Wildland Fire Detection and Monitoring using a Drone-collected RGB/IR Image Dataset

Xiwen Chen[1], Bryce Hopkins[2], Hao Wang[1], Leo O'Neill[3], Fatemeh Afghah[2], Abolfazl Razi[1], Peter Fulé[3], Janice Coen[4], Eric Rowell[5], and Adam Watts[6]

[1]School of Computing, Clemson University, Clemson, SC 29634
[2]Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634
[3]School of Forestry, Northern Arizona University, Flagstaff, AZ 84001
[4]National Center for Atmospheric Research, Boulder, Colorado 80301
[5]Desert Research Institute, Reno, NV 89512
[6]US Forest Service, Pacific Wildland Fire Science Lab, Seattle, WA 98103

2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) | 978-1-6654-7729-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/AIPR57179.2022.10092208

Drone-based Unmanned Aerial Systems (UAS) provide an efficient means for early detection and monitoring of remote wildland fires due to their rapid deployment, low flight altitudes, high 3D maneuverability, and ever-expanding sensor capabilities. Recent sensor advancements have made side-by-side RGB/IR sensing feasible for UASs. The aggregation of optical and thermal images enables robust environmental observation, as the thermal feed provides information that would otherwise be obscured in a purely RGB setup, effectively "seeing through" thick smoke and tree occlusion. In this work, we present Fire detection and modeling: Aerial Multi-spectral image dataset (FLAME 2) [1], the first ever labeled collection of UAS-collected side-by-side RGB/IR aerial imagery of prescribed burns. Using FLAME 2, we then present two image-processing methodologies with Multi-modal Learning on our new dataset: (1) Deep Learning (DL)-based benchmarks for detecting fire and smoke frames with Transfer Learning and Feature Fusion. (2) an exemplary image-processing system cascaded in the DL-based classifier to perform fire localization. We show these two techniques achieve reasonable gains than either single-domain video inputs or training models from scratch in the fire detection task.

I. Introduction

Even though techniques of rapid public reporting systems, including geostationary satellites and network of optical smoke observation cameras [2], have greatly improved, there is still a need to quickly identify, map and monitor the specific location, extent and progress of fires. With their features of low flight altitudes, robust 3D maneuverability, and ever expanding sensor capability, Unmanned aerial systems (UAS) are a valuable tool for initial fire detection, monitoring, and management. These features enable the collection of rapid, high-resolution maps of vast areas of wildlands.

New generations of hardware have greatly expanded UAS' onboard computation and communication capabilities. This expanded edge computing, combined with the unprecedented performance of Deep Learning (DL) models, enables sophisticated UAS-based wildfire detection and monitoring models to run in real time. To configure a drone fire detection system embedded with data-driven algorithms, a dataset of aerial imagery of wildfires and prescribed burning is required, preferably with a high revisit rate.

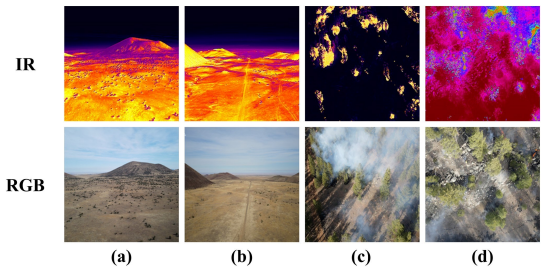

Figure 1. Handpicked frame pairs from FLAME 2 Dataset [1]. (a)(b) No Flame with No Smoke, (c) Flame with Smoke, (d) Flame with No Smoke.

Considering this open niche in fire detection datasets, we collected and published the "Fire detection and modeLing: Aerial Multi-spectral imagE" dataset (FLAME2) [1]. FLAME 2 provides a collection of side-by-side RGB and IR drone-collected videos and images taken during a prescribed burn in northern Arizona in November of 2021. Some sample frame pairs from the 254p set are presented in Fig 1. A detail of the labels[1] annotated by the human experts is presented in Table I.

Table I
FLAME 2 Dataset [1] frame pair label breakdown.

Label	Number of Image Pairs
Fire, Smoke	25,434
Fire, No Smoke	14,317
No Fire, No Smoke	13,700
No Fire, Smoke	0
Total Number	53,451
Resolution (for CNN input)	254 254

We examine different DL-based methods on the collected dataset FLAME 2 for fire detection (i.e., frame-by-frame fire classification). Recently, a number of Convolutional Neural

Corresponding author: Fatemeh Afghah (email: fafghah@clemson.edu). This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-20-1-0090 and the National Science Foundation under Grant Numbers CNS-2232048, CNS-2204445, CNS-2038741 and CNS-2038759.

[1]The "Fire" vs "No-Fire" label indicates whether fire is observed in either the RGB/IR frame in each pair. The "Smoke" vs "No-Smoke" label indicates whether smoke obscures at least 50% of the RGB frame in each pair.

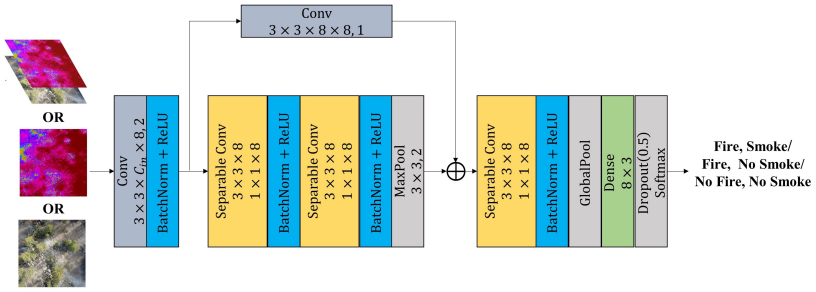

Figure 2. The architecture of the Flame network introduced in [3]. For the regular convolutional layer, the form of parameters is k k C $_{in}$Cout; stride. For separable convolutional layers, the form of parameters is k k C $_{in}$ and k k Cout. For the max pooling layer, the form of parameters is k k; stride. For the dense layer, the form of parameters is C Cout$_{in}$k denotes the kernel size, C $_{in}$ and Cout denote the number of input channels and output channels, respectively.

Networks (CNN) models have demonstrated outstanding performance on the vision-based classification task, which is often known as the most upstream task. An example of a network proposed in [3] for this purpose is shown in Fig. 2. This kind of model can further be fine-tuned to a wide variety of downstream tasks such as object detection, semantic segmentation, and instance segmentation. It is noted that new fire detection tasks or data often require the time-consuming annotation of new task data and the high computational cost of training a model from scratch. To the authors' knowledge, Transfer Learning is a strong approach to compromise this issue, whose concept is to employ prior knowledge transferred from a related domain to accelerate and enhance the new model.

Generally, We name the data from the related domain as source data and the data from the current task as target data. In fire detection, a common practice (e.g., used in [4], [5]) is to utilize feature spaces of related source data (i.e., images in other fields) and target data. This is known as Homogeneous transfer learning. The basic operation is to fine-tune the most recent state-of-the-art CNN models of some general tasks (often very large). They are pre-trained on some large natural image datasets, such as ImageNet-1K (1.28M images with 1,000 classes), which leads them to have enough capacity to extract different levels of representations from the imagery signals. Then we revise the classifier (generally, some last layers) of the model based on the purpose of the wildfire task and retrain these layers. In summary, Transfer Learning can offer such advantages: 1) fast employment since few parameters need to be re-trained on the new task; 2) rich experience in low-level feature extraction often boosts the model performance. A typical training strategy is presented in Fig.3.

In the fire domain, thermal cameras expand data redundancy, preserving feature information that is occluded to shorter, visible spectrum wavelengths. Medium and long wavelength thermal infrared cameras are able to penetrate dense smoke and foliage, providing information that would otherwise be lost in a purely visual spectrum setup [6]. Thus, some DL models use RGB-thermal image pairs as the input. This technique is often named Multi-modal learning. The essential steps of this technique include: i) in different layers, the model learns features from different domains separately. ii) at some layers, the features from two domains will be mixed, which is known as Feature Fusion. This procedure is important and is well-studied in some works [7], [8]. The most fundamental operations of this procedure include concatenation, weighted addition, etc.. It is noteworthy that different tasks with different domains may require a different strategy for appropriate Feature Fusion.

The following content is organized as follows: in Section II-B, we present some popular models and our proposed network in [3] perform on the FLAME 2 dataset, in Section II-B, we present a fast fire localization framework using multi-modal data, in Section III, we discuss current challenge regard to the fire detection task, and in Section IV, we conclude our paper.

II. Case Studies using FLAME2 Dataset

A. DL-based fire classification

We define the type of wildfire as four types: Flame with Smoke (YY), Flame with No Smoke (YN), Smoke with No Flame (NY), and No Flame with No Smoke (NN). Something noteworthy is that the class 'Smoke with No Flame' does not exist in the FLAME 2 dataset, but we retain this type for future research. Additionally, the "Smoke" class indicates whether smoke is observed to fill at least 50% of the frame, as per visual estimate by human experts [1].

In this work, we evaluate some widely used machine learning and deep learning classification models (i.e. "benchmarks"), including Logistic Regression, LetNet(1989) [9], Vgg(2014) [10], MobileNet(2017) [11], and ResNet(2016) [12], on the

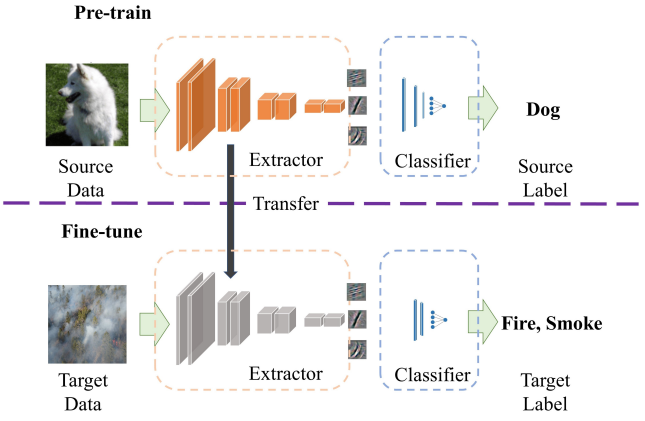

Figure 3. A typical training strategy of transfer learning. The gray layers are frozen, meaning no need to update when training in target task.

new dataset, as well as our method "Flame" [3]). Note that Vgg, MobileNet, and ResNet are pre-trained, which leverage Transfer Learning as discussed in Section I. Some of the models that used Multi-modal Learning are shown in Table II. Here, we only consider two simple methods to perform feature fusion (shown in Fig 4(a-b)), named Early Fusion and Late Fusion. Specifically, in Early Fusion, we just concatenate the paired images and modify the number of channels of the first layer input from 3 to 6. In Late Fusion, either RGB or IR will be fed to models with the same architecture (i.e., the upper stream learns from the RGB domain while the lower stream learns from the IR domain). We then concatenate the features extracted from each stream and feed the fused features to a fully connected layer to perform classification. Hence, subsequent layers can learn high-level representations from the RGB and IR domains.

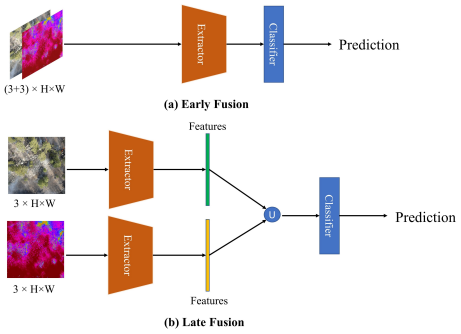

Figure 4. Two approaches for feature fusion.

In order to accelerate the testing process, for each experiment, we only used 1% randomly sampled data and split it into 80% to train (500 pairs) and 20% to test (120 pairs). Sampling also enhances the reliability of each model's performance, which alleviates the similarity of training and test datasets as consecutive frames in a video are very similar. Each model was

Table II
COMPARISON OF AVERAGED PERFORMANCE OF DIFFERENT MODELS WITH DIFFERENT MODE. THE MODELS WITH THE SIGN '*' DENOTE THE FINE-TUNED PRE-TRAINED MODELS.

Model	Mode	F1 Score	Precision	Recall	Accuracy
Logistic	RGB	89.57	90.99	89.48	90.37
Logistic	IR	92.61	92.94	92.43	92.43
Logistic	Early Fusion	96.71	96.92	96.65	96.54
LeNet5	RGB	95.39	95.86	95.12	95.33
LeNet5	IR	92.3	92.19	92.79	92.15
LeNet5	Early Fusion	97.16	97.35	97.1	97.01
Flame	RGB	94.53	95.18	94.38	94.86
Flame	IR	86.81	87.47	86.91	85.79
Flame	Early Fusion	94.88	96.01	94.95	94.86
Flame	Late Fusion	95.24	95.84	95.61	94.95
VGG16*	RGB	99.92	99.9	99.93	99.91
VGG16*	IR	97.35	97.57	97.26	97.29
MobileNetV2*	RGB	99.36	99.33	99.42	99.35
MobileNetV2*	IR	97.51	97.65	97.43	97.38
MobileNetV2*	Late Fusion	99.82	99.78	99.87	99.81
ResNet18*	RGB	98.46	98.57	98.37	98.32
ResNet18*	IR	96.54	96.97	96.27	96.26
ResNet18*	Late Fusion	99.5	99.46	99.56	99.44

evaluated ten times using the above 1% sampling approach. ADAM [13] optimizer with 1e 3 learning rate is used for the Flame network, and 1e 4 for the other models. The batch size is set to 64. Also, the label smoothing with probability 0.2 is applied in the training phase. For a fair comparison, we train the models that learn from scratch with 50 epochs and the pre-trained models with 30 epochs. We are more interested in the macro-level metrics, such as macro F1 score, macro recall, and macro precision, rather than only accuracy. This is because in the real world, wildfire is occasional, and the model's performance cannot be simply demonstrated by the classification accuracy (i.e., For instance, one may have only one wildfire sample in a total of 1,000 images. If the model labels all samples as no-fire, the accuracy would be 99.9%.

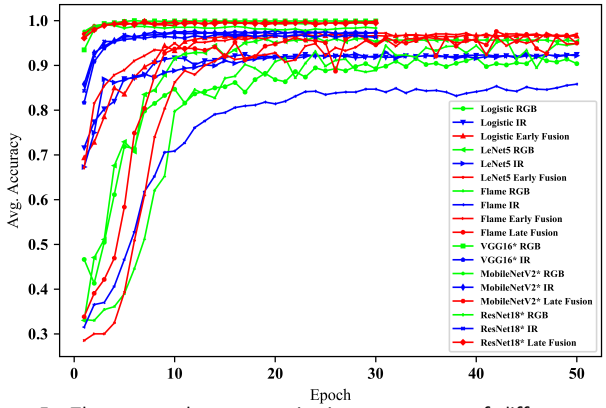

Figure 5. The averaged accuracy via time on test set of different models. TThe models with the sign '*' denote the fine-tuned pre-trained models, while the remaining models denote the models training from scratch.

The results are shown in Table II and Fig. 5. Generally, models that learn from multi-modality exhibit improved performance as compared to models that learn from a single domain. This is consistent with our discussion before. Similarly, the pre-trained models generally outperform our customized models, as they are pre-trained on large datasets and have a good understanding of the different features of fire imagery in our task. Our customized model, on the other hand, is trained using only a few hundred images with fewer parameters (Flame only has 7003,000 parameters) and can achieve reasonable results.

B. Image-processing-based fire localization

After classification, fire localization can perform by using the multi-modal data. We use Maximally Stable Extremal Regions (MSER) method to detect the image blob features and then generate bounding boxes based on these detected features. Specifically, we first convert the image to gray-scale, in which an extremal region R is defined as a contiguous subset of the image D which satisfies, for all p 2 R; q 2 @R : I(p) > I(q) or I(p) < I(q), where @R denotes the boundary of the region R, and I() denotes the intensity of the pixel. Suppose an extremal region R_i denotes the intensity of every pixel in the region is smaller than i. We define q(i) as

$$q(i) = \frac{jR_{i+} \quad R_ij}{jR_ij} \tag{1}$$

where denotes a small positive number, R_i R_{i+} always holds, and jj denotes cardinality. When q(i) is a local minimal, R_i is a maximally stable extremal region. Then bounding box is generated based on the maximally stable extremal regions. As a conventional image processing method, it does not require any data for training purposes. Moreover, this approach is stable and can perform multi-scale detection without any smoothing.

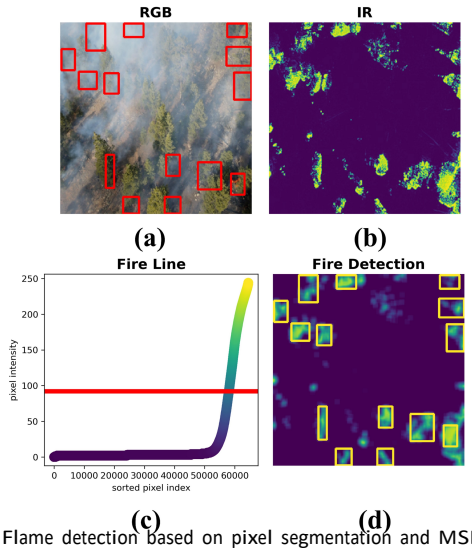

Figure 6. Flame detection based on pixel segmentation and MSER-NMS. (a) RGB image with detected flame location (red bounding box), (b) IR image, (c) Fire line which applied for pixel segmentation, pixels below the fire line will be suppressed, (d) Image processed IR image with detected flame location (yellow bounding box).

As the MSER method usually generates many partially overlapping bounding boxes for the same object. In order to avoid unnecessary calls and to provide a more precise localization of the flames, we use a Non-Maximum Suppression (NMS) method to eliminate the overlapping bounding boxes in favor of the strongest one. By fine-tuning the suppression processes of non-maximum parameters and the threshold of pixel intensity (fire-line) on IR images, the algorithm identifies areas with higher fire probability.

Figure 6 shows the result of flame detection, where the flame detection's accuracy is not affected by smoke. Thus, our proposed framework is simple, stable, computation friendly, and labor-free.

III. Challenge

Fire detection tasks often suffer from the lack of generalization as a result of cross-dataset domain shift. Each dataset has its specific underlying characteristics, such as camera angle, image scale, terrain, etc. This issue often results in poor transferable performance on the new task or catastrophic forgetting of the old task. This can be eliminated by training all data simultaneously or with Multi-task Learning; however, this is not practically feasible. By guiding model adaptations based on relations of domain knowledge between tasks, continuous learning provides a more efficient, middle-ground solution for sequential task learning. Some classic solutions included: i)regularization-based [14] and ii) replay-based [15]. It is noteworthy that the former is privacy-preserving which does not require access to the old data, while the latter often can reach a better performance. From another perspective, even if the data in the new task is insufficient annotated, Domain Adaption can alleviate the problem, which aims to leverage knowledge learned by the model from another related domain with adequate labeled data [16].

To the authors' knowledge, these paradigms for wildfire should attract the attention of the community, but only limited works focus on it. This may be because of the lack of a standardized benchmark.

IV. Conclusion

This work presents two image-processing-based methodologies, showcased on our newly released FLAME 2 dataset. The first methodology investigated multiple DL models with different training strategies, including training from scratch, Transfer Learning, and Multi-modal learning. We exhibit the respective strengths of Transfer Learning and Multi-modal learning to accelerate and enhance the detection model. We then demonstrate the fire localization with smoke occlusion based on conventional methods, which are fast, stable, and, more importantly, do not require pixel-level annotated data for training purposes. Our goal is to develop a real-time wildfire detection system for compute-limited edge devices based on our image processing methods. We also hope the community can improve our fundamental approaches and explore more tasks using the FLAME 2 dataset.

References

[1] B. Hopkins, L. O'Neill, F. Afghah, A. Razi, A. Watts, P. Fule, and J. Coen, "FLAME 2: Fire detection and modeLing: Aerial Multi-spectral imagE dataset," 2022.
[2] "ALERT Wildfire." https://www.alertwildfire.org/, 2018.
[3] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, "Aerial imagery pile burn detection using deep learning: The flame dataset," Computer Networks, vol. 193, p. 108001, 2021.
[4] H. Wu, H. Li, A. Shamsoshoara, A. Razi, and F. Afghah, "Transfer learning for wildfire identification in UAV imagery," in 2020 54th Annual Conference on Information Sciences and Systems (CISS), pp. 1–6, IEEE, 2020.
[5] M. J. Sousa, A. Moutinho, and M. Almeida, "Wildfire detection using transfer learning on augmented datasets," Expert Systems with Applications, vol. 142, p. 112975, 2020.
[6] A. Bouguettaya, H. Zarzour, A. M. Taberkit, and A. Kechida, "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms," Signal Processing, vol. 190, p. 108309, 2022.
[7] S. Guo, B. Hu, and R. Huang, "Real-time flame segmentation based on rgb-thermal fusion," in 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1435–1440, IEEE, 2021.
[8] S. T. Seydi, V. Saeidi, B. Kalantar, N. Ueda, and A. A. Halin, "Fire-net: a deep learning framework for active forest fire detection," Journal of Sensors, vol. 2022, 2022.
[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, no. 4, pp. 541–551, 1989.
[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[11] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
[12] 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, pp. 770–778, 2016.
[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[14] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al., "Overcoming catastrophic forgetting in neural networks," Proceedings of the national academy of sciences, vol. 114, no. 13, pp. 3521–3526, 2017.
[15] Z. Li and D. Hoiem, "Learning without forgetting," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 12, pp. 2935–2947, 2017.
[16] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," Neurocomputing, vol. 312, pp. 135–153, 2018.