Data augmentation

A guide to using the data augmentation features in Edge Impulse.

What is data augmentation?

Data augmentation is a method that can help improve the accuracy of machine learning models. A data augmentation system makes small, random changes to your training data during the training process.

Being exposed to these variations during training can help prevent your model from taking shortcuts by "memorizing" superficial clues in your training data, meaning it may better reflect the deep underlying patterns in your dataset.

> ### Data augmentation will not work with every dataset
>
> As with most things in machine learning, data augmentation is effective for some datasets and models but not for others. While experimenting with data augmentation, bear in mind that it is not guaranteed to provide results.
>
> Data augmentation is likely to make the biggest difference when used with **small datasets**. Large datasets may already contain enough variation that the model is able to identify the true underlying patterns and avoid overfitting to the training data.

Enabling data augmentation

Edge Impulse provides easy to use data augmentation options for several types of data. It is currently available for audio spectrogram data (generated by the MFCC and MFE blocks) and image data when used with Transfer Learning blocks.

If data augmentation is available for the data type you are working with, you can enable it via a checkbox in the Neural Network block. For example, here are the data augmentation options for audio data:

An example of data augmentation options for an audio classification model.

Workflow for using data augmentation

Here is a step-by-step guide to making the most out of the data augmentation features in Edge Impulse.

1. Train a Neural Network block without data augmentation

As mentioned above, there's no guarantee that data augmentation will improve the performance of your model. Before you start experimenting, it's important to train a model without data augmentation. You can use this model's performance as a baseline to understand whether data augmentation improves your model.

First up, create a Neural Network block, and then try to get the best possible performance from your model without enabling augmentation.

2. Train a second Neural Network block with the same settings

It's helpful to be able to compare both models side by side. To allow this, create a second neural network block with the same settings as the first.

3. Enable data augmentation

In your new Neural Network block, check the **Data augmentation** checkbox. If there are options, leave the defaults in place.

4. Increase the number of training cycles

Often, the beneficial effects of data augmentation are only seen after training a network for longer. Increase the number of training cycles your network trains for. A good rule of thumb might be to double the number of training cycles compared to your baseline model. You can look at the training output after your first run to determine if the model still seems to be improving and can be trained longer.

best possible version.

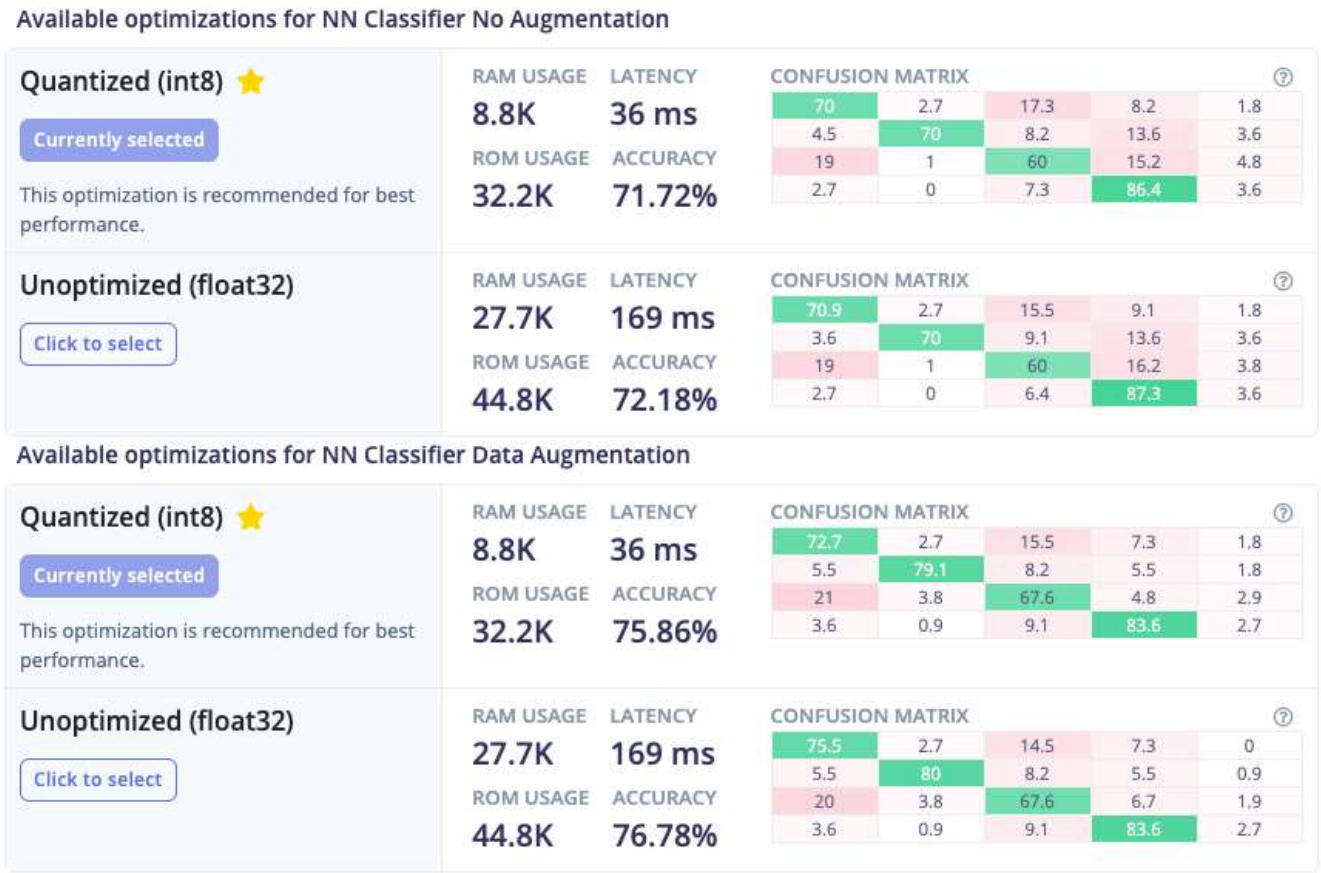
5. Compare and iterate

Now that you've trained a model with data augmentation, compare it to your baseline model by checking the numbers in the Neural Network block. If it's more accurate or has a lower loss value, augmentation was successful.

Whether it was successful or not, you may be able to find settings that work better. If available, you can try other combinations of data augmentation options. You can also try adjusting the architecture of your model. Since data augmentation can help prevent overfitting, you may be able to improve accuracy by increasing the size of your model while applying augmentation.

6. Check the models using your test dataset

Once you have a couple of model variants, you can use the **Analyze optimizations** tool on the **Deployment** tab to apply them to your test dataset and examine them side by side. In the following example, the model with data augmentation was >4% more accurate on the test dataset:



A performance comparison between models trained both with and without data augmentation.

It's also worth comparing the confusion matrices for each model. Data augmentation may affect your model's performance on different labels in different ways. For example, precision may improve for one pair of classes but be reduced for another.

Once you've compared models and found the one that works best, delete the others from your Impulse before deploying.

> **Versioning**
>
> You can use Edge Impulse's Versioning feature to save the state of your Impulse at this point, so that you can restore it if you wish to perform more experimentation. You can find it in the Versioning tab in the left menu.

Data augmentation and deployment

Data augmentation occurs only during training. It will have no impact on the memory usage or latency of your model once it has been deployed.

Updated 10 months ago