
ΔnalyticsPro Bono Data Consulting for the Social Sector

Delta fills the skill gap and enables non-profits to accelerate their impact.

Who is Delta Analytics?

26 projects with non-profits and social impact organizations

80+ Fellows volunteering part-time over 3 years

Over **15,000** hours donated

$0.00 charged for services

16 US and 10 International projects (Tanzania, UK, Kenya, and more)

Which sectors do we serve?

Analytics

Community Engagement

United Way
UNITED WAY OF THE BAY AREA

TENANTS TOGETHER

foundation for sustainable development

ASHBY VILLAGE

Education

ROOTS OF SUCCESS

eneza education
spreading education everywhere

KUOW.ORG

CITY YEAR

Economic Development

Village Enterprise

TECHNOSERVE
BUSINESS SOLUTIONS TO POVERTY

BUILD
In Business to Learn

sama

Environmental

LIVELYHOODS

BioLite

KARIBU

BGCI
Plants for the Planet

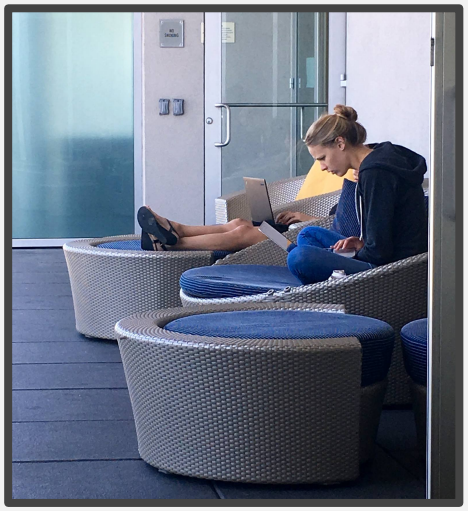
Where do Delta Fellows Work?

And many more!

What does the fellowship look like?

Monthly program-wide **hackathons** and ongoing social events

6 month engagement between non-profit and teams of 3 to 4 full-time data professionals



External **speakers and trainings** for ongoing technical growth and skill development

Delta teaching fellows make machine learning knowledge accessible by building technical capacity around the world.

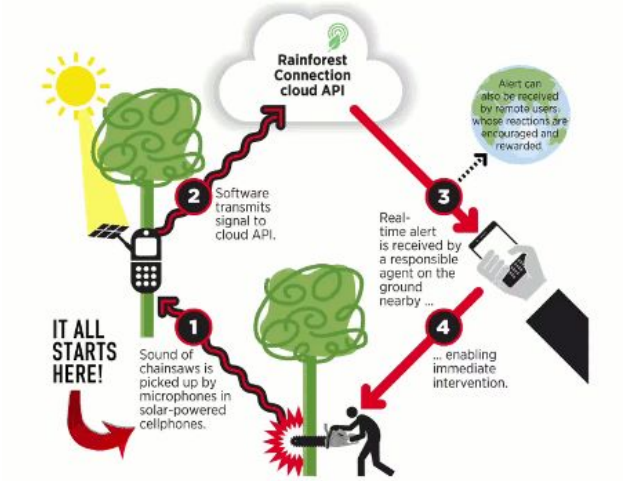
moringa SCHOOL

BRAVE academy

Andela

Deep dive into our work with Rainforest Connection

Rainforest Connection - RFCX



- Guardians - recycled cell phones powered by solar panels
- Audio and metadata streamed from rainforests in Ecuador, Peru and Brazil
- Alerts sent to conservation partners who go out into the field to catch illegal deforestation

Rainforest Team

Sara Hooker
Delta Analytics

Sean Mcpherson
Data Analyst at
Northrop Grumman

Steven Troxler
Data Scientist at
Stitch Fix

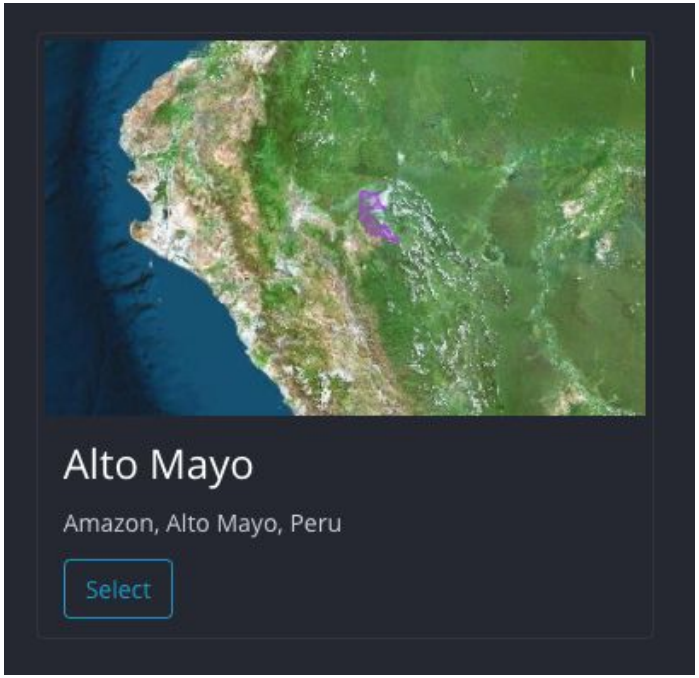
Cassandra Jacobs
Data Scientist at
Stitch Fix

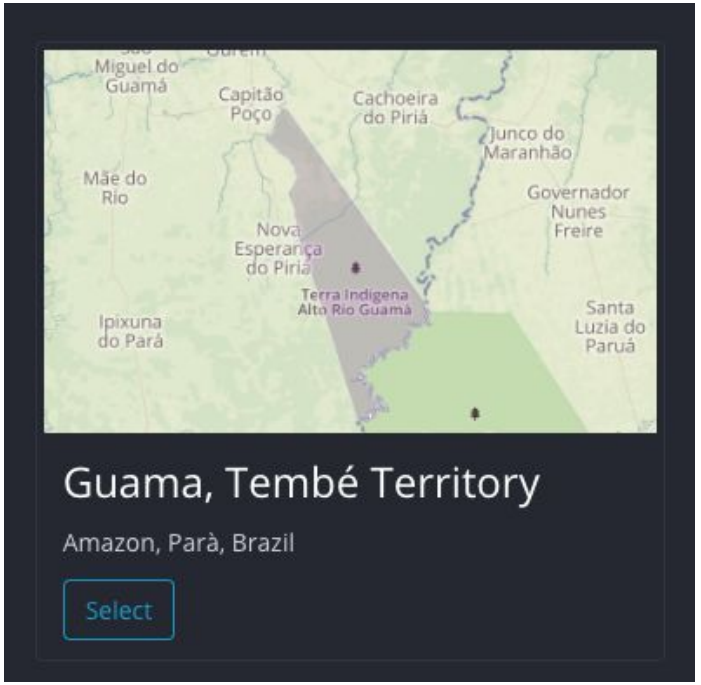
Deep Dive - Rainforest Connection Project

3 goals we are collaborating with Rainforest Connection achieve:

1. Improve the accuracy of chainsaw detection model.
2. Provide conservationists the direction and distance of the sound.
3. Contribute production code to train, score data and deploy alerts.

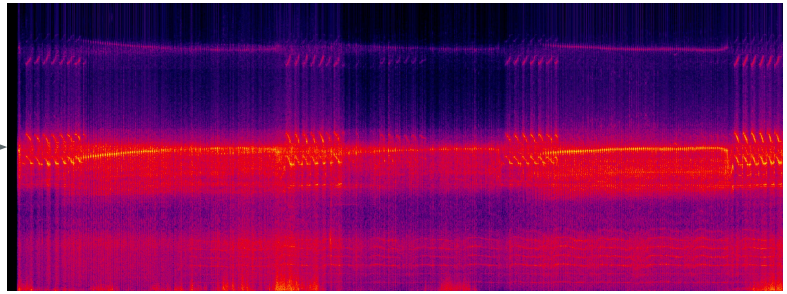
Goal 1: Improve accuracy of chainsaw detection model





- Improved accuracy will:
  - Increase ranger trust
  - Decrease time responding via dashboard or in-person

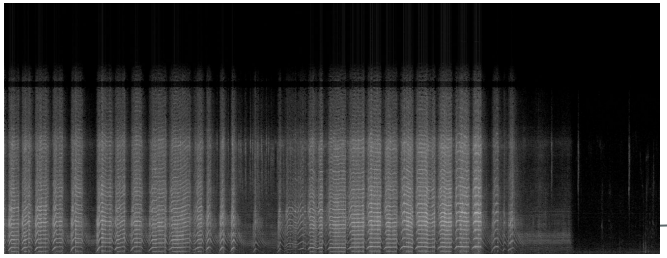
Approach: Turn audio detection into an image classification problem.

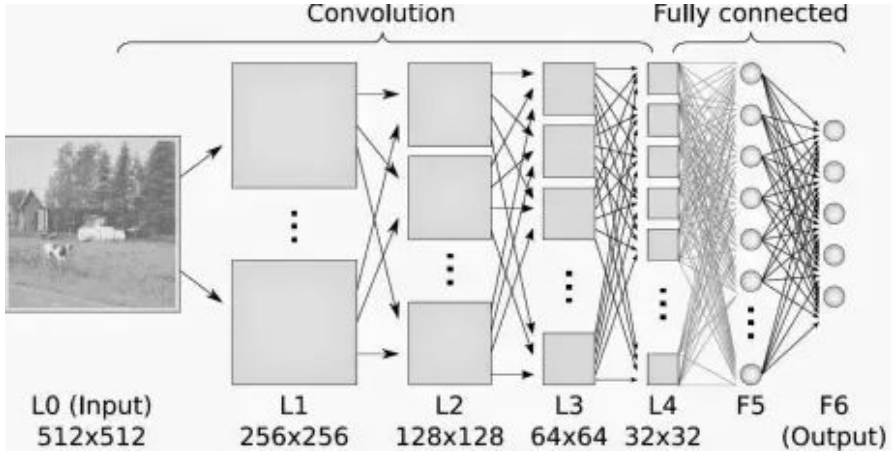


Audio streamed from conservation partners in Ecuador, Peru and Brazil.

Convert audio to a spectrogram (visual way to represent the signal strength of a sound).

Use CNN framework for image classification

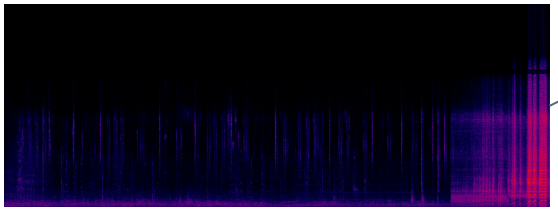




Balanced labelled spectrogram is our training set**. Labelled by team and through pseudo labelling.**

Convolutional neural network is a special type of deep learning architecture used to detect objects in images.

For each labelled example, we output a probability that the image is a chainsaw:



Positive example	Negative example
0.09271322	0.90728676
0.12525368	0.87474632
0.12527105	0.87472898
...	
0.10612514	0.89387488
0.1254736	0.87452638
0.12120702	0.878793

Once we have trained our model, we score unlabelled data to gain additional examples (pseudo labelling). We confirm whether the model is correct using human validation.

Threshold set high at .9 for classification due to the high cost of false negatives (conservationists actually go into the field based upon the alerts)..

Goal 2: provide direction and distance of sound

Providing direction and distance of sound reduces the area that rainforests conservationists have to search when they receive an alert.

RFCX Delta Analytics Test Site

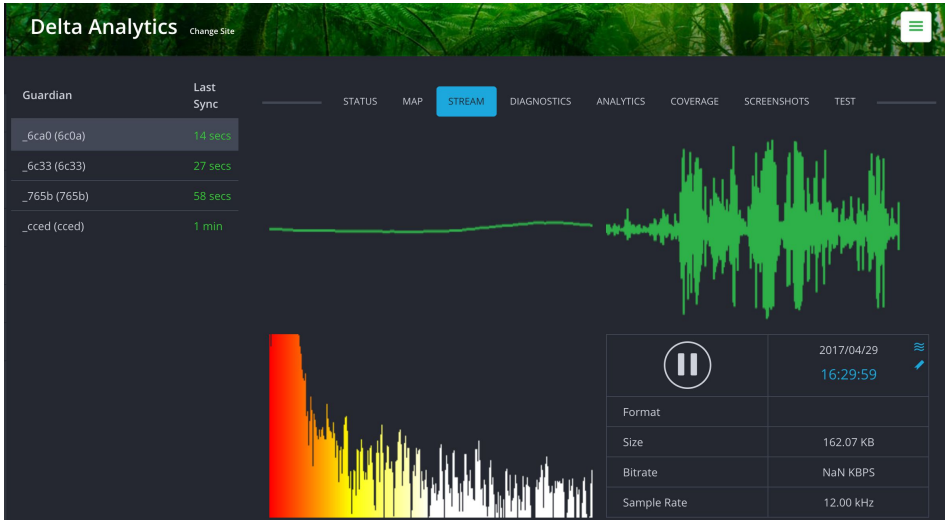
We have been given four Guardians and our own site

Goals of Delta site are to generate training data to support two research areas:

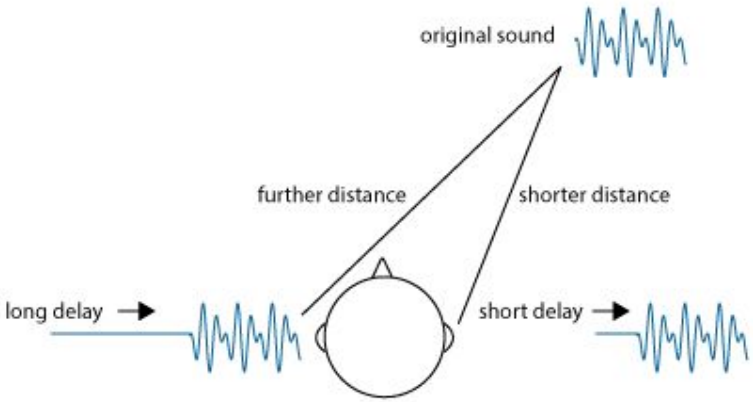
1. Estimate direction of arrival of sounds
2. Estimate distance of sound from Guardian array

Guardians are not attached, vary configuration

- Shape of array (square, triangular)
- Distance between Guardians



Approach: Source localization based on time delay of arrival



Time Delay

- Sound reaches further away sensor later than closer sensors
- Delay time + geometry => localize audio source

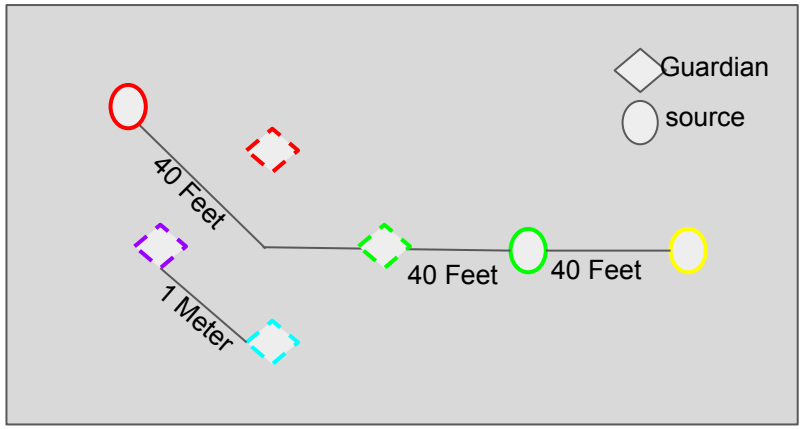
Image Source: http://gureckislab.org/courses/spring13/robots/SoundLocalization-2.html

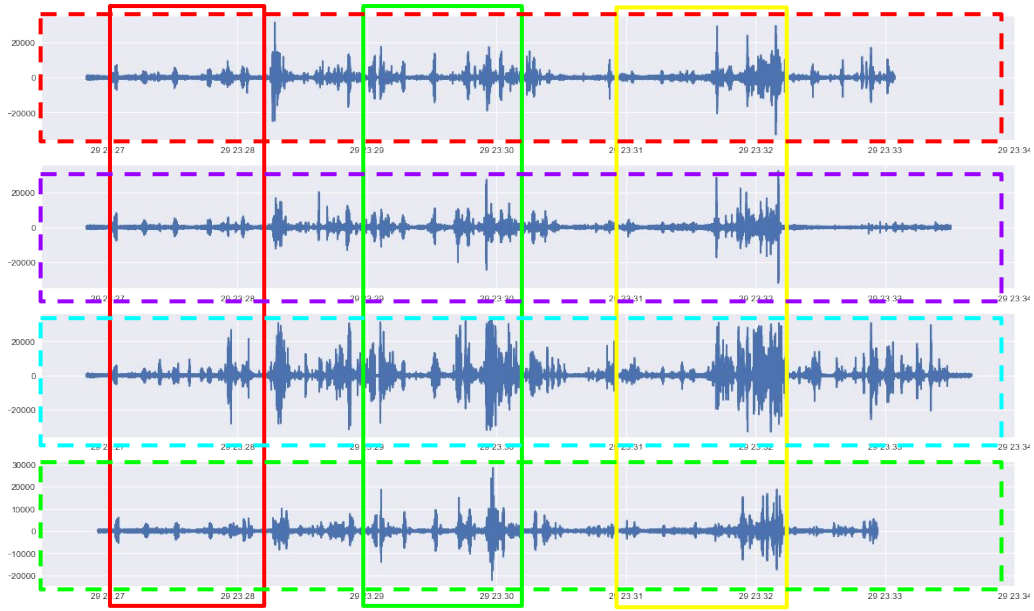
Unique Challenges

- Relative position of Guardians unknown
  - Source + Sensor Localization
- Large coverage area
  - Optimal placement/spacing of Guardians
- Guardians not synced in time

Initial Experiments

- Walk circularly around Guardian array and generate multiple impulsive sounds (e.g., whistle) and/or harmonic sounds (e.g., chainsaw, birdsong) at given locations (i.e., hours on a clock-face)
- Walk in straight line away from Guardian array and generate multiple impulsive sounds every N feet





Questions?

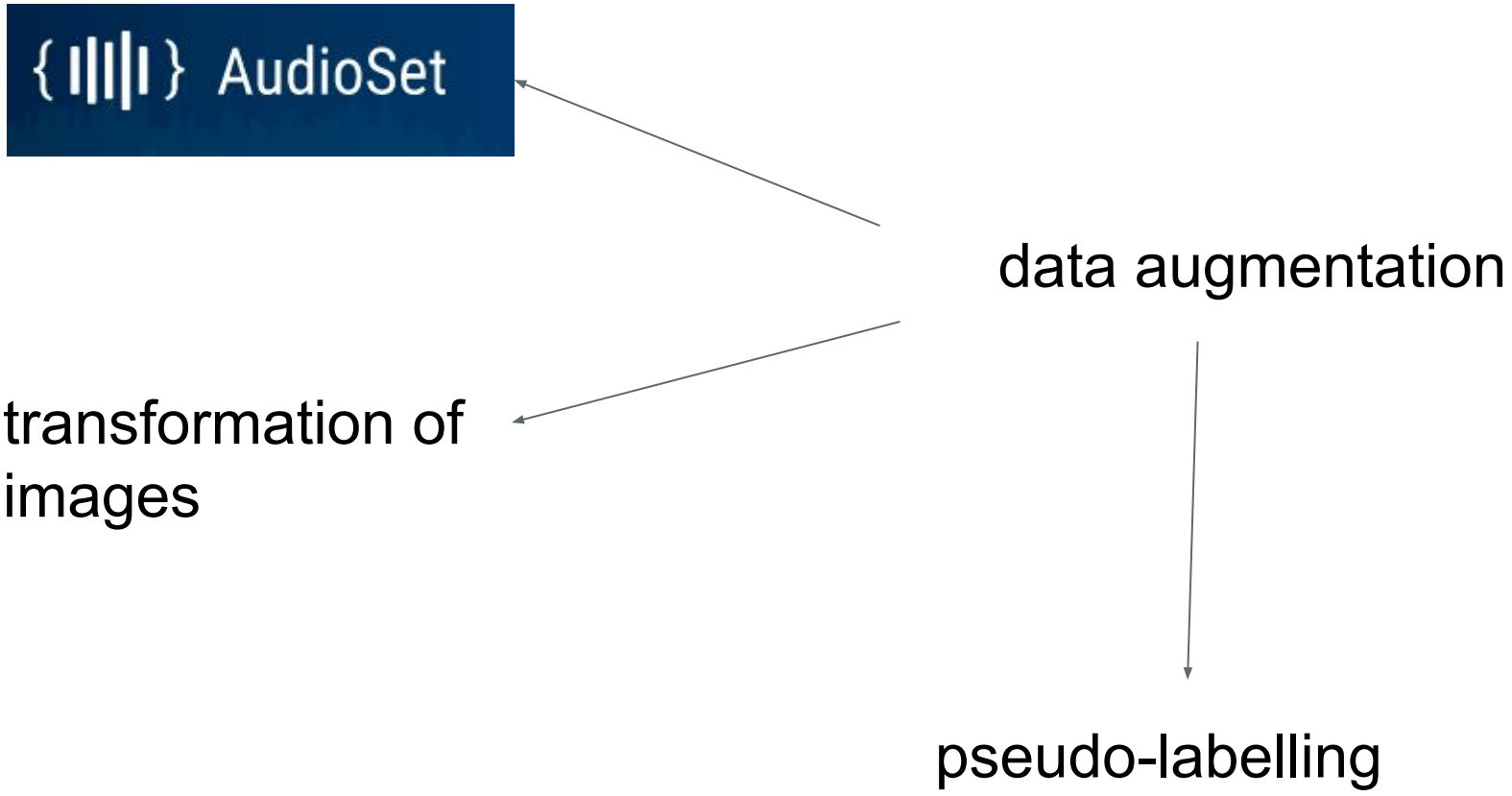
Find out more about our grant recipients here: http://www.deltanalytics.org/

Key questions:

- true ground truth
- limited labelled data set
- different data distributions at each test site
- moving towards incorporating metadata
- accounting for different test conditions (rain, mosquito season)

Limited labelled data set:

- initial data set only has 2424 spectrograms



Different data distributions at each test site:

Alto Mayo

Amazon, Alto Mayo, Peru

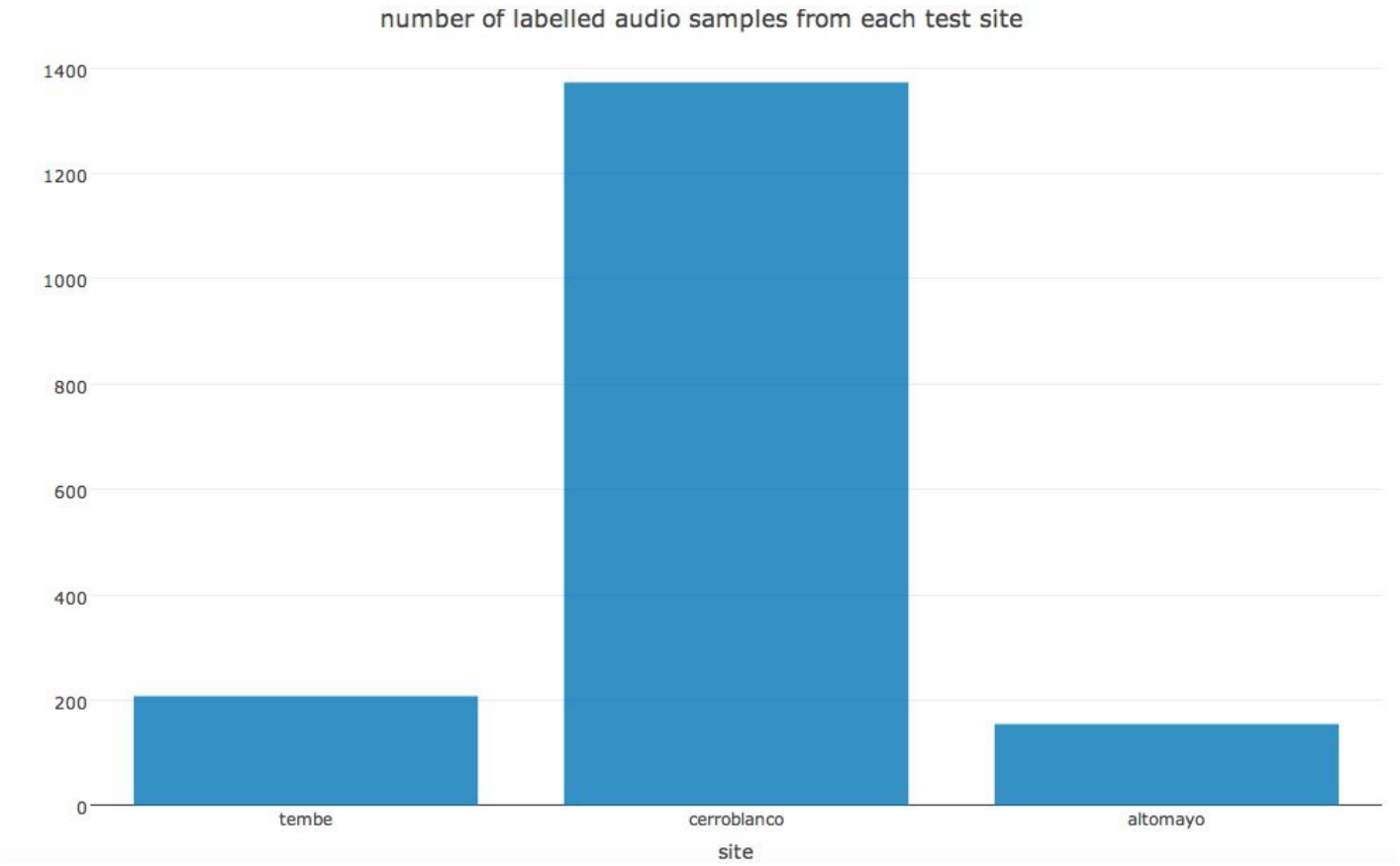
Select

Guama, Tembé Territory

Amazon, Parà, Brazil

Select

Different amounts of labelled data available at each site:



RFCX Experimental Setup

We have been given four Guardians and our own site

Goals of test site are to generate training data to support two research areas:

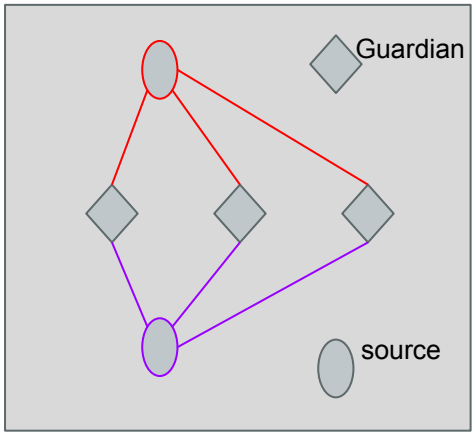
1. Estimate direction of arrival of sounds
2. Estimate distance of sound from Guardian array

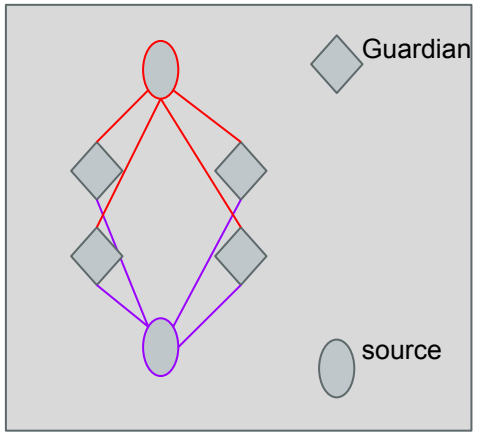
Current test site setup is a linear array with Guardians ~300 m apart

- This will likely lead to left/right confusion

We will be given our own Guardians for testing

- Square or Triangular arrangement of Guardians should prevent left/right confusion
- Shorten distance between Guardians as well

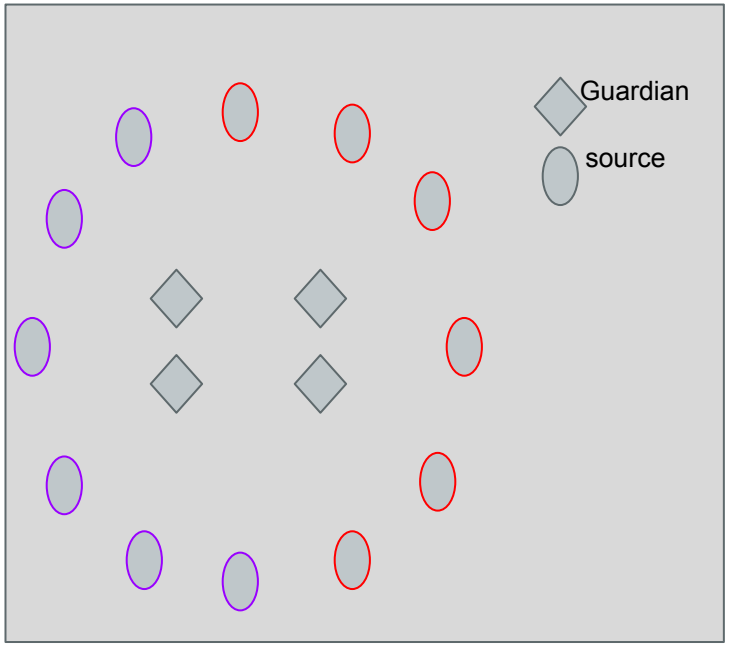


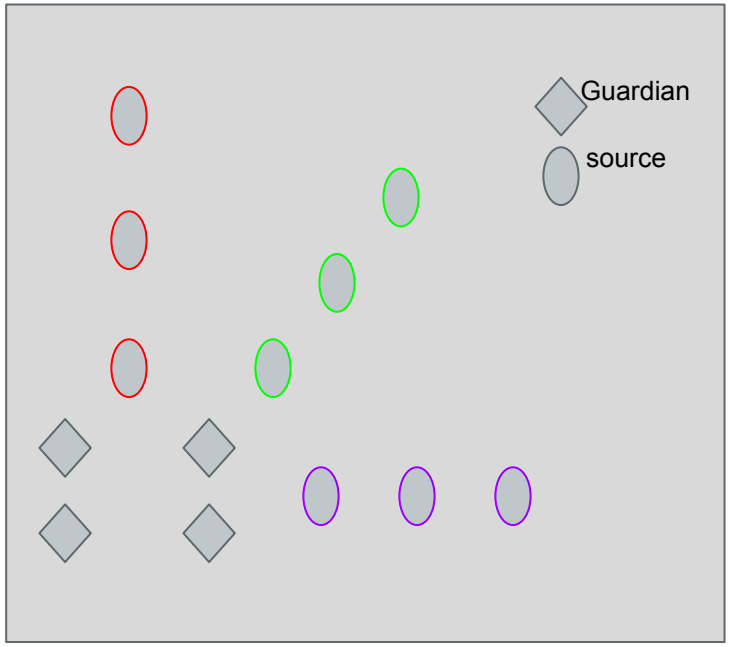


RFCX Test Site Experimental Setup

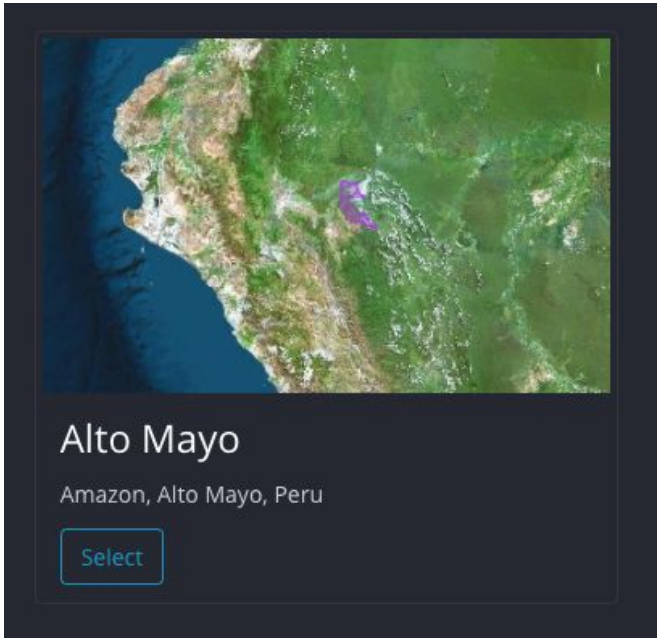
Recommended Experiments:

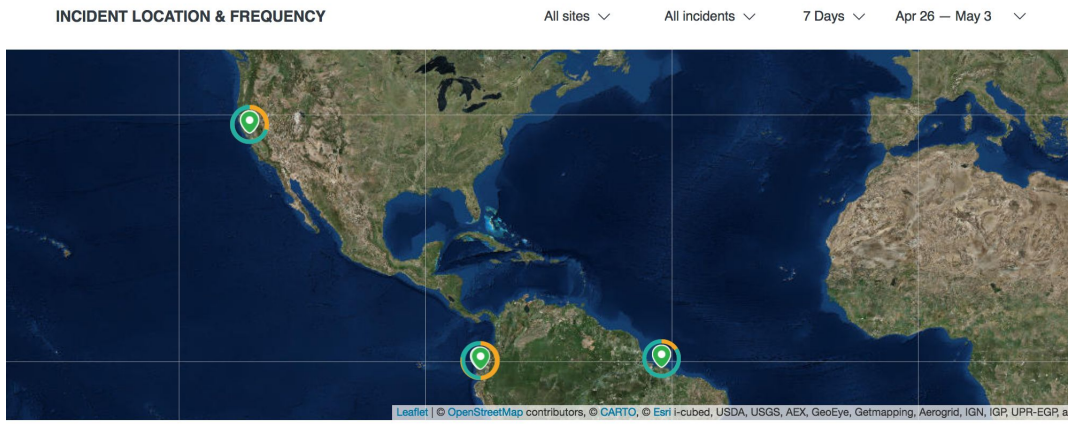
- Walk circularly around Guardian array and generate multiple impulsive sounds (e.g., hand claps) and/or harmonic sounds (e.g., chainsaw, birdsong) at given locations (i.e., hours on a clock-face)
  - Ensure all Guardians are recording, signal level is sufficient
  - Note which location correspond to which recording time
- Walk in straight line (probably perpendicular to Guardian array) and generate multiple impulsive sounds every N feet

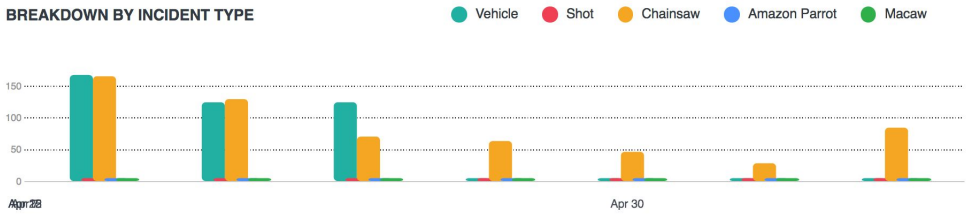




Goal 1: Improve accuracy of chainsaw detection model







- Improved accuracy will:
  - Increase ranger trust
  - Decrease time responding via dashboard or in-person

Key contributions:

Improve model:
* Add batch norm
* Increase dropout
* Reduce number of Conv layers
  * Due to small training set
* MaxPool layer between each Conv layer

Expand Training Data:
* Data Augmentation
  * Pitch Shift audio waveform
  * Translate/rotate spectrogram
* Pseudo Labeling
  * Discover more positive samples
* Google AudioSet

Enhance Data Quality:
* Chainsaw rarely only sound
* Obscure chainsaw in spectrogram
* Independent component analysis
  * Separate chainsaw and background

Utilize Metadata:
* Site
* Weather
* Time of day