Outline

- Brief introduction to Deep Neural Networks
  - Examples of applications (image recognition, natural speech processing, etc)
  - Why is deep learning successful now?
  - Feature extraction, transfer learning
  - Basic network architectures (Multilayer perceptron, CNN, RNN)
  - Acoustic data representations

- Projects
  - Detecting arctic cod grunts (UVic data)
  - Classifying killer and pilot whales (WHOI)
  - Matching individual Killer whale calls (WHOI data)
  - Detecting Baleen whales (ONC Barclay Canyon data)

- Training Datasets
  - HF5 standard

- Data augmentation strategies
  - Image manipulation
  - Sound propagation modelling
  - Deep Generative models

- Ketos library
  - HDF5 database

- Workflow visions
  - Interactive app
  - Data augmentation tool

- Your inputs
Brief introduction to neural networks
Examples of applications - Computer vision

Facial Recognition

Object detection
Examples of applications - Natural language processing

Speech recognition/synthesis

Translation

Ocean Networks Canada, Victoria, BC. 2019.04.29
Deep Learning systems aim to be end-to-end, although in practice there’s usually some level of input (signal) processing left.
Why is deep learning successful now?

- Software improvements (algorithms + frameworks)
- Hardware Improvements (Larger storage. Faster processors: cpus, gpus, tpus, etc.)
- Data improvements (More and better data)

16,000 + papers on ArXiv.org

MIT Technology review
(https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/)
Why is deep learning successful now?

- Software improvements (algorithms + frameworks)
- Hardware Improvements (Larger storage. Faster processors: cpus, gpus, tpus, etc.)
- Data improvements (More and better data)

![VGGFace2](image)

A large scale image dataset for face recognition

3 million images, 9 thousand individuals

![ImageNet](image)

14 million images, 921 thousand synsets (categories)
Transfer learning can drastically reduce the training time and amount of data required. Make models more adaptable and reusable.
Transfer learning

- Transfer learning can drastically reduce the training time and amount of data required
- Make models more adaptable and reusable

**FaceNet: A Unified Embedding for Face Recognition and Clustering**

Florian Schroff, Dmitry Kalenichenko, James Philbin

*(Submitted on 12 Mar 2015 (v1), last revised 17 Jun 2015 (this version, v3))*
Basic network architectures

Multilayer perceptron

- One of the foundational architectures for modern Deep Learning
- Rarely used on its own nowadays
Basic network architectures

Recurrent Neural Networks (RNNs)

- The output from previous cells is used by the next cells in addition to the inputs ("memory")
- Works very well for sequences/time-series
- Very useful for language modelling
Convolutional neural network (CNN)

- Creates feature maps that summarize information from one layer to the next
- Works very well for images/spatially related features
- Very useful for dealing with images; the building block of modern computer vision
Hybrid architectures

Neural Networks are usually built by combining basic architectures
Acoustic data representations
Transfer learning

- **Waveforms**
- **Spectral representations**
  - Constant Q Transform
  - Fourier transform
  - Morlet wavelet transform
- **Scalar features**

<table>
<thead>
<tr>
<th>Input signal</th>
<th>Model</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Waveform image]</td>
<td><a href="#">Pre-Processing</a> <img src="#" alt="neural-network" /> <a href="#">Neural Network</a></td>
<td>(True/False) <img src="#" alt="whale" /></td>
</tr>
</tbody>
</table>
Projects
Arctic cod grunts

Using a CNN (convolutional neural network) to detect fish sounds
Arctic cod grunts

Using a CNN (convolutional neural network) to detect fish sounds

INPUT
2D Convolution
ReLU activation
Max pooling
Dropout
2D Convolution
ReLU activation
Max pooling
Dropout
2D Convolution
ReLU activation
Max pooling
Dropout
Dense
ReLU activation
Dense
ReLU
Dense
Softmax
Output

0/1
Arctic cod grunts

Using a CNN (convolutional neural network) and a to detect fish sounds

Accuracy: 98.4%
Precision: 72.7%
Recall: 66.7%
Differentiating between killer and pilot whales using ResNETs

Accuracy: 98.44%

Ocean Networks Canada, Victoria, BC. 2019.04.29
Differentiating between killer and pilot whales using ResNETs
Matching individual killer whale calls with Siamese Networks

15 individuals

Accuracy: 94.6%
Matching individual killer whale calls with Siamese Networks

Accuracy: 94.6%

---

**INPUT 1**
- 2D Convolution
- ReLU activation
- Max pooling
- 2D Convolution
- ReLU activation
- Max pooling
- 2D Convolution
- ReLU activation
- Max pooling
- Dense
- Sigmoid

**INPUT 2**
- 2D Convolution
- ReLU activation
- Max pooling
- 2D Convolution
- ReLU activation
- Max pooling
- 2D Convolution
- ReLU activation
- Max pooling
- Dense
- Sigmoid

**L1 distance**
- Dense
- Sigmoid

**Output**
Using a sequence to sequence model to detect humpback whales
Using a sequence to sequence model to detect humpback whales
Using a sequence to sequence model to detect humpback whales

Accuracy: 87%
Precision: 72%
Recall: 36%
ONC Barclay Canyon

INPUT
- 1D Convolution
- Batch Normalization
- ReLU activation
- Dropout

1D Convolution
- Batch Normalization
- ReLU activation
- Dropout

Feature matrix

1D Convolution
- Batch Normalization
- ReLU activation
- Dropout

GRU
- Batch Normalization
- Dropout
- ML Perceptron
- Sigmoid activation

GRU
- Batch Normalization
- Dropout
- ML Perceptron
- Sigmoid activation

...
Training datasets
Datasets are organized in hierarchical format (using HDF5)

<table>
<thead>
<tr>
<th>id</th>
<th>data</th>
<th>labels</th>
<th>boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>uvic_23</td>
<td><img src="image.png" alt="3s_spectrograms image" /></td>
<td>1</td>
<td>(1.1,1.6,80,600)</td>
</tr>
</tbody>
</table>
Data Augmentation
Simple signal manipulation tricks

- Extract the signal of interest and place it in a different background
- Change amplitude
Sound propagation modeling

- Pass a signal through a sound propagation model, to simulate what it would sound like in different environments.
Deep generative models

Generative Adversarial Networks

- WaveNETs
- Variational Autoencoders
- Recurrent Neural Networks
- etc
Deep generative models

Generative Adversarial Networks

Sageev Oore

Canada CIFAR AI Chair,
Associate Professor, Faculty of Computer Science, Dalhousie University
Research Faculty Member, The Vector Institute

Previously:

Visiting Research Scientist, Google Brain
Associate Professor & Chairperson (on leave)
Department of Mathematics & Computer Science, SMU
Ketos library
Ketos library

- Open-source python package (GPL3 license)
- Available on PyPI:
  ```
  pip install ketos
  ```
- Built on top of Numpy, Tensorflow and HDF5/PyTables
- Provides:
  - Data handling tools (including for larger than memory datasets)
  - Signal processing methods
  - Useful network architectures with a common interface
Welcome to Ketos's documentation!

Introduction

Ketos is a software package for acoustic data analysis with neural networks. It was developed with a particular eye to detection and classification tasks in underwater acoustics. Ketos is written in Python and utilizes a number of powerful software packages including NumPy, HDF5, and TensorFlow. It is licensed under the GNU GPL v3 license and hence freely available for anyone to use and modify. The project is hosted on GitLab at https://gitlab.meridian.cs.dal.ca/public_projects/ketos.

Ketos was developed by the MERIDIAN Data Analytics Team at the Institute for Big Data Analytics at Dalhousie University. We are grateful to Amalie Riera and Francis Jaques at the University of Victoria, Kim Davies and Chris Taggart at Dalhousie University, and Kristen Kanes at Ocean Networks Canada for providing us with annotated acoustic data sets, which played a key role in the development work. The first version of Ketos was released in April 2019.

The intended users of Ketos are primarily researchers and data scientists working with (underwater) acoustics data. While Ketos comes with complete documentation and comprehensive step-by-step tutorials, some familiarity with Python and especially the NumPy package would be beneficial. A basic understanding of the fundamentals of machine learning and neural networks would also be an advantage.

The name Ketos was chosen to highlight the package’s main intended application, underwater acoustics. In Ancient Greek, the word ketos denotes a large fish, whale, shark, or sea monster. The word ketos is also the origin of the scientific term for whales, cetacean.
Workflow vision
Workflow visions

MERIDIAN’s library of pre-trained models

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
</tbody>
</table>

- **Scenario**
- **Performance**

- **right whales**
- **right whales + ships**
- **right whales + humpback whales**
Workflow visions

MERIDIAN's library of pre-trained models

- model 1
- model 2
- model 3

Performance

Scenario

right whales
right whales + ships
right whales + humpback whales

User's scenario

right whales + humpback whales
Workflow visions

MERIDIAN's library of pre-trained models

- model 1
- model 2
- model 3

Scenarios and performance:

- right whales
- right whales + ships
- right whales + humpback whales

User's scenario:
- right whales + humpback whales
Workflow visions

MERIDIAN’s library of pre-trained models

- **model 1**
- **model 2**
- **model 3**

---

**Scenario**

- **Performance**
  - right whales
  - right whales + ships
  - right whales + humpback whales

**User’s scenario**

- right whales + seismic noise
Workflow visions

MERIDIAN’s library of pre-trained models

- Model 1
- Model 2
- Model 3

User’s scenario

- Right whales
- Right whales + ships
- Right whales + humpback whales
- Right whales + seismic noise
Workflow visions

MERIDIAN’s library of pre-trained models

model 1
model 2
model 3

Performance
Scenario

Performance
Scenario

Performance
Scenario

right whales
right whales + ships
right whales + humpback whales

User’s scenario

right whales + seismic noise

Ocean Networks Canada, Victoria, BC. 2019.04.29
Workflow visions

Transfer learning

model 2 + Data + User's scenario = custom model

User's scenario

right whales + seismic noise

Performance

Scenario

Ocean Networks Canada, Victoria, BC. 2019.04.29
Interactive training tool

- visual interactive learning application
- human analyst and neural network work together
- human analyst inspects (and corrects) classifications proposed by neural network.
- neural network improves its performance in response to feedback from analyst
Workflow visions

(Inter)active learning

model 2 + Data

(custom model)
Workflow visions

Data augmentation tool

- A large database of annotated underwater acoustic samples (multiple species/call types)
- A variety of background noises
- Users can mix and match
- Data augmentation, signal processing and other methods also available
Your inputs!
How can we best serve the community?

What kind of interface to build for trained detectors/classifiers?

- Desktop App
- We App
- Embedded system
- CLI
Thank you!