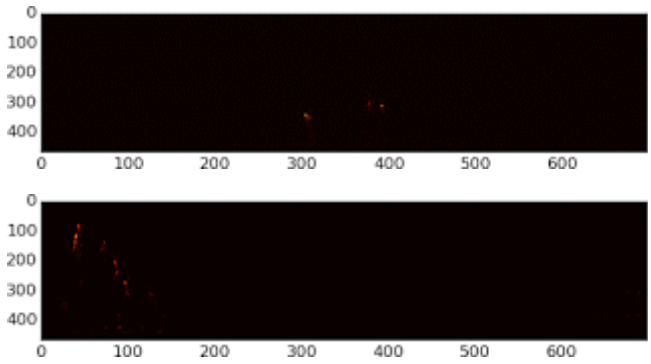
Bonus: Bird extraction

Using guided backpropagation (https://github.com/Lasagne/Recipes/blob/master/examples/Saliency%20Maps%20and%20Guided%20Backpropagation.ipynb), we can actually ask the network where in the spectrogram it found a bird call -- i.e., which time-frequency bins (or pixels, if seen as an image) have the largest influence on the detection result. For the two files in my previous post, we get the following:



So we see where in the noisy spectrogram it detected a bird (first picture), and where it found a real bird in the fake bird recording (second image). Technical details: This is the gradient of the sum of all local predictions that exceed 0.0 after applying the running average, but before the sigmoid (after the sigmoid, the gradient for the most confident predictions is very small as the sigmoid saturates, so this would give a less clear picture), and I use guided backpropagation instead of the true gradient.

We can also multiply these maps by the original complex-valued spectrograms and listen to the results, see "bird_hidden.mp3" and "bird_fake.mp3" in the attachment. Compare to the original files in the dataset (testdata/a235ab95-9878-437b-8ed4.wav and testdata/375bf073-e669-46b9-b6cf.wav, respectively) to verify that these are indeed bird calls.