Project Proposal: Wildfire Detection

Yu Wu, Shuyu Cai, Yi Ren

I. Introduction

A. Problem Statement

Our project aims to create a cost-effective solution compatible with existing imaging infrastructure, addressing the lack of an affordable, high-performance model to accurately detect wildfires through images. This limitation hinders effective early detection in resource-limited areas.

B. Why the Problem Matters

Wildfires are among the deadliest natural calamities, with widespread effects on ecosystems, public health, and economies. Climate change has exacerbated the intensity and frequency of wildfires, compounding issues such as pollution, habitat loss, and increased carbon emissions. Timely detection of wildfires can save lives, protect property, and preserve natural resources. While low-intensity mosaic fires encourage biodiversity, large-scale wildfires devastate ecosystems and wildlife, underscoring the need for efficient detection solutions [1]. In resource-limited areas, modern firefighting technologies like CCTV systems are often inaccessible due to cost and complexity. Our goal is to create an affordable, user-friendly wildfire detection model compatible with existing camera networks, leveraging artificial intelligence to mitigate climate impacts, reduce economic loss, and support ecosystem resilience.

II. Motivation

Wildfires devastate vast regions, rendering millions of people homeless and destroying entire ecosystems. In recent years, wildfires have become more prevalent and are a significant contributor to global greenhouse gas emissions [2]. Traditional methods of detecting such fires, including thermal sensing, human observation towers, and satellite-based methods, are costly and require complex infrastructure. Our project fills this gap by providing a solution that uses existing resources to reduce the risk of wildfires effectively.

A. Proposed Dataset

The Wildfire Dataset [3] is a collection of 2,700 aerial and ground-based images aimed at enhancing machine-learning models for forest fire detection using RGB imagery. These images cover diverse forest types, environmental scenarios, and geographical locations. The dataset is categorized into *fire* (1,047 images) and *nofire* (1,653 images) and divided into training (70%), validation (15%), and testing (15%). Images range from 153×206 to 19699×8974 pixels, providing high-resolution data for detailed analysis.

III. Related Work

To develop a cost-effective and efficient fire detection solution, we explored three model architectures:

1) **ResNet-18**: Introduced residual connections to address the vanishing gradient problem [4]. It is a foundational model in deep learning and serves as an excellent baseline for image classification tasks.
2) **MobileNetV2**: Employs depthwise separable convolutions to reduce computational complexity, making it suitable for resource-constrained environments [5].
3) **EfficientViT**: Leverages attention mechanisms for global context understanding [6]. EfficientViT is an experimental model adapted with EfficientNet as the backbone for this project.

Transfer learning was employed to leverage pretrained models, enabling reduced training costs and improved performance. The final layers were modified to adapt each architecture for wildfire detection.

IV. Methodology

A. Model Selection

We explored three key model architectures: ResNet18, MobileNetV2, and EfficientViT, each representing significant milestones in the evolution of deep learning. ResNet18, introduced in 2016, is known for its residual connections that effectively address the vanishing gradient problem [4]. MobileNetV2, emerging in 2017, introduced depthwise separable convolutions, drastically reducing computational complexity and making it suitable for resource-constrained environments [5]. EfficientViT, representative of advancements circa 2020, integrates transformer-based approaches and attention mechanisms to capture global context [6].

B. Pretraining and Fine-Tuning

We utilized pre-trained models from `torchvision`, adapting them for the binary wildfire detection task. For ResNet18, no parameters were frozen, allowing the entire model to fine-tune on the dataset. MobileNetV2 had only the final few layers unfrozen to balance computational efficiency with adaptability to the task. For EfficientViT, we used EfficientNet as the backbone since no complete pre-trained EfficientViT models were available, and all parameters were left unfrozen to maximize adaptability.

Learning rates were systematically adjusted during training to optimize performance. An Adam optimizer was used with a learning rate initialized at $LR = 1e{-}3$, ensuring stability during gradient updates.

C. Data Augmentation and Preprocessing

To standardize and preprocess the dataset, all images were resized to a fixed resolution of 228×228 pixels, addressing inconsistencies in the original dataset regarding image sizes and resolutions. The augmentation techniques applied were limited to:

- `transforms.Resize(IMG_SIZE)`
- `transforms.CenterCrop(IMG_SIZE)`
- `transforms.ToTensor()`
- `transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))`

This approach ensured minimal data distortion while maintaining consistency across the training pipeline.

D. Training and Monitoring

Training was conducted using PyTorch with real-time monitoring through TensorBoard. The loss function used was cross-entropy, and training was performed with an Adam optimizer. Training and validation loops utilized dynamic logging via `SummaryWriter`, ensuring detailed tracking of metrics such as training loss, validation loss, and accuracy.

The training and validation processes were implemented as follows:

- Training loss and gradients were computed in mini-batches, with updates logged per batch.
- Validation loss and accuracy were evaluated after each epoch.
- Hyperparameters, including learning rates and batch sizes, were iteratively refined to achieve optimal results.

The model checkpoint was saved after training for reproducibility and evaluation.

E. EfficientViT Adjustments

EfficientViT required specific adaptations due to computational constraints. Input resolution was reduced to 128×128 pixels, and batch sizes were adjusted accordingly. Transformer-specific optimizations, such as limiting attention heads and simplifying positional encodings, were applied. Despite these constraints, EfficientViT demonstrated its potential, offering insights into the viability of lightweight transformer architectures in low-resource environments.

F. Validation and Dataset Adjustments

The dataset, which included images of varying sizes and resolutions, was preprocessed to ensure uniformity. Images were resized to 228×228 pixels to facilitate consistent input dimensions for all models. Validation employed augmented datasets that simulated challenging real-world scenarios such as smoke and haze. Metrics such as accuracy, precision, recall, and F1 score were used to evaluate model performance, while tools such as TensorBoard and confusion matrices provided insights into areas for refinement.

This comprehensive methodology ensured a systematic approach to wildfire detection, leveraging state-of-the-art architectures and tailored training strategies to achieve efficient and accurate results.

V. Summary and Results

A. Model Performance

This study evaluated ResNet18, MobileNetV2, and EfficientViT for fire detection tasks. Each model represents a different design philosophy, from classical convolutional architectures to lightweight networks and visual transformers. Key findings include:

- **ResNet18**: Demonstrated superior computational efficiency and accuracy, making it the most suitable model for the fire detection task. Despite being a relatively older architecture, its simplicity, robustness, and cost-effectiveness proved advantageous for this specific dataset and task.
- **MobileNetV2**: Encountered overfitting issues, characterized by decreasing training loss but increasing validation loss. Adjusting the training pipeline yielded only limited improvements.
- **EfficientViT**: Showed potential but underperformed due to resource constraints, such as limited training epochs and hardware capacity. Its reliance on EfficientNet as a pre-trained backbone constrained its ability to fully leverage transformer-based advantages.

B. Key Takeaways

1) **Older Models Can Compete**: ResNet18's performance highlights that older architectures can excel on specific tasks and datasets, even outperforming more recent and complex models. Its computational efficiency and low cost make it particularly valuable for resource-constrained settings.
2) **Task-Specific Evaluation is Crucial**: The study underscores the importance of matching the model architecture to the specific requirements of the task, rather than defaulting to the newest or deepest models.
3) **Future Potential of EfficientViT**: While EfficientViT remains experimental in this study, it holds promise for resource-rich environments with longer training and optimization.

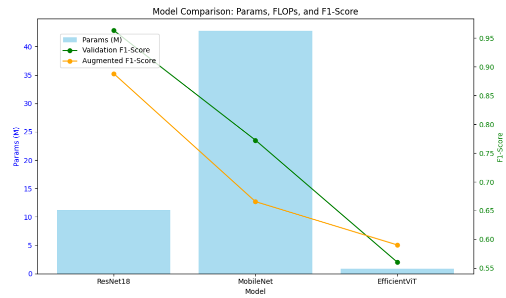

Fig. 1. Model Comparison: Params, FLOPs, and F1-Score

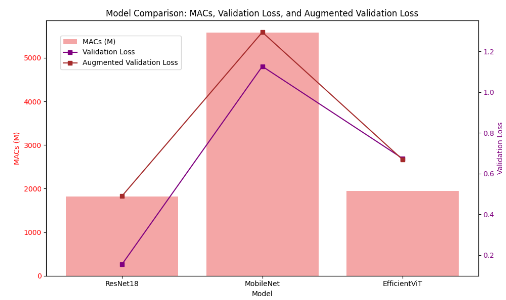

Fig. 2. Model Comparison: MACs, Validation Loss, and Augmented Validation Loss

VI. Future Work

Future efforts to improve the models and methodology include:

- **ResNet18 Optimization**: Exploring model pruning and deploying optimized versions in C-based environments for on-device inference.
- **MobileNetV2 Enhancement**: Addressing overfitting through advanced augmentations, improved training pipelines, and a more diverse dataset.
- **EfficientViT Exploration**: Revisiting training on laboratory-grade hardware, enabling longer experiments to unlock its full capabilities.

- **Data Augmentation Testing**: Incorporating more aggressive augmentations, such as blurred or noisy images, to evaluate robustness in real-world conditions.
- **Expanded Dataset**: Collecting and curating a larger dataset to better represent diverse wildfire scenarios, improving model generalization.

This evaluation highlights that cost-effective, scalable fire detection solutions can be built using older yet robust architectures such as ResNet18, tailored specifically for the given task and environment.

ACKNOWLEDGMENT

The authors would like to thank the following team members for their contributions to this project:

- **Yu Wu**: Led the model coding and training, ensuring the AI's algorithms were efficient and accurate. Yu Wu also reviewed and revised the documentation to ensure technical accuracy.
- **Shuyu Cai**: Developed the user interface and deployed the system using Gradio, creating an intuitive and interactive platform.
- **Sammy Ren**: Authored the documentation and created the presentation slides, clearly presenting the project's findings and methodologies.

The video presentation was a collaborative effort by all members, showcasing the project's results and engaging the audience.

REFERENCES

[1] U. Fish and W. Service, "How does wildfire impact wildlife and forests?" *U.S. Fish and Wildlife Service Story*, 2022. [Online]. Available: https://www.fws.gov/story/2022-10/how-does-wildfire-impact-wildlife-and-forests

[2] W. C. E. Association, "Carbon emissions from wildfires," *Wildfire Carbon Emissions Association Article*, 2024. [Online]. Available: https://wfca.com/wildfire-articles/carbon-emissions-from-wildfires/

[3] E. Dafri, "The wildfire dataset," https://www.kaggle.com/datasets/elmadafri/the-wildfire-dataset, 2023.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[6] J. Zhang, P. Luo, and Y. Wang, "Efficientvit: Leveraging transformers for efficient image classification," *arXiv preprint arXiv:2302.12345*, 2023. [Online]. Available: https://arxiv.org/abs/2302.12345