Supplementary Materials: Deep Learning vs. Standard Machine Learning in Audio Beehive Monitoring

Vladimir Kulyukin, Sarbajit Mukherjee and Prakhar Amlathe

1. Training and Validation Losses of RawConvNet with Different Receptive Field Sizes

Figures S1, S2, S3, S4 and S5 show the graphs for the training and validation loss curves of RawConvNet with different receptive field sizes in the first convolution layer. The train loss and validation loss curves gradually reduce with training. The shapes of the training and validation loss curves are almost identical, which indicates that there is no overfitting.

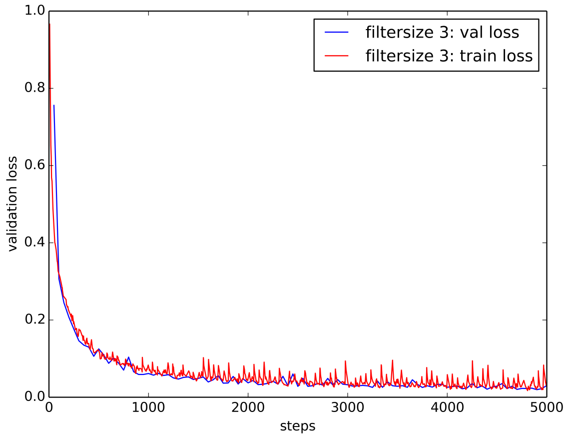

Figure S1. Training and validation losses of RawnConvNet with receptive field size $n = 3$.

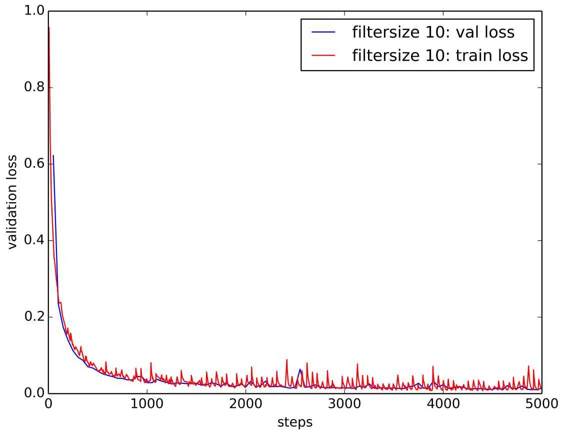

Figure S2. Training and validation losses of RawConvNet with receptive field $n = 10$.

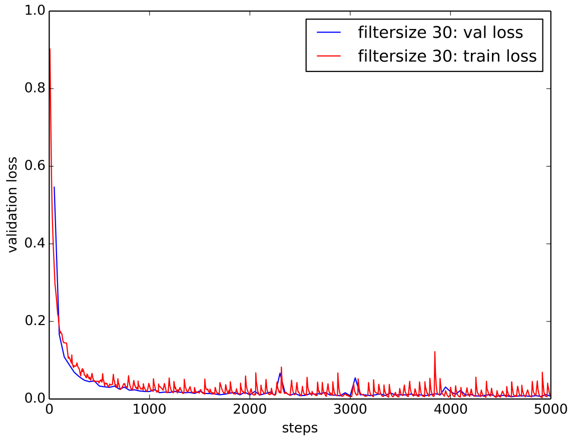

Figure S3. Training and validation losses of RawConvNet with receptive field $n = 30$.

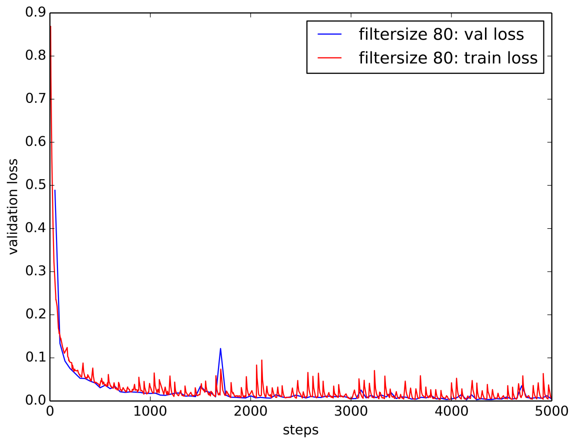

Figure S4. Training and validation losses of RawConvNet with receptive field $n = 80$.

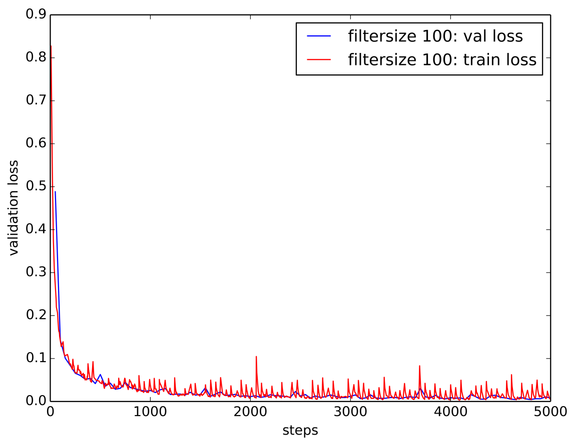

Figure S5. Training and validation losses of RawConvNet with receptive field $n = 100$.

2. Gradient Distributions in the Final FC Layers of Raw Audio Conv Nets

We analyzed the contribution of the custom layer in RawConvNet by generating the plots of the gradient weight distributions in the final FC layers of RawConvNet, ConvNet 1, ConvNet 2, and ConvNet 3. The plots are given in Figures S6, S7, S8 and S9. These plots consist of the time stamped histograms of gradients for ConvNet 1, ConvNet 2, ConvNet 3, and RawConvNet, respectively. Each figure shows temporal slices of data over different steps during training with each slice being a gradient weight histogram in the FC softmax layer of an appropriate ConvNet with the oldest time slice in the back and the most recent one in the front. In all figures, the y-axis represents the step count during training and the x-axis represents the histogram bins. The step count on the y-axis starts from the back and moves to the front with the training process.

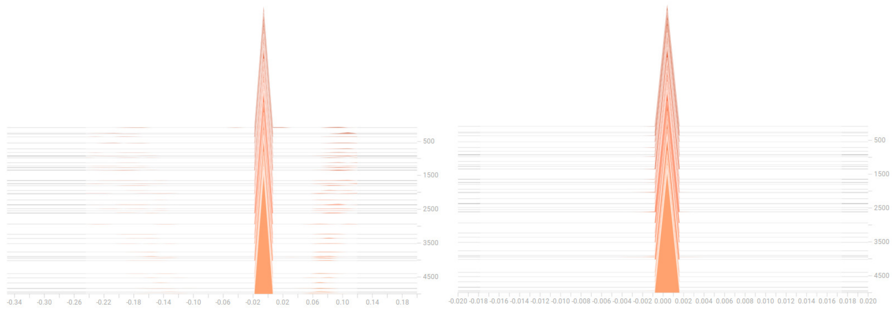

Figure S6. Histograms of the gradients for layers 4 and 5 in ConvNet 1. The histograms on the left show the gradient distribution in the FC softmax layer in layer 4; the histograms on the right show the gradient distribution for the FC softmax layer in layer 5.

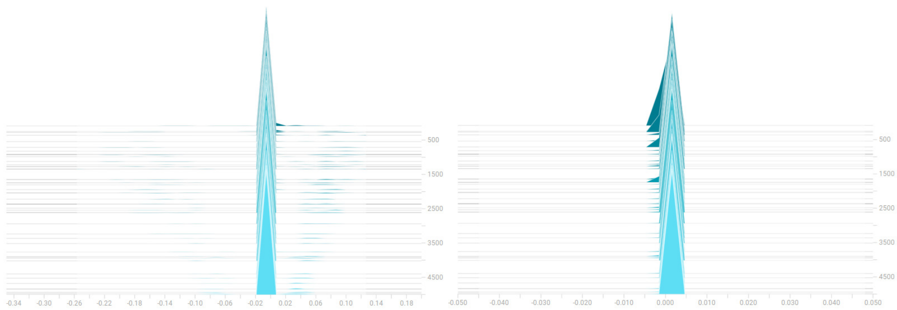

Figure S7. Histograms of the gradients for layers 5 and 6 in ConvNet 2. The histograms on the left show the distribution of gradients for the FC softmax layer in layer 5; the histograms on the right show the gradient distribution for the FC softmax layer in layer 6.

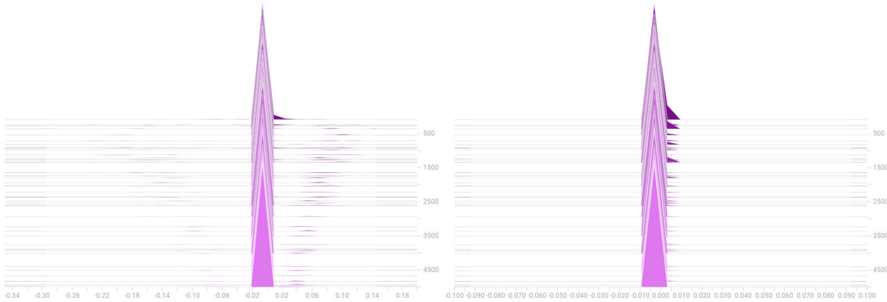

Figure S8. Histograms of the gradients for layers 6 and 7 in ConvNet 3. The histograms on the left show the distribution of the gradients for the FC softmax layer in layer 6; the histograms on the right show the gradient distribution for the FC softmax layer in layer 7.

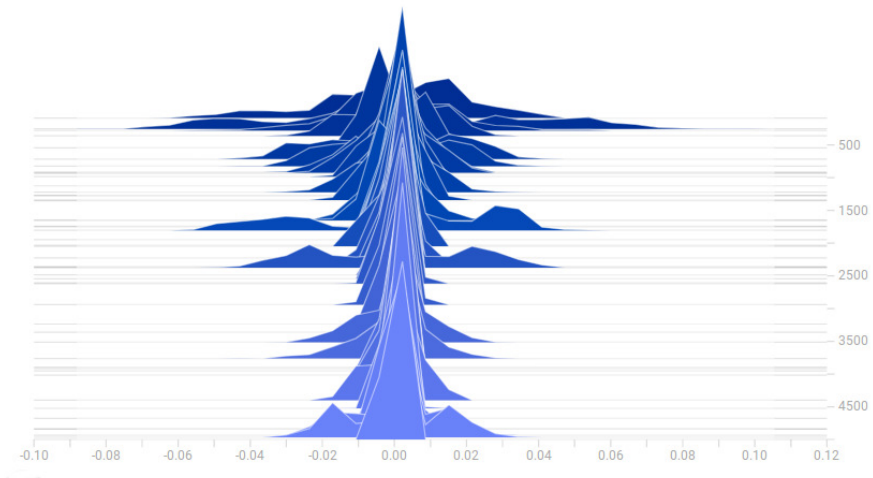

Figure S9. Histograms of the gradients in the FC softmax layer in layer 5 in RawConvNet.

As can be seen in Figures S6, S7, and S8, ConvNets 1, 2, and 3 did not learn much in their FC layers as the shapes of the curves remain almost identical between the consecutive histograms, which suggests that the gradients in the FC softmax layers of these ConvNets were changing rather slowly. In ConvNets, such gradients are used to update the weights in a way that minimizes the cost function during backpropagation. Mathematically, the weight update is modeled by Equation 1, where $\mathrm{w_i}$ is the weight matrix between the layers $i-1$ and $i+1$, and $\mathrm{w_i^*}$ is the updated weight matrix. The term $\partial loss/\partial \mathrm{w_i}$ is the gradient of the corresponding weight matrix $\mathrm{w_i}$.

$$\mathrm{w_i^*} = \mathrm{w_i} - \eta \cdot \frac{\partial loss}{\partial \mathrm{w_i}}, \text{ where } \eta \text{ is the learning rate} \tag{1}$$

From Equation 1 and the figures one can observe that since the spread of the weight gradients is small, it suggests that ConvNets 1, 2, and 3 networks were learning slowly as weight updates $\mathrm{w_i^*}$ are rather small. Consequently, these networks made fewer correct predictions for the validation data. Specifically, Table 4 in Section 3.1 shows that the validation losses of ConvNets 1,2, and 3 were higher than the validation losses of RawConvNet although the accuracies of all four ConvNets were high. The graph in Figure S9 shows that the histogram of the weight gradient in the FC softmax layer of RawConvNet is more distributed over the entire training timeline. In RawConvNet, the gradients changed over different steps in the learning process and were much more spread than the gradients of

ConvNets 1, 2, and 3. RawConvNet appeared to learn continuously during training and made more correct predictions for the validation data.

3. ROC Curve

ROC curves are graphs that summarize a binary classifier's performance over all possible thresholds. They are generated by plotting true positive rates on the y-axis against false positive rates on the x-axis. Since SVM OVR is a binary classifier, the ROC curves were computed for the bee, noise, and cricket classes. These curves are shown in Figure S10. As the figures show, the areas under all three ROC curves are equal to 1.0.

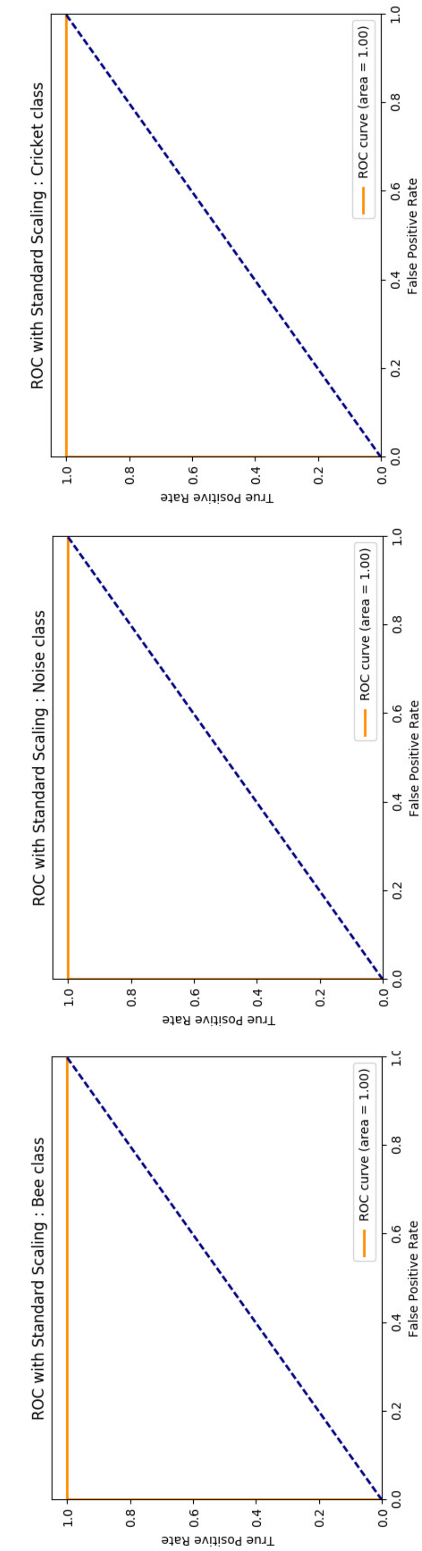

Figure S10. ROC of SVM OVR with Standard Scaling for Bee, Noise and Cricket classes.