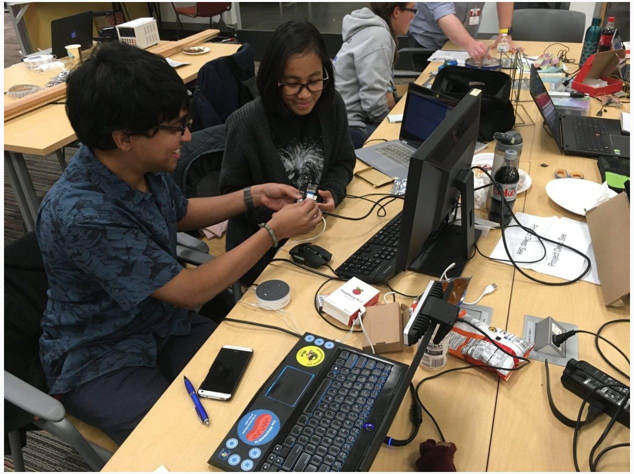
Deep See

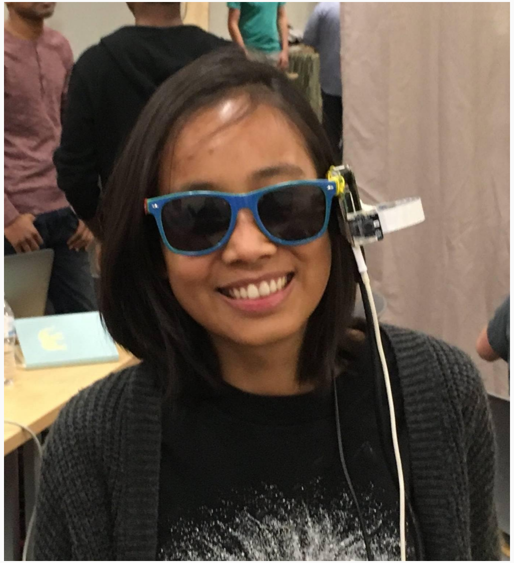
Presented by Madhu Tennakoon & Myat Mo

Introductions: Project Deep See



- IoTHackDay 2017 project
- Creative runner-up award
- Highlight of the night…

...we got hired!



Introductions

75F®

Check us out at www.75f.io

Motivation

- Assistive Technology has come a long way since the Braille typewriter.
- In today's world of intelligent voice assistants, smart homes and gesture detection wearables, we know that technology will always make life easier.
- However, there is a long way to go.

Motivation (contd.)

- The beautiful world around us - with all its captivating visual stimuli - is out of some people's reach.
- How do we use the technology available to us today to make their lives not just simpler but also more meaningful?
- Maybe the answer lies in AI.
- Maybe we can use machine vision to help others to see and understand the world around us.

Where do we start?

amazon alexa

- Let’s start with Alexa.
- Alexa is an AI system designed to engage with one of the world's biggest and most tangled data sets: human speech.
- Alexa Voice Service focuses on 3 important markets: home automation, home entertainment, and shopping.
- But at its core, Alexa is an accessibility tool, and that’s what we focused on.

The Plan: *make Alexa smarter*

- Our hope is to use deep learning and voice recognition tools to help the visually impaired see and understand the world around them.
- We plan to make Amazon Alexa a lot smarter by adding machine vision capabilities to her ‘*skill set*’
- We use Google's deep learning libraries for brain power and AWS Lambda for compute power.

“See without looking”

- Bob Dylan

Sidenote: Other applications

- Smarter virtual assistants
- Industry insight
- Remote activity monitoring

Cookbook

h/w

- Amazon Echo Dot
- Raspberry Pi
- RPi Camera

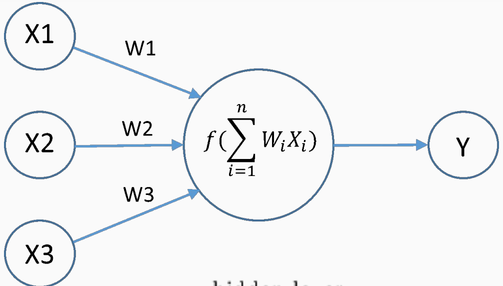
s/w

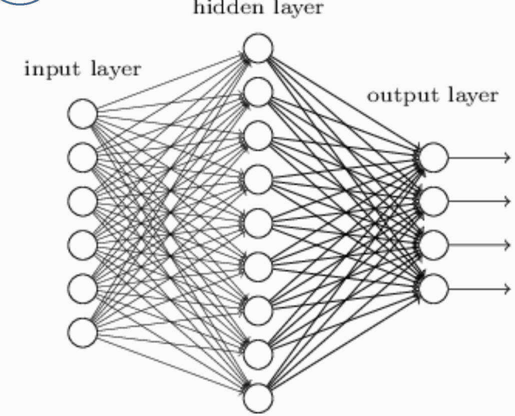
- Google TensorFlow
- AWS IoT
- AWS Lambda
- AWS Rekognition

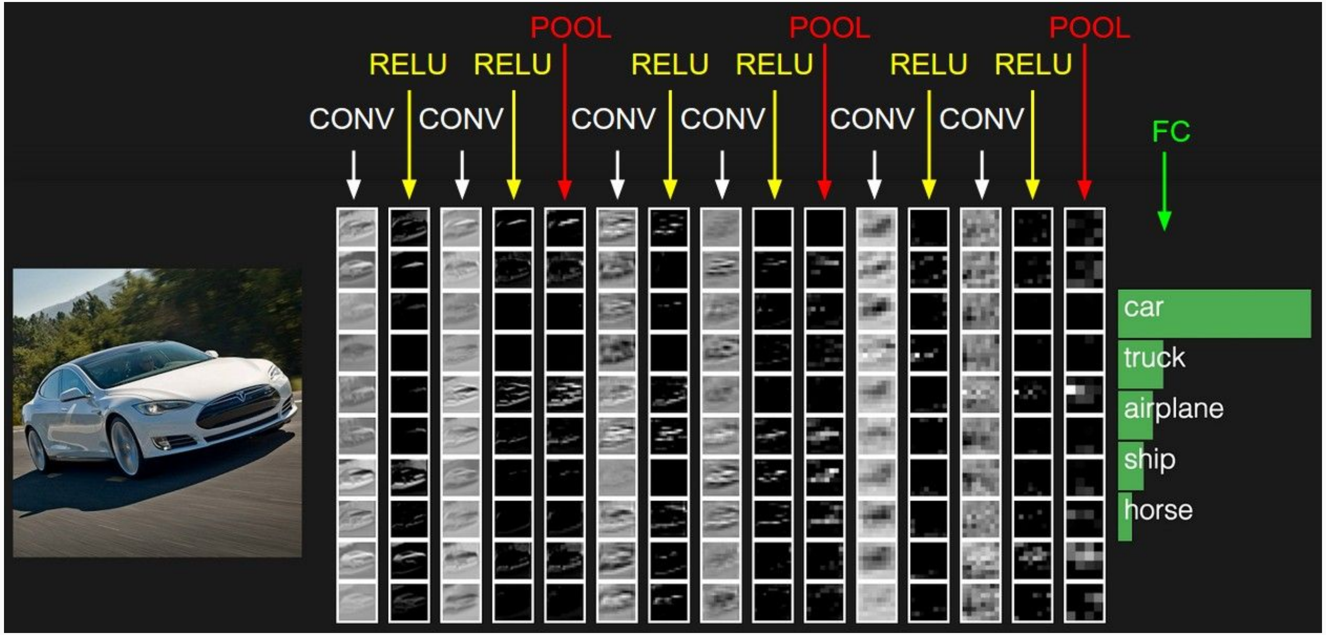
Background

Part I: Image Classification

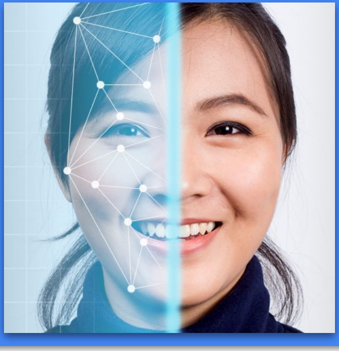
Deep Convolutional Neural Networks

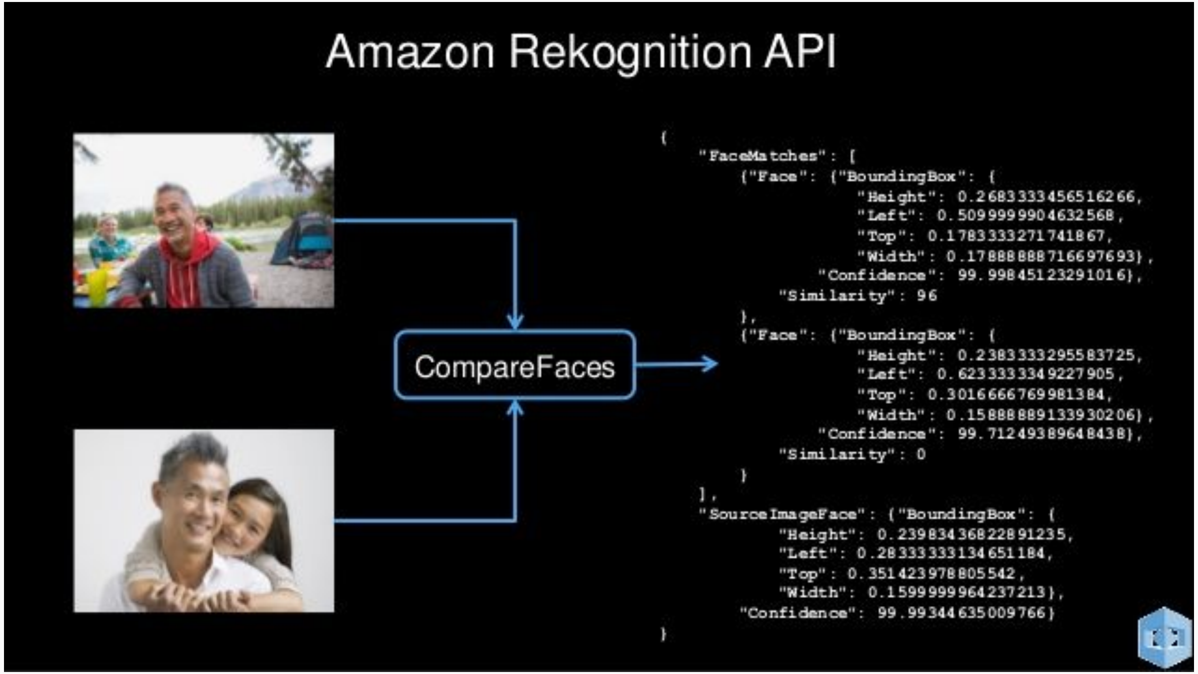






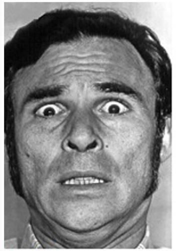
Real-time facial recognition with Alexa





Amazon Rekognition API

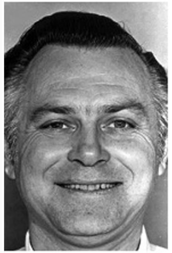
```
{
    "FaceMatches": [
        {"Face": {"BoundingBox": {
                    "Height": 0.2683333456516266,
                    "Left": 0.5099999904632568,
                    "Top": 0.1783333271741867,
                    "Width": 0.1788888871669769З},
                "Confidence": 99.9984512329101б},
            "Similarity": 96
        },
        {"Face": {"BoundingBox": {
                    "Height": 0.2383333295583725,
                    "Left": 0.6233333349227905,
                    "Top": 0.3016666769981384,
                    "Width": 0.1588888913393020б},
                "Confidence": 99.7124938964843８},
            "Similarity": 0
        }
    ],
    "SourceImageFace": {"BoundingBox": {
            "Height": 0.2398343682289123５,
            "Left": 0.2833333313465118４,
            "Top": 0.351423978805542,
            "Width": 0.1599999964237213},
        "Confidence": 99.99344635009766}
}
```



Fearful

Angry

Sad



Happy

Disgusted

Surprised

Background

Part II: More neural networks

Long-Short Term Memory (LSTM)

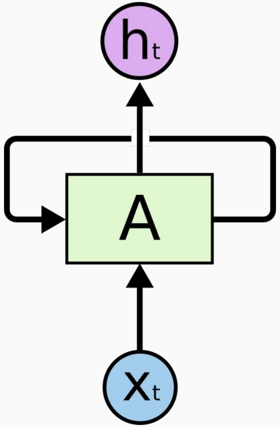


Fig. 1: A single RNN cell

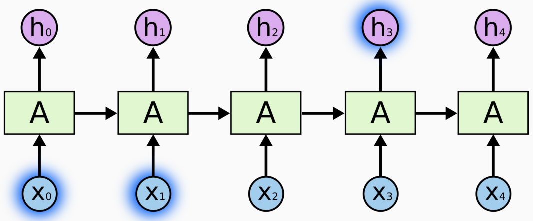


Fig. 2: An RNN network

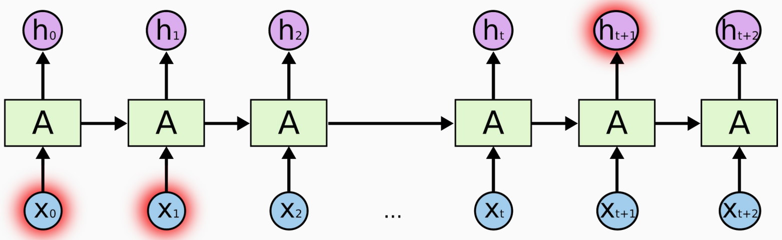


Fig. 3: RNN’s have trouble recalling much older data

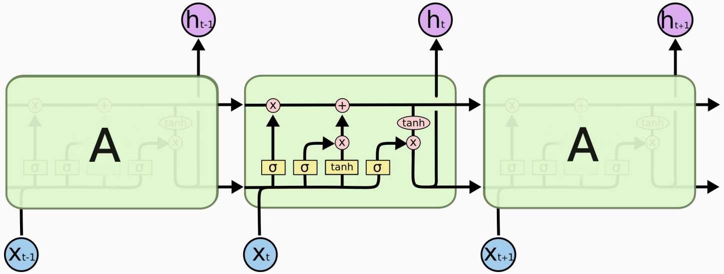
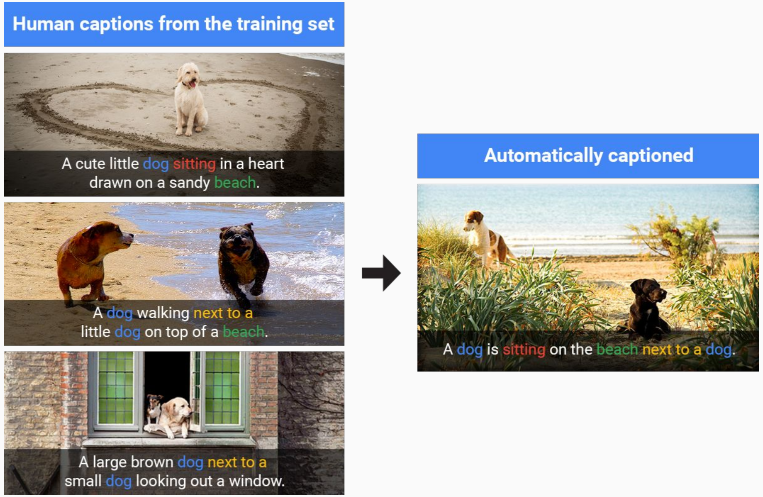


Fig. 3: Close-up of a single LSTM cell, a special kind of RNN

Deep Learning with TensorFlow

- TensorFlow is an open source software library for numerical computation
- In 2014, the Google Brain team trained a ML program to automatically produce captions that accurately describe images
- In 2016, they released an open-source model in TensorFlow
- It was pre-trained using tonnes of human-captioned data
- The result was a model that generates natural language descriptions of images and their regions



Human captions from the training set

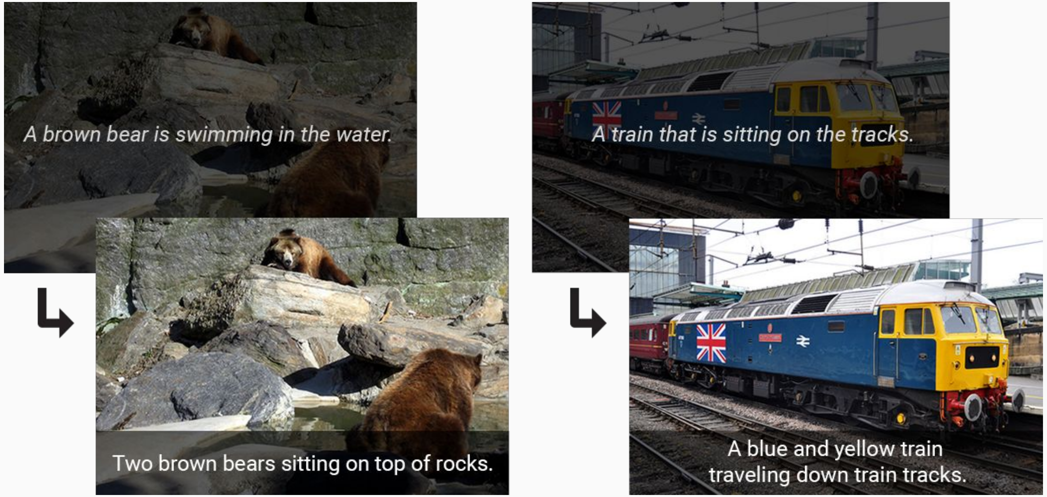
A cute little dog sitting in a heart drawn on a sandy beach.

A dog walking next to a little dog on top of a beach.

A large brown dog next to a small dog looking out a window.

Automatically captioned

A dog is sitting on the beach next to a dog.



A brown bear is swimming in the water.

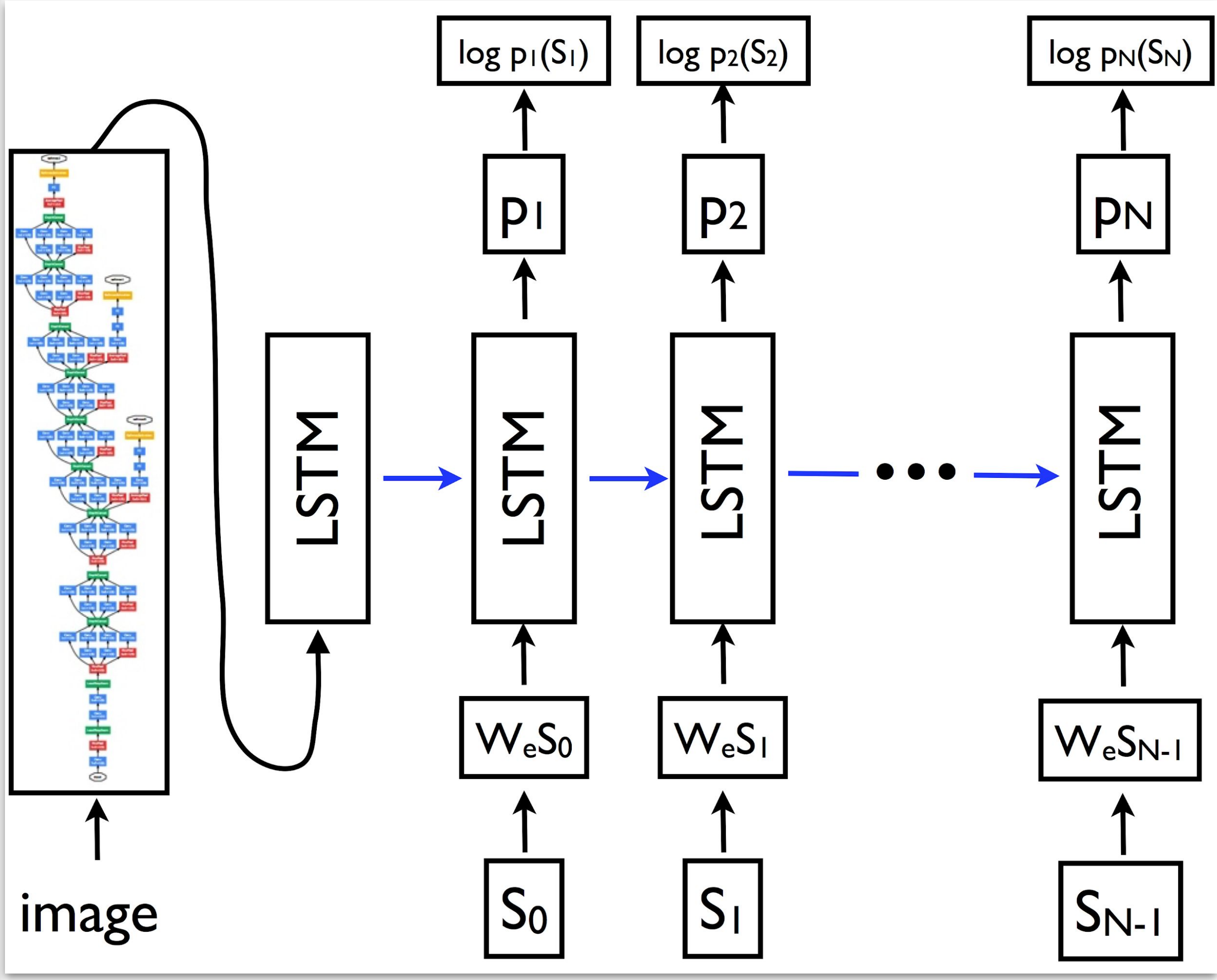
Two brown bears sitting on top of rocks.

A train that is sitting on the tracks.

A blue and yellow train traveling down train tracks.

Deep Image-Captioning

- CNN + LSTM
- CNN → does the classification
- LSTM → does the captioning
- Uses Inception v3
  - Trained on 1000 classes
  - 93.9% classification accuracy
- Extremely deep CNN layers



Putting it all together...

alexa + TensorFlow AWS =

AWS Lambda, IoT, S3, Rekognition

Time for a demo.

2A
MONOCLE
2A
THE MONOCLE SHOP — LONDON, NEW YORK, TOKYO & ZÜRICH
ART PRINTS
BOOKS
FRAGRANCES
FURNITURE
PERIODICALS
KNITWEAR
MUSIC
LUGGAGE
POSTERS
STATIONERY
ART PRINTS
BOOKS
FRAGRANCES
FURNITURE
PERIODICALS
KNITWEAR
MUSIC
LUGGAGE
POSTERS
STATIONERY

Robert Berdan ©

Future Improvements

- **Region-based CNN’s:** CNN’s tells us *‘who’* and *‘what’*. RCNN’s also tell us *‘where’*
- **Transfer learning:** *what else can we ask Alexa to see?*

References

- https://ai.googleblog.com/2016/09/show-and-tell-image-captioning-open.html
- *http://karpathy.github.io/2015/05/21/rnn-effectiveness/*
- *http://colah.github.io/posts/2015-08-Understanding-LSTMs/*

Thank you!

Any questions?