ICIUS-2022-P220076

U-net-based Wildfire Segmentation with Fastai

Linhan Qiao[1], Youmin Zhang[2*], Yaohong Qu[3]

[1,2] *Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, Canada*, H3G 1M8

[3] *School of Automation, Northwestern Polytechnical University, Xi'an, China*, 710012

Corresponding author. E-mail: ymzhang@encs.concordia.ca

Abstract

This paper purposes a scheme of building U-net with Fastai framework for early wildfire detection, which could be deployed on a ground workstation or an on-board computer of an unmanned aerial vehicle if the computation power is sufficient. The designed model in this paper works especially for smoke detection. This U-net model is based on Resnet34, and attention gates are applied on skip connections through Fastai. Different from Tensorflow or Pytorch, Fastai is a higher level application programming interface, by which the transfer learning could be easier and faster deployed for network tuning. The data used in this paper is mixed by data-sets from Kaggle and experiment captured images. The experiment segmentation results show that, after unfreezing the entire encoder-decoder structure, the accuracy is acceptable with a level of 94.9%, the attention gates helped reducing the impacts of noisy data and improved the generalization ability of this model.

Keywords wildfire detection, U-net, Fastai, Attention gate.

1 Introduction

Wildfire is threatening forest resources a lot and it is also considered as a disaster to the life safety of animals and humans who live in or nearby forests. Commonly, the wildfire could be considered as an abnormal state of forest or grass environment system. Because of the complexity of decoupling the flame action and environmental varieties, such as vegetation, wind action, topography during the process of wildfire spreading, it is necessary to detect and manage the wildfire at its early period. Discovering and detecting wildfire early is demonstrated as one of the most efficient and demanded schemes for preventing or managing wildfire to decrease losses. Comparing with filtering schemes or other kinds of sensors, the RGB cameras cooperate with infrared camera could capture the image or heating information in forest more timely for recognising wildfire. Encouraged by the development of the study of neural network models and computer vision (CV) technology, computational intelligent image-based or video-based wildfire detection is now being widely accepted and applied in forest surveillance and forestry study area[2].

U-net, one of these typical NN-based segmentation models, is an encoder-decoder structured model for image segmentation or object detection [12]. Especially, for some medical applications, U-net is considered as one of the most efficient model to use the graphics processing units (GPUs) memory. Also, different level features are extracted from multiple stages of the encoder part of U-net, which means that smaller dataset could be used to train an acceptable segmentation model. Commonly, in wildfire detection, the data of forest environment is not easy to acquire, and the available data are less. Therefore, U-net is recommended to detect wildfire in early period. To be more specific, U-net is appropriate to detect or segment smoke during early wildfire as the non-regular shape smoke raise before flames are visible [9].

Based on the analysis above, this paper proposes a ResNet34-based U-net architecture smoke detection scheme using Fastai framework. The rest part of this paper is arranged as follows: In Section 2, related works of U-net and smoke segmentation are briefly reviewed. The methodology of U-net based wildfire detection is stated in Section 3. Details of the design of experiments and tests are arranged in Section 4. Finally, conclusions and potential future works are stated in Section 5.

2 Related Works

With the great development of the study of neural network (NN) or even deep neural network (DNN) models, there are many studies considering NN for smoke segmentation because of their faster convergence and deployment through the acceleration of matrix computation (development of GPU). Commonly, there are three typical models could perform well for segmentation tasks: non-end-to-end model, end-to-end model and sequence-to-sequence model.

Non-End-to-End Model: Region-based convolutional neural networks (R-CNNs) are considered as non-end-to-end model for detection tasks. At first, this kind of model extracts region-proposal areas where the targets are suspected to be located before input. Then, these region-proposal areas are loaded into CNNs to detect whether the targets are contained to finish the detection missions. Faster R-CNN is one typical and popular non-end-to-end model for wildfire detection [18]. The advantages of such schemes are that the parameters could be shared and the detection accuracy could be increased significantly. Based on the concept of region proposal, a bounding box of the target area will be output as the region-proposal area are loaded into the model [10]. You-look-only-once model (YOLO) achieved faster and lighter-weight for simple detection tasks [11]. However, the performance of YOLO for multiple no-rigid bodies, such as multiple smoke areas in forest scene, is still challenging.

End-to-End Model: Compared with the region-proposal scheme of non-end-to-end models, end-to-end models could be considered as a process in which the neural network model learns to extract feature itself. Because the feature extraction or feature learning all happen inside the model after masks or labels are supplied, one of the salient advantage of end-to-end scheme is less emphasis on the feature itself, which decreases the faults of manual feature extraction. A fully convolutional network (FCN) is considered for wildfire detection early [14]. Based on these models, some schemes are developed by focusing on increasing the accuracy or training speed [15]. And in recent years, different structured models are designed for better wildfire detection performance. A 3D-fully connected pyramid classification is designed in [6] to deal with the false-positive problem when detecting wildfire smoke. U-net could also be considered as a kind of FCN for wildfire detection. It could be transferred to the wildfire detection or forestry applications. In [17], attention gate (AG) enhanced U-net and squeezeNet modules [9] are applied for fire or flame detection. However, that model only classifies flame, does not consider the smoke of early wildfire period, and the detection accuracy is around 81%, which still need to be improved.

Sequence-to-Sequence Model: Some of the recurrent neural networks (RNNs), such as long-short-term memory (LSTM) model [5], could be defined as Sequence-to-Sequence model and have proved their performance in wildfire detection. Because of the re-hot of attention mechanism, there are also many schemes preferring to apply attention layers with their model for potential future works [1]. However, performance of models with pure attention mechanism and some attention enhanced network models such as vision transformer [13] still need more time and demonstration.

3 Methodologies

Fastai is a Pytorch-based application programming interface (API) which provides high-level components that can quickly and easily provide state-of-the-art results in standard deep learning domains, and provides low-level components that can be mixed and matched to build new approaches [4]. Therefore, some convenient methods are supported in Fastai to help build neural networks for specific missions.

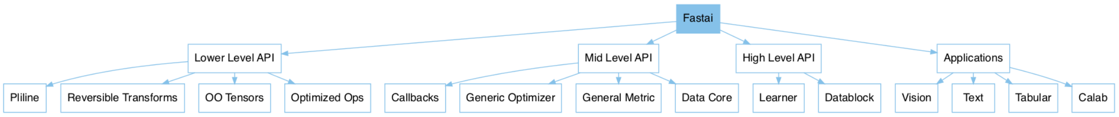


Figure 1. The layered API of Fastai. [4]

It is illustrated in Fig. 1 that Fastai consists different level of APIs. Low level APIs and high level applications are all contained in Fastai to help meeting design goals. For the objective of this paper, it is possible to create a state-of-the-art vision model using transfer learning to segment smoke in forest scene so that the wildfire could be detected early through forestry monitoring systems, such as watch towers and especially the unmanned aerial vehicles (UAVs).

Because of the advantage of U-net that it does not require large data set. And the short skip-connections of U-net insured to capture enough original features with limited number of data. U-net is appropriate to detect wildfire smoke and flame. There are also some optimization for U-net to improve its performance. In recent research, attention mechanism is getting hot to optimize or develop the CNN models [17], attention gate could be designed on skip-connections to deal with the false-positive detection problem, and the model is also possible to be squeezed as a smaller model to work on-board.

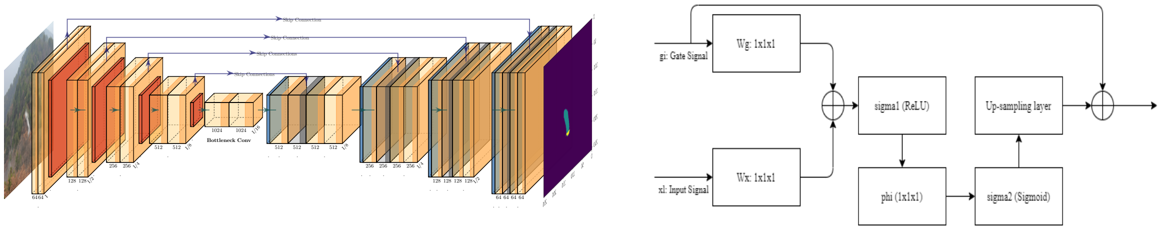
As analyzed above, smoke image segmentation is considered as an efficient scheme for early wildfire detection, because the smoke could be viewed much earlier than flame. By analyzing these evidences, some characters of U-net make it more suitable for early wildfire detection or wildfire smoke segmentation task:

Encoder-Decoder Structure: The U-net is typical encoder-decoder structure network which connects sub-sampling processes to up-sampling processes through skip connections. The skip-connection of U-net helps to avoid the size of image shrink too much. It could also be thought as a process of multi-scaled features mixture, where the features are added pixel by pixel, and the concatenation of feature maps [7].

Multi-scaled Features Extraction: The real wildfire smoke are often not that easy to get, the size of data set for training is often limited. This reason caused the low level features of original images very important to train the network. Multiple layers of down-sampling of U-net could ensure to capture most low-level features. On the other hand, the flame and smoke are non-rigid body, they do not have stable and clear edges. The low resolution and high resolution process of U-net can help to understand the edge of smoke to segment them earlier.

Therefore, U-net is one of the most appropriate NN-based schemes for wildfire detection. With the help of Fastai, it is simpler to build up a U-net and optimize it with attention gate to finish the smoke detection task.

The U-net model in this paper is based on Resnet34, it could be illustrated in Fig. 2a. The proposed structure in this paper is based on the first stream FCN of the 'two-stream' model which is stated in [15]. It ensures the model extracts enough features from inputs. Between the Convolution and ReLU activation, there is a Batch-Norm to averaging the batch height and weight by channel [16].



(a) Schematic diagram of the structure of U-net. **(b) Brief process of attention gate.**

Figure 2. Brief architecture of the designed Attention U-net for wildfire detection.

Pure U-net is not enough for the wildfire segmentation. The skip connections in U-net are concatenating the cropping results of the down-sampling part layers to up-sampling part layers, where crop functions are resizing the encoder part feature maps into the same size so that they could match the decoder part ones. Because of the challenge of smoke detection that small objects with large shape variability lead to higher false-positive predictions, the attention gates (AGs) are applied in this work to replace original skip-connections. The architecture of attention gates as skip connections could be briefly illustrated in Fig. 2b. As shown in Fig. 2b, gating vector $g_i \in \mathbf{R}^{F_g}$ for each pixel i contains contextual information to prune lower-level feature responses, and additive attention could be applied to get the gating coefficient. Balanced the squeeze module significantly decreased the computing complexity, addictive attention sacrificed a small computing resource to achieve higher accuracy. The additive attention could be formulated as Eq. (1):

$$q^l_{attention} = \phi^T(\sigma_1(W^T_x x^l_i + W^T_g g_i + b_g)) + b_\phi; \qquad \alpha^l_i = \sigma_2(q^l_{attention}(x^l_i, g_i; \bigtriangledown)) \tag{1}$$

where σ_1 and σ_2 separately remain the activation of ReLU and sigmoid function; $\bigtriangledown$ consists of a set of AG parameters: the linear transformations $W_x \in \mathbf{R}^{F_l \times F_{int}}$, $W_g \in \mathbf{R}^{F_g \times f_{int}}$, $\phi \in \mathbf{R}^{F_{int} \times 1}$, and the bias terms $b_g \in \mathbf{R}$, $b_\phi \in \mathbf{R}$; α^l_i is acquired by linearly mapping concatenated features x and g to a $\mathbf{R}^{F_{int}}$ dimensional space.

Information extracted from coarse scales (Encoders) are used as gating signals to disambiguate irrelevant and noisy responses in skip connections. And AGs filter the activation of neurons during forward pass and backward pass. It enhanced the parameters of shallow layers to update based on special regions. It could also be said that only crucial features are paid attention (Why it is called attention gate). The update rule in layer $l-1$ could be formulated as Eq. (2)

$$\frac{\partial(\hat{x}^l_i)}{\partial(\Phi^{l-1})} = \frac{\partial(a^l_i f(x^{l-1}_i; \Phi^{l-1}))}{\partial(\Phi^{l-1})} = a^l_i \frac{\partial(f(x^{l-1}_i; \Phi^{l-1}))}{\partial(\Phi^{l-1})} + \frac{\partial(a^l_i)}{\partial(\Phi^{l-1})} x^l_i \tag{2}$$

where x^{l-1} remains the feature map in layer $l-1$; i is the spatial subscript, the function $f(x^l - 1; \Phi^l - 1) = x^l$ applied in convolution layer l is characterised by trainable kernel parameter $\Phi^l - 1$; a_i remains the attention coefficients which identity salient image regions and prune feature responses to preserve only the activation (ReLU, Sigmoid) relevant to the specific task.

The negative cross entropy is chosen as the loss function of U-net,

$$E = -\sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)) \tag{3}$$

where x is the positions of pixels in a whole image; $p_{l(x)}(x)$ represent the possibility of real mask label output on feature channels (final output 3); $w(x)$ is the weight to emphasis more on the boundaries between objects to detect. Therefore, the output pixel-wise softmax is

$$p_k(x) = \exp{(a_k(x))}/(\sum_{k'=1}^{K} \exp{(a_{k'}(x))}) \tag{4}$$

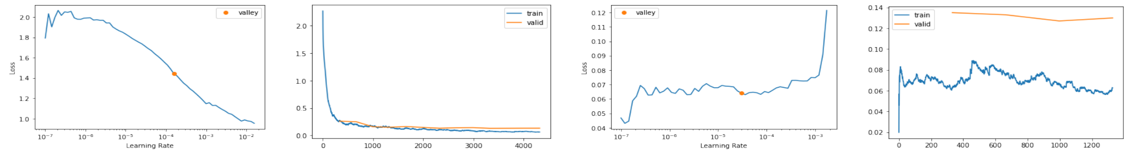
where $a_k(x)$ is the value of pixel x on channel k of output layer; K is the number of classes. Then, the output $p_k(x)$ of Eq. (4) is the possibility of x belonging to class k.

4 Experiments

The data-set used in this paper consists of an open data source from ‘Kaggle’, where these smoking images are captured by watchtower or UAVs. And the data comes from authors’ experiment. In this experiment, the smoking images are captured by DJI Phantom 4. These two parts data are shuffled and labelled through pixel region of interest (RoI) labels by MATLAB Image-Labeller. By which, masking labels are handle-made. The U-net in this work is deployed through the `unet_learner` of `learner` of Fastai. The input to the `unet_learner` includes data-loader, built U-net model, and evaluation ratios.

Training: Simulated fire annealing is applied for finding the learning rate. With the support of CUDA, half-float precision (FP16) is applied for saving computational resource or Ram [8]. With the help of Fastai, the learning rate could be found through simulated annealing algorithm (SAA) [4]. As SAA is applied, the learning rate will be increased in a slope of 1, and then getting down slowly. As shown in Fig. 3a, it finds the appropriate learning rate where the convergence happens most quickly. Based on experiences, the learning rates could often be chosen from $1e^{-6}$ to $1e^{-3}$, where it is assigned as $1.58e^{-4}$. 415 images are used as data set for the training, and the model is trained for no more than 20 epochs to avoid over-fitting. Fig. 3b shows the changing of the training loss and validation loss. It could be seen that the losses are both decreasing and the slope of validation loss is finally near zero, which means the model is trained enough, it is time to stop the iteration, or there would be over-fitting.

After unfreezing the model to use entire Encoder-Decoder structure for predicting, a trick of half-float computing [3] could be applied, which is helpful in saving the computation resource. Finally, the fine-tuned U-net parameters could be saved for testing.



(a) (b) (c) (d)

Figure 3. Training U-net process. (a) Finding learning rate through the convergence of loss using Fastai. (b) The trend of the training loss and validation loss when Decoder part is locked (frozen). (c) Finding learning rate through the convergence of loss as Decoder is opened. (d) The trend of the training loss and validation loss when entire Encoder-Decoder structure is opened (unfreeze).

With the help of the GPU GTX 1660s (6GB), it takes about 35 seconds for every epoch in frozen training part and 41 seconds for every epoch in the unfreeze training part. The final accuracy is 94.916%, which is acceptable for wildfire smoke and flame detection. So, such a trained model is saved for testing.

Testing and Results: The prediction performance comparison between our model and pure U-net is shown in Fig. 4. The last two columns are experiment images and prediction results, where a smoke cake ignited. However, in the third column, it shows that both our model and pure U-net predicted a suspect flame area, which means the feature learning of the model is not only limited in color feature. In the last column result, the pure U-net prefers to advocate the yellow color leaves under daylight to be flame, which is false-positive prediction. The prediction of our model shows the correct result. Such a result is a testimony of the statement that the self-attention mechanism helps to reduce the false-positive predictions for smoke and flame. It means our model is performing better on the aspect of generalization and robustness. The prediction result is acceptable even there are some noisy images where might be the flame like colors or shapes, such as the yellow color leaves under daylight in the last column.

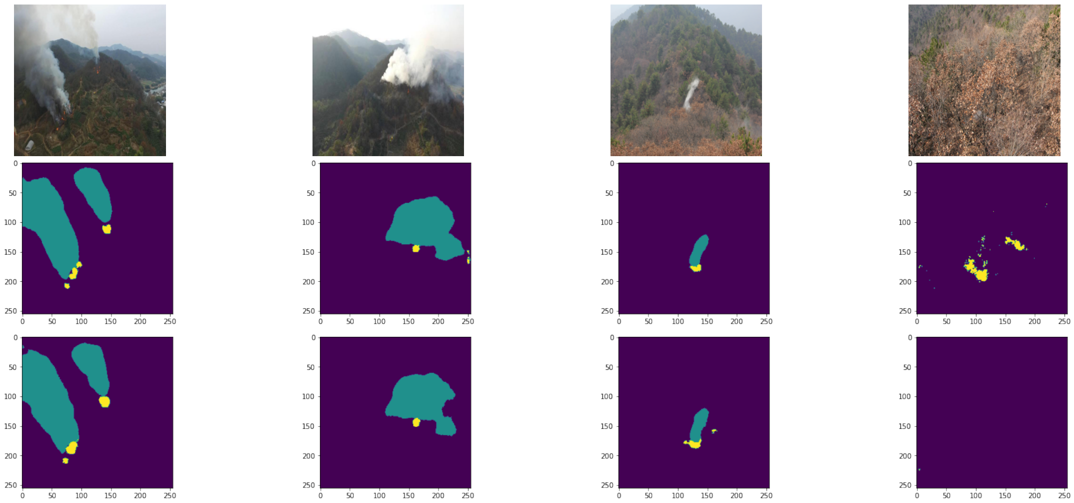


Figure 4. Prediction results. The figures in top row are the original input images; The mid row are the predictions of pure U-net; The bottom row are the predictions of our model.

5 Conclusions

In this paper, a U-net enhanced with self-attention layer using Fastai is proposed for early wildfire smoke and suspected flame area segmentation. The segmentation result of this model demonstrated its robustness generalization ability after adding the attention gate to reduce the false-positive predictions for smoke detection. However, in this paper, the video-based data is theoretically working with this model but not tested. The schemes of geolocation of the wildfire points is the most interested part in potential works.

Acknowledgement

The work of this paper is partially supported by the Natural Sciences and Engineering Research Council of Canada, National Natural Science Foundation of China (No. 61833013), and the Key Research and Development Program of Shaanxi Province of China (No. 2020SF-376). Thanks to the selfless contributions of Pytorch, Fastai, and Kaggle teams. Specially, thanks to the hard work of third author's team for the experiment support and experimental data. The original data set from Kaggle is available at the link: https://www.kaggle.com/kutaykutlu/forest-fire. The processed data set, mask label and coding of this paper are open sourced at Github link: https://github.com/qiaolinhan/Wildfire_Segmentation_Resnet34.

References

[1] Cao, Y. *et al.* (2019), "An attention enhanced bidirectional LSTM for early forest fire smoke recognition", IEEE Access, Vol. 7, pp. 154732–154742.

[2] Cruz, H. *et al.* (2020), "Machine learning and color treatment for the forest fire and smoke detection systems and algorithms, a recent literature review", in XV Multidisciplinary International Congress on Science and Technology, Quito, Ecuador, pp. 109–120.

[3] He, S. *et al.* (2021), "An efficient GPU-accelerated inference engine for binary neural network on mobile phones", Journal of Systems Architecture, Vol. 117, p. 102156.

[4] Howard, J. and Gugger, S. (2020), "Fastai: a layered API for deep learning", Information, Vol. 11 No. 2, p. 108.

[5] Jeong, M. *et al.* (2020), "Light-weight student LSTM for real-time wildfire smoke detection", Sensors, Vol. 20 No. 19, p. 5508.

[6] Li, X. *et al.* (2018), "3D parallel fully convolutional networks for real-time video wildfire smoke detection", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30 No. 1, pp. 89–103.

[7] Long, J. *et al.* (2015), "Fully convolutional networks for semantic segmentation", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, US, pp. 3431–3440.

[8] Micikevicius, P. *et al.* (2018), "Mixed precision training", in International Conference on Learning Representations (ICLR), Vancover, Canada, pp. 1–10.

[9] Oktay, O. *et al.* (2018), "Attention U-net: Learning where to look for the pancreas", in Medical Imaging with Deep Learning, Amsterdam, Netherlands, pp. 1–10.

[10] Redmon, J. (2016), "Darknet: Open source neural networks in C", available at: http://pjreddie.com/darknet/.

[11] Redmon, J. *et al.* (2016), "You only look once: Unified, real-time object detection", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, US, pp. 779–788.

[12] Ronneberger, O. *et al.* (2015), "U-net: Convolutional networks for biomedical image segmentation", in International Conference on Medical Image Computing and Computer-assisted Intervention, Lima, Peru, pp. 234–241.

[13] Shahid, M. and Hua, K. l. (2021), "Fire detection using transformer network", in Proceedings of International Conference on Multimedia Retrieval, New York, US, pp. 627–630.

[14] Yuan, F. (2008), "A fast accumulative motion orientation model based on integral image for video smoke detection", Pattern Recognition Letters, Vol. 29 No. 7, pp. 925–932.

[15] Yuan, F. *et al.* (2019), "Deep smoke segmentation", Neurocomputing, Vol. 357, pp. 248–260.

[16] Zeiler, M. D. and Fergus, R. (2014), "Visualizing and understanding convolutional networks", in European Conference on Computer Vision, (ECCV), Zurich, Switzerland, pp. 818–833.

[17] Zhang, J. *et al.* (2021), "ATT squeeze U-Net: A lightweight network for forest fire detection and recognition", IEEE Access, Vol. 9, pp. 10858–10870.

[18] Zhang, Q. *et al.* (2018), "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images", Procedia Engineering, Vol. 211, pp. 441–446.