Gurukul International Multidisciplinary Research Journal (GIMRJ)*with*
International Impact Factor 8.357
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ/2503I01S01V13P0009

e-ISSN No. 2394-8426
Special Issue on Scientific Research Mar'25
Issue–III(I), Volume–XIII

Unsupervised Classification and Detection of Forest Fire Images using Convolutional Autoencoder

V Tharun
Department of NWC
Faculty of Engineering and Technology, SRM Institute of Science and Technology, KTR campus, Chennai, India
vv3078@srmist.edu.in
S Navin Kumar
Department of NWC
Faculty of Engineering and Technology, SRM Institute of Science and Technology, KTR campus, Chennai, India
ns8889@srmist.edu.in
C Hemanth Varma
Department of NWC
Faculty of Engineering and Technology, SRM Institute of Science and Technology, KTR campus, Chennai, India
cc8658@srmist.edu.in
Dr P Visalakshi
Associate Professor
Department of NWC
Faculty of Engineering and Technology, SRM Institute of Science and Technology, KTR campus, Chennai, India
visalakp@srmist.edu.in

Abstract— Forests are essential ecosystems, providing habitats and resources for diverse species. Existing research has largely relied on supervised methods for classification and detection, but annotating large datasets remains a major challenge. Traditional forest fire detection methods, primarily reliant on sensors such as infrared and smoke detectors, often exhibit limited performance in accurately identifying fire events. In this study, we propose a novel approach leveraging the "ClusterMask Enhancer" to effectively segment and localize forest fires in aerial imagery. This method enables precise identification of fire-affected regions, empowering firefighters to strategize their response and mitigate fire spread more efficiently. This project addresses this gap by employing unsupervised autoencoder-based learning and segmentation, utilizing large sets of unlabeled data to minimize the need for extensive human labeling. The results show that the proposed model outperforms state-of-the-art models, showcasing its superior performance.

Key Words — *Unsupervised Learning, Autoencoder, Data Annotation, Segmentation*

1.Introduction

Wildfires refer to large, uncontrolled fires that spread rapidly across natural landscapes such as forests, grasslands, or shrublands. These fires are caused by a combination of natural factors, such as lightning strikes, and human-induced activities, including deforestation and agricultural burning. They pose significant challenges, such as loss of biodiversity, increased carbon emissions, and risks to human life and infrastructure. Addressing these impacts requires accurate and timely detection methods [3] to identify wildfires in their early stages.

Current studies primarily rely on supervised learning techniques for classification and detection tasks [1]. These methods depend heavily on labelled datasets, making them less effective in scenarios where labelled data is scarce or unavailable. Furthermore, supervised approaches [6] are often designed for specific datasets, limiting their ability to generalize to new or evolving wildfire patterns. By focusing on classification and detection, these studies [1] are constrained to identifying predefined categories, which can result in suboptimal performance due to the unavailability of labelled data

The unsupervised segmentation [7] approach uses convolutional autoencoders (CAE), a widely used method in unsupervised learning due to their ability to extract meaningful features in the latent space. In segmenting input images [3], the most important factor is the latent space, which is generated by the encoder. This latent space captures the key features of the input image, providing a compressed representation. These extracted features are then used to reconstruct and analyze the image during the decoding phase, allowing for accurate segmentation without the need for labeled data

To facilitate the aforementioned rationale, a novel context-driven unsupervised approach for forest fire segmentation via convolutional autoencoders (CAE) [2] has been proposed in this work. Various clustering methods are employed to segment the latent representations of the input images. The unsupervised segmentation [6] approach utilizes two key models: the convolutional autoencoder (CAE) model and the Cluster Mask Enhancer model, which is specifically used for analyzing and refining the clustered images. By combining these two models, the approach focuses on extracting the most relevant features from the spatially inputted data using an unsupervised method [2], thus eliminating the need for labelled data. Moreover, the unsupervised segmentation technique [7] for forest fire detection, via convolutional autoencoders, has been successfully employed to enhance segmentation performance. The major contributions of this paper are as follows:

The major contributions of the paper are as follows:

- Proposal of an unsupervised approach for segmentation using convolutional autoencoders, which leverages large unlabeled datasets without the need for extensive human labelling.
- Development of the Cluster Mask Enhancer model, which is used for clustering the images and performing in-depth analysis of the segmented images, helping to achieve high accuracy in the unsupervised approach.
- Comprehensive analysis of the Corsican dataset, which contains spatial images of forest fires. The analysis, conducted using the Cluster Mask Enhancer model, enables accurate segmentation of the input images.

The rest of the paper is organized as follows. The related works are described in Section 2. methodology in Section 3. In Section 4 and Section 5, the experimental setup and the results are discussed in detail. Finally, the summary of the paper and some future plans are enumerated in Section 6.

2.Related Works

Previous research and studies on forest fire analysis have predominantly relied on supervised learning approaches [1], such as classification and detection methods. These supervised techniques, while effective, demand extensive labelled datasets, which are often time-consuming, expensive, and challenging to acquire, particularly for large-scale or remote forest environments. To address some of these challenges, unsupervised methods have been explored as alternatives. One notable approach involves the use of GK-RGB-based clustering [6], which operates on visual spectrum data. This method applies the Gustafson-Kessel (GK) [8] clustering algorithm for segmentation, leveraging the RGB color space to distinguish fire regions. However, its reliance on static images and sensitivity to complex backgrounds can limit its robustness in dynamic or noisy environments.

Another approach uses infrared (IR) video data for segmentation [4], capitalizing on thermal imaging to identify fire regions. This method is designed to process sequential frames from video data, enabling it to capture temporal dynamics of fire spread. While effective in certain contexts, it is restricted by its dependency on thermal imaging, which requires specialized equipment and is not always feasible in cost-sensitive or large-scale scenarios.

Semi-supervised approaches [1] have been explored in the domain of forest fire classification and segmentation to reduce the reliance on extensive labelled datasets. One such method involves the use of autoencoders, which are particularly effective for feature extraction and representation learning. By leveraging autoencoders, these approaches minimize the need for large amounts of annotated data while maintaining robust performance. Additionally, models like the KV framework have been utilized for segmenting thermal infrared images [4]. These models not only perform segmentation but also help in verifying the segmented outputs, ensuring higher accuracy in identifying fire regions. Furthermore, several fully convolutional models, such as U-Net and PSPNet, have been extensively implemented in image segmentation tasks [5]. These architectures are well-suited for extracting spatial features and generating detailed segmentation maps. Their ability to capture both global and local context makes them ideal for tasks involving complex environmental data, such as identifying fire-prone regions or delineating fire boundaries in various conditions.

In our proposed work, we adopted an autoencoder-based [2] unsupervised approach for the segmentation of forest fires. To enhance feature extraction, we incorporated a channel attention mechanism within the encoder, enabling the model to learn the most informative latent representations from the input images and additionally, we developed a cluster-enhancer model where the latent representations are segmented by applying various clustering methods, such as Gaussian Mixture Models (GMM) and K-Means. The resulting clustered image is then masked to facilitate in-depth analysis. This approach allows for comprehensive evaluation and comparison with state-of-the-art methods

3. Methodology: convolutional Autoencoder & Cluster Mask Enhancer

In this section, the convolutional autoencoder with unsupervised segmentation is explained, and all the various clustering methods and data pre-processing techniques are described.

3.1 Convolutional Autoencoders (CAE):

An Autoencoder is a type of neural network that contains 2 parts-encoder and decoder [2]. The encoder compresses the input into a lower dimensional space called the latent space. The decoder reconstructs the input images from the latent space

The Autoencoder uses the convolutional layers for extracting the features and correlating the spatial information within the image. Convolutional neural networks (CNNs) are commonly used in computer vision for image processing tasks.

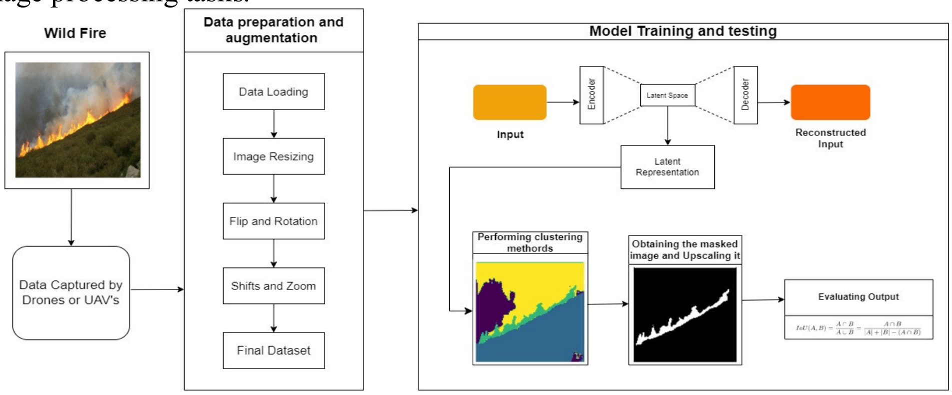


Figure 1: workflow diagram

Gurukul International Multidisciplinary Research Journal (GIMRJ)*with*
International Impact Factor 8.357
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ/2503I01S01V13P0009
e-ISSN No. 2394-8426
Special Issue on Scientific Research Mar'25
Issue–III(I), Volume–XIII

This hierarchical structure enables the CAE to capture complex patterns and spatial hierarchies, providing a better representation of the input data. By using convolutional layers, the CAE [3] can learn a more robust set of features for image segmentation.

The encoder function maps the original data to a latent space, located in the bottleneck layer. The decoder function then maps the latent space back to the output, which aims to replicate the original input. The goal of the algorithm is to reconstruct the original image after applying a generalized nonlinear compression. While training, we used color aware regularization [4] which focuses on the parts of the images that contain the characteristic red-amber color that fires have. Autoencoders have proven effective in various image processing tasks, such as image denoising and restoration.

3.1.1 Channel Attention

channel attention was implemented [8] within the encoder of the convolutional autoencoder to prioritize the most significant channels across the frames. This mechanism allows the model to focus on channels that carry critical and common information, enhancing the quality of the learned latent representations. Given input features as follows

$$x \in R^{C \times H \times W} \qquad (1)$$

Where C, H and W denote the number of channels, height and width and channel attention was implemented using a squeezed-and-Excitation (SE) mechanism to achieve this, **Global Average Pooling (GAP)** [8] was applied to the feature maps derived from the convolutional layers. GAP effectively summarizes each channel by averaging its spatial information, reducing the feature maps to a feature vector $z \in R^{c}$ where each element is computed as follows:

$$z_C = \frac{1}{H \times \omega} \sum_{I=1}^{H} \sum_{j=1}^{W} X_{c,i,j} \qquad (2)$$

This feature vector was then passed through a series of linear layers, with a Rectified Linear unit (ReLU) activation followed by a sigmoid activation function [3]. The sigmoid function scaled the values, producing a channel importance score for each channel. The channel importance scores were subsequently used to weight the original feature maps, emphasizing the most relevant channels while suppressing less informative ones. To calculate the channel weight $s \in R^{c}$ using:

$$s = \sigma\big(W_2 \,.\, ReLU(W_1 . 2)\big) \qquad (3)$$

Where W_1 and. $W_2$are the weight matrices of the fully connected layers. This resulting weighted channel features were then propagated to the next layer of the encoder, ensuring that the model retained the most meaningful information for subsequent processing and with each channel X_cScaled by its corresponding importance score $s_{c'}$ resulting in the refined feature maps

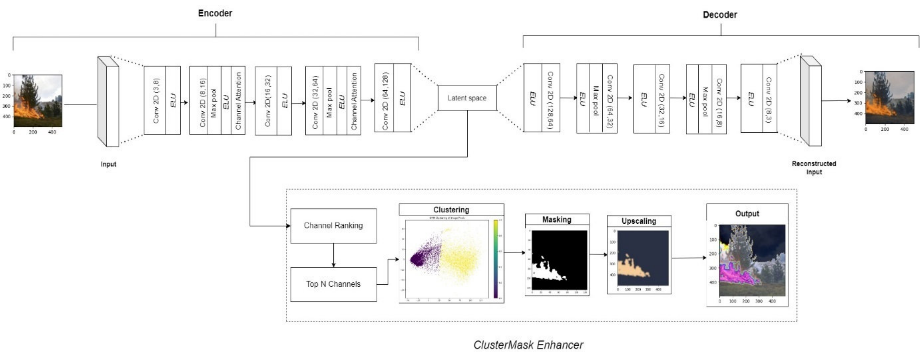


Figure - 2 Architecture of CAE model with cluster mask enhancer

3.2 Cluster Mask Enhancer

This module is developed for rigorous, in-depth analysis by implementing a range of advanced clustering methods. It utilizes scatter plot visualizations to uncover and highlight intricate data patterns, enabling a thorough examination of relationships and trends within the dataset. By applying diverse clustering approaches, this module facilitates enhanced data segmentation and provides valuable insights for comprehensive analytical exploration.

3.2.1 Latent Representation

The latent representation refers to the compact, abstracted features extracted from the input data during the encoding process of a model such as a convolutional autoencoder. These latent representations [2] serve as a condensed version of the input, preserving the most critical information while discarding redundant or less significant details. They are designed to encapsulate the essential features of the original images in a highly efficient and smaller dimensional format, enabling downstream tasks like segmentation or clustering to focus on the core patterns in the data and Channel attention plays a pivotal role in refining these latent representations by selectively emphasizing the most important aspects of the encoded features. This mechanism evaluates the relevance of each feature channel in the latent space and assigns varying levels of importance accordingly. By prioritizing the most critical channels, the model can effectively focus on key patterns that are highly relevant prioritizing the most critical channels, the model can effectively focus on key patterns that are highly relevant to the task at hand, such as fire edges, intense heat zones, or smoke formations in wildfire images. This selective attention allows the model to minimize the influence of less informative features, thereby enhancing the overall representation quality and task performance.

3.3 Clustering Methods

Various clustering methods were applied here for accurately identifying fire-segmented regions, we use clustering techniques on the top n ranked latent representations derived from the initial feature extraction. Specifically, K-means, Gaussian Mixture Model (GMM) clustering algorithms [2] are applied to these latent features to cluster the channels. These two algorithms are configured with k – meant with 2 clusters and GMM was with 4 cluster

3.3.1 K-means clustering

K-means clustering, which is one of the most widely used clustering algorithm and it aims to partition data into k clusters based on feature similarity. In our work, K-Means clustering was applied to the latent representations in order to effectively segment the fire-related regions. The algorithm minimized the within-cluster variance, quantified by the objective function

$$J = \sum_{K=1}^{K} \sum_{i \in c_k} \|x_i - \mu_k\|^2 \qquad (5)$$

where $x_i$represents a data point, $\mu_k$is the centroid of cluster k, and C_k is the set of points in the k-th cluster. The algorithm iteratively assigns data points to the nearest centroid and recalculates centroids until convergence. In this work, k=2k = 2k=2 was selected to separate fire and non-fire regions. This approach is computationally efficient and ensures a clear demarcation between fire and background regions.

3.3.2 Gaussian Mixture Model (GMM) Clustering

GMM is a probabilistic clustering approach that models the data as a mixture of Gaussian distributions. Each cluster is represented as a Gaussian distribution defined by its mean, covariance, and weight. For a data point xi the likelihood of belonging to cluster k is given by:

Gurukul International Multidisciplinary Research Journal (GIMRJ)*with*
International Impact Factor 8.357
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ/2503I01S01V13P0009
e-ISSN No. 2394-8426
Special Issue on Scientific Research Mar'25
Issue–III(I), Volume–XIII

where is the dimensionality of GMM computes the posterior probabilities of each data point belonging to each cluster and assigns points accordingly. For our work, GMM was configured with k=4 clusters to capture more nuanced details in fire-segmented regions, such as varying intensity levels of fire and smoke. This method provides a more flexible clustering mechanism compared to K-Means, as it accounts for covariance structures within the data, enabling finer segmentation

3.4 Image Masking

After accurately identifying and segmenting the fire regions, the next phase is to extract a masked image from these segmented regions. This masked image specifically highlights the fire-affected areas, isolating them from other parts of the image and creating a clear focus on critical regions. To achieve this, the clustered image is converted into four different binary images that correspond to each cluster. Each binary image is used for filtering the red channel of the input image and the mean red intensity is measured. The cluster having the highest mean red intensity is taken as the final masked image.

This masked image is then utilized to assess the segmentation accuracy, where it is compared with state-of-the-art segmentation models. These comparisons allow us to gauge the model's performance, specifically examining how accurately it identifies fire regions compared to established methods. This final assessment step ensures that the segmentation approach is both precise and effective, meeting the rigorous standards required for fire detection and analysis.


(a) (b)

Figure – 3 (a)was original and (b) was the masked image

4. Experimental Setup

4.1 Datasets used and evaluation protocols

Corsican Fire Dataset is an open fire database that contains images of wildfires and controlled fires. It consists of 595 RGB images of forest fires captured at a close range, including some with sequences of frames of different fires. The images have dimensions of 1024 × 768.

4.2 Data Augmentation

In this process, the dataset of 595 images (each with dimensions 500x500x3) is augmented to 1785 images using image augmentation techniques. We utilized the ImageDataGenerator class from Keras to apply various transformations, such as rotation (up to 40 degrees), width and height shifts (20%), shearing, zooming (10%), and horizontal flipping. Each original image undergoes three different augmentations, resulting in a tripling of the dataset. The augmentation helps improve model performance by simulating data variability, reducing overfitting, and enhancing the model's ability to generalize. The final shape of the augmented dataset reflects the increased size, ensuring more robust training data.

4.3 Implementation Details

The neural network models are built in Python version 3.9 and trained on Visual studio code and Google colab. For speeding up the training and working on images, we used NVIDIA RTX 3050 GPU. Also, for building the neural network, we used Pytorch version 2.4.0+cu118.

For training the model, we augmented the original dataset thereby increasing its size to 3 times the original size. The number of channels is increased from 3 to 128 as the images go through the encoder.

***Table 1**: Architecture of CAE*

Layer	No. of channels	Output Shape	Activation
Input	3	(500,500, 3)	
Conv2D+BN+ELU	(3,8)	(500,500,8)	ELU
Conv2D+ MP+BN+ ELU+ Channel Attention	(8,16)	(250,250,16)	ELU
Conv2D+BN+ELU	(16,32)	(250,250,32)	ELU
Conv2D+MP+BN+ ELU+ Channel Attention	(32,64)	(125,125,64)	ELU
Conv2D+BN+ELU	(64,128)	(125,125,128)	ELU
ConvTranspose2D + ELU	(128,64)	(125,125,64)	ELU
ConvTranspose2D + ELU	(64,32)	(250,250,32)	ELU
ConvTranspose2D + ELU	(32,16)	(250,250,16)	ELU
ConvTranspose2D + ELU	(16,8)	(500,500,8)	ELU
ConvTranspose2D + ELU	(8,3)	(500,500,3)	ELU
Reshape	3	(500,500,3)	

The neural network models are built in Python version 3.9 and trained on Visual studio code and Google colab. For speeding up the training and working on images, we used NVIDIA RTX 3050 GPU. Also, for building the neural network, we used Pytorch version 2.4.0+cu118.

For training the model, we augmented the original dataset thereby increasing its size to 3 times the original size. The number of channels is increased from 3 to 128 as the images go through the encoder.

4.4 Channel Attention

An attention block was implemented to enhance the model's focus on relevant spatial features. We have implemented Squeeze and Excitation model where the squeeze block consists of an adaptive average pooling layer to reduce spatial dimensions, followed by a flattening operation. A linear transformation reduces the feature dimensions by half, followed by a ReLU [3] activation for non-linearity. Excitation layer consists of a linear layer which restores the original dimensions, and finally, a sigmoid activation outputs attention weights. These weights are multiplied with the channels so that the next layer can selectively emphasize important features and suppress irrelevant ones.

4.5 Evaluation Protocols

This protocol includes multiple stages to validate the performance of the Convolutional Autoencoder (CAE) model, including training, validation, and testing phases. During each phase, specific metrics and methodologies are employed to evaluate the model's reconstruction quality, clustering accuracy, and overall robustness.

4.5.1 Autoencoder training:

The training process of the CAE (Convolutional Autoencoder) model employs the Mean Squared Error (MSE) loss function as a critical evaluation metric to assess the model's performance. During training, the primary objective of the CAE is to minimize the difference between the input images and the reconstructed images produced by the model. MSE is chosen as the loss function for this purpose because

Gurukul International Multidisciplinary Research Journal (GIMRJ)*with*
International Impact Factor 8.357
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ/2503I01S01V13P0009

e-ISSN No. 2394-8426
Special Issue on Scientific Research Mar'25
Issue–III(I), Volume–XIII

it effectively quantifies the average squared differences between the predicted outputs (reconstructed images) and the actual inputs (original images). Along with MSE loss, we have used color aware regularization that takes the red channel of the input image and output image and calculates MSE loss of those 2 values. This helps the model to emphasize the red tones of the image. Further, this loss is added with training loss with a regularization parameter to control the strength of this regularization.

The loss is then backpropagated through the network, allowing the optimizer to adjust the model parameters based on the calculated gradients. This iterative process continues across multiple epochs, enabling the model to progressively minimize

4.5.2 Segmentation Evaluation Metrics:

The Intersection over Union (IoU) and Dice Coefficient are two common metrics used to measure how well a segmentation model is performing. IoU checks how much the predicted mask overlaps with the ground truth, giving an idea of how accurate the segmentation is. The Dice Coefficient, meanwhile, looks at the similarity between the predicted and actual masks, focusing on their shared areas. Both of these metrics help us understand how well the model is doing, with higher values meaning better performance.

Intersection over Union:

$$\mathrm{IoU} = \frac{|P_i \cap G_i|}{|P_i \cup G_i|} \qquad (9)$$

P_i = *Predicted mask for image*

G_i = *Ground truth mask for image*

DICE:

$$DICE = \frac{2|P \cap G|}{|P|+|G|} \qquad (10)$$

P = Predicted mask

G = Ground truth mask

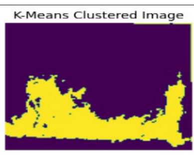
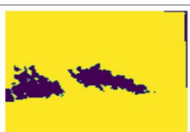
5 Results

To verify the effectiveness of our proposed approach, various experiments using multiple clustering methods such as K-means and GMM are carried out on the latent space. The results are shown in Tables 3 and 4.

5.1 Segmentation

After clustering the data into 4 clusters, the binary masks corresponding to each cluster were extracted and upscaled using linear interpolation. Subsequently, the red channel of the input image was isolated and filtered using the four binary masks. The mean intensity was computed for each filtered image, and the image with the highest intensity was identified as the correctly segmented mask.

With k-Means:

Input image	Clustered image	Masked image
		
		

Gurukul International Multidisciplinary Research Journal (GIMRJ)*with*
International Impact Factor 8.357
Peer Reviewed Journal
https://doi.org/10.69758/GIMRJ/2503I01S01V13P0009

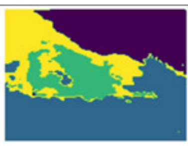
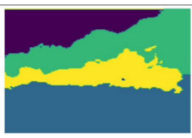
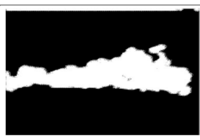
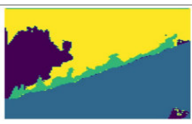
e-ISSN No. 2394-8426
Special Issue on Scientific Research Mar'25
Issue–III(I), Volume–XIII

The series of images showcases the process of segmenting forest fire regions step by step. The **first image** is the original input, capturing a real-world forest fire scene with flames, smoke, and the surrounding environment. This serves as the raw data for feature extraction. Moving to the **second image**, we see the result of applying a clustering algorithm used was K-Means or, on the extracted features. The clusters are represented in distinct colors, separating fire regions from the background based on their characteristics. Finally, the **third image** presents a masked output, where the fire regions are highlighted in white, and the rest of the background is suppressed in black. This simplified binary representation isolates the fire-affected areas, making them easier to analyze and compare with other segmentation techniques. This process demonstrates how the model identifies and focuses on critical fire regions in a systematic and interpretable way.

With GMM:

Table 4: Segmentation and masked images implemented with GMM

Input image	Clustered image	Masked image
		
		
		

The images illustrate the process of identifying forest fire regions using a clustering approach with GMM (Gaussian Mixture Model), configured for four clusters. The **first image** shows the original input, a forest fire scene with visible flames, smoke, and surrounding vegetation. This image serves as the starting point for extracting meaningful features. In the **second image**, the GMM algorithm has been applied, dividing the features into four clusters. Each cluster captures distinct characteristics, such as fire intensity, smoke, and background elements, which are represented in different colors. Finally, the **third image** shows the masked output, where the fire regions are clearly highlighted in white while non-fire areas are suppressed in black. By leveraging GMM's ability to model data probabilistically, this process ensures finer segmentation, capturing varying intensities of fire and smoke. The result provides a focused and interpretable representation of fire-affected areas for further analysis.

5.2 Performance of the segmentation

The performance of our model, with and without channel attention, was evaluated using the Mean Squared Error (MSE) loss. Incorporating channel attention significantly improved the model's accuracy, as it helped to focus on the most relevant features. When using clustering methods like GMM and K-Means with channel attention, the MSE loss was reduced to 0.004 for both methods. In contrast, without channel attention, the loss increased to 0.006. This highlights the critical role of channel attention in achieving better accuracy by effectively enhancing the important features in the latent representations

5.3 State of the art Comparison

The proposed approach was evaluated against several recent state-of-the-art methods to assess its performance. A comparison of Intersection over Union (IoU) scores for different techniques is summarized as follows:

We evaluated our proposed approach by comparing it with several recent state-of-the-art methods to see how well it performs in forest fire segmentation. The results, based on the Intersection over Union (IoU) scores, show a clear progression in performance across different techniques. The Semi-Supervised Classification and Segmentation method achieved an IoU of 75.62, while the Unsupervised Segmentation Based on GK-RGB performed slightly better with a score of 81.34. Moving further, the Segmentation Algorithms Based on Fully Convolutional Networks improved the IoU to 87.52. However, our approach, which uses Unsupervised Learning with a ClusterMask Enhancer, stood out by achieving the highest IoU score of 89.98. This demonstrates how effective our method is, combining unsupervised learning with advanced clustering and masking to deliver accurate and reliable segmentation, even in complex scenarios.

6. Conclusion and future works

In this study, we analyzed the potential of utilizing autoencoders for unsupervised learning in the context of forest fire segmentation. We employed a Convolutional Autoencoder (CAE) based unsupervised approach, leveraging the ClusterMask Enhancer for effective segmentation. The latent space of the autoencoder was utilized to perform clustering using both k-means and Gaussian Mixture Models (GMM). Additionally, we incorporated channel attention blocks within the CAE and implemented color-aware regularization techniques to enhance segmentation accuracy. Our experiments were conducted using the Corsican dataset, where we investigated visualization results from model training, clustering feature space plots, and segmentation outcomes. The proposed framework yielded promising results, achieving an Intersection over Union (IoU) of 0.8398 and a DICE score of 0.8998, thereby outperforming state-of-the-art models in this domain.

For future work, we plan to extend our approach to video-based segmentation, allowing for better adaptation to dynamic environments. Additionally, we aim to incorporate contrastive learning techniques to generate more accurate feature representations by pulling similar samples closer together in the latent space and pushing dissimilar ones apart. We also intend to introduce spatial attention layers in the encoder to focus on the most relevant areas of the image, like fire features, and combine these attention mechanisms with convolutional layers to boost feature extraction and segmentation accuracy. To improve the model's generalizability, we will explore the inclusion of diverse datasets, such as FLAME and real-world wildfire images, to account for various environmental conditions. Furthermore, we plan to integrate advanced image processing techniques like edge detection and region-based smoothing to refine the extracted features, enhancing the model's robustness and accuracy.

6. References:

1. Akash Koottungal, Shailesh Pandey, and Athira Nambiar: Semi-supervised Classification and Segmentation of Forest Fire Using Autoencoders
2. Thomas Di Martino, Regis Guinvarc'H, Laetitia Thirion-Lefevre, Elise Koeniguer. Deep Temporal Convolutional Autoencoder for Unsupervised Representation Learning of Incoherent Polsar Time- Series. IGARSS 2021, Jul 2021, Bruxelles (virtual), Belgium. hal-03345533

3. Citation: Wang, Z.; Peng, T.; Lu, Z. Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks. Forests 2022, 13, 1133. https://doi.org/10.3390/f13071133
4. Bin Yang, Haiwei Pan, Shuning He, Kun Han, Xuecheng Zhao. Forest Fire Thermal Infrared Image Segmentation Based on K-V Model
5. Ziqi Wang , Tao Peng, Zhaoyou Lu. Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks
6. Shen, X.; Liu, Z.; Xu, Z. Flame Segmentation Method Based on GK-RGB in Complex Background. Fire 2023, 6, 384. https://doi.org/10.3390/ fire6100384
7. Unsupervised Method for Wildfire Flame Segmentation and Detection EMIR BUZA AND AMILA AKAGIC, (Member, IEEE)
8. Jiang, X.; Wang, Z.; Li, J.; Zhang, L. "Real-Time Detection of Forest Fires in Remote Sensing Images Using YOLOv4." *International Journal of Applied Earth Observation and Geoinformation* 2024, 115, 102541.https://doi.org/10.1016/j.jag.2022.102541
9. Khan, I.; Hashmi, M. S.; Jan, U. A.; Baig, M. H.; Wang, J. "Comparative Study of Deep Learning Techniques for Fire Detection in Forest Regions." *Fire Technology* 2023, 59, 919-934. https://doi.org/10.1007/s10694-023-01311-5
10. MEENU AJITH , MANEL MARTÍNEZ-RAMÓN: Unsupervised Segmentation of Fire and Smoke From Infra-Red Videos
11. Xuejie Shen, Zhihuan Liu and Zhuonong Xu: Unsupervised Flame Segmentation Method Based on GK-RGB in Complex Background