Reading: Deep Smoke Segmentation

Feiniu Yuan, etc.

January 30 2020

1 Abstract

Fully convolution neural network in semantic segmentation → deep smoke segmentation network to infer high quality segmentation mask from blurry smoke images.
Overcomes:
Large variations in texture, color and shape of smoke appearance.
Avoid the great difficulty in manually labelling fuzzy smoke boundaries.
Methods:
1. Divide the proposed network into coarse and fine path. → coarse path: encoder-decoder FCN with skip structure. → fine path: encoder-decoder FCN with skip structure, it is shallower than the coarse path work. (it is shallower than the coarse path network). → a very small network containing only addition, convolution and activation layers to fuse the results of the two paths.
2. Propose a method to generate synthetic smoke images.
Results:
1. can easily train the proposed network end to end for simultaneous optimization of network parameters.
2. We can easily and accurately perform smoke detection on videos. This is proved much better performance than state-of-art segmentation algorithms.

2 Introduction

• Smoke can provide earlier clues for fire alarms than the flame.

There are three smoke detection categories:
1. merely judge, whole image smoke recognition.
2. not only to identify whether there is smoke, but also indicate rough locations of smoke by bounding boxes, smoke detection.
3. densely classify each pixel, smoke segmentation

Smoke segmentation:
• Requires accurate separation of smoke components from background scenes at pixel levels.

• Often outputs a mask with detailed edges.

Traditional smoke segmentation methods, hand-crafted features: smoke color, shape, motion.→ very difficult to design, define and choose. **Due to:**large variations of smoke appearance, resulting in very poor segmentation.
Dynamic features extracted from videos. But unstable

Deep learning based methods:
1. CNN, can not directly applied to dense classification of each pixel.
2. FCN, accepts an input image of arbitrary size, utilize a set of convolution layers, non-linear activation functions, pooling and up-sampling layers → predicted map for input image.
Major operation:
• Convolution, has the advantage of computational simplicity, effectiveness and local weight sharing.
• Perform an end-to-end semantic segmentation.

We present: Two-path encoder-decoder network based on FCNs.
Each pixel is densely classified as smoke or non-smoke labels. How to label them?
• Fuzzy, translucent, non-rigid issue.→ two paths: coarse and fine segmentation.
• Multi-scale problem (due to the lens of camera) → 1. feed images with different size. 2. fuse feature maps from different layers using skip and concatenation operations.
• Speed up training → VGG16 network in first path.

Difference:
• Contain skip structures in order to reuse more abundant features.
• Can be trained end-to-end instead of three-stage.
• Simultaneously optimize the parameters.

2.1 FCN for Object Segmentation

Semantic segmentation – an important foundation for image understanding, widely used in autonomous driving.
For end-to-end semantic segmentation.
Covering the last fully connected layers into 1×1 convolutions.
Added skip structures in the network ← skip structures provide finer spatial information since posterior layers of an encoding network lose a lot of spatial details due to multiple pooling operations.

SegNet:
Utilizes a stage-wise training technique.
Adds batch normalization after each convolutional layer
A smaller network and easier to train by removing the fully connected layers of VGG16.

The decoding stage of SegNet contains a series of convolution and upsampling. There is a consistent one-to-one match between encoding and decoding stages. Saves the max-pooling indices or unpooling operations.

U-Net:
Like the combination, including both skip structures and decoding stage.
Transfers the feature maps of an encoder to its corresponding decoder and then concatenates them by copying and cropping operations.

2.2 Smoke Segmentation

Mainly based on color characteristics of smoke.
Rough sets and color features combined. Eg. Gaussian mixture model (GMM).
Utilized Kalman filter (UKF) to update background to exclude objects with colors similar to smoke, then separate smoke by roughness distribution of smoke colors.
Not much literature use deep learning for smoke segmentation. Main reason: Lack of adequate labelled smoke images for training.
Inspired by the achievement of FCNs. Most algorithms widely use pre-trained parameters of VGG16 and skip structures for performance improvement.

3 Data-set generation

• Obstacle of deep learning based segmentation:
1. Lack of adequate annotated training data.
• **Solution:**
1.used computer graphics to virtually generate 8162 pure smoke images, denoted as a PureSmoke data-set.
2. linearly blend a pure smoke image and a background image.

Pure smoke sample: 256×256 RGBA, RGB channel S, alpha channel α. Use the linear color composition and a background RGB image B to generate new smoke RGB image I:

$$I(x) = (1 - \alpha(x))B(x) + \alpha(x)S(x) \tag{1}$$

Where $\alpha(x)$ denotes the value of an alpha channel α at pixel x.
We randomly select the background image from Places365-Standard data-set. Then combine with the pure smoke image to generate a synthesized smoke image.
To further augment smoke images, we change α values to change the concen-

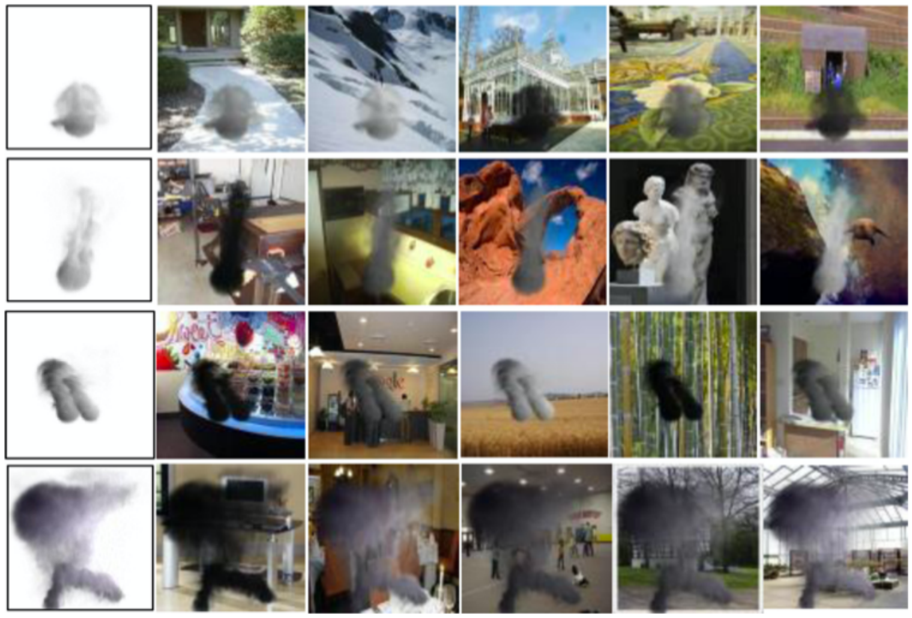

Figure 1: Pure smoke + background image

tration of smoke images by a parameter β.

$$I_R(x) = (1-\alpha(x))B_R(x) + \alpha(x)\beta S_R(x) \tag{2}$$
$$I_G(x) = (1-\alpha(x))B_G(x) + \alpha(x)\beta S_G(x) \tag{3}$$
$$I_B(x) = (1-\alpha(x))B_B(x) + \alpha(x)\beta S_B(x) \tag{4}$$

Where β is a global coefficient with range (0,1) used to control the color of smoke.
This synthesizing method avoids the difficulty in manually annotating ground truths of smoke images.

4 Two-path fully convolutional network

FCN path: FCN with asymmetrical structures, used for global segmentation prediction.
Refinement path: produces more detailed prediction on the basis of global prediction.

4.1 global context information

FCN path:
- Aims: gaining global context information for generation of a coarse smoke

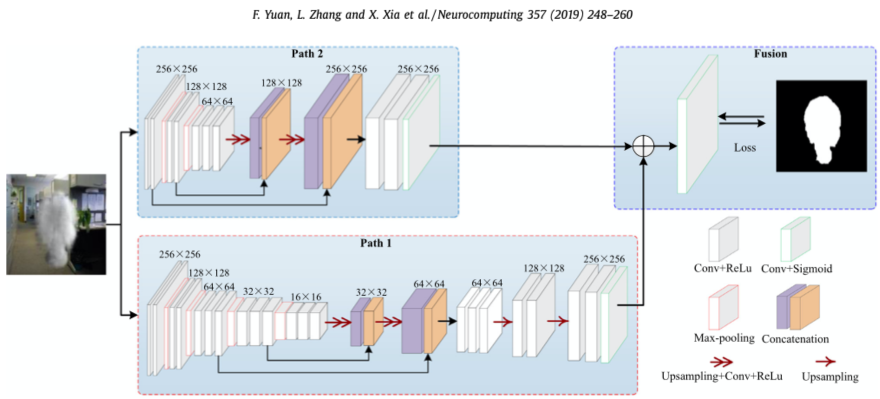

Figure 2: TPFCN structure

segmentation map.
• input: single RGB image. produces: a prediction map with the same size of the input.
• Removed fully connected layers to reduce the numbers of trainable parameters. Use an asymmetric structure in decoding phase • increase the depth and add the skip structure.
• leverages a binary cross-entropy loss.

$$\begin{aligned} L(P_i, G_i) =& \lambda \|\mathbf{W}\|_2^2 - \frac{1}{N \times h_i \times w_i} \\ & \sum_{i=1,k=1}^{N,h_i \times w_i} \times (g_i \log p_i^k + (1 - g_i) \log(1 - p_i)) \end{aligned} \tag{5}$$

Where p_i^k, g_i^k are the values of the kth pixel in the ith predicted image $\mathbf{P}_i$ and the ith ground truth $\mathbf{G}_i$.

4.2 A shallow network for local fine information

Smoke: blurry edges, translucent property. → first path: coarse segmentation. → second path: capture details of smoke. Shallower layer
Using the first three blocks of VGG16 with 2 max-pooling to remain more details of the input.
7 conv, 2 max in encoding phase, 4 conv, 2 upsampling in decoding phase.

4.3 A fusion network of two path

Goal: To gain an accurate smoke segmentation map.
1. Add up the network outputs of the two paths, then feed the summed seg-

mentation map into a 1×1 conv+sigmoid to produce the final prediction map.
2. feature informations are from the last three blocks in VGG16.