Voice and Image Recognition AI

Team 13
Tyler Thompson
Kieran Harrigan
Matthew Wong
Patrick Liao

What is Image Recognition?

Image recognition is a discipline in Computer Vision that uses specially designed algorithms to extract patterns and data out of images and videos.

The idea of image recognition is to design a system that can deduce information from an image or video at the same level or detail a person can.

Millions of images and videos are uploaded daily, which allow image recognition technologies to learn and identify a large amount objects in a short amount of time.

What is Image Recognition?

Some applications of Image Recognition include

Object Recognition

Action Recognition in Surveillance and Security

Medical Image Analysis

Self-Driving Vehicles

Face Recognition and Tracking

Augmented Reality

CamFind Search and Image Recognition API

Neural Networks

Most applications of Image Recognition use Neural Networks

Neural Networks are systems of nodes that take a number of inputs and produce a number of outputs.

- These Neural Networks produce these outputs based on the inputs given and the weights associated with the connections between nodes
- A Neural Network learns by iterating through training data, and altering the weights until the results match the already known target values
- This is known as Backpropagation

Neural Networks

The nodes in the Neural Network are organized into 3 effective layers

The Input Layer is a layer that contains as many nodes as there are inputs

- The Input Layer just feeds input to the next layer

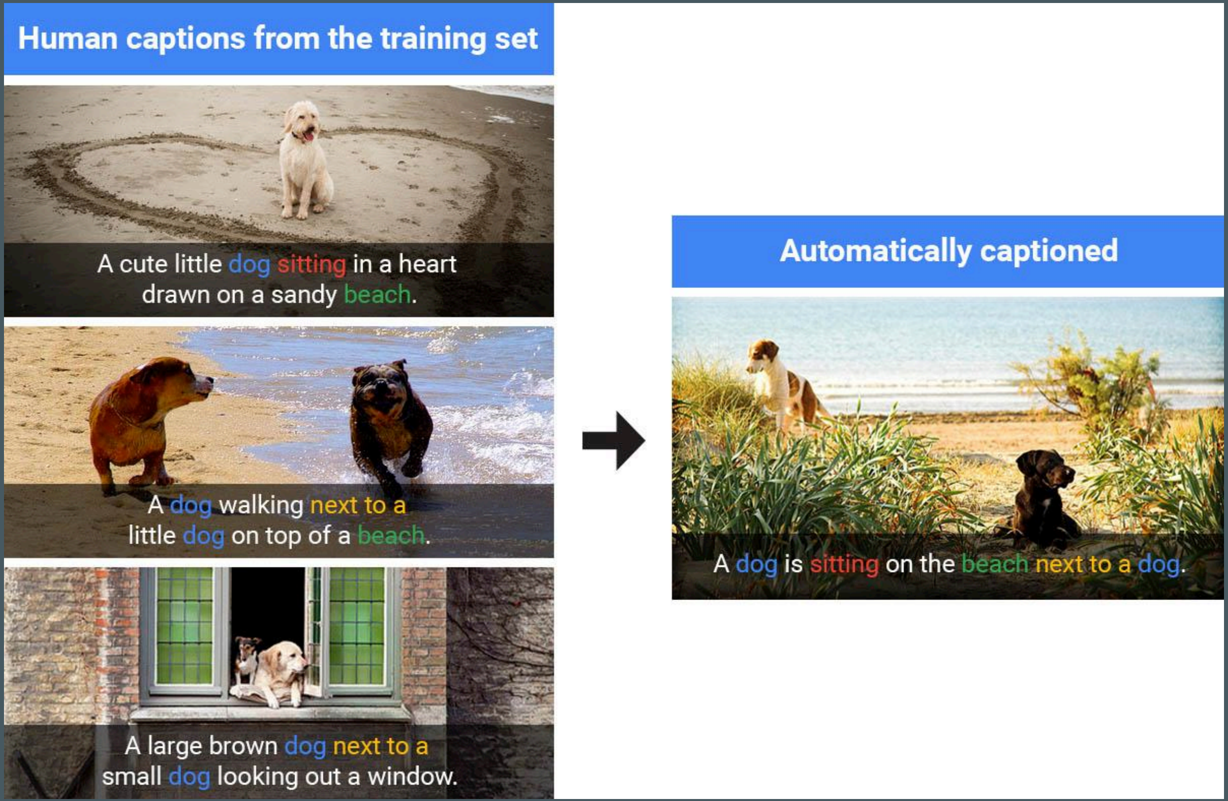
The Hidden Layer's node count depends on the implementation and serves as a filter for the data

- The Hidden Layer may contain more than one layer of nodes, but typically it does not exceed two

The Output Layer's node count is the number of classes, and the output of the Output Layer is the network's overall response or prediction of the input

The weights that are associated with each connection are first random, then

Neural Networks

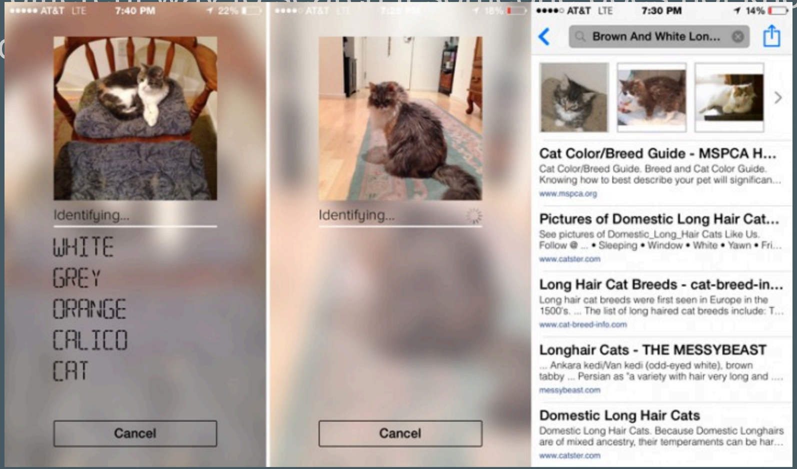

Human captions from the training set
A cute little dog sitting in a heart drawn on a sandy beach.
A dog walking next to a little dog on top of a beach.
A large brown dog next to a small dog looking out a window.
Automatically captioned
A dog is sitting on the beach next to a dog.

Why is Image Recognition Important?

Search Convenience

If a person wants to know about a certain book or breed of a dog, he or she can take a picture of it and get detailed information about it almost immediately.

This also provides a different way to search if someone does not know how to describe something in wor



Cat Color/Breed Guide - MSPCA H...
Cat Color/Breed Guide. Breed and Cat Color Guide. Knowing how to best describe your pet will significan...
www.mspca.org

Pictures of Domestic Long Hair Cat...
See pictures of Domestic_Long_Hair Cats Like Us. Follow @ ... • Sleeping • Window • White • Yawn • Fri...
www.catster.com

Long Hair Cat Breeds - cat-breed-in...
Long hair cat breeds were first seen in Europe in the 1500's. ... The list of long haired cat breeds include: T...
www.cat-breed-info.com

Longhair Cats - THE MESSYBEAST
... Ankara kedi/Van kedi (odd-eyed white), brown tabby ... Persian as "a variety with hair very long and ....
messybeast.com

Domestic Long Hair Cats
Domestic Long Hair Cats. Because Domestic Longhairs are of mixed ancestry, their temperaments can be har...
www.catster.com

Why is Image Recognition Important?

Makes Online Shopping Easier

Someone can take a picture of a product, and a website/app can track that product and let that person know when and where that product is being sold.

Neiman Marcus launched an app that lets a user take a picture of an item they want, and the app uses an image recognition feature that will find similar items and allow the user to search and shop through these items.

A platform in China is working on a feature that lets mobile users pay for products with a selfie.

Why is Image Recognition Important?

Companies can Analyze Visual Content

Images and videos are very popular today, with social media sites/apps leaning towards communication via photos and short videos.

Companies can use image recognition to analyze these photos and videos to see which brands are popular and well-liked, and how their products are being used every day.

With the data retrieved, companies can change their marketing plan on certain products, and may create new products that they feel would be popular with the targeted audience

Case Study: Neuroph

Neuroph is a Java framework that allows programmers to develop software that require or utilize image recognition

Images are represented as three 2D arrays that represent a pixel’s red, green, and blue values.

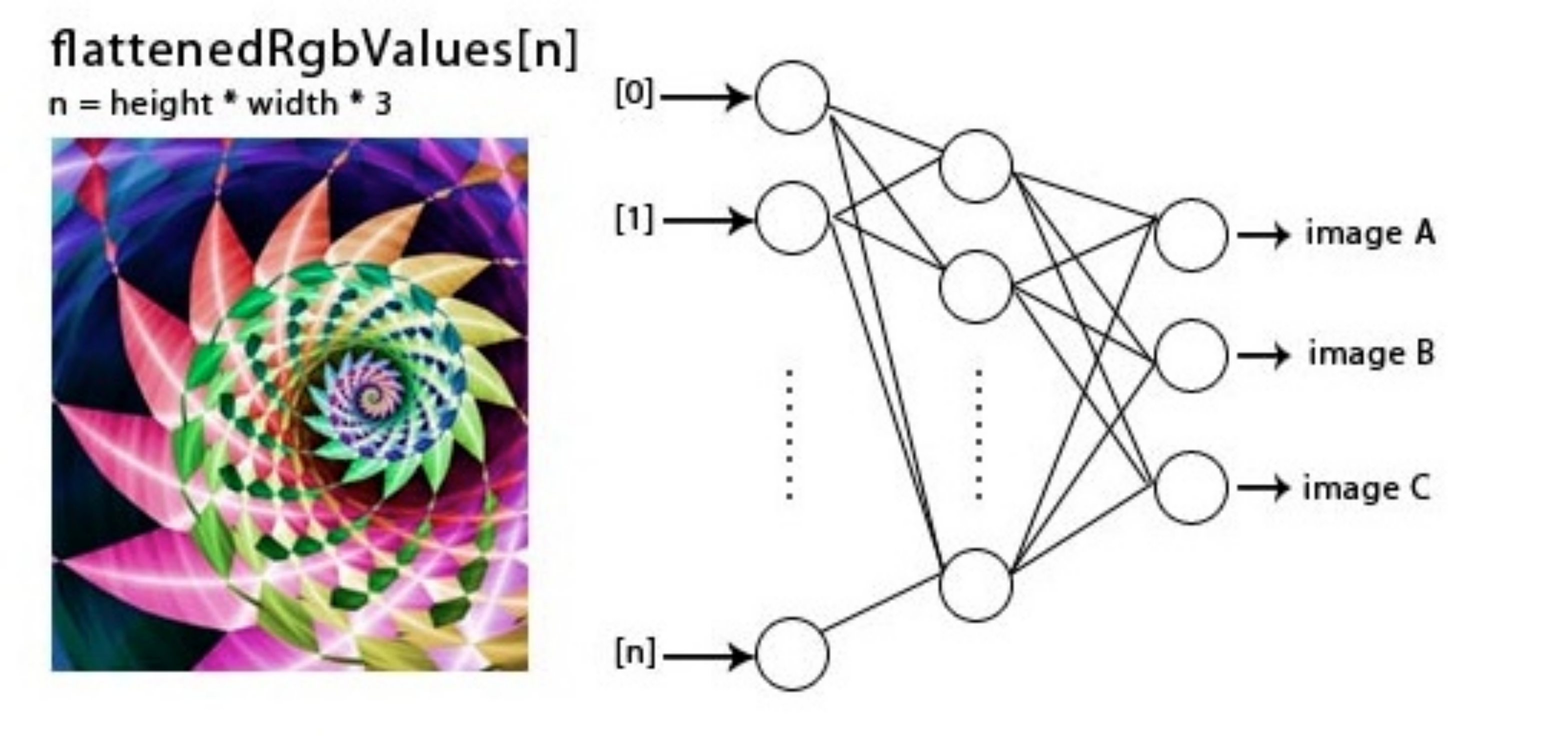
A red pixel in the top left corner of a picture will have the following values

redValues[0][0] = 255, greenValues[0][0] = 0, blueValues[0][0] = 0

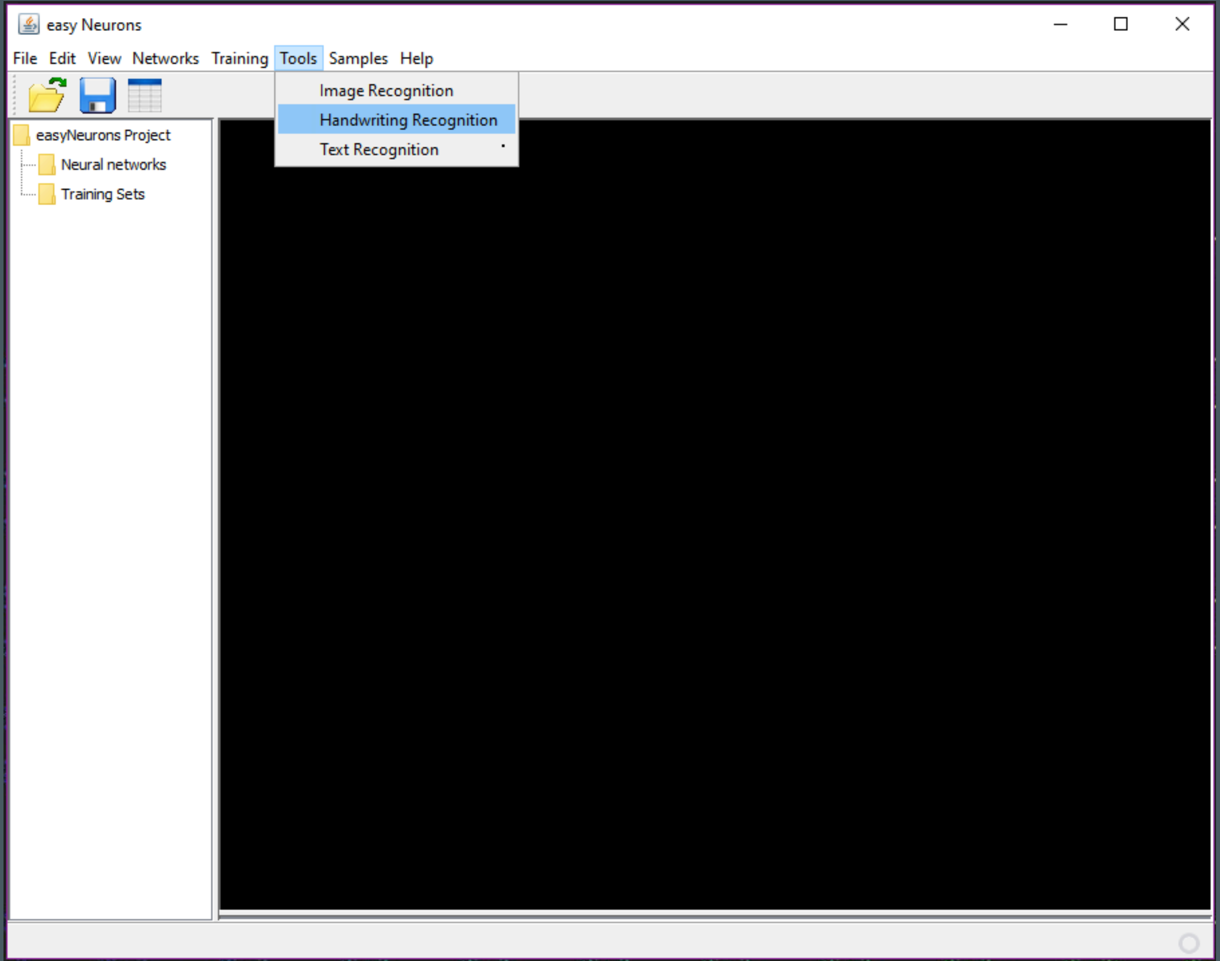
The three 2D arrays are then combined to create one flat array.

The array can then be used as input for a neural network to determine patterns in images

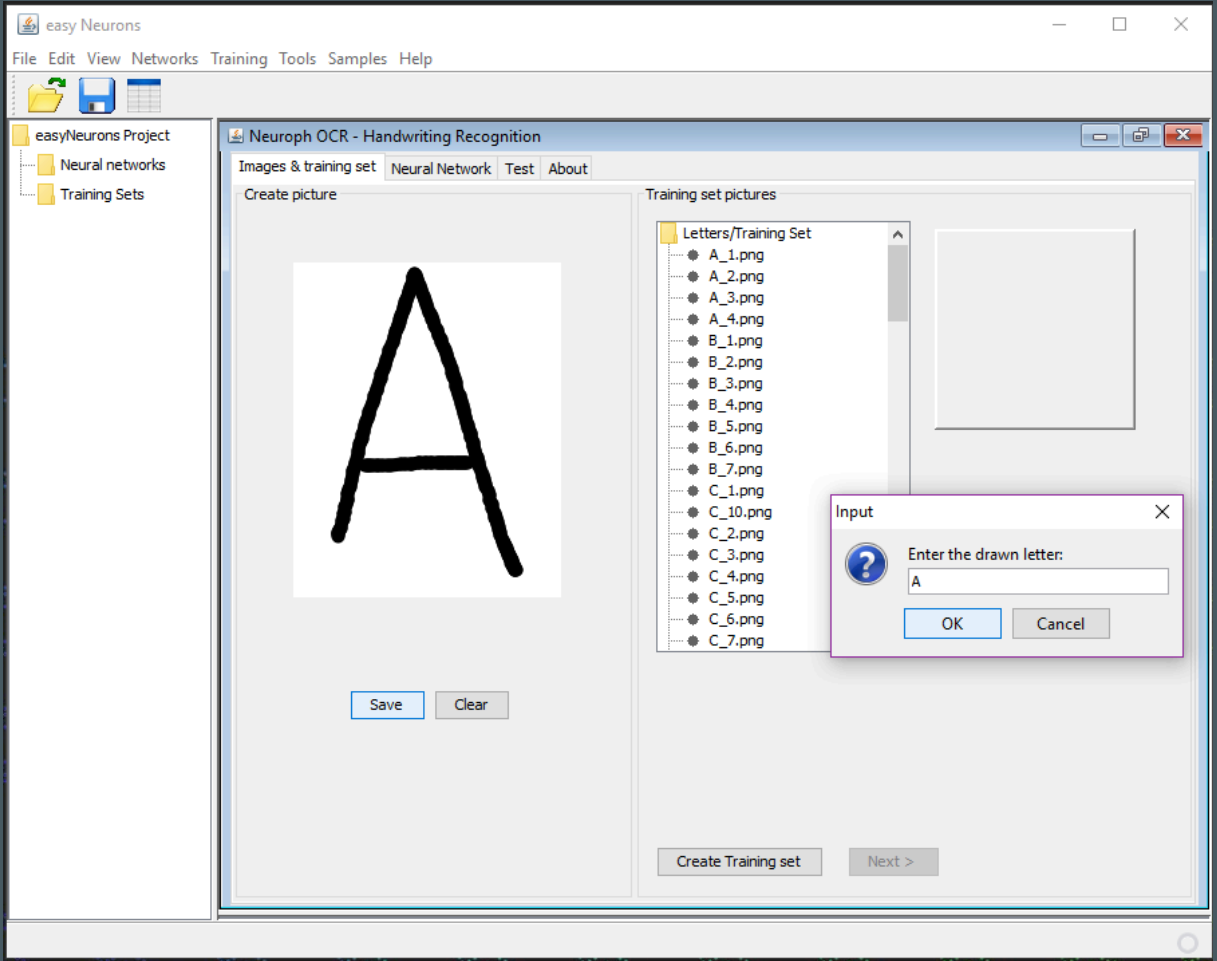
Case Study: Neuroph



Neuroph Demo

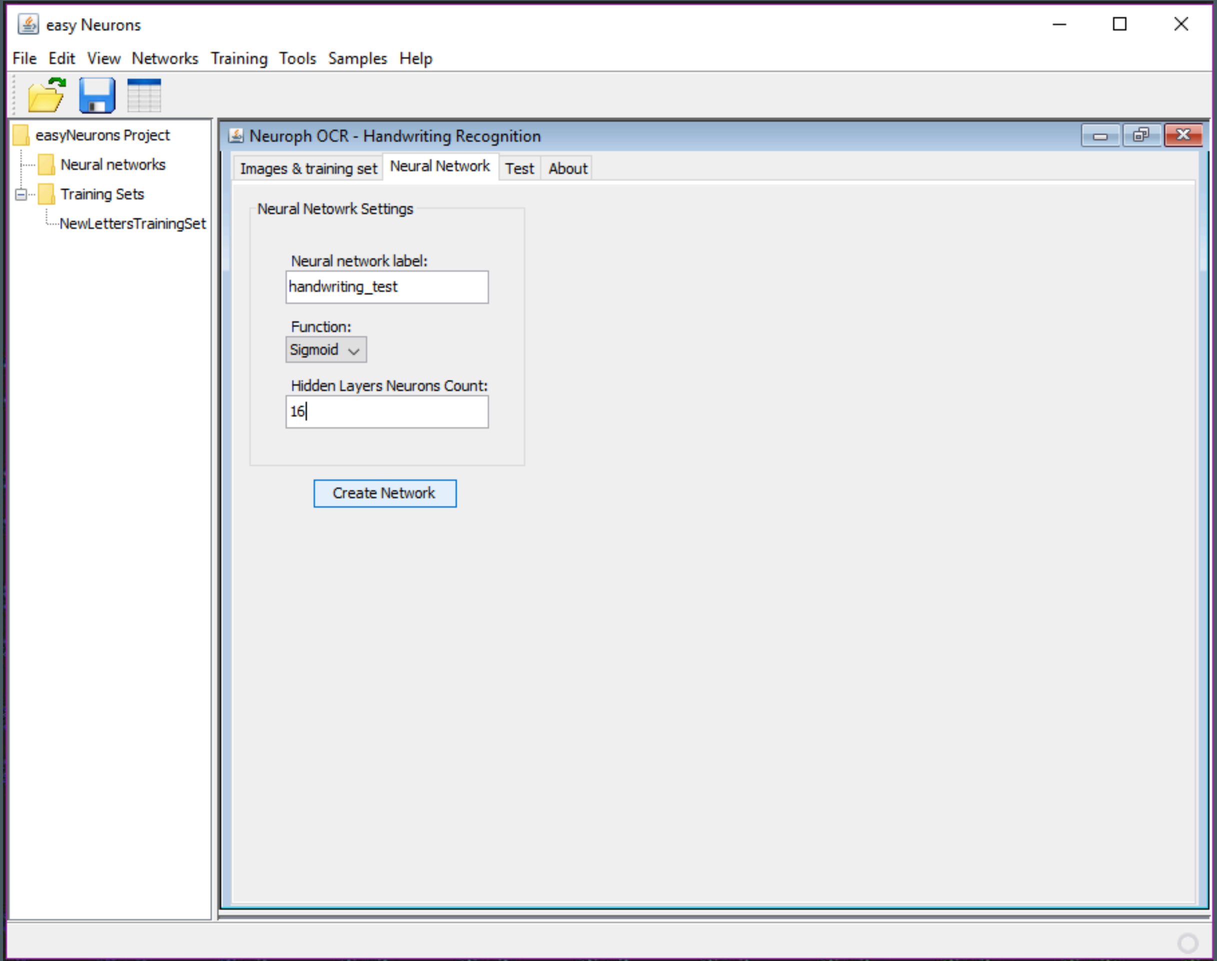


Neuroph Demo



Enter the drawn letter: A

Neuroph Demo



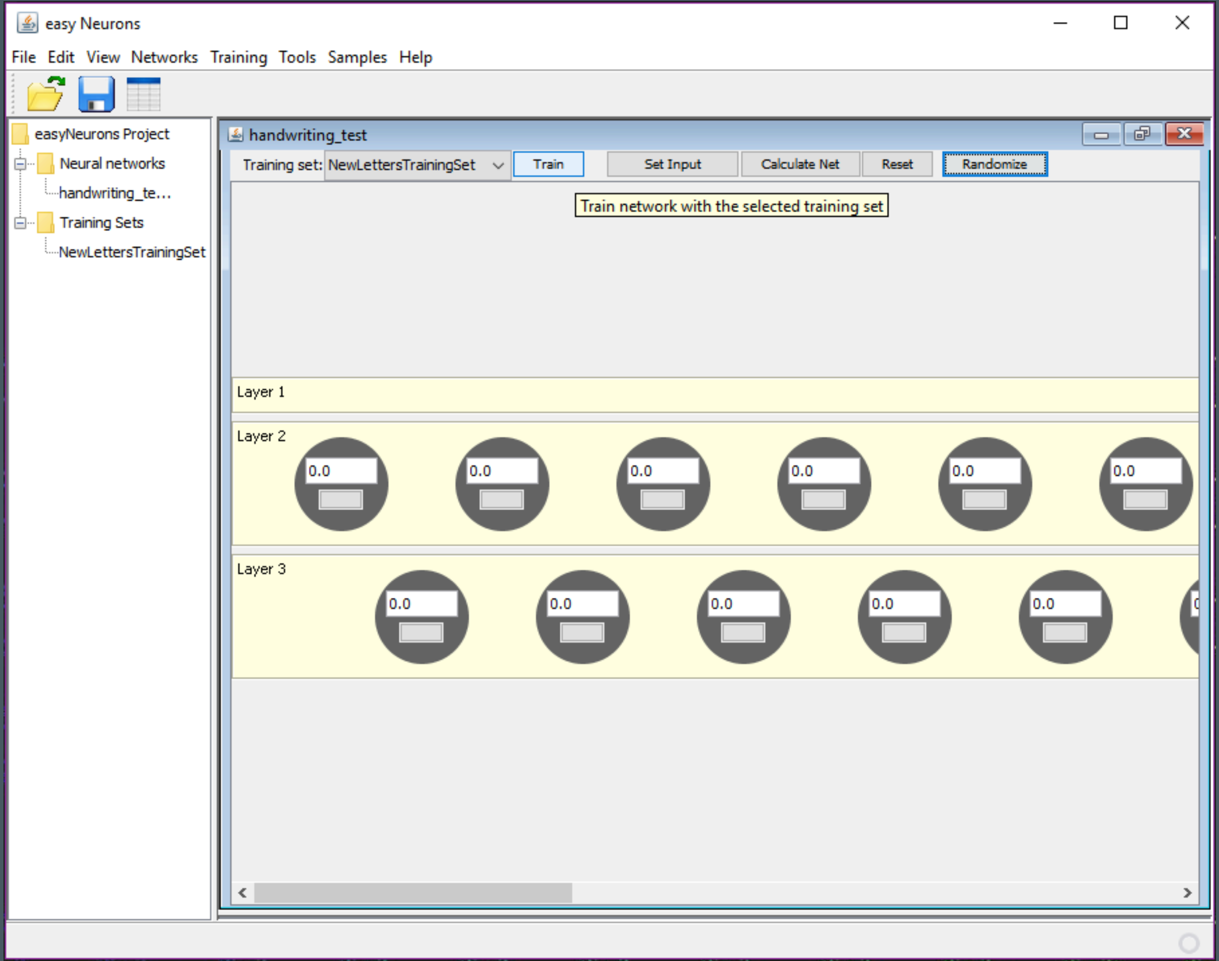
Neural Netowrk Settings

Neural network label: handwriting_test

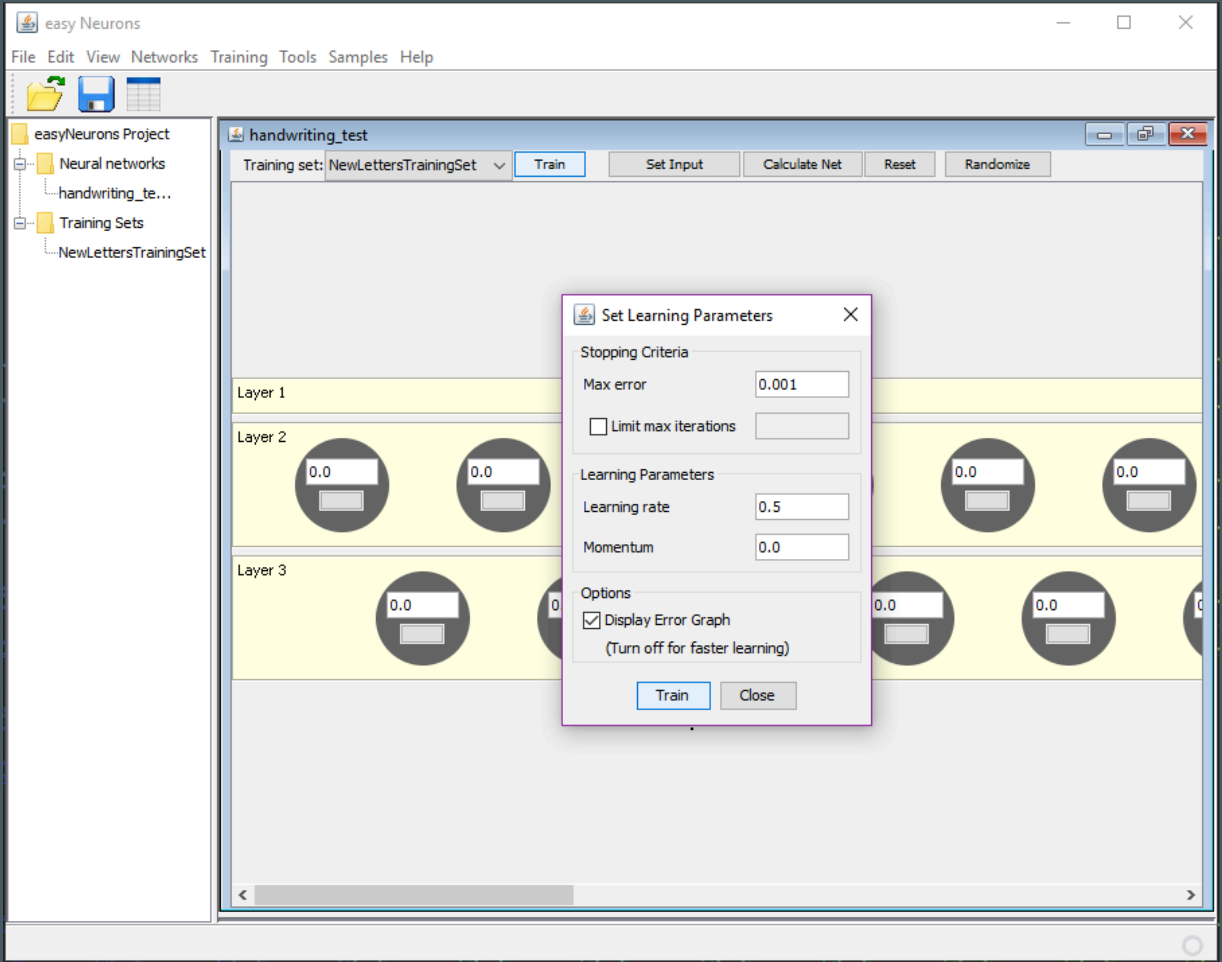
Function: Sigmoid

Hidden Layers Neurons Count: 16

Neuroph Demo



Neuroph Demo



Training set: NewLettersTrainingSet

Set Learning Parameters

Stopping Criteria

Max error: 0.001

□ Limit max iterations

Learning Parameters

Learning rate: 0.5

Momentum: 0.0

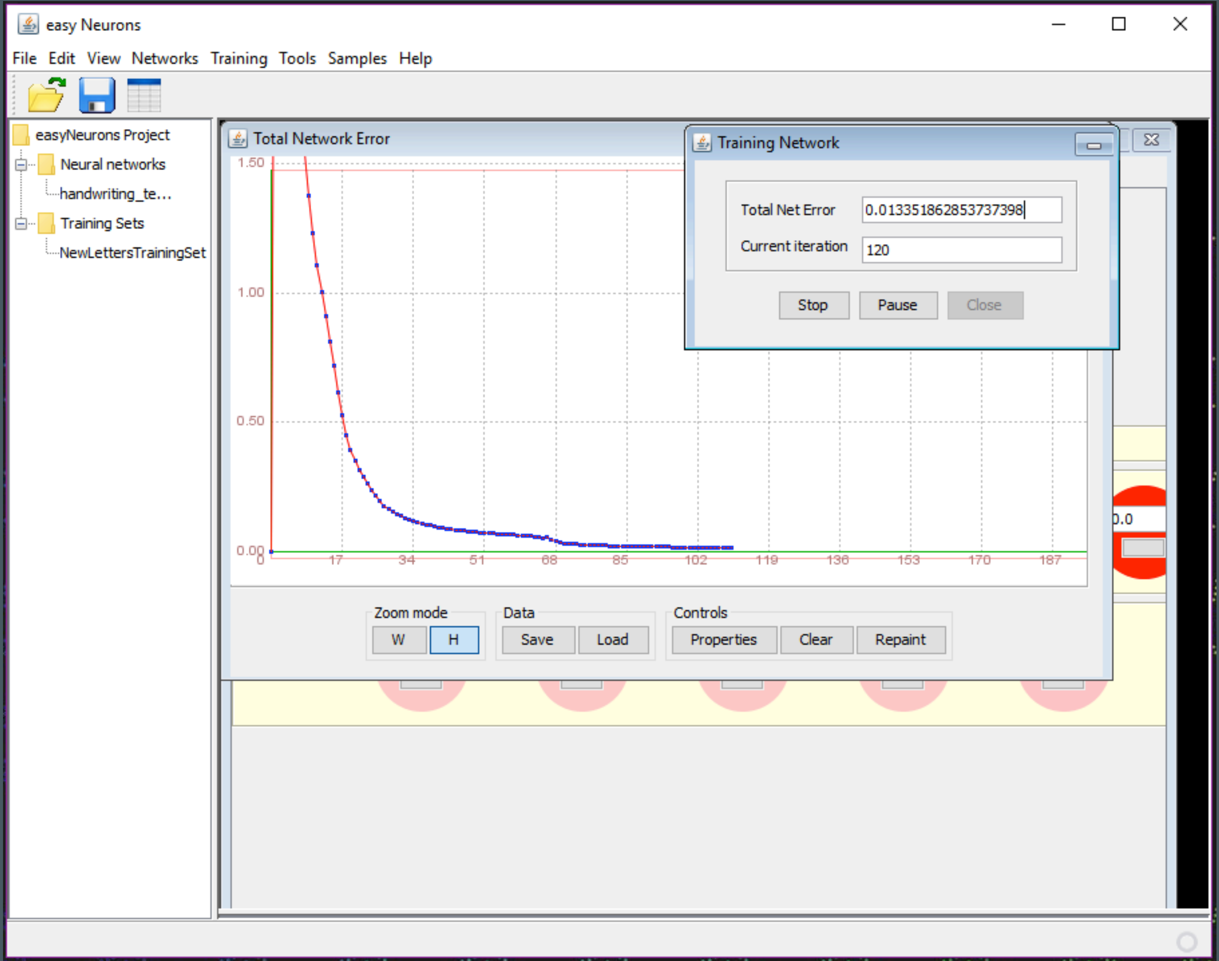
Options

☑ Display Error Graph

(Turn off for faster learning)

Train | Close

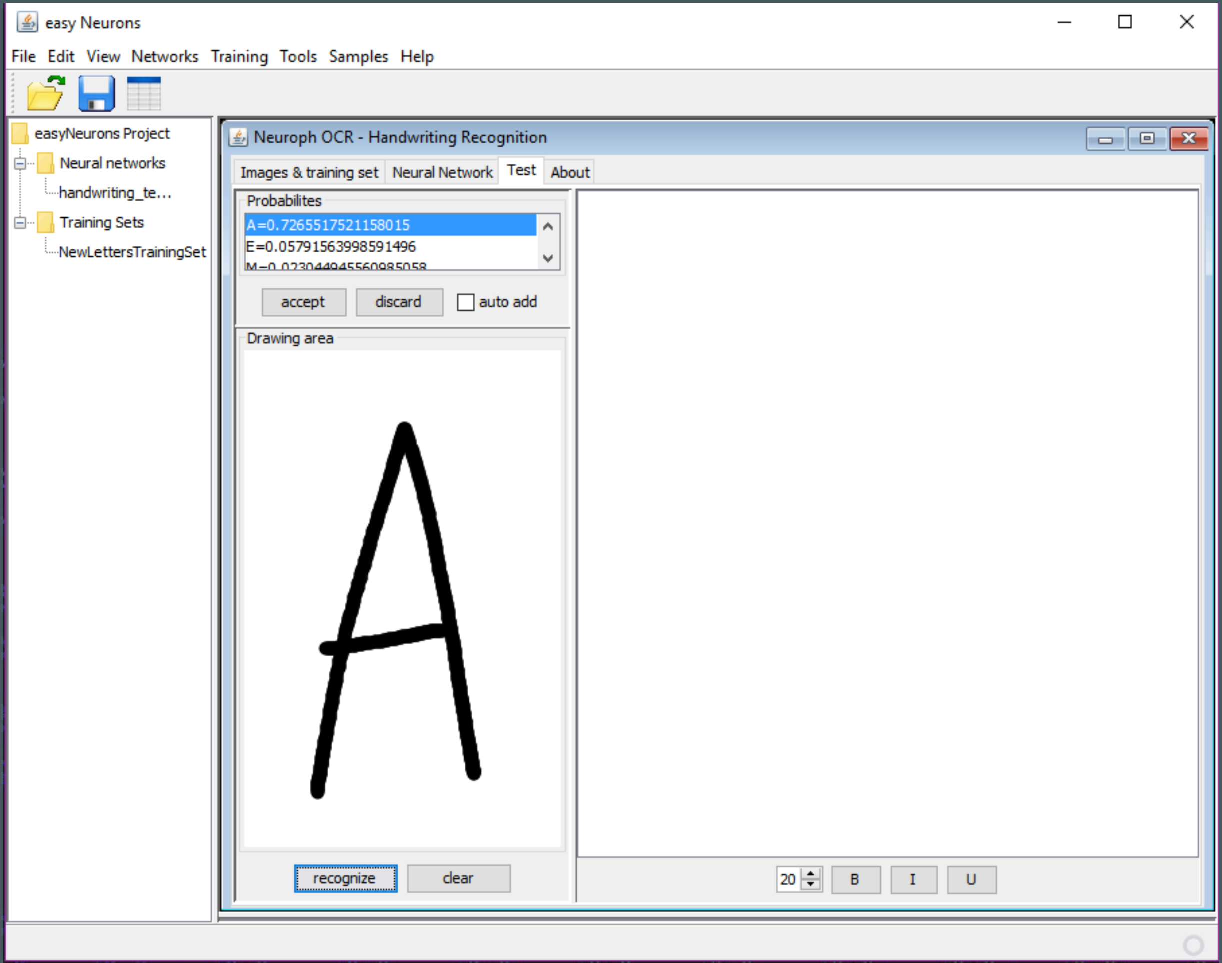
Neuroph Demo



Total Net Error 0.013351862853737398

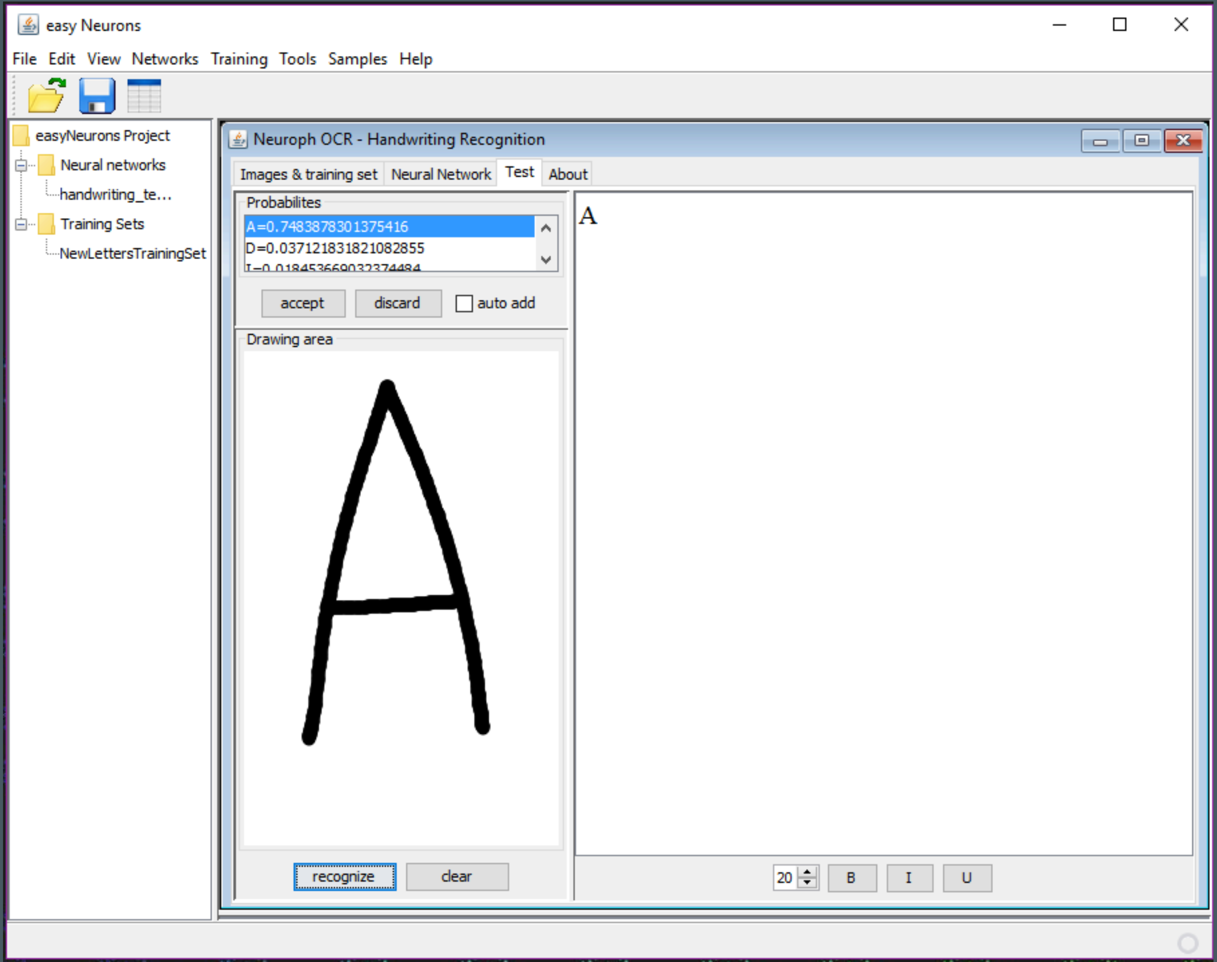
Current iteration 120

Neuroph Demo



A=0.7265517521158015
E=0.05791563998591496
M=0.02304494556098505 8

Neuroph Demo

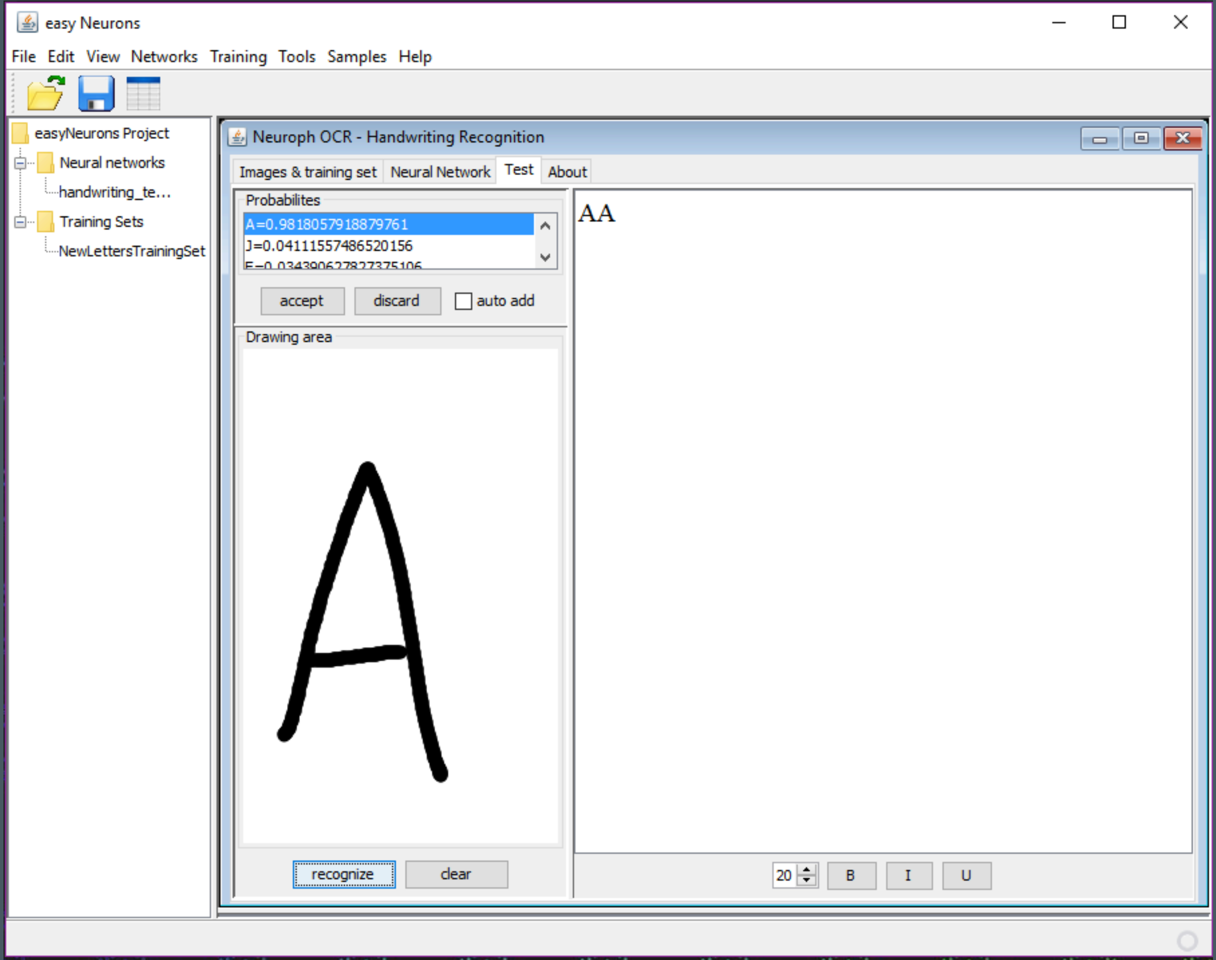


Probabilites

```
A=0.7483878301375416
D=0.037121831821082855
I=0.018453669032374484
```

A

Neuroph Demo



A=0.9818057918879761
J=0.04111557486520156
E=0.034390627827375106

AA

What is Voice Recognition?

Voice recognition also commonly called speech recognition is a Computer Science discipline that focuses on getting computers to translate spoken words into text.

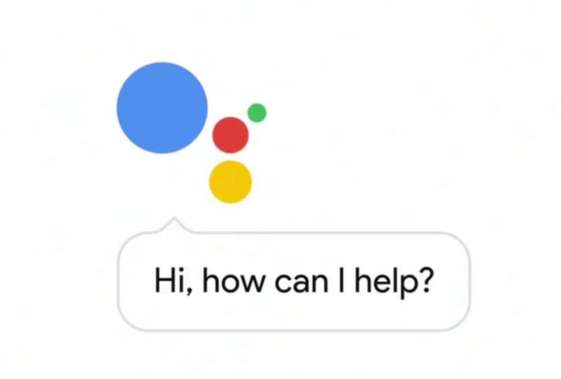
Modern voice recognition processes involve the use of two key methods: **acoustic modeling** and **language modeling**.

Both acoustic and language models take advantage of Neural Networks in order to produce the most accurate results.

What is Voice Recognition?

Some applications of Voice Recognition include

- Intelligent Personal Assistants (Google Assistant, Siri, Cortana, etc…)
- Automated Phone Systems
- Live Subtitling
- Voice Biometrics
- Vehicle Systems
- Video Games (*There Came an Echo*)



Why is it Important?

- Convenience
- Never having to type, makes it easier to send messages or search for something on the go when you don’t have both hands free
- Easier to multitask

Why is it important?

- Language Translations
- Recently, Waverlylabs created earbuds that translate a foreign language to their native tongue as someone speaks
- Bridges the gap in language barriers
- Allows communication between people wit an translator to be actively there.

History of Voice Recognition

-First Speech Recognition developed by Bell Laboratories in 1952

-Only recognized digits spoke by one person

-1970s The Harpy Speech Recognition System was founded by Carnegie Mellon

-Recognized 1011 words, the average word span of a three year old.

-1980s Introduction of the Hidden Markov Model

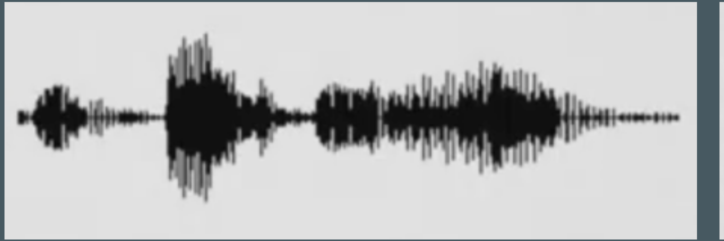
-The main model used in today’s recognition software.

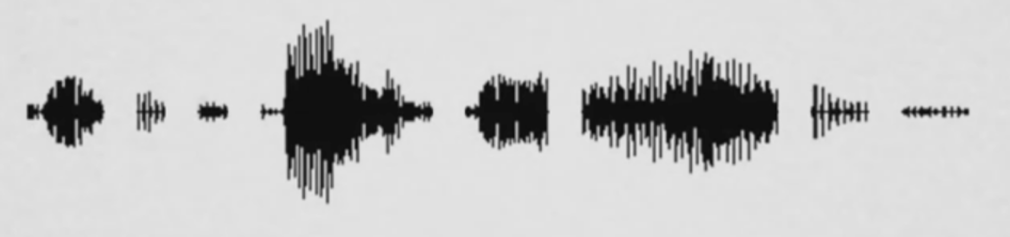
How does Voice Recognition Work

-Raw audio waveform is taken and chopped up into different segments

-From the different segments, we identify which phoneme is spoken in the small segments

-Phoneme is the primitive unit for expressing words(eg: bat = /b/+/ae/+/t/)





How Voice Recognition works cont.

-The possibilities of what the word is

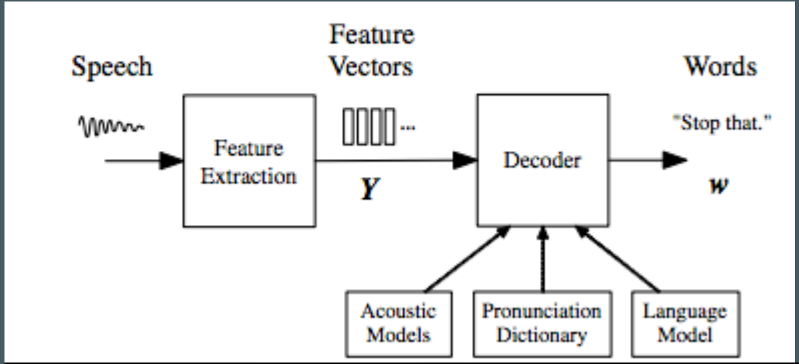
Decoded through Acoustic modeling

-Then the probability of the word is

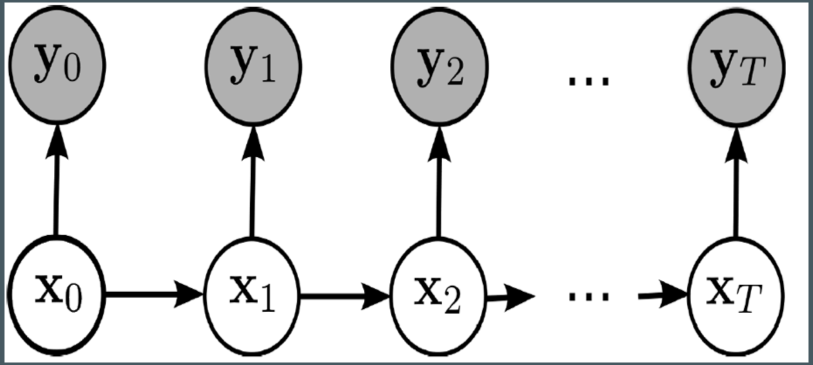
Determined through the language model

-It is like the software is going through

a series of spelling bee questions.



Hidden Markov Model



-Paradigm shift in recognition

Software, made huge advancements

-Still used to this very day in different

Variations or versions.

-Acoustic Model mainly uses HMM

-The Y variables are the observed variables(raw chopped audio segment) and the X variables are the actual phonemes.

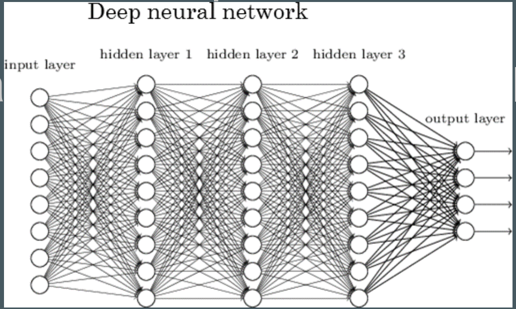
Acoustic Modeling

An acoustic model’s job is to represent the physical sounds produced by a language.

The most effective way to build an acoustic model is by using a Deep Neural Network.

A **Deep Neural Network** is a kind of feedforward neural network, except it contains many hidden layers between the input and output.

To build our Deep Neural Network, we must “tra les of sounds that correspond to given text.



Language Modeling

A language model’s job is to generate a probability distribution for sequences of words in a given language.

In every language, certain words are more likely to appear next to each other than others. For example, “he was billed $100” is a more likely phrase than “he was build $100”.

A language model needs something on which to base its distribution on. Similar to acoustic models, language models are “trained”.

The training of a language model requires feeding large sets of text into a neural network.

Acoustic and Language Modeling

Let’s look at the example on the previous slide: “he was billed $100” vs “he was build $100”.

A well-trained acoustic model will be able to recognize the waveforms that match to the words: he, was, billed/build, one/won, hundred, dollars

The problem that arises is the distinction between **homophones** or words that sound the same (i.e billed/build and one/won).

This is where a language model is necessary because it is able to take a probabilistic approach and say that billed and one are more likely to fit this sentence than build and won.

Review of Literature

"Acoustic Modeling - Microsoft Research." Microsoft Research. N.p., n.d. Web. 31 Oct. 2016.

Albane. "Why Brands Need Image Recognition to Understand Visual Content – The Experts' View." Talkwalker. Talkwalker, 15 Mar. 2016. Web. 29 Oct. 2016.

Chen, Yuyu. "Why Should Marketers Care about Image Recognition Technology?" ClickZ Why Should Marketers Care about Image Recognition Technology Comments. ClickZ, 19 Nov. 2015. Web. 29 Oct. 2016.

Collins, Michael. Language Modeling. N.p.: n.p., n.d. N. pag. Web. 29 Oct. 2016. <http://www.cs.columbia.edu/~mcollins/lm-spring2013.pdf>.

Dent, Steve. "Google's AI Is Getting Really Good at Captioning Photos." Engadget. AOL, 23 Sept. 2016. Web. 29 Oct. 2016.

Dove, Jackie. "CamFind Launches CloudSight API to Advance Visual Search." The Next Web RSS. N.p., 24 Feb. 2015. Web. 29 Oct. 2016.

Flattened RGB values to Neural Network. Digital image. Image Recognition with Neural Networks. Neuroph, n.d. Web. 29 Oct. 2016.

Gao, Z., Y. Zhang, H. Zhang, Y. B. Xue, and G. P. Xu. "Multi-dimensional Human Action Recognition Model Based on Image Set and Group Sparsity."
Neurocomputing 215 (2016): 138-49. ScienceDirect. Web. 29 Oct. 2016.

Review of Literature

Hinton, Geoffrey, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." IEEE Signal Processing Magazine (n.d.): n. pag. Web. 29 Oct. 2016.

"Image Recognition That Actually Works." *CloudSight*. CloudSight, n.d. Web. 29 Oct. 2016.

Mazur, Dominik. "How Businesses Use Image Recognition to Understand Digital Data." ProgrammableWeb. N.p., 21 May 2014. Web. 29 Oct. 2016.