

## Improving computer vision with DeepLab

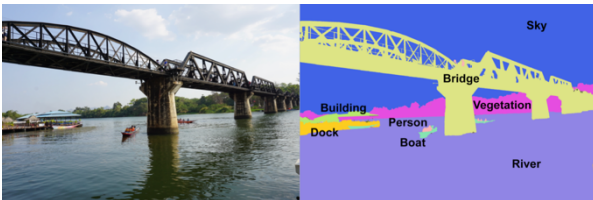
Editorial submission by Jonas N. Søndergaard @thefairjournal

### Original title

DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.<sup>1</sup>

### Introduction

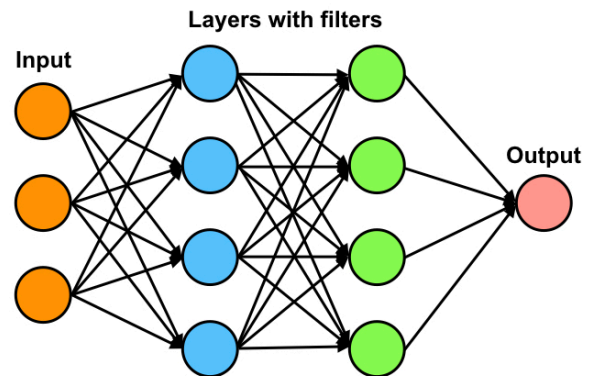
Semantic segmentation is when an image is divided into different components/segments, such as identifying a boat separately from the human that is sailing it, or the sky separately from trees and water. This is a very simple task for any human, but a very difficult one for computers. So why do computers even need to segment an image? Some of the most well-known uses for computational image segmentation are for self-driving cars, medical imaging, and face recognition.



**Example of semantic segmentation.** On the left we see the original image, and on the right we see the image divided into its different components.

Before a computer can recognize different parts of a picture it needs to be trained. This is done by showing the computer images that has already been divided into components such as person, boat, and river. Using this input, the computer builds a new “neural network”, called like this because it resembles the connections between neurons that humans have in their brains. Using what the computer have learned from the images it has already seen, it will then be able to determine

what is what on a new image that it has never seen before.



**Example of a neural network.** The network receives input, which is processed through a set of layers with specific filters, producing the output. The connections resemble the neurons in the human brain.

Researchers are constantly trying to improve how good the computer can identify images, which usually is done by improving the layers inside the neural network (the layers between the input and the output). Specifically, for image segmentation “convolutional neural networks” are used, which has layers that contains filters able to pick up patterns and make sense of them (such as the border between the river and the trees, or the shape of the boat).

### Findings

The authors of “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”<sup>1</sup> has improved previous image segmentation by using three new strategies. The first one is to up-sample the filters in the network, meaning that each filter can see more of the

image and thereby increase the speed at which an image is processed. This step is called “atrous convolution”. The second strategy is to use multiple filters at the same time to analyze one specific feature in the image, and thereby getting more information simultaneously at multiple levels. This step is called “atrous spatial pyramid pooling”. The final strategy is to capture fine details on the location of the objects’ boundaries, using a structure predictor called a “conditional random field”. Together the authors call their new way of segmenting an image for DeepLab.

In order to test how well DeepLab performs compared to other methods, it was tested in the “PASCAL visual object classes 2012 test set”, which is a set of images of showing e.g. humans, dogs, sheep, airplanes, and motorcycles on different backgrounds. DeepLab scored 79.7%, which is better than any other neural network has done to date.

## Conclusions

The combination of the three strategies used in DeepLab increased the overall speed and accuracy for segmenting an image. With DeepLab we are thus getting closer to perfect computer vision than ever before. There is of course still room for improvement, as there is still some way from 79.7% to 100% accurate (human) vision.

## Article info

Editorial submission by Jonas N. Søndergaard @thefairjournal. ID: 2019.04.25. Please refer to the original article<sup>1</sup> for more details.

## References

1. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 834–848 (2018).