

BLOG POST

Make the Most of Limited Datasets Using Audio Data Augmentation

MACHINE LEARNING

Daniel Situnayake

October 7, 2020

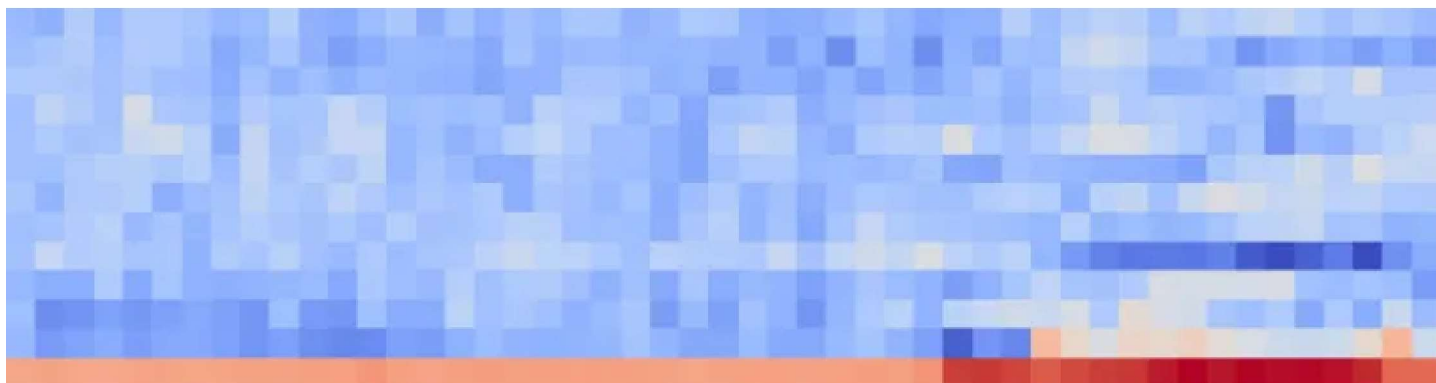


Table of contents

- Why is data so important?
- Introducing data augmentation for audio
- Getting started

Often, the most challenging portion of a machine learning project is obtaining a large



We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

Functional only**Accept all**

you feed your model during training, the real world. Unfortunately, high quality data is a precious resource.

Edge Impulse includes data augmentation tools to help you make the most of your available data. In this blog post, we'll explore

why data is so important to machine learning models, and how our new spectrogram data augmentation features can help you train more accurate audio classification models.

Why is data so important?

A common task for deep learning models is classification. This is when a model learns to distinguish one type of input from another. For example, a model might learn to recognize the sound of running water and tell it apart from generic background noise (as in our [Recognize sounds from audio](#) tutorial).

To train an audio classification model, we feed it many samples of each class of audio. For each sample, the model generates a set of probabilities that indicate the classes it is predicted to match with. For example:

```
running water: 0.75, background noise: 0.25
```

The training code compares these numbers with the sample's actual class (also expressed as probabilities, for example

```
running water: 1.0, background noise: 0
```

), and then tweaks the model's internal state so that the probabilities will be slightly more correct the next time the model is run.

Since the model gets a little better each time a sample is fed through, it's good to have as much data as possible! While a typical training process will involve feeding each sample of data into the model multiple times, this isn't quite as good as adding more data.

For example: imagine you capture the sound of your bathroom faucet running and use it to train a model. You then test the model on your kitchen faucet—but it doesn't work! Since the model has only been trained on the particular sounds of the bathroom faucet, it doesn't recognize the sound of a different type.



We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

Functional only

Accept all

model with sounds captured from both. More importantly, the model should—be more common between all faucets, rather than just like. And the more faucets we add to

This is why it's always good to collect as much data as possible. The larger and more diverse your dataset, the better your model will be able to generalize, and understand variations of input that it hasn't seen before.

🔗 Introducing data augmentation for audio

It's great to have a ton of data, but there's a problem. Data is expensive and time consuming to collect and label (meaning to sort it into the correct categories, such as "running water" and "background noise"). This means that we often need to do everything we can to make the most of the data that we have.

As we just learned, adding more data is helpful because the natural variations in the samples help the model learn to generalize. For example, the sounds of any two running faucets will always be slightly different, so being trained on both helps the model learn what is common between all faucets.

So if variations are good, what if we were to introduce some artificial variations on top of our existing dataset? This is the idea behind data augmentation: by slightly modifying each sample of data before feeding it into the network during training, we can gain some of the benefits of having a larger dataset than we actually do.

In Edge Impulse, users who are training audio classification models can now choose to enable several types of data augmentation. It looks like this:



We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

Functional only

Accept all

Audio training options

Data augmentation ?



Add noise ?

None Low High

Mask time bands ?

None Low High

Mask frequency bands ?

None Low High

Warp time axis ?



The new audio data augmentation options in Edge Impulse

During training, each time a sample of data is fed into the network, it will be randomly transformed. Since each sample is fed in dozens of times during a typical training run, it's as if your dataset has been expanded to include dozens of variations of each sample.

We currently support the following transformations, which include a state-of-the-art technique known as **SpecAugment**:

Adding **gaussian noise**

Masking random frequency bands

Masking random time bands

Warping randomly along the time axis

Finding the best combination of settings can take some trial and error, so we've made



We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

Functional only

Accept all

but if you want to dig in deep, you can fine-tune our Expert Mode code editor:

```
can be configured in visual mode.  
_length / 13), 13],
```

```
mF_num_freq_masks=3, F_freq_mask_max_consecutive=4,  
mT_num_time_masks=3, T_time_mask_max_consecutive=2,  
enable_time_warp=True, W_time_warp_max_distance=6,  
mask_with_mean=False)  
train_dataset = train_dataset.map(sa.mapper(), tf.data.experimental.AUTOTUNE)
```

Data augmentation can have a real impact on the accuracy of your models. For example, with a model trained on a challenging keyword spotting reference task using a limited dataset[1], adding data augmentation in Edge Impulse boosted the model's test accuracy by 4.14%. When you're trying to squeeze as much performance as possible from a tiny model, every little helps!

Getting started

We hope that support for data augmentation makes it even easier to get started with Edge Impulse, even if your training dataset is still small. Edge Impulse now supports data augmentation for both audio and image datasets, and we're working on adding augmentation for time series data. You can enable data augmentation in the Neural Network block.

To learn the best practices for using data augmentation in Edge Impulse, see our [Data augmentation](#) documentation page.

To learn how to train an audio classification model with Edge Impulse, try our [Recognize sounds from audio](#) tutorial.

And if you have questions or suggestions, please let us know on the [Edge Impulse forum](#)!



We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

Functional only

Accept all

[1] The task was to distinguish between “yes”, “no”, background noise, and a combined set of all of the other words from a subset of 5% of the samples in the Speech Commands dataset.

[Start Discussion](#)

0 replies

Subscribe

Are you interested in bringing machine learning intelligence to your devices? We're happy to help.

Subscribe to our newsletter

Edge Impulse

[Home](#)[Blog](#)[Jobs](#)[About Us](#)

Studio

[Login](#)[Documentation](#)

Community

[Forum](#)[!\[\]\(1f56542a42e2413e44a2b2023033aa2e_img.jpg\) Twitter](#)[!\[\]\(19d44b37fb4fa155bf9d60c77a3d3cb2_img.jpg\) LinkedIn](#)[!\[\]\(5a351309c3b87e4420622c1f0e57efc0_img.jpg\) GitHub](#)

Contact

hello@edgeimpulse.com

Edge Impulse
3031 Tisch Way
110 Plaza West
San Jose, CA 95128
USA

[Privacy Policy](#)[Terms of Service](#)

We use cookies to personalize and improve your experience. See our [Privacy Policy](#) to learn more.

[Functional only](#)[Accept all](#)

marks of Edgelpulse Inc. All rights reserved.