Patty Ryan, Chenhao Yang, Nile Wilson
Commercial Software Engineering
Applied Machine Learning Group

Industrial Audio Classification with Music Domain Features

Listening to Machine Audio to Characterize Quality or Distinguish between Machines

Motivation

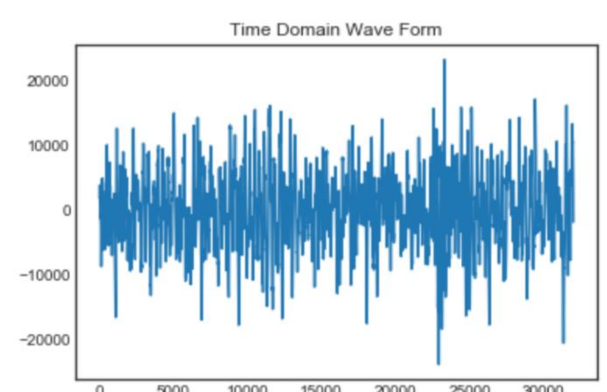
How do we scale scarce human experts who can diagnose machine quality by ear?

In a manufacturing setting, sounds that are heard on the factory floor provide unmistakable information about whether a machine or tool is working correctly or not. Domain experts working in manufacturing environments rely on their sense of hearing, and their knowledge of what sounds right or wrong to help them identify quality problems based on the sound pitch, rhythm, timbre and other characteristics. Using machine learning to classify these sounds has broad applications to automate the manual quality recognition work currently being done. In application, we found we could use the same techniques to distinguish between models of motors or between specific motors.

Characterize 'Music' in Machine Audio

Recording and Preprocessing

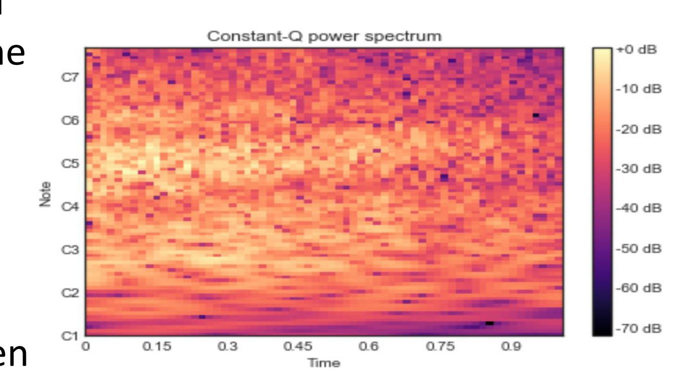
- Industrial audio captured @>32000 Hz
- Resample to 32000 Hz
- Split source recordings to sample groups
- Chunk into 1 Second intervals



1 Second Wave Form

Featurization of Pitch and Timbre

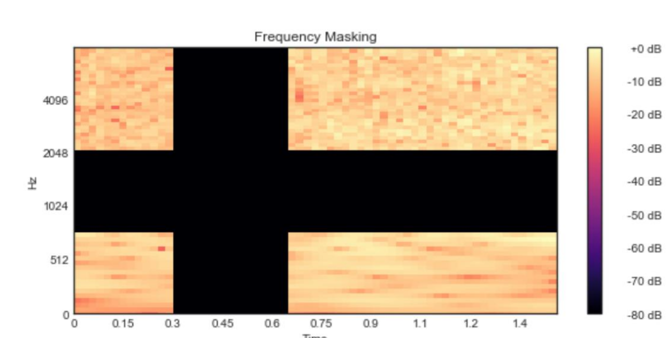
- We characterize pitch and timbre through frequency domain changes over time, represented by 2D spectrograms based on either the Constant Q Transform (CQT) or the Short-Time Fourier Transform (STFT)
- CQT is a time-frequency analysis method with higher frequency resolution at lower frequencies and higher time resolution towards higher frequencies3 better capturing both pitch and timbre*
- STFT determines the sinusoidal frequency and phase conten of local sections of the signal
- CREPE transfer learns fundamental frequency using a pre-trained pitch model



Constant Q Transform

Augmentation

- Time and frequency masking of CQT
- Added white noise and volume reduction
- Masking stabilized learning and increased generalizability
- Explored generating synthesized sample, revealed unstable pitch



Masked Time and Frequency

'VROOM' Industrial Audio Dataset

'VROOM': Various Recordings of Other Motors

Complement to the Urban Sound Dataset (8)

Dataset to date includes:

- Machine tooling at correct and incorrect settings
- Commercial ferry boats on Puget Sound
- Commercial aircraft landings at LAX and SEA
- Boat yard metal grinding operations at Lake Union

https://github.com/SingingData/Vroom

References and Inspiration

(1) CREPE: A Convolutional Representation for Pitch Estimation, Jong Wook Kim, Justin Salamon, Peter Li, Juan Pablo Bello, Feb 17, 2018

(2) Echolocation: Measurement of Pitch versus Distance for Sounds Reflected on a Flat Surface, The Journal of the Acoustical Society of Amer, May 1964, Bassett and Eastmond

(3) TimbreTron: A WaveNet(CycleGAN(CQT(Audio))) Pipeline for Musical Timbre Transfer, Huang, Li, Anil, Bao, Oort, Grosse, May 2, 2019

(4) GANSynth: Adversarial Neural Audio Synthesis, Jesse Engel, Kumar Krishna Agrawal, Shuvo Chen, Ishaan Gulrajani, Chris Donahue, Adam Roberts

(5) Air filter particulate loading detection using smartphone audio and optimized ensemble classification, Engineering Applications of Artificial Intelligence, Siegel, Bhattacharyyam, Sumeet, Sanjay, and Sarma

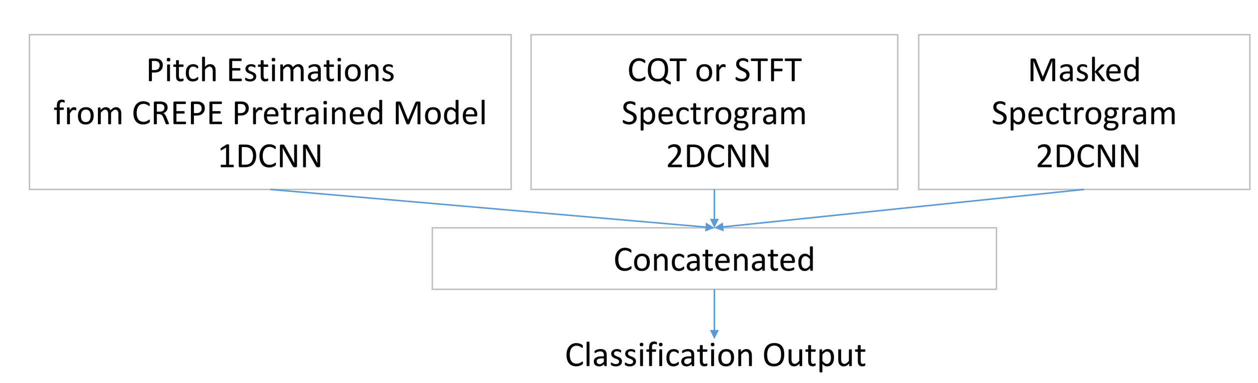
(6) Makcedward: makcedward GitHub repo and python package nlpaug and SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition, Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin Cubuk, Quoc Le Apr, 2019

(7) Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks, Muhammad Huzaifah, June, 2017

(8) A Dataset and Taxonomy for Urban Sound Research by Justin Salamon and Juan Pablo Bello, Nov 2014

Multi-input Model Architecture

- We demonstrate an applied method for classifying industrial audio using approaches developed in the music domain
- We use the Keras functional API with multi-input convnet model combining 1D representations of pitch estimations using pretrained model 'CREPE' using time domain waveform, and CQT or STFT 2D transforms
- We have applied this method to distinguishes motorized tool quality classes in industrial settings to classify welding quality and machining settings.
- The approach also works to classify a variety of motorized sounds that we've gathered and tested thus far
- We have also applied this method to distinguish between motor sounds as varied as models of aircrafts landing at LAX and specific ferry boats in operation on Puget Sound.
- Scarce labeled data of expert quality, and variation between human experts, led us to explore alternative data augmentation techniques



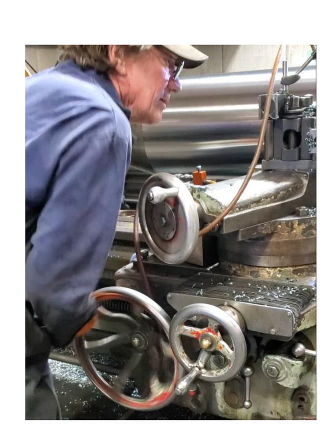
Results

Accuracy	Crepe Only	Crepe + STFT + SpecAugment	Crepe + CQT + SpecAugment
Machining Tool Operation: Correct vs. Incorrect	94.74%	95.61%	100.00%
Washington State Ferries: Wenatchee vs. Tacoma	53.09%	64.42%	82.21%
Airplane Landing: Embraer vs. Others	48.45%	84.53%	76.80%

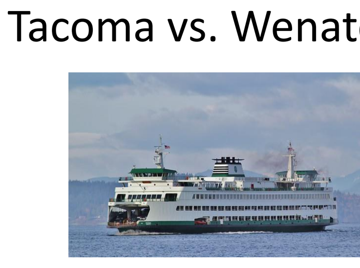
Notes:

- The machining sample includes homogenous machines with the same operator
- Other machine tool classifications we've measured have had less strong pitch component
- LAX landings with wind <15mpg
- Models with STFT and CQT included masked time and frequency data augmentation

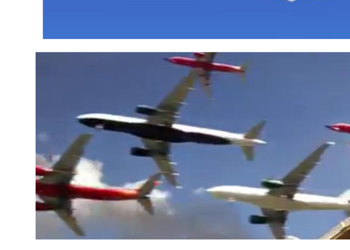
Machining Speed
Correct vs. Incorrect




Ferry Boat Classifier
Tacoma vs. Wenatchee




Aircraft Model Classifier
Embraer vs. Others

Future Direction

- Build out the VROOM Dataset with additional instances of industrial tool quality classification
- Create pre-trained model of industrial audio classification for re-use
- Extend validation of approach across industrial machine types
- Experiment with depthwise separable convolutions architecture

For more information:
pattyry@microsoft.com