future smoke-related computer vision tasks that are specifically prone to conditions of hazy environments. Researchers that work on wildfire smoke detection can improve their systems if they employ the proposed method presented in this thesis.

6.2 Societal Impact and Contribution to Global Sustainability

Wildfires are a growing threat throughout the world and among the foremost devastating natural disasters that can have immediate and long term effects on environment and population, consequently an immense impact on global economy. This thesis will help to reduce the adversarial effect of wildfires by developing an early forest fire detection system that is fast, feasible, accurate, and versatile.

Forests are a part of life. Yet, fire can be fatal, burning buildings, forests, and habitat while also contaminating the air with hazardous fumes to people's health. Carbon dioxide, a significant greenhouse gas, is also released into the atmosphere by fire. Fire effects can be long-lasting and are impacted by forest conditions prior to the fire as well as management actions.

Moreover, commercially 50% of forests are for energy and 28% for construction; only 13% is used for paper production, which is known as the most popular usage area. Therefore, forests, beyond their environmental impact, have an important socio-economic importance with the ecosystem services and wave effect they create, and it is critical that they be sustainable.

In this point, developing an effective forest fire detection system at an early stage benefit many problems such as global warming, greenhouse gas effects, drought and degradation of air quality, ecosystem, biodiversity. Sustainable use of forests provides shelter, fuel, medicine and other services for people who depend on this environment. It is a habitat for all plants and animals and helps against climate change.

This thesis is related to the fifteenth United Nation's Sustainable Development Goal titled "Life on Land", and target (15.1) corresponds to conservation, "By 2020, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements", determination of suitable forest fire detection structure will contribute the this goal.

As the second target (15.2) is corresponds to "By 2020, promote the implementation of sustainable management of all types of forests, halt deforestation, restore degraded forests and substantially increase afforestation and reforestation globally", our proposed study support this goal.

As the second goal (13.1) is titled "Climate Action", and corresponds to "Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries", the thesis will provide robust model for natural disasters like forest fires.

6.3 Future Prospects

The current work can be expanded by using suitable future research. In recent years, natural language processing (NLP) based methods have been quite remarkable due to their success. Some of them have also started to be used in image processing fields.

Several state-of-the-art fine-tuned CNN structures and transformer based classification structures were implemented for the smoke detection tasks in this study. NLP based structures can be implemented for smoke detection tasks such as Transformer based models.

In this thesis, image dehazing based pre-processing techniques were presented to obtain the clear dataset images. Therefore, different image pre-processing techniques can be examined to improve the dataset.

This research supports an idea that design the hybrid structures. Other DL based video smoke detection structures can be combined to improve the accuracy of fire safety model.

BIBLIOGRAPHY

[1] K. Hoover, and L. Hanson, Wildfire Statistics Congressional Research Service, 2018.

[2] H. Clarke, B. Cirulis, T. Penman, O. Price, M.M. Boer, and R. Bradstock, “The 2019–2020 Australian forest fires are a harbinger of decreased prescribed burning effectiveness under rising extreme conditions. ” *Scientific reports*, 2022. 12(1): p. 1-10.

[3] T.C Tarım Ve Orman Bakanlığı Orman Genel Müdürlüğü. Orman Genel Müdürlüğü İstatikler. Available from: https://www.ogm.gov.tr/tr/e-kutuphane/resmi-istatistikler (October 2022).

[4] F. Yuan, J. Shi, X. Xia, Y. Fang, Z. Fang, and T. Mei, "High-order local ternary patterns with locality preserving projection for smoke detection and image classification." *Information Sciences*, 2016. 372: p. 225-240.

[5] A.E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y.H. Habiboğlu, B.U. Töreyin, and S. Verstockt, "Video fire detection–Review." *Digital Signal Processing*, 2013. 23(6): p. 1827-1843.

[6] J. Schmidhuber, "Deep learning in neural networks: An overview." *Neural networks*, 2015. 61: p. 85-117.

[7] S. Chaturvedi, P. Khanna, and A. Ojha, "A survey on vision-based outdoor smoke detection techniques for environmental safety." *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022. 185: p. 158-187.

[8] F. Yuan, J. Shi, X. Xia, Y. Yang, Y. Fang, and R. Wang, "Sub oriented histograms of local binary patterns for smoke detection and texture classification." *KSII Transactions on Internet and Information Systems (TIIS)*, 2016. **10**(4): p. 1807-1823.

[9] Chunyu, Y., F. Jun, W. Jinjun, and Z. Yongming, *Video fire smoke detection using motion and color features*. Fire technology, 2010. 46(3): p. 651-663.

[10] M.R. Islam, M. Amiruzzaman, S. Nasim, and J. Shin, "Smoke object segmentation and the dynamic growth feature model for video-based smoke detection systems." *Symmetry*, 2020. 12(7): p. 1075.

[11] P. Barmpoutis, K. Dimitropoulos, and N. Grammalidis. "Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition." in *2014 22nd European Signal Processing Conference (EUSIPCO)*. 2014. IEEE.

[12] H.Tian, W. Li, P.O. Ogunbona, and L. Wang, "Detection and separation of smoke from single image frames." I*EEE Transactions on Image Processing*, 2017. 27(3): p. 1164-1177.

[13] M. Torabnezhad, and A. Aghagolzadeh. "Visible and IR image fusion algorithm for short range smoke detection." in *2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*. 2013. IEEE.

[14] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, "Higher order linear dynamical systems for smoke detection in video surveillance applications." *IEEE Transactions on Circuits and Systems for Video Technology*, 2016. 27(5): p. 1143-1154.

[15] Y. Cui, H. Dong, and E. Zhou. "An early fire detection method based on smoke texture analysis and discrimination." in *2008 Congress on Image and Signal Processing*. 2008. IEEE.

[16] F. Yuan, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids." *Fire safety journal*, 2011. 46(3): p. 132-139.

[17] T.A. Vidal-Calleja, and G. Agammenoni. "Integrated probabilistic generative model for detecting smoke on visual images." in *2012 IEEE International Conference on Robotics and Automation*. 2012. IEEE.

[18] C.-C Ho, "Nighttime fire/smoke detection system based on a support vector machine." *Mathematical Problems in Engineering*, 2013.

[19] P. Brodatz, "Textures: A Photographic Album for Artists and Designers, Dover Pubns." http://www. ux. uis. no/tranden/brodatz. html, 1999.

[20] X. Li, W. Song, L. Lian, and X. Wei, "Forest fire smoke detection using back-propagation neural network based on MODIS data." *Remote Sensing*, 2015. **7**(4): p. 4473-4498.

[21] Z. Li, A. Khananian, R.H. Fraser, and J. Cihlar, "Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery." *IEEE Transactions on geoscience and remote sensing*, 2001. 39(9): p. 1859-1870.

[22] F.A. Hossain, Y. Zhang, C. Yuan, and C.-Y. Su. "Wildfire flame and smoke detection using static image features and artificial neural network." in *2019 1st international conference on industrial artificial intelligence (iai)*. 2019. IEEE.

[23] A. Gagliardi, F. de Gioia, and S. Saponara, "A real-time video smoke detection algorithm based on Kalman filter and CNN." *Journal of Real-Time Image Processing*, 2021. 18(6): p. 2085-2095.

[24] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, "Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection." *IEEE transactions on circuits and systems for video technology*, 2014. 25(2): p. 339-351.

[25] A. Gagliardi, and S. Saponara, "Advised: Advanced video smoke detection for real-time measurements in antifire indoor and outdoor systems." *Energies*, 2020. 13(8): p. 2098.

[26] F.A. Hossain, Y.M. Zhang, and M.A. Tonima, "Forest fire flame and smoke detection from UAV-captured images using fire-specific color features and multi-color space local binary pattern." *Journal of Unmanned Vehicle Systems*, 2020. 8(4): p. 285-309.

[27] C. Long, J. Zhao, S. Han, L. Xiong, Z. Yuan, J. Huang, and W. Gao. "Transmission: a new feature for computer vision based smoke detection." in *International Conference on Artificial Intelligence and Computational Intelligence*. 2010. Springer.

[28] X. Wang, A. Jiang, and Y. Wang. "A segmentation method of smoke in forest-fire image based on fbm and region growing." in *2011 Fourth International Workshop on Chaos-Fractals Theories and Applications*. 2011. IEEE.

[29] Y. Li, Y. Zhu, and A. Vodacek, "An unsupervised statistical segmentation algorithm for fire and smoke regions extraction." 2005.

[30] D. Xing, Y. Zhongming, W. Lin, and L. Jinlan, "Smoke image segmentation based on color model." *Journal on Innovation and Sustainability RISUS*, 2015. **6**(2): p. 130-138.

[31] D. Xiong, and L. Yan, "Early smoke detection of forest fires based on SVM image segmentation." *Journal of Forest Science*, 2019. 65(4): p. 150-159.

[32] B.U. Töreyin, Y. Dedeoğlu, and A.E. Cetin. "Wavelet based real-time smoke detection in video." in *2005 13th European signal processing conference*. 2005. IEEE.

[33] T.-H. Chen, Y.-H. Yin, S.-F. Huang, and Y.-T. Ye. "The smoke detection for early fire-alarming system base on video processing." in *2006 international conference on intelligent information hiding and multimedia*. 2006. IEEE.

[34] Z. Xu, and J. Xu. "Automatic fire smoke detection based on image visual features." in *2007 International Conference on Computational Intelligence and Security Workshops (CISW 2007)*. 2007. IEEE.

[35] Y. Chunyu, Z. Yongming, F. Jun, and W. Jinjun. "Texture analysis of smoke for real-time fire detection." in *2009 second international workshop on computer science and engineering*. 2009. IEEE.

[36] computer vision based fire detection software. *fire detection system*. 2019, september 12; Available from: http://signal.ee.bilkent.edu.tr/VisiFire/.

[37] B.U. Toreyin, and A.E. Cetin. "Wildfire detection using LMS based active learning." in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009. IEEE.

[38] T.X. Tung, and J.-M. Kim, "An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems." *Fire Safety Journal*, 2011. 46(5): p. 276-282.

[39] R.D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2013. 43(4): p. 1003-1012.

[40] F. Yuan, L. Zhang, X. Xia, Q. Huang, and X. Li, "A gated recurrent network with dual classification assistance for smoke semantic segmentation." *IEEE Transactions on Image Processing*, 2021. 30: p. 4409-4422.

[41] VISOR, https://aimagelab.ing.unimore.it/visor/ (October 2022).

[42] N. Alamgir, K. Nguyen, V. Chandran, and W. Boles, "Combining multi-channel color space with local binary co-occurrence feature descriptors for accurate smoke detection from surveillance videos." *Fire safety journal*, 2018. 102: p. 1-10.

[43] B.C. Ko, K.-H. Cheong, and J.-Y. Nam, "Fire detection based on vision sensor and support vector machines." *Fire Safety Journal*, 2009. 44(3): p. 322-329.

[44] B.U. Toreyin, Y. Dedeoglu, and A.E. Cetin. "Contour based smoke detection in video using wavelets." in *2006 14th European signal processing conference*. 2006. IEEE.

[45] X. Wu, Y. Cao, X. Lu, and H. Leung, "Patchwise dictionary learning for video forest fire smoke detection in wavelet domain." *Neural Computing and Applications*, 2021. 33(13): p. 7965-7977.

[46] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, 2012. 25: p. 1097-1105.

[47] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[50] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger. "Densely connected convolutional networks." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[52] Z. Yin, B. Wan, F. Yuan, X. Xia, and J. Shi, "A deep normalization and convolutional neural network for image smoke detection." *IEEE Access*, 2017. **5**: p. 18429-18438.

[53] A. Filonenko, L. Kurnianggoro, and K.-H. Jo. "Comparative study of modern convolutional neural networks for smoke detection on image data. in 2017 10th international conference on human system interactions (HSI)." 2017. *IEEE*.

[54] C. Tao, J. Zhang, and P. Wang. "Smoke detection based on deep convolutional neural networks." in *2016 International conference on industrial informatics-computing technology, intelligent technology, industrial information integration (ICIICII)*. 2016. IEEE.

[55] A. Namozov, and Y. Im Cho, "An efficient deep learning algorithm for fire and smoke detection with limited data*." Advances in Electrical and Computer Engineering*, 2018. 18(4): p. 121-128.

[56] 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*, 2020.

[57] K. Gu, Z. Xia, J. Qiao, and W. Lin, "Deep dual-channel neural network for image-based smoke detection." *IEEE Transactions on Multimedia*, 2019. 22(2): p. 311-323.

[58] M. Liu, X. Xie, G. Ke, and J. Qiao, "Simple and efficient smoke segmentation based on fully convolutional network*." DEStech Trans. Comput. Sci. Eng.(ica)*. https://doi.org/10.12783/dtcse/ica2019/30707, 2019.

[59] F. Zhang, W. Qin, Y. Liu, Z. Xiao, J. Liu, Q. Wang, and K. Liu, "A Dual-Channel convolution neural network for image smoke detection." *Multimedia Tools and Applications*, 2020. 79(45): p. 34587-34603.

[60] S. Khan, K. Muhammad, S. Mumtaz, S.W. Baik, and V.H.C. de Albuquerque, "Energy-efficient deep CNN for smoke detection in foggy IoT environment." *IEEE Internet of Things Journal*, 2019. 6(6): p. 9237-9245.

[61] Muhammad, K., S. Khan, V. Palade, I. Mehmood, and V.H.C. De Albuquerque, *Edge intelligence-assisted smoke detection in foggy surveillance environments.* IEEE Transactions on Industrial Informatics, 2019. 16(2): p. 1067-1075.

[62] L. He, X. Gong, S. Zhang, L. Wang, and F. Li, "Efficient attention based deep fusion CNN for smoke detection in fog environment*." Neurocomputing*, 2021. 434: p. 224-238.

[63] R. Ba, C. Chen, J. Yuan, W. Song, and S. Lo, "*S*mokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention." *Remote Sensing*, 2019. 11(14): p. 1702.

[64] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[65] J. Redmon, and A. Farhadi. "YOLO9000: better, faster, stronger." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[66] S. Ren, K. He, R.B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks.*" CoRR abs/1506.01497 (2015).* arXiv preprint arXiv:1506.01497, 2015.

[67] J. Redmon, and A. Farhadi, "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767, 2018.

[68] J. Zeng, Z. Lin, C. Qi, X. Zhao, and F. Wang. "An improved object detection method based on deep convolution neural network for smoke detection." in *2018 international conference on machine learning and cybernetics (ICMLC)*. 2018. IEEE.

[69] S. Wu, and L. Zhang. "Using popular object detection methods for real time forest fire detection." in *2018 11th International symposium on computational intelligence and design (ISCID)*. 2018. IEEE.

[70] G. Xu, Q. Zhang, D. Liu, G. Lin, J. Wang, and Y. Zhang, "Adversarial adaptation from synthesis to reality in fast detector for smoke detection." *IEEE Access,* 2019. **7**: p. 29471-29483.

[71] Z., Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, and D. Liu. "A deep learning based forest fire detection approach using UAV and YOLOv3." in *2019 1st International conference on industrial artificial intelligence (IAI)*. 2019. IEEE.

[72] J. Shi, W. Wang, Y. Gao, and N. Yu, "Optimal placement and intelligent smoke detection algorithm for wildfire-monitoring cameras.*" IEEE Access,* 2020. **8**: p. 72326-72339.

[73] P. Li, and W. Zhao, "Image fire detection algorithms based on convolutional neural networks.*" Case Studies in Thermal Engineering,* 2020. **19**: p. 100625.

[74] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems.*" Journal of Real-Time Image Processing,* 2021. 18(3): p. 889-900.

[75] S. Frizzi, M. Bouchouicha, J.M. Ginoux, E. Moreau, and M. Sayadi, "Convolutional neural network for smoke and fire semantic segmentation." *IET Image Processing,* 2021. 15(3): p. 634-647.

[76] S. Khan, K. Muhammad, T. Hussain, J. Del Ser, F. Cuzzolin, S. Bhattacharyya, Z. Akhtar, and V.H.C. de Albuquerque, "Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments.*" Expert Systems with Applications,* 2021. 182: p. 115125.

[77] A. Larsen, I. Hanigan, B.J. Reich, Y. Qin, M. Cope, G. Morgan, and A.G. Rappold, "A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication." *Journal of exposure science & environmental epidemiology,* 2021. 31(1): p. 170-176.

[78] C. Hu, P. Tang, W. Jin, Z. He, and W. Li. "Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method." in *2018 37th Chinese Control Conference (CCC)*. 2018. IEEE.

[79] M.D. Nguyen, D. Kim, and S. Ro, "A video smoke detection algorithm based on cascade classification and deep learning." *KSII Transactions on Internet and Information Systems (TIIS),* 2018. 12(12): p. 6018-6033.

[80] S. Aslan, U. Güdükbay, B.U. Töreyin, and A.E. 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. IEEE.

[81] X. Yang, and Y. Sun. "Research on smoke detection based on densenet." in *Proceedings of the 2019 ACM Southeast Conference*. 2019.

[82] F. Shi, H. Qian, W. Chen, M. Huang, and Z. Wan. "A fire monitoring and alarm system based on YOLOv3 with OHEM." in *2020 39th Chinese Control Conference (CCC)*. 2020. IEEE.

[83] H. Pan, D. Badawi, and A.E. Cetin. "Fourier domain pruning of mobilenet-v2 with application to video based wildfire detection." in *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021. IEEE.

[84] H. Pan, D. Badawi, and A.E. Cetin, "Computationally efficient wildfire detection method using a deep convolutional network pruned via fourier analysis." *Sensors*, 2020. 20(10): p. 2891.

[85] B. Kim, and J. Lee, "A video-based fire detection using deep learning models." *Applied Sciences*, 2019. 9(14): p. 2862.

[86] Y. Zhao, Y. Ban, and A. Nascetti. "Early Detection of Wildfires with GOES-R Time-Series and Deep GRU Network." in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 2021. IEEE.

[87] A. Çetin, Computer vision based fire detection software, 2007.

[88] M. Tan, and Q. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks. in International conference on machine learning." 2019. PMLR.

[89] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929, 2020.

[90] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. "Swin transformer: Hierarchical vision transformer using shifted windows." in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

[91] D. Berman, T. Treibitz, and S. Avidan. "Air-light estimation using haze-lines." in *2017 IEEE International Conference on Computational Photography (ICCP)*. 2017. IEEE.

[92] J. Sumitha, J. Miruthula, and N. Nikhila, "Haze Removal Techniques in Image Processing." *International Journal for Scientific Research & Development*|, 2021. 9(2): p. 144-146.

[93] R. Mondal, S. Santra, and B. Chanda. "Image dehazing by joint estimation of transmittance and airlight using bi-directional consistency loss minimized FCN." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

[94] W. Wang, F. Chang, T. Ji, and X. Wu, "A fast single-image dehazing method based on a physical model and gray projection." *IEEE Access*, 2018. **6:** p. 5641-5653.

[95] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. "Efficient image dehazing with boundary constraint and contextual regularization." in *Proceedings of the IEEE international conference on computer vision*. 2013.

[96] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior." I*EEE transactions on image processing*, 2015. 24(11): p. 3522-3533.

[97] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. "Single image dehazing via multi-scale convolutional neural networks." in *European conference on computer vision*. 2016. Springer.

[98] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal." *IEEE Transactions on Image Processing*, 2016. 25(11): p. 5187-5198.

[99] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "An all-in-one network for dehazing and beyond." arXiv preprint arXiv:1707.06543, 2017.

[100] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang. "Gated fusion network for single image dehazing." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[101] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing*, 2004. 13(4): p. 600-612.

[102] N. Zhang, L. Zhang, and Z. Cheng. "Towards simulating foggy and hazy images and evaluating their authenticity." in *International Conference on Neural Information Processing*. 2017. Springer.

[103] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A.C. Berg. "Ssd: Single shot multibox detector." in *European conference on computer vision*. 2016. Springer.

[104] M. Taş, and B. Yılmaz, "Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images." *Computers & Electrical Engineering*, 2021. 90: p. 106959.

[105] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. "Feature pyramid networks for object detection." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[106] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo algorithm developments." *Procedia Computer Science*, 2022. 199: p. 1066-1073.

[107] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick. "Microsoft coco: Common objects in context. in European conference on computer vision." 2014. Springer.

[108] facebook research. About fire detection 2020, November 14, ; Available from: https://github.com/facebookresearch/detectron2.

[109] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN", *in: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy.* 2017.

[110] A. Graves, A.-R. Mohamed, and G. Hinton. "Speech recognition with deep recurrent neural networks." in *2013 IEEE international conference on acoustics, speech and signal processing*. 2013. Ieee.

[111] K. Soomro, A.R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild." arXiv preprint arXiv:1212.0402, 2012.

[112] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "Imagenet large scale visual recognition challenge." *International journal of computer vision*, 2015. 115(3): p. 211-252.

CURRICULUM VITAE

2009 – 2013	B.Sc., Electrical-Electronics Engineering, Erciyes University, Kayseri, TURKEY
2013 – 2016	M.Sc., Electrical-Electronics, Erciyes University, Kayseri, TURKEY
2017–2022	Ph.D., Electrical and Computer Engineering, Abdullah Gul University, Kayseri, TURKEY.

SELECTED PUBLICATIONS

J1) M. Taş and B. Yılmaz, Super Resolution Convolutional Neural Network Based Pre-Processing for Automatic Polyp Detection in Colonoscopy Images, published in Journal of Computers & Electrical Engineering (March 2021).

J2) M. Taş, Y. Taş, O. Balki, Z. Aydın and K. Taşdemir, Camera Based Wildfire Smoke Detection for Foggy Environments, published in Journal of Electronic Images (October 2022).