SEGMENTATION OF INDUSTRIAL BURNER FLAMES: A COMPARATIVE STUDY FROM TRADITIONAL IMAGE PROCESSING TO MACHINE AND DEEP LEARNING

S. Landgraf[1,*], M. Hillemann[1], M. Aberle[2], V. Jung[2], M. Ulrich[1]

[1, 2]Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), Germany -
[1] (steven.landgraf, markus.hillemann, markus.ulrich)@kit.edu
[2] (moritz.aberle, valentin.jung)@student.kit.edu

KEY WORDS: Segmentation, Industrial Burner Flames, Image Processing, Machine Learning, Deep Learning.

ABSTRACT:

In many industrial processes, such as power generation, chemical production, and waste management, accurately monitoring industrial burner flame characteristics is crucial for safe and efficient operation. A key step involves separating the flames from the background through binary segmentation. Decades of machine vision research have produced a wide range of possible solutions, from traditional image processing to traditional machine learning and modern deep learning methods. In this work, we present a comparative study of multiple segmentation approaches, namely Global Thresholding, Region Growing, Support Vector Machines, Random Forest, Multilayer Perceptron, U-Net, and DeepLabV3+, that are evaluated on a public benchmark dataset of industrial burner flames. We provide helpful insights and guidance for researchers and practitioners aiming to select an appropriate approach for the binary segmentation of industrial burner flames and beyond. For the highest accuracy, deep learning is the leading approach, while for fast and simple solutions, traditional image processing techniques remain a viable option.[†]

1. INTRODUCTION

Industrial combustion processes play a critical role in many industrial processes, including power generation, chemical production, and waste management. To ensure safe and efficient operation, it is essential to accurately monitor the characteristics of the burner flames, such as their shape, size, and temperature. A key step in camera-based monitoring involves separating the flames from the background through segmentation (Großkopf et al., 2021; Landgraf et al., 2022).

Traditionally, binary segmentation has been performed using image processing techniques such as thresholding, edge detection, region growing, and morphological operations (Steger et al., 2018). While these techniques can be effective, they often require deep knowledge about the application, significant manual tuning, and may not generalize well to new data.

In recent years, deep learning has emerged as a powerful data-driven approach for segmentation tasks. These methods can learn meaningful features and patterns from training data, which generalize well to the real application, and therefore achieve higher accuracy than traditional techniques (Long et al., 2015). However, they require large amounts of training data, which are costly to label, and are typically computationally more expensive.

In this work, we present a comparative study of traditional image processing techniques, traditional machine learning-based methods, and deep learning approaches for binary segmentation of industrial burner flames in grayscale images. Our study provides insights into the strengths and limitations of each approach and can help guide researchers and practitioners in selecting the most appropriate method for their specific application.

---

[*]Corresponding author
[†]Our modified version of the used dataset:
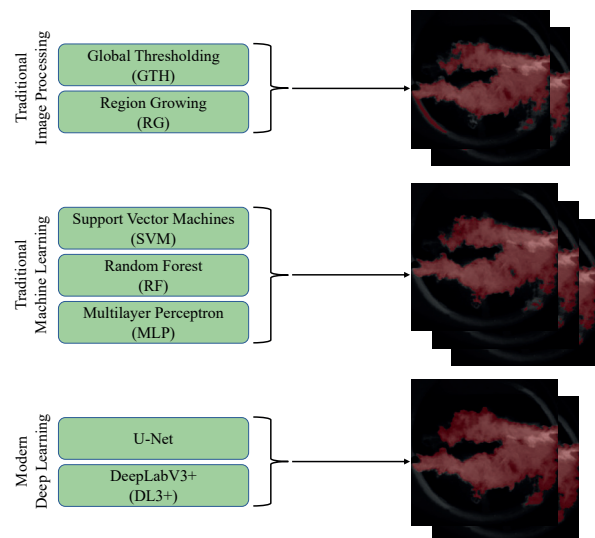https://publikationen.bibliothek.kit.edu/1000159497

Figure 1. An overview of our comparative study. We compare seven different approaches for the binary segmentation of industrial burner flames: Global Thresholding, Region Growing, Support Vector Machines, Random Forest, Multilayer Perceptron, U-Net, and DeepLabV3+.

2. RELATED WORK

The following related work focuses on flame segmentation and is divided into three categories: i) Traditional image processing methods, ii) traditional machine learning methods, and iii) deep learning methods. With traditional image processing methods, we refer to methods that use hand-crafted features and rule-based decisions to distinguish between flame and background. Traditional machine learning methods, on the other hand, refer to data-driven methods like Support Vector Machines (SVM), Random Forest (RF), and Multilayer Perceptron (MLP) which work with hand-crafted features. Finally, by deep learning meth-

ods, we mean state-of-the art methods for image segmentation with modern deep neural network-based architectures. This section only covers related work regarding the segmentation of flames in images. Classification and detection approaches are not covered to limit the scope.

Traditional Image Processing. Early approaches for flame segmentation use traditional image processing techniques (Celik et al., 2007; Zhang et al., 2009; Fan et al., 2014; Wang et al., 2014; Matthes et al., 2019; Li and Wang, 2020). These methods rely on different hand-crafted features such as color (Celik et al., 2007; Zhang et al., 2009; Wang et al., 2014; Li and Wang, 2020) and geometrical characteristics like area, roundness and contour fluctuation (Zhang et al., 2009). Several of these methods use image sequences or videos of the flames to determine time-dependent features like color and area changes (Zhang et al., 2009; Wang et al., 2014; Matthes et al., 2019; Li and Wang, 2020). Some methods distinguish between flame and background by using empirically determined thresholds for combinations of these features (Zhang et al., 2009; Wang et al., 2014) or determine the thresholds automatically using Otsu's method (Matthes et al., 2019). Others combine the features with fuzzy logic (Celik et al., 2007) or a Bayesian model (Li and Wang, 2020). Fan et al. (2014) choose a different approach and use a level set method to detect the contour of the flame.

Traditional Machine Learning. Many of the traditional machine learning methods for flame segmentation also use hand-crafted features like those described above. However, these methods do not decide based on rules created by the human developer, but create their own set of rules based on the data. In addition to some of the features described above, Borges and Izquierdo (2010) use surface coarseness, boundary roughness and skewness as geometrical characteristics and a Bayesian Classifier flame segmentation. Zhao et al. (2011) and Jamali et al. (2013) also use image sequences and time-dependent features and then perform segmentation with a SVM, where Zhao et al. (2011) takes into account other texture features such as entropy and contrast. Recently, Liang et al. (2022) proposed a two step segmentation method. In the first step, they determine the feature importance of different geometric, texture and color features with a RF, and in the second step they use the five most important features as input neurons of a MLP that distinguishes between flame and background.

Deep Learning. Since the breakthrough of deep learning methods in many areas of image processing, they have also been applied for flame segmentation. These methods represent the current state-of-the-art in terms of segmentation quality. In comparison to traditional machine learning methods, they no longer rely on hand-crafted features, they learn the features best suited to separate foreground and background through training. In this regard, Zhong et al. (2018) is a hybrid method, since it first determines candidate regions for flames based on a threshold on color values and then uses a Convolutional Neural Network (CNN) to classify the candidate regions into flame or background. Großkopf et al. (2021) introduce a pure CNN-based method to segment industrial burner flames in images. They not only focus on binary segmentation of flames, but also perform multi-class segmentation. They also provide a public image dataset with ground truth labels. Recently, Landgraf et al. (2022) analyzed several self-supervised learning methods as pre-training for the segmentation of flames. They show how the required amount of training data can be significantly reduced without decreasing the quality of the segmentation. Most recently, Wang et al. (2022) compared four different deep learn-

ing methods, namely Fully Convolutional Network (Long et al., 2015), U-Net (Ronneberger et al., 2015), PSPNet (Zhao et al., 2017) and DeepLabV3+ (Chen et al., 2018a) on the segmentation of forest fires in images captured by UAVs. Their analysis shows that U-Net achieves the best performance, though also having the slowest inference time. Compared to that, DeepLabV3+ is slightly faster but also yields slightly worse results. In contrast, the FCN and PSPNet models have shown to be less suitable because of lower scores for the performance metrics.

3. METHODOLOGY

In the following, we provide an overview of our comparative study, as well as explain the examined traditional image processing, traditional machine learning and deep learning methods in more detail.

3.1 Overview

In contrast to the related work on flame segmentation and as Figure 1 shows, we provide a comprehensive study that covers methods from traditional image processing to traditional machine learning and modern deep learning. Our goal is to provide helpful insights and guidance for researchers and practitioners aiming to select an appropriate approach for the binary segmentation of industrial burner flames and similar applications.

For our study, we compare seven popular binary segmentation approaches:

1. Global Thresholding (GTH) (Lee et al., 1990),

2. Region Growing (RG) (Adams and Bischof, 1994),

3. Support Vector Machines (SVM) (Hearst et al., 1998),

4. Random Forest (RF) (Breiman, 2001),

5. Multilayer Perceptron (MLP) (McClelland et al., 1987),

6. U-Net (Ronneberger et al., 2015),

7. DeepLabV3+ (DL3+) (Chen et al., 2018b).

The hyperparameters of the traditional image processing techniques, GTH and RG, as well as the traditional machine learning methods RF, SVM, and MLP, are tuned on the training dataset. On the other hand, all of the deep learning models are trained on the same fixed set of hyperparameters to limit the computational cost. In return, we evaluate the effect of training from scratch in comparison to fine-tuning from ImageNet pre-training and estimate the impact of the model size, through training both U-Net and DL3+ with three different backbones each.

3.2 Traditional Image Processing

Traditional image processing offers an easy and effective approach for binary segmentation if they are adapted for the underlying application. This is why we explore GTH and RG as a baseline for the more sophisticated data-driven methods. We implemented both GTH and RG with HALCON (MVTec Software GmbH, 2023).

Global Thresholding. GTH is a widely-used method for performing binary image segmentation, where two threshold values are selected to separate the foreground objects from the
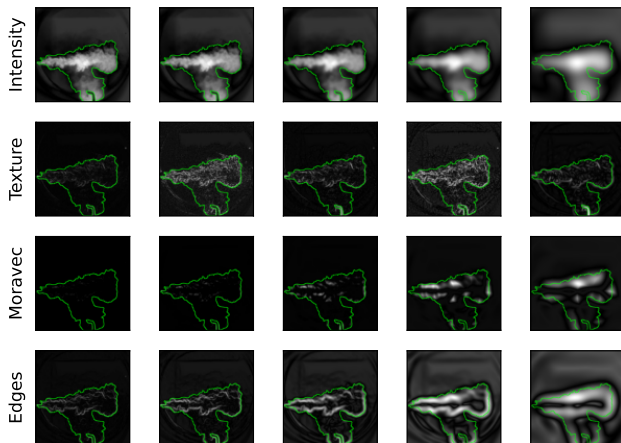
Figure 2. Example of the used basic multiscale features and Moravec corners from Scikit-image for a single image with the ground truth label's contour overlaid in green.

background. In our case, each pixel is classified as industrial burner flame based on the following condition:

$$T_L \leq g \leq T_U, \tag{1}$$

where g is the gray value of the respective pixel between 0 and 255, T_L is the lower, and T_U is the upper threshold. To determine the optimal values for the lower and upper thresholds, we performed a grid search on the training dataset.

Region Growing. RG is another traditional image processing technique for binary segmentation, which can start at any pixel in the image. It then grows the segmented region by including neighboring pixels that satisfy certain criteria. In our case, a pixel is added to the region if the following criterion is met:

$$|r_{\bar{g}} - g| \leq T, \tag{2}$$

where $r_{\bar{g}}$ is the mean gray value of the region, g is the gray value of the respective pixel, and T is the chosen threshold. We segment the background with the RG and invert the segmentation results instead of segmenting the flames directly. This results in a better performance because industrial burner flames are not always coherent. For the starting pixel, we choose the lowest gray value in every given image and to determine the optimal value of the threshold, we performed a grid search on the training dataset.

3.3 Traditional Machine Learning

In addition to the aforementioned traditional image processing techniques, we also explored three traditional machine learning approaches: SVM, RF, and MLP. We implemented all of them in Python with Scikit-learn (Pedregosa et al., 2011), Scikit-image (Van der Walt et al., 2014) and OpenCV (Bradski et al., 2000).

Features. To improve the performance of the traditional machine learning classifiers, we extract 23 hand-crafted features from the images. Like Figure 2 shows, these include 20 intensity, texture, edge, and corner features obtained using basic multiscale features and Moravec corners from Scikit-image. For smoothing the image with a Gaussian kernel, we used standard deviations from 1.0 to 30.0. We also use OpenCV to build three median features with square filters of sizes 51, 101, and

151. All of the features are extracted pixelwise and as a consequence, the classification is performed on each pixel individually.

Support Vector Machines. SVM is a popular algorithm for binary segmentation that works by finding the optimal hyperplane that separates the positive and negative examples with the largest possible margin. For hyperparameter optimization, we performed a grid search on the training dataset to determine the optimal values for the regularization parameter and the kernel function.

Random Forest. RF is an ensemble learning method that combines multiple decision trees to improve the performance of the classifier. Thereby, each decision tree is trained on a random subset of the dataset. For hyperparameter optimization, we performed a grid search on the training dataset to determine the optimal values for the number of decision trees, the maximum depth of the trees, and the maximum number of training images for each tree.

Multilayer Perceptron. MLP is a basic neural network that consists of multiple layers of interconnected nodes that can be utilized for binary segmentation. Hereby, each node applies a nonlinear activation function to the weighted sum of its inputs. Following the previous machine learning classifiers, we used the 23 extracted features as input for the MLP. To optimize the hyperparameters of the MLP, we performed a grid search on the training dataset using a range of possible values for the number of hidden layers, the number of nodes in each layer and the learning rate.

3.4 Deep Learning

Aside from traditional image processing techniques and traditional machine learning methods, we also explore deep learning approaches for binary segmentation of industrial burner flames. Both U-Net and DL3+ were implemented with PyTorch (Paszke et al., 2019).

U-Net. U-Net is a fully convolutional neural network that uses a contracting path to capture the image context and a symmetric expanding path to achieve precise localization. The resulting encoder-decoder architecture has a U-shape, which utilizes skip connections that allow information to be propagated from the encoder to the decoder. The U-Net was first introduced for medical image segmentation (Ronneberger et al., 2015).

DeepLabV3+. DeepLabV3+ is a fully convolutional neural network that uses atrous (or dilated) convolutions within the atrous spatial pyramid pooling (ASPP) module to capture multiscale contextual information. It builds upon the encoder-decoder architecture by fusing high-level ASPP features with low-level features from earlier layers in the network (Chen et al., 2018b).

Backbones and Initialization. In order to estimate the impact of the model size, we train both U-Net and DL3+ with three different backbones:

1. MobileNetV3 Small (MN) (Howard et al., 2019),

2. ResNet-18 (RN18) (He et al., 2016),

3. ResNet-101 (RN101) (He et al., 2016).

Additionally, we evaluate the impact of the initialization for every model by:
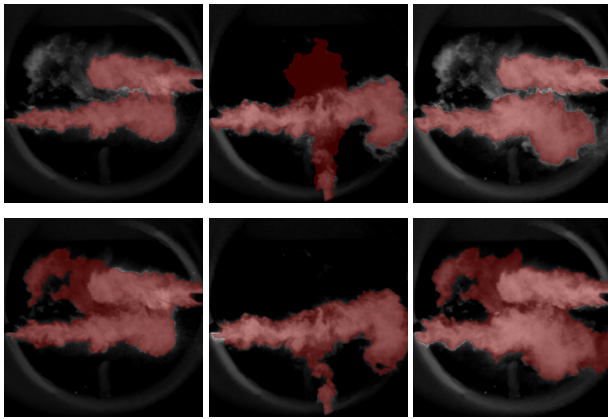
Figure 3. Example images of the industrial burner flames dataset with the ground truth labels overlaid as transparent red regions. The upper row displays the original labels by Großkopf et al. (2021), whereas the lower three images represent the labels that we created.

1. Training from scratch, i.e. with random weights (R),

2. Training from ImageNet (I) (Deng et al., 2009).

Implementation Details. We train all the deep learning models with a binary cross-entropy loss and employ a Stochastic Gradient Descent (SGD) optimizer based on Robbins and Monro (1951) with an initial learning rate of 0.01, momentum of 0.9, and weight decay of 0.0005 as optimizer-specific hyperparameters. During training, the learning rate decays based on:

$$lr = lr_{initial} \cdot (1 - \frac{iteration}{total\ iterations})^{0.9}, \qquad (3)$$

where lr is the current learning rate, and $lr_{initial}$ is the initial learning rate. All models are trained for 25 epochs with a batch size of 8 and without any data augmentations to ensure a fair comparison.

4. EXPERIMENTS

In this section, we share a variety of experiments conducted on the basis of a modified public dataset to gather helpful insights about the strengths and limitations of all the methods in our study.

4.1 Dataset

All of our experiments are based on a modified version of the publicly available industrial burner flames dataset provided by Großkopf et al. (2021). The original dataset consists of 3000 labeled grayscale images of two industial burner flames with 552×552 pixels in size.

In Figure 3, the upper row shows that some labels in the original dataset are questionable. We suspect that these labels were at least in part created automatically. To address this, we randomly selected 200 images and relabeled them by hand, creating a new dataset of 160 training images and 40 test images. As shown in the lower row of Figure 3, this process resulted in an improved label quality. In the original labels, flames represented 23.8% of the dataset, whereas in our labels the portion of flames is 26.5%. Finally, we computed the Intersection over Union (IoU) between the original labels and our labels, which yielded a score of 80.8%.

	IoU [%] ↑	Inference Time [ms] ↓
GTH	80.3	0.1
RG	77.0	13.4
SVM	82.3	8.2
RF	87.0	124.8
MLP	86.6	106.3
U-Net (MN)	92.6 (91.9)	106.7 (7.6)
U-Net (RN18)	92.9 (92.6)	138.4 (5.6)
U-Net (RN101)	93.1 (91.9)	436.9 (16.6)
DL3+ (MN)	92.6 (91.9)	81.2 (6.5)
DL3+ (RN18)	93.2 (92.9)	117.8 (4.6)
DL3+ (RN101)	92.9 (92.3)	394.3 (15.5)

Table 1. Quantitative comparison between the traditional image processing, traditional machine learning, and deep learning Intersection over Union (IoU) and inference time per image. IoU scores in parantheses refer to training from scratch, whereas the regular scores depict fine-tuning from ImageNet. Inference times in parantheses were computed on a GPU, while the regular values are CPU times.

4.2 Quantitative Evaluation

Table 1 presents a quantitative comparison between the best segmentation results obtained by each method in our study, along with their respective inference time per image. Unless otherwise noted, we used an AMD EPYC 7502 32-Core CPU with 1 TB of RAM. The results demonstrate that there are clear improvements going from traditional image processing (GTH: 80.3%) over traditional machine learning (RF: 87.0%) to deep learning (DL3+ (RN18-I): 93.2%). There is, however, a corresponding increase in CPU time going from 0.1 ms per image with GTH up to 436.9 ms with U-Net (RN101-I).

Traditional Image Processing. Looking at the traditional image processing techniques in particular, GTH outperforms RG in both the segmentation task and the needed inference time. GTH yields an IoU score of 80.3% which is 3.3% higher than what RG achieves. On top of that, GTH only needs 0.1 ms to process an image, whereas RG needs 13.4 ms.

Traditional Machine Learning. In comparison to the traditional image processing techniques, the machine learning approaches, SVM, RF, and MLP, yield a significant performance improvement in the segmentation task. SVM achieved an IoU score of 82.3%, outperforming both GTH and RF by noteworthy margin. In addition, SVM only requires 8.2 ms to process an image, making it faster than RF but slower than GTH. With 87.0% RF achieved the highest IoU score among the traditional machine learning methods. However, it also requires 124.8 ms to process an image. MLP achieved a very similar IoU score of 86.6%, but with a slightly faster inference time of 106.3 ms.

Deep Learning. Judging by the results laid out in Table 1, all of the deep learning approaches outperform the traditional image processing and machine learning methods on the segmentation task. Depending on the configuration, the models achieve IoU scores between 91.9% and 93.2%. Table 1 also shows that training from scratch in comparison to fine-tuning from ImageNet pre-training deteriorates performance by 0.3% to 1.2%. As Table 1 displays, deep learning comes at a high computational cost. Inference times on the CPU range from 81.2 ms for DL3+ (MN) to a maximum of 436.9 ms for U-Net
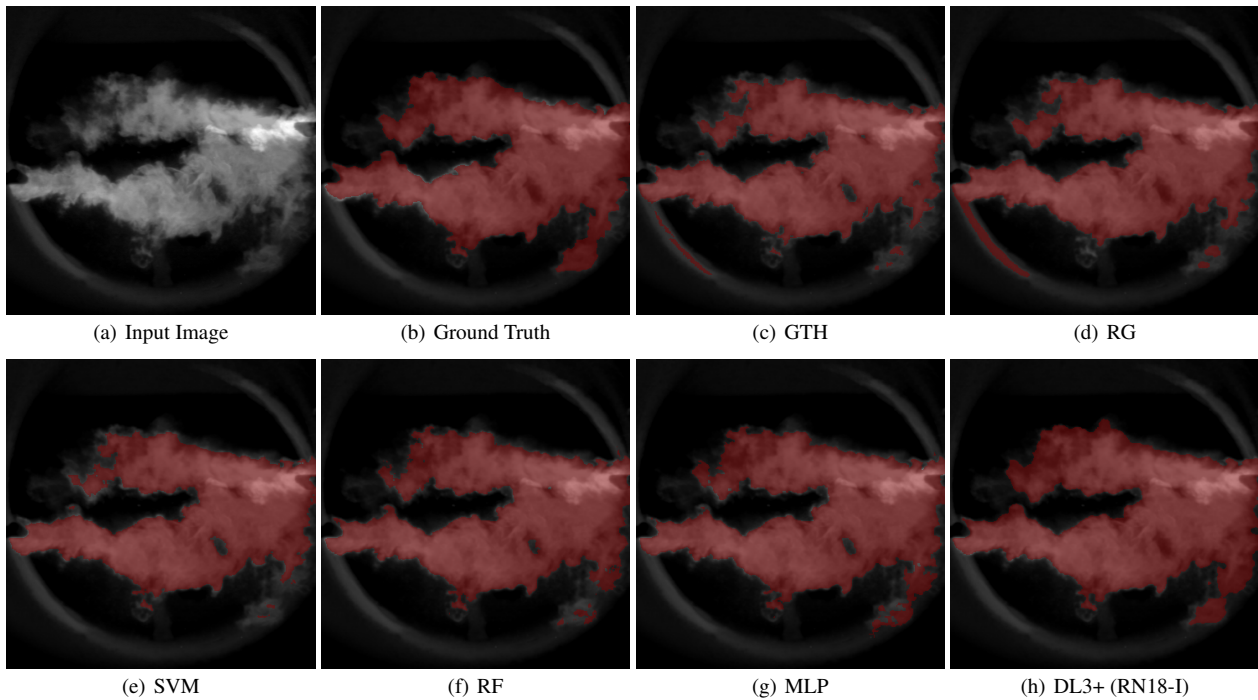
Figure 4. Qualitative comparison between the input image (a) overlaid in transparent red color with the ground truth (b), the segmentation results of Global Thresholding (c), Region Growing (d), Support Vector Machines (e), Random Forest (f), Multilayer Perceptron (g), and DeepLabV3+ with a ResNet-18 backbone pre-trained on ImageNet (h).

(RN101). However, inference times can be decreased greatly if inference is performed on a GPU. In our case, using a common NVIDIA GeForce RTX 3090 GPU with 24 GB of memory reduced inference times to a range from 4.6 to 16.6 ms depending on the architecture. Among all of the deep learning models, DL3+ (RN18-I) achieved the highest IoU score of 93.2% while simultaneously having the fastest inference time of 4.6 ms with the GPU.

4.3 Qualitative Evaluation

Figure 4 shows a visual comparison between the traditional image processing techniques, the traditional machine learning approaches, and the best deep learning model. Overall, the qualitative evaluation of the segmentation results confirms the observations of the quantitative analysis: There are clear improvements going from traditional image processing to traditional machine learning to deep learning. Figure 5 in the Appendix shows more qualitative examples, which corroborate this claim.

Traditional Image Processing. The two traditional image processing techniques, GTH and RG, largely suffer from the same shortcoming: Both oversegment the industrial burner chamber and undersegment the flame. As both methods solely rely on the underlying gray values, there is no way for these methods to distinguish between bright areas of the image that belong to the chamber or darker parts of the flame. Hence, the segmentation results are systematically flawed in these situations.

Traditional Machine Learning. In comparison to the traditional image processing techniques, the machine learning approaches, SVM, RF, and MLP, do not suffer from oversegmentation of the chamber. However, they undersegment darker parts of the flames too, which is especially noticeable for SVM. Overall, all of the traditional machine learning methods can improve upon the traditional image processing techniques due to the extra information added by the hand-crafted features.

	IoU [%] ↑	Performance Loss [%] ↓
GTH	78.6	1.7
RG	74.3	2.7
SVM	80.6	1.7
RF	84.7	2.3
MLP	84.9	1.7
U-Net (MN)	91.7 (90.5)	0.9 (1.4)
U-Net (RN18)	91.6 (91.1)	1.3 (1.5)
U-Net (RN101)	92.1 (88.1)	1.0 (3.8)
DL3+ (MN)	91.4 (90.6)	1.2 (1.3)
DL3+ (RN18)	91.8 (91.6)	1.4 (1.3)
DL3+ (RN101)	91.8 (90.7)	1.1 (1.6)

Table 2. Quantitative comparison between the traditional image processing, traditional machine learning, and deep learning IoU with just 40 training images and 160 test images. Additionally, the performance loss compared to the regular dataset with 160 training images is displayed. The IoU scores in parantheses refer to training from scratch, whereas the regular scores depict fine-tuning from ImageNet.

Deep Learning. Visually, an even better segmentation result is achieved by the best deep learning model DL3+ (RN18-I). The segmentation result suffers from no apparent systematic shortcomings. Through the training process, the deep learning model learns meaningful features and patterns that seem to generalize very well to the test data. This also holds for the other deep learning models not shown in Figure 4.

4.4 Impact of Training Dataset Size

In order to evaluate the effect of the number of training images, we retuned all of the traditional image processing and traditional machine learning methods and retrained all deep learn-

ing models with an inverse ratio of training and test images, i.e. we use just 40 randomly selected images for training and the remaining 160 images for testing.

As Table 2 shows, the observations from the previous experiments remain unchanged: There are significant segmentation improvements going from traditional image processing (GTH: 78.6%) to traditional machine learning (MLP: 84.9%) to deep learning (U-Net (RN101-I): 92.1%). To our surprise, however, the average performance loss was larger for the traditional image processing techniques and traditional machine learning approaches.

GTH and RG suffered a performance loss of 1.7% and 2.7%, respectively. The IoU of the traditional machine learning-based methods deteriorated by 1.9% on average. Whereas the average performance loss for the deep learning models pre-trained on ImageNet was only 1.2% and 1.8% for the models trained from scratch.

5. DISCUSSION

We conducted extensive experiments to compare traditional image processing techniques with traditional machine learning and deep learning methods. In the following, we discuss our key insights into the strengths and limitations of each approach.

Traditional Image Processing. GTH and RG are both effective for creating a baseline binary segmentation of industrial burner flames. They are computationally efficient, require little tuning and offer fast inference times, making them a good choice when speed is a priority. In our experiments, GTH outperformed RG on the segmentation task, with the fastest inference time of all examined methods with just 0.1 ms per image. However, these techniques may not be suitable if the quality of the segmentation results is more important, as they rely solely on gray values and can suffer from under- and oversegmentation.

Traditional Machine Learning. SVM, RF, and MLP significantly improve segmentation results compared to GTH by using hand-crafted features. For example, RF achieved a 6.7% relative IoU score improvement over GTH. However, these methods are computationally more expensive and require careful engineering and feature selection. It is worth noting that adding more features could improve the performance. On the other hand, it can also introduce noise and lead to overfitting, aside from increasing computation time. As a possible solution, SVM and RF allow the computation of feature importances to help select appropriate features.

Deep Learning. Our experiments show that the deep learning models U-Net and DL3+ achieve the best segmentation performance overall with a maximum IoU score of 93.2%. Surprisingly, the choice of architecture or pre-training on ImageNet had very little impact on the performance, although the latter helped in all experiments. Another revealing advantage of deep learning is its ability to generalize from small datasets, as the models suffered the smallest performance loss when trained on less data. The only apparent downside to the deep learning methods is their high computational cost. However, training and inference times can greatly benefit from using a GPU. Overall, deep learning is the best choice for binary segmentation of industrial burner flames, particularly if a GPU is available.

6. CONCLUSION

In this work, we conducted a comparative study of traditional image processing techniques, traditional machine learning methods, and deep learning approaches for the binary segmentation of industrial burner flames. Extensive experimentation on a modified version of a public dataset (Großkopf et al., 2021) revealed key insights into the strengths and limitations of each approach.

Traditional image processing techniques like Global Thresholding and Region Growing offer a fast and simple solution but suffer from systematic under- and oversegmentation. Traditional Machine learning methods, such as Support Vector Machines, Random Forests, and Multilayer Perceptrons, improve the segmentation performance in exchange for increased computational cost and feature engineering effort. Deep learning models, like U-Net and DeepLabV3+, achieve the best segmentation performance in our study while showing remarkable ability to generalize even from small datasets. Despite their high computational requirements, they are the best choice for binary segmentation of industrial burner flames, especially if a GPU is available.

In summary, our study provides helpful insights and guidance for researchers and practitioners aiming to select an appropriate approach for the binary segmentation of industrial burner flames and beyond. For the highest accuracy, deep learning is the leading approach, while for fast and simple solutions, traditional image processing techniques remain a viable option. With the continuing progress in deep learning and the increasing availability of compute power, we expect these methods to become even more capable and efficient in the future.

References

Adams, R., Bischof, L., 1994. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641–647.

Borges, P. V. K., Izquierdo, E., 2010. A probabilistic approach for vision-based fire detection in videos. *IEEE transactions on circuits and systems for video technology*, 20(5), 721–731.

Bradski, G., Kaehler, A. et al., 2000. OpenCV. *Dr. Dobb's journal of software tools*, 3(2).

Breiman, L., 2001. Random forests. *Machine learning*, 45, 5–32.

Celik, T., Ozkaramanli, H., Demirel, H., 2007. Fire pixel classification using fuzzy logic and statistical color model. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, 1, IEEE, I–1205.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei, 2009. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, FL, 248–255.

Fan, H., Cong, Y., Xia, Y., Shao, W., Wang, Y., 2014. Flame front detection and curvature calculation using level set. *Proceeding of the 11th World Congress on Intelligent Control and Automation*, IEEE, 2918–2922.

Großkopf, J., Matthes, J., Vogelbacher, M., Waibel, P., 2021. Evaluation of deep learning-based segmentation methods for industrial burner flames. *Energies*, 14(6), 1716.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18–28.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V. et al., 2019. Searching for mobilenetv3. *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324.

Jamali, M., Samavi, S., Nejati, M., Mirmahboub, B., 2013. Outdoor fire detection based on color and motion characteristics. *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, IEEE, 1–6.

Landgraf, S., Kühnlein, L., Hillemann, M., Hoyer, M., Keller, S., Ulrich, M., 2022. Evaluation of self-supervised learning approaches for semantic segmentation of industrial burner flames. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 601–607.

Lee, S. U., Chung, S. Y., Park, R. H., 1990. A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics, and Image Processing*, 52(2), 171–190.

Li, Y., Wang, L., 2020. Flame image segmentation algorithm based on motion and color saliency. *Communications, Signal Processing, and Systems: Proceedings of the 2018 CSPS Volume II: Signal Processing 7th*, Springer, 184–191.

Liang, J.-X., Zhao, J.-F., Sun, N., Shi, B.-J., 2022. Random forest feature selection and back propagation neural network to detect fire using video. *Journal of Sensors*, 2022, 1–10.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Matthes, J., Waibel, P., Vogelbacher, M., Gehrmann, H.-J., Keller, H., 2019. A new camera-based method for measuring the flame stability of non-oscillating and oscillating combustions. *Experimental Thermal and Fluid Science*, 105, 27–34.

McClelland, J. L., Rumelhart, D. E., Group, P. R. et al., 1987. *Parallel Distributed Processing, Volume 2: Explorations in the Microstructure of Cognition: Psychological and Biological Models*. 2, MIT press.

MVTec Software GmbH, 2023. HALCON operator reference.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv:1912.01703*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825–2830.

Robbins, H., Monro, S., 1951. A stochastic approximation method. *The annals of mathematical statistics*, 400–407.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 234–241.

Steger, C., Ulrich, M., Wiedemann, C., 2018. *Machine Vision Algorithms and Applications*. John Wiley & Sons.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., 2014. scikit-image: image processing in Python. *PeerJ*, 2, e453.

Wang, S., He, Y., Zou, J., Duan, B., Wang, J., 2014. A flame detection synthesis algorithm. *Fire Technology*, 50, 959–975.

Wang, Z., Peng, T., Lu, Z., 2022. Comparative research on forest fire image segmentation algorithms based on fully convolutional neural networks. *Forests*, 13(7), 1133.

Zhang, Z., Zhao, J., Yuan, Z., Zhang, D., Han, S., Qu, C., 2009. Color based segmentation and shape based matching of forest flames from monocular images. *2009 International Conference on Multimedia Information Networking and Security*, 1, IEEE, 625–628.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2881–2890.

Zhao, J., Zhang, Z., Han, S., Qu, C., Yuan, Z., Zhang, D., 2011. SVM based forest fire detection using static and dynamic features. *Computer Science and Information Systems*, 8(3), 821–841.

Zhong, Z., Wang, M., Shi, Y., Gao, W., 2018. A convolutional neural network-based flame detection method in video sequence. *Signal, Image and Video Processing*, 12, 1619–1627.
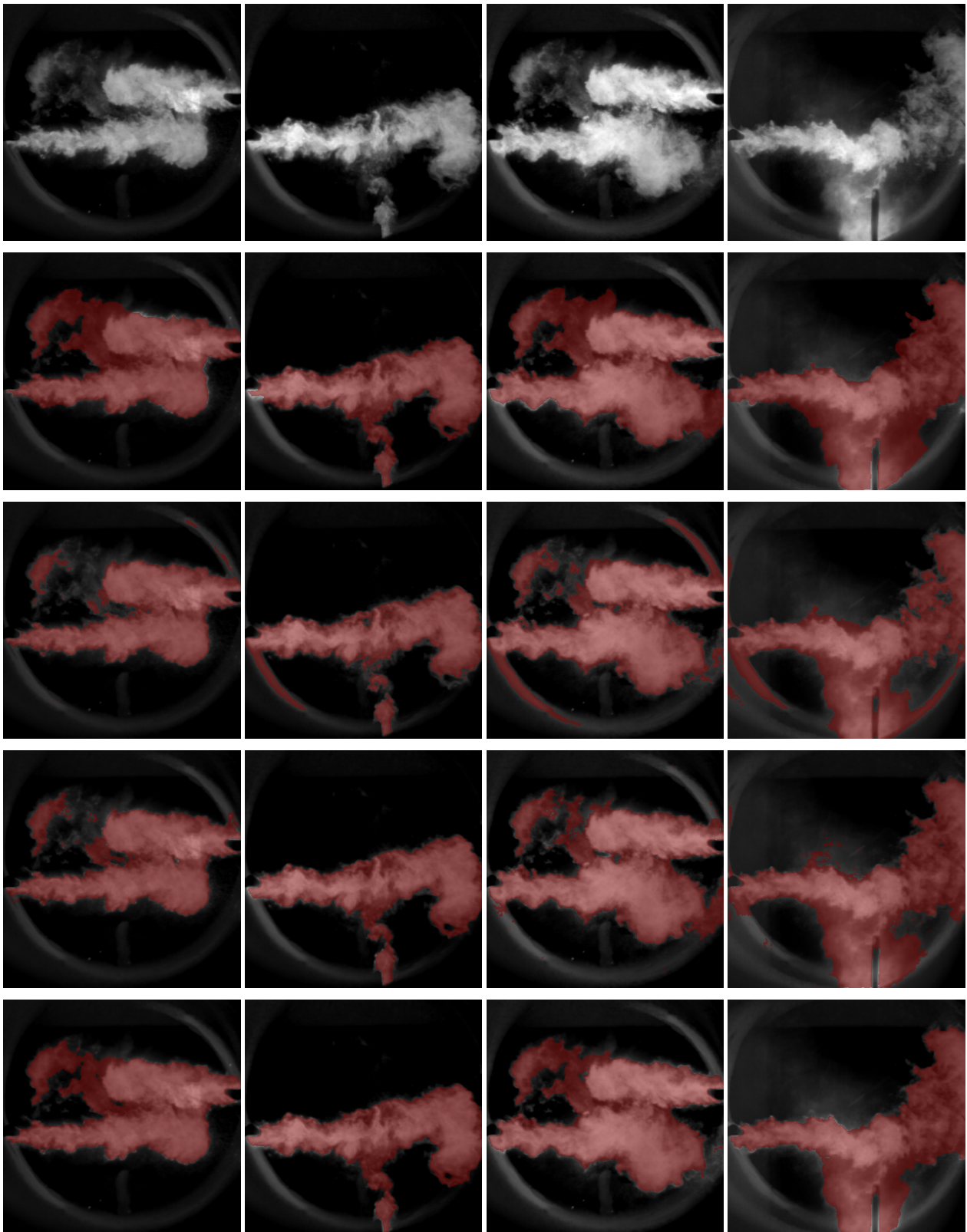
APPENDIX

Figure 5. More qualitative comparisons between the input image in the first row, the ground truth in the second row, the segmentation results of GTH in the third row, RF in the fourth row, and DL3+ (RN18-I) in the last row.