PAPER • **OPEN ACCESS**

Pyramid Attention Based Early Forest Fire Detection Using UAV Imagery

To cite this article: Yongtuo Zhang *et al* 2022 *J. Phys.: Conf. Ser.* **2363** 012021

View the article online for updates and enhancements.

Pyramid Attention Based Early Forest Fire Detection Using UAV Imagery

Yongtuo Zhang[1], Sheng Chen[1], Weiyi Wang[2], Wei Zhang[1] and Li Zhang[1*]

[1] Tianjin College, University of Science and Technology Beijing, Tianjin, China
[2] Complex Systems Research Group, Faculty of Engineering, The University of Sydney, Darlington, New South Wales, Australia
Email: vic.ytzhang@outlook.com, chenshenguav@outlook.com, weiyi.wang@sydney.edu.au, kyoosan450@163.com, lee.lizhang@outlook.com

Abstract. Early forest fire detection using unmanned aerial vehicle (UAV) imagery is widely implemented in recent years. However, as the size of moving object in UAV imagery changes dramatically, the issues of focusing more informative areas of an image, large-scale variation as well as temporal consistency preservation have yet to be resolved due to the wide field of vision of UAVs in flight. To address these issues, in this work, we concerned the visual characteristics of early stage forest fire in UAV imagery and employ pyramid attention mechanism. We engineered a novel convolutional neural network (CNN) based on the state-of-the-art backbone by stacking our proposed PyrAtten blocks. We built a large dataset of early stage forest fire, including real world aerial photographs shot using our UAV prototype. Extensive evaluation of PyrAtten demonstrates its efficiency and effectiveness in detecting multi-scale objects of fire and smoke in UAV imagery, even if the object takes up a very small proportion of the whole image. Compared with existing similarly-sized networks, average detecting accuracy of PyrAtten-ResNet-50 reaches 97.5%, with negligible increase in computational overhead.
Keywords. Forest fire detection; UAV imagery; deep learning; pyramid attention mechanism.

1. Introduction

Forests always play a significant role in stabilising climate, regulating the water cycle, and providing habitat to thousands of life forms. Forests are also of vast economic benefits to human beings. However, in recent years, due to abnormal global climate change and continual human activities in forests, forest fires occur more frequently, which poses a huge threat to the earth's ecological environment and the safety of human life and property. As wood is inflammable, forest fire can spread incredibly quickly, and the latter it is found and controlled, the harder it is to turn it off. Hence, if the forest fire or its smoke is accurately and timely detected, localised and reported in the early stage of occurrence, it will help fire departments to take measures to minimise the spread of forest fire in time, so as to avoid causing disastrous consequences.

Conventionally, forest fires are detected basically via human patrol following fixed routes and observation from watchtowers located across the monitored area. However, human patrol is not only of low efficiency but also labour-intensive. Watchtowers have limited sight that can barely cover the whole area of interest. Although satellite remote sensing is capable of detecting large-scale forest fires, the cost of start-up and maintenance is pretty high, and the monitoring effect is often affected by weather, cloud thickness, orbital period etc., let alone sizes of early forest fires [1]. As flight control technology improves by leaps and bounds, unmanned aerial vehicles (UAVs) have been widely used in the field of forest fire detection due to flexibility and low cost. UAVs flying at a low altitude above the forest can detect forest fires, especially in their early stage, more efficiently and localise them more precisely [2]. Figure 1 shows the UAV prototype we assemble for the mission of early forest fire detection. As the embedded processor on it is computationally resource-restricted, the UAV in our system is used as the data collector, and transmits image streams to ground control station where forest fire and smoke detection works. Practically, a swarm of UAVs whose detection missions are planned in advance can sufficiently cover the task forest area searching for fire and smoke round the clock [1].

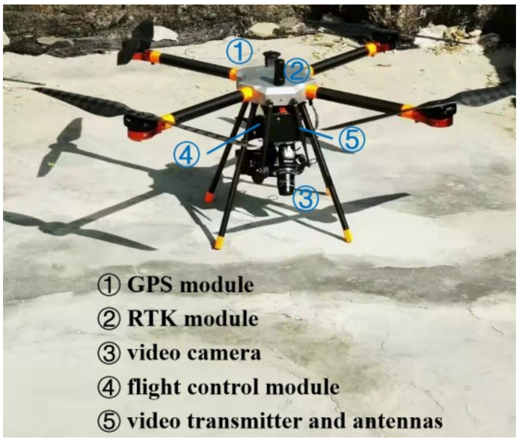

Figure 1. The UAV prototype for early forest fire detection.

Stimulated by the rocketing progress of deep learning in computer vision in the last decade, forest fire detection based on convolutional neural networks (CNNs) using UAV imagery is getting widespread as well. We therefore consider the characteristics of early stage forest fire and network complexity for low-cost UAV systems. In this work, we propose a novel UAV system, which utilise our proposed Pyramid Attention (PyrAtten) mechanism for early stage forest fire detection. We implement our pyramid attention blocks for channel-wise attention module and replace the 3×3 convolutions in the bottleneck blocks of the classic ResNet [3] with our PyrAtten modules. Extensive experiments demonstrate that our method achieves promising improvement with an acceptable computational cost increment for UAVs, remaining better detection accuracy and efficiency.

2. Related Work

2.1. Vision Based Fire Detection Using UAV
UAV based fire detection has been widely implemented in the world with the aid of computer vision technology. Hossain et al. [4] study and leverage features of forest fire and smoke extracted in colour space, setting a remarkable example of using feature engineering in computer vision to optimise forest fire detection algorithms. Sudhakar et al. [5] concern the false forest fire alert resulting from the characteristics of flying UAVs, considering the instability of both UAV and smoke recognition as challenges for detection algorithm researchers. CNN based forest fire detection using UAV has recently been widely deployed [6, 7], and the trade-off between high-performance detection and restricted computation resources of UAVs is taken into account as a practical issue to iron out.

2.2. Attention Mechanism
Attention mechanism has been proved to be efficient in exploiting discriminative areas of the image while mitigating the misalignment [8], making the model under certain context attending to the more informative parts adaptively.

SENet [9] extracts channel-wise attention by conducting two major operations on a feature map: "squeeze" it with global average pooling in channel dimension and then "excite" it to compute recalibrated feature map. CBAM [10] is another network module focusing on self-attention. Motivated by SENet, CBAM considers both channel attention and spatial attention which is computed by 2D convolutions with large kernel sizes, and finally combine them. ECA-Net [11] re-weights the feature map using a 1D convolution of kernel size k to locally capture inter-channel relations of each channel and its k neighbours. ECA-Net also alleviates the use of fully connected layer, making its implementation more light-weight. Coordinate attention [12] is designed to establish long-range dependencies for vision tasks and easy to be plugged into mobile networks, which makes it running efficiently on mobile devices without more notable computational cost. As early stage forest fire and smoke in UAV imagery are multi-scale and frequently change their sizes, we piggyback the idea of pyramid attention used in semantic segmentation [13] and propose a novel design of PyrAtten block, which can be easily plugged into existing network backbones and improve detection performance significantly with little model complexity increment.

3. Methodology

This section describes how our UAV system inspects the forest area of interest autonomously, and how the proposed PyrAtten works and is used on one of the state-of-the-art backbones.

3.1. Workflow of Forest Fire Inspection Using UAVs

Figure 2 shows the workflow of the task of fire inspection using UAVs in our system.

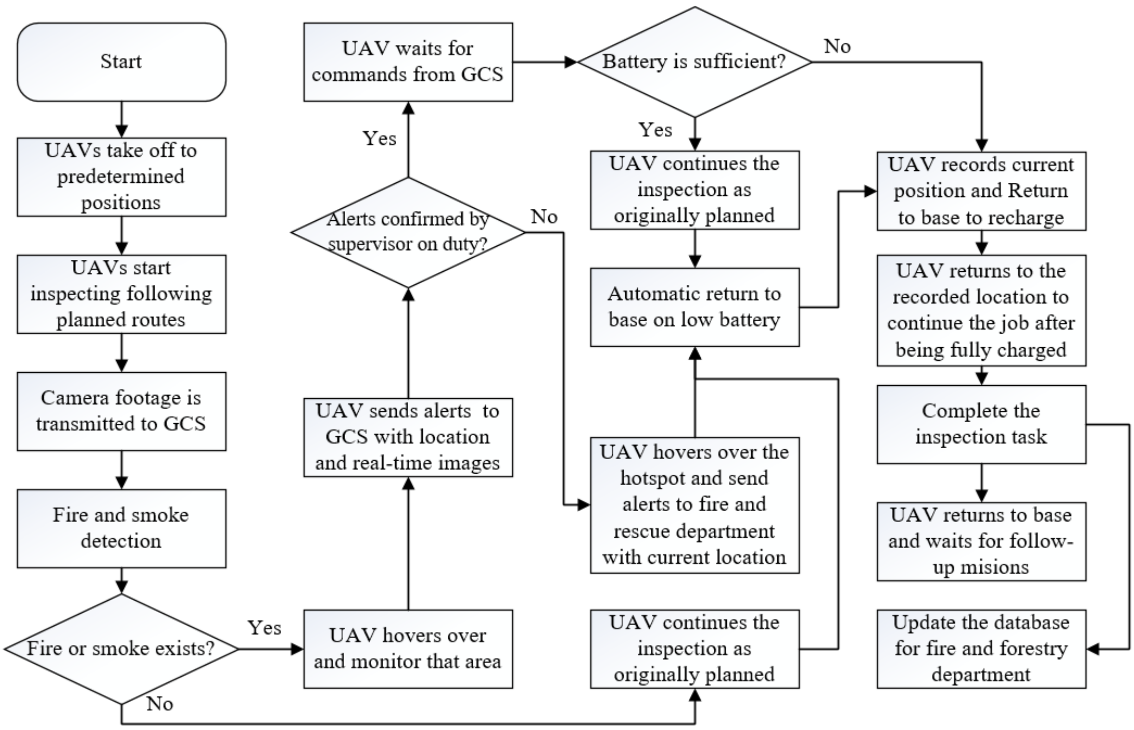

Figure 2. Workflow of fire inspection using UAVs in our system.

In our system, after mission planning on the ground control station, a UAV takes off from its automatic charging station for fire inspection, and can fly back to the charging station to get charged autonomously. The drone's thermal imager and visible-light camera capture images of the forest in real time. The data transmission module transmits the infrared image stream and the visible light image stream to the data processing module implemented on the ground control station, which uses the flame and smoke detector based on deep learning to detect flame and smoke. The inspection is conducted following the planned inspection task path until flame or smoke is detected. If an anomaly is found in the received infrared frame and is recognised as flames or smoke in a certain number of frames in the image stream that last for several seconds, a pop-up alert is continuously sent to the forest department personnel on duty until the manual confirmation is made. If it is artificially identified as a fire, the system automatically reports to the fire department that there is a fire alarm at the location of the drone, and sends a command for the drone to continuously hover around and monitor the fire location, so as to provide real-time pictures and geographic information for the firefighters to carry out rescue. The alarm will be cancelled if it is artificially determined that it is not a fire. If there is no manual confirmation for 5 minutes from the start of the alarm, the fire alarm will be reported directly to the fire department, and the drone will be to monitoring the location of the fire continuously. The data of the daily inspection of the UAV is compiled into a database of fire risk based on a geographic information system, which reflects the historical data and incidence of fires in the forest areas of interest on the monthly bases, which can provide a reference for forestry, fire protection and other departments to implement fire risk assessment and fire prevention. For very vast forest areas, the system can adopt UAV swarms. We use a mission planning software to divide the inspection area and assign them to UAVs, and the UAVs will conduct uninterrupted inspection of the allocated area according to the planned path.

3.2. Early Stage Fire Detection Motivated Design of Pyramid Attention

Pyramid attention mechanism is widely used in semantic segmentation and is proved effective in object detection. Empirically, early stage forest fire is commonly on different scales in UAV imagery, as the height at which a UAV flies varies with the changes in terrain to avoid collisions with trees or rocks. Performance of state-of-the-art CNN architectures nowadays is shown significantly improved, and thereby engineering a new backbone is not what we concern in this work. Therefore, piggybacking on previous high-performance deep neural networks, we investigate the focus on the structure of pyramid attention designated to suppress or emphasize feature maps for effective early stage forest fire detection.

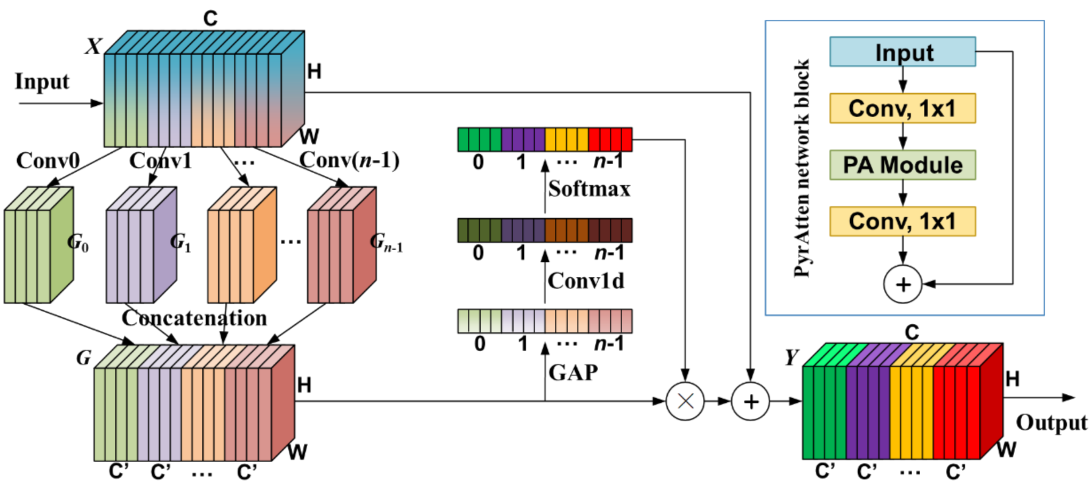

Figure 3. Illustration of the proposed PyrAtten module and PyrAtten network block.

We employ ResNet as the backbone of the network. As illustrated in figure 3, in our system, we engineer a novel PyrAtten module to generate pyramid attention with both channel and spatial information, and integrate them in ResNet as PyrAtten network blocks.

In a PyrAtten module, an input feature map $X \in R^{C \times W \times H}$ with C channels is first processed at n different scales in parallel to obtain it spatial information. Then a feature map pyramid can be calculated using group convolution method, from $Conv0$ to $Conv(n-1)$, of correspondingly n different kernels. G_i in the pyramid has the same number of channels, notated as C'. The group size of G_i is set to $2^{(k_i-1)/2}$, where k_i is the i-th kernel size, and G_i is therefore calculated by

$$G_i = \mathrm{Conv}(i)(X) = \mathrm{Conv}(k_i \times k_i, 2^{(k_i-1)/2})(X) \quad i = 0,\ 1,\ ...,\ n-1 \tag{1}$$

where $G_i \in R^{C' \times W \times H}$. Using a concatenation operation in the channel dimension, the feature map G with multi-scale spatial information is then aggregated as

$$G = \mathrm{Concat}([G_0,\ G_1,\ ...,\ G_{n-1}]) \tag{2}$$

where $G \in R^{C \times W \times H}$. Then we start to get the channel-wise information with a global average pooling, which reshape the feature map into $C' \times 1 \times 1$. A 1D convolution of an adaptively determined size is then used to make sure the interaction of neighbouring channels. We then apply a softmax function on it, thereby obtaining the attention vector, and an element-wise multiplication is performed to get the re-weighted version of G, which is later added by X to obtain the final output feature map $Y \in R^{C \times W \times H}$. Y is the emphasized tensor with both spatial and channel-wise information considered, and in our evaluation of the system, the network with PyrAtten network blocks is verified to be effective and efficient.

Table 1 shows the design of the proposed PyrAtten-ResNet-50, together with vanilla ResNet-50 and SE-ResNet-50. We engineer PyrAtten-ResNet-50 by stacking PyrAtten blocks like ResNet-50. Note that we elaborate kernel size with which group size of group convolution can be calculated.

Table 1. Network design of ResNet-50, SE-ResNet-50 and our proposed PyrAtten-ResNet-50.

Output size	ResNet-50	SE-ResNet-50	PyrAtten-ResNet-50
112×112	7×7, 64, stride 2		
56×56	3×3, max pool, stride 2		
56×56	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix} \times 3$ w/ *fc*, [16, 256]	$\begin{bmatrix} 1\times1, & 64 \\ PyrAtten, & 64 \\ 1\times1, & 256 \end{bmatrix} \times 3$ w/ $k = 3$
28×28	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix} \times 4$ w/ *fc*, [32, 512]	$\begin{bmatrix} 1\times1, & 128 \\ PyrAtten, & 128 \\ 1\times1, & 512 \end{bmatrix} \times 4$ w/ $k = 5$
14×14	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix} \times 6$ w/ *fc*, [64, 1024]	$\begin{bmatrix} 1\times1, & 256 \\ PyrAtten, & 256 \\ 1\times1, & 1024 \end{bmatrix} \times 6$ w/ $k = 7$
7×8	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix} \times 3$ w/ *fc*, [128, 2048]	$\begin{bmatrix} 1\times1, & 512 \\ PyrAtten, & 512 \\ 1\times1, & 2048 \end{bmatrix} \times 3$ w/ $k = 9$
1×1	GAP, 1000d-*fc*, softmax		

4. Experiments

In this section, our system is evaluated using a benchmark dataset of more than 8000 images of early stage forest fire and smoke. A UAV prototype we assembled is also used in real world experiments to collect images of forest fire and smoke. Fire and smoke objects are of different sizes in images as they are in the real world. All networks we test are implemented using PyTorch on a PC with one Intel(R) Xeon Silver 4208 CPU, one NVIDIA GeForce RTX 3070Ti GPU and 32GB memory.

Figure 4 gives some detecting results using PyrAtten integrated networks, in which from figures 4a to 4c, the size of the object is increasing while the image resolutions remain unchanged. Although the size of early stage fire and smoke in UAV imagery changes significantly, ResNet-50 with the PyrAtten blocks integrated is competent to extract multi-scale representation of feature maps layer by layer.

(a) (b) (c)

Figure 4. Detecting results of multi-scale smoke of early stage forest fire using PyrAtten blocks.

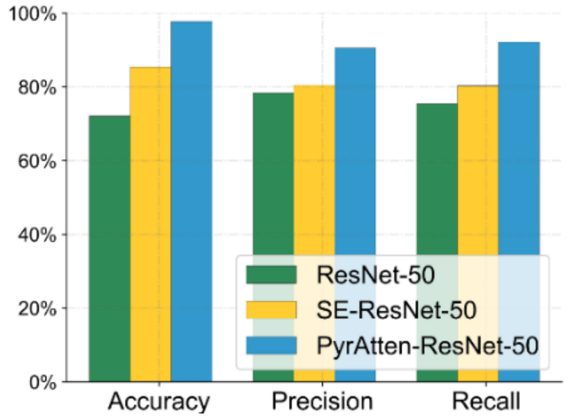

Figure 5. Comparison of accuracy, precision and recall for early forest fire detection in our dataset.

PyrAtten-ResNet-50 also outperforms in accuracy, precision and recall compared with ResNet-50 baseline and SE-ResNet-50 which consider only channel-wise attention, as shown in figure 5. The average accuracy for the early stage forest fire (smoke included in certain cases) detection of PyrAtten is 97.5%, 25.2% and 12.3% higher than that of ResNet and SE-ResNet, respectively. Besides, the number of SE-ResNet-50's parameters is 29.54M in our setting, 5.24% and 15.57% more than ResNet-50 and SE-ResNet-50, respectively, indicating that we manage to improve detecting performance eminently at the cost of minor increase in computational overhead.

5. Conclusion

The ability to focus on important areas of an image while discarding irrelevant ones is a hotspot worth studying, which can facilitate a wide range of applications s in the fields of deep learning based object detection. As early stage forest fire, especially the smoke caused by the fire, in UAV imagery is multi-scale and frequently changes their sizes. In this paper, we utilise pyramid attention combined with global context information to learn better feature representation and build long-range dependencies for the model. Taking into account the visual characteristics of early stage forest fire in UAV imagery, we engineer a novel CNN architecture equipped with our proposed PyrAtten blocks. Extensive evaluation of PyrAtten demonstrates its effectiveness and efficiency in detecting multi-scale objects of early stage forest fire and smoke in UAV imagery, even though the object takes up merely a small proportion of the whole image.

References

[1] Sherstjuk V, Zharikova M and Sokol I 2018 Forest fire monitoring system based on UAV team, remote sensing, and image processing *Proc. IEEE 2nd Inter. Conf. on Data Stream Mining Processing* pp 590-594.

[2] Hossain F A, Zhang Y M and Yuan C 2019 A survey on forest fire monitoring using unmanned aerial vehicles *Proc. 3rd Int. Symp. on Autonomous Systems* pp 484-489.

[3] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770-778.

[4] Hossain F A, Zhang Y M and Tonima M A 2020 Forest fire flame and smoke detection from UAV-captured images using fire-specific color features and multi-color space local binary pattern *J. Unmanned Veh. Syst.* **8** 285-309.

[5] Sudhakar S, Vijayakumar V, Kumar C S, Priya V, Ravi L and Subramaniyaswamy V 2020 Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires *Comput. Commun.* **149** 1-16.

[6] Zhao Y, Ma J, Li X and Zhang J 2018 Saliency detection and deep learning-based wildfire identification in UAV imagery *Sensors* (Basel) **18** (3) 712.

[7] Chen Y, Zhang Y, Xin J, et al. 2018 A UAV-based forest fire detection algorithm using convolutional neural network *Proc. IEEE 37th Chinese Control Conf.* pp 10305-10310.

[8] Chen G, Gu T, Lu J, Bao J A, Zhou J 2021 Person re-identification via attention pyramid *IEEE Trans. Image Process* **30** 7663-7676.

[9] Hu J, Shen L, Albanie S, Sun G and Wu E 2020 Squeeze-and-excitation networks *IEEE Trans Pattern Anal Mach Intell.* **42** (8) 2011-2023.

[10] Woo S, Park J, Lee J Y and Kweon I S 2018 CBAM: Convolutional block attention module *Proc. European Conf. on Computer Vision* pp 3-19.

[11] Wang Q, Wu B, Zhu P, Li P, Zuo W and Hu Q 2020 ECA-Net: Efficient channel attention for deep convolutional neural networks *IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 11531-11539.

[12] Hou Q, Zhou D and Feng J 2021 Coordinate attention for efficient mobile network design *IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 13708-13717.

[13] Sang H, Zhou Q and Zhao Y 2020 PCANet: Pyramid convolutional attention network for semantic segmentation *Image and Vision Computing* **103** 103997.