Diabetes Diagnostic Imaging

Machine Learning Undergraduate Research

Walker Christensen & Mitch Maegaard

Problem Statement

Project objective

Company from China

Database of tongue images and personal health questions

Inspiration

Ancient Chinese medicine, doctors could diagnose diabetes by looking at the tongue

Build algorithm for app

User can take a picture of their tongue, answer a few health-related questions, then receive a real-time diabetes diagnosis

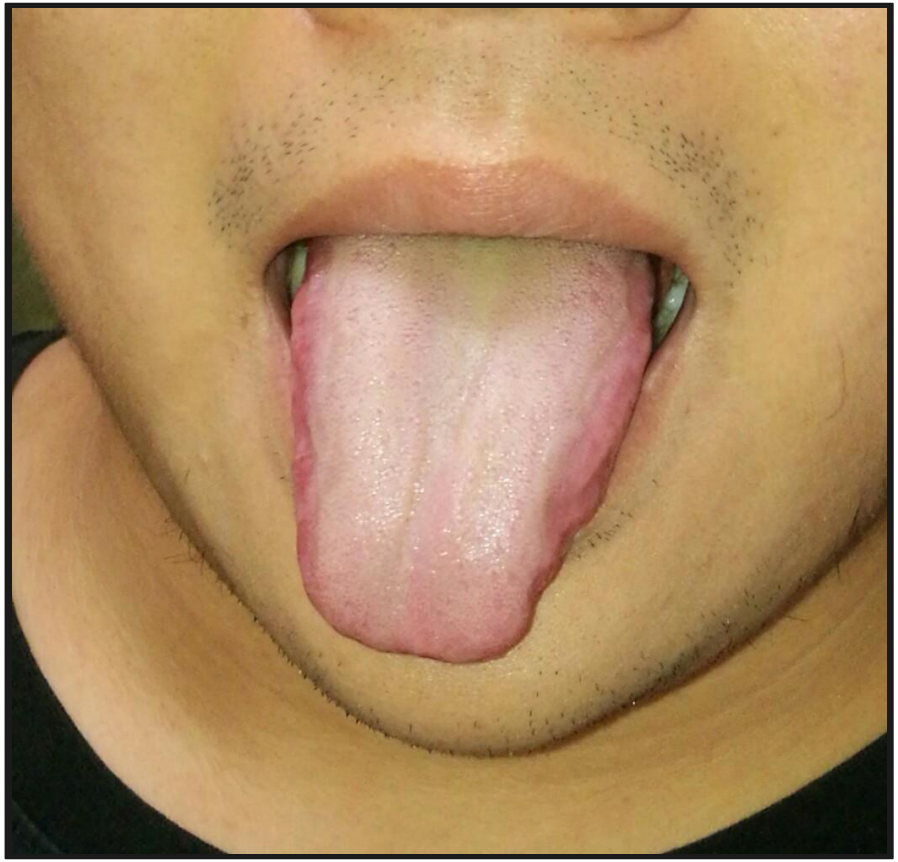
Understanding the problem

Step 1	Can we diagnose *diabetes* using only the **picture** of a tongue?
Step 2	Can we diagnose the *stage* of diabetes with the **picture** of a tongue?
Step 3	Can we diagnose diabetes using **health survey** questions?
Step 4	Can we improve diagnostic accuracy by combining **picture and survey**?

Data Introduction

Images

- 517 healthy
- 224 diabete



Health Survey

- 57 questions
- 164 respondents

Demographics

- Age
- Gender
- Height
- Weight

Questions

- Are you pregnant?
- Do you have unexplained weight loss?
- Do you feel hungry/thirsty?
- Do you have insomnia?

Labels

- Identification Code
- Diabetes Status

Machine Learning Techniques

Image processing

- Images are made up of pixels (a single color)

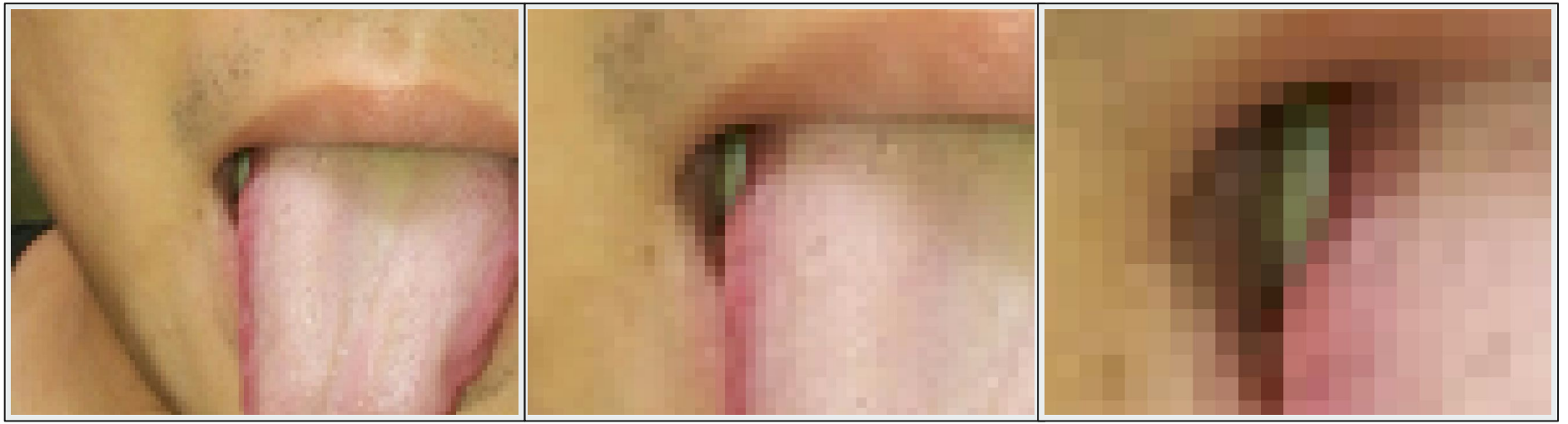


Image processing

- Each pixel has value range:
  0 (black) to 255 (white)
- (5x5) = 25 data points

5x5 grayscale image

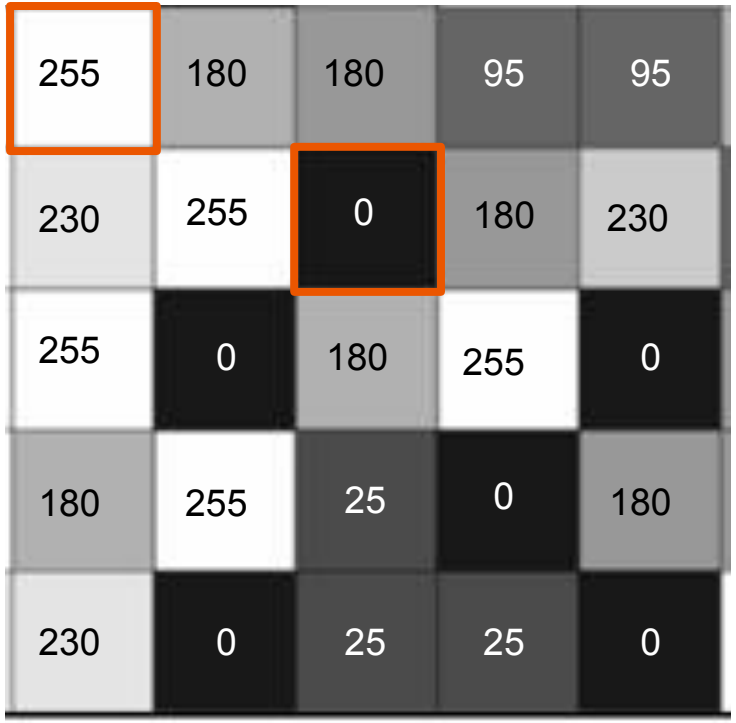


Image processing

- Each pixel has value range: 0 (dark) to 255 (light)
- **R**ed, **G**reen, **B**lue “channels”
- (5x5x3) = 75 data points

5x5 colored image

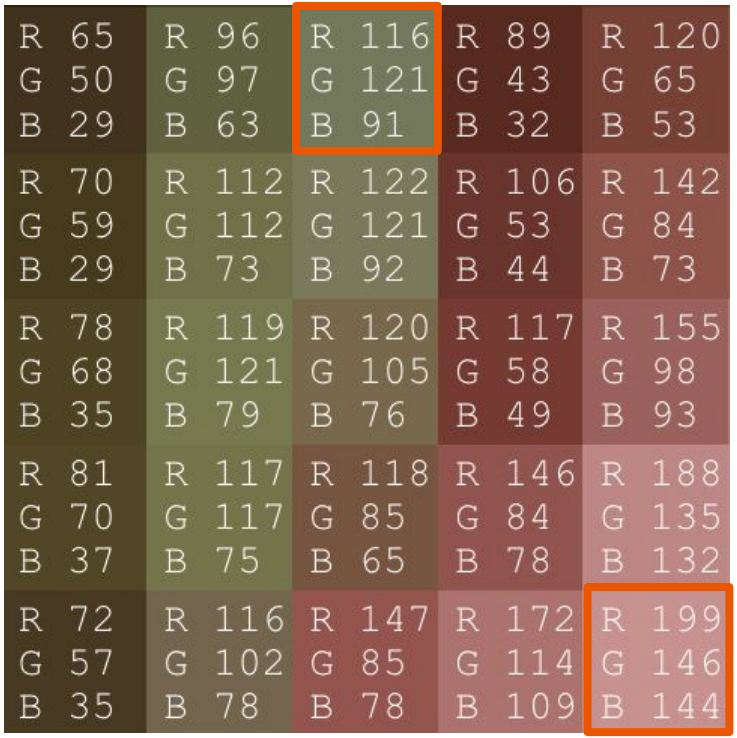


Image processing

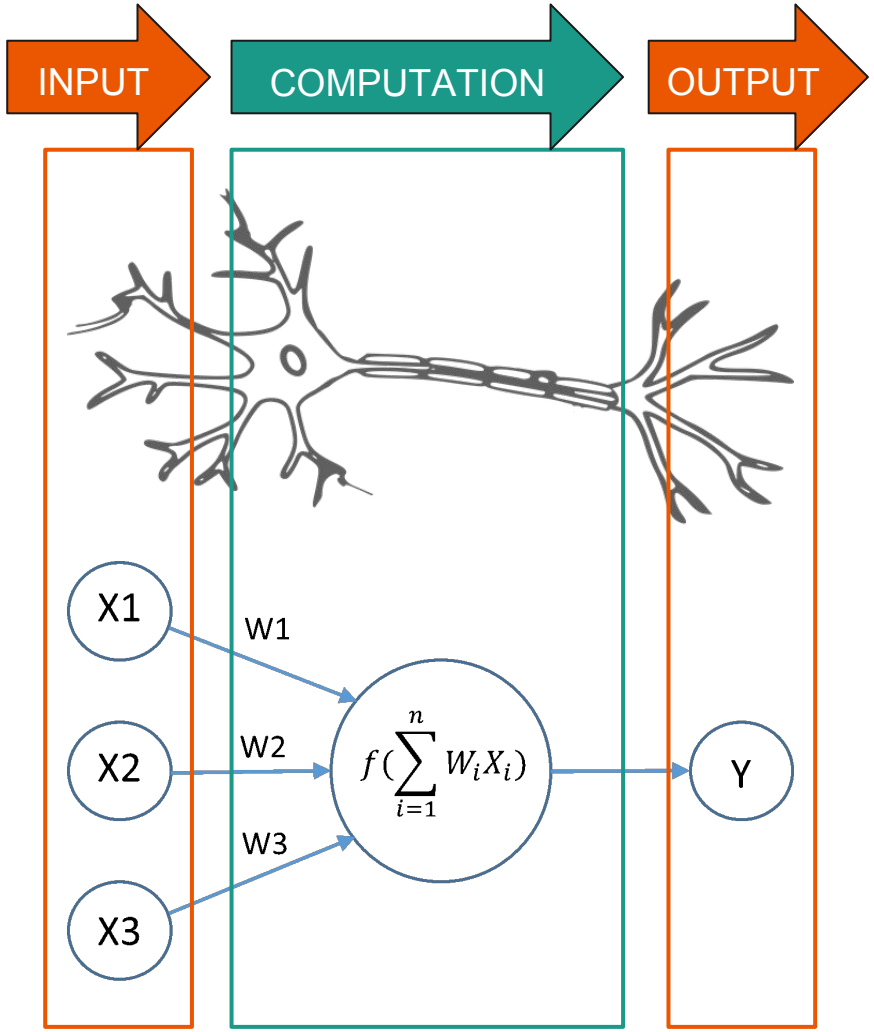
Balance	too *many* pixels vs. too *few* pixels → **128x128** pixel images
Normalize	divide each point by 255 → data range **{0.0, 1.0}**
Apply	128x128x3 = (49,152) x (741 images) → **34.5 million** data points
Algorithm	how do we utilize these numbers? → **Convolutional Neural Network**

Convolutional Neural Network

(CNN, ConvNet)

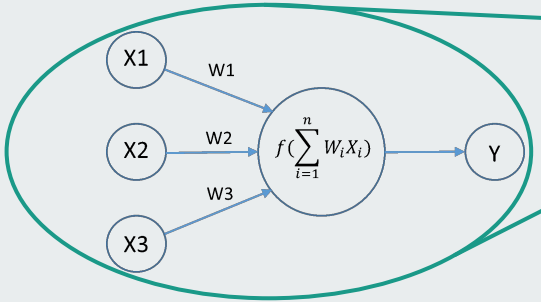
What is a Neural Network?

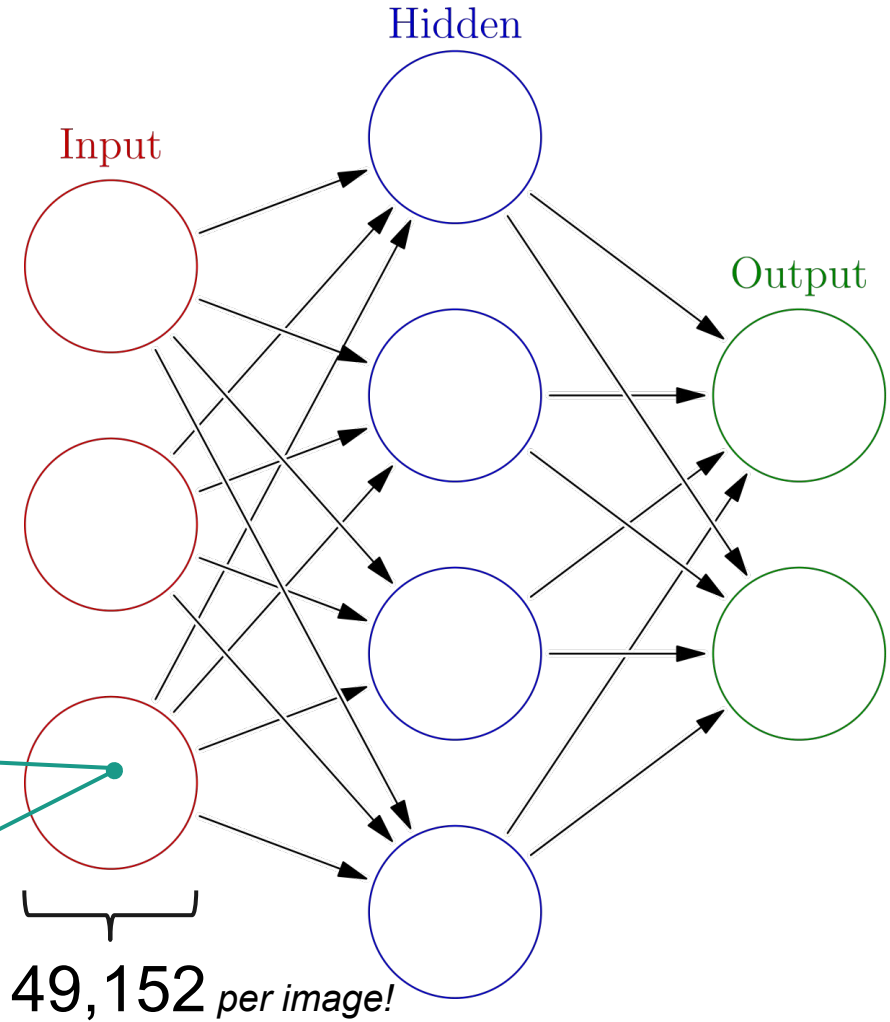
- Want to classify images as diabetic or healthy
- Inspired by neurons in the brain



What is a Neural Network?

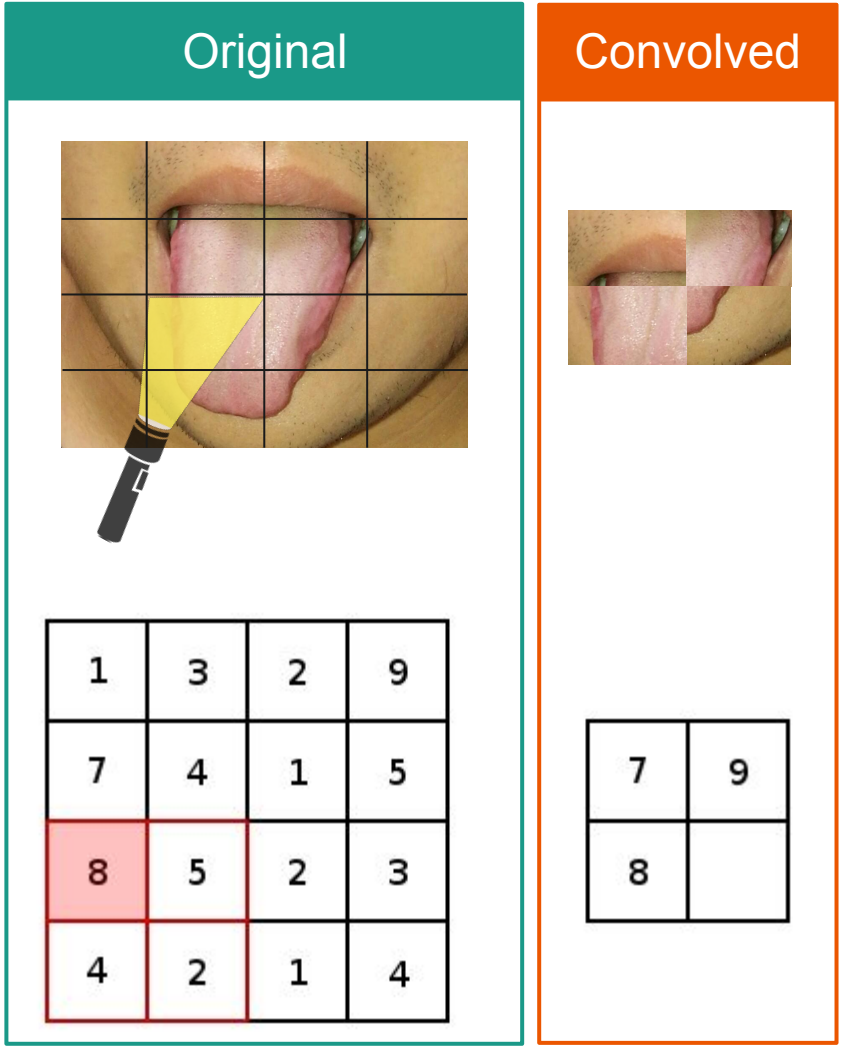
- Neurons working together create a network





ConvNet approach

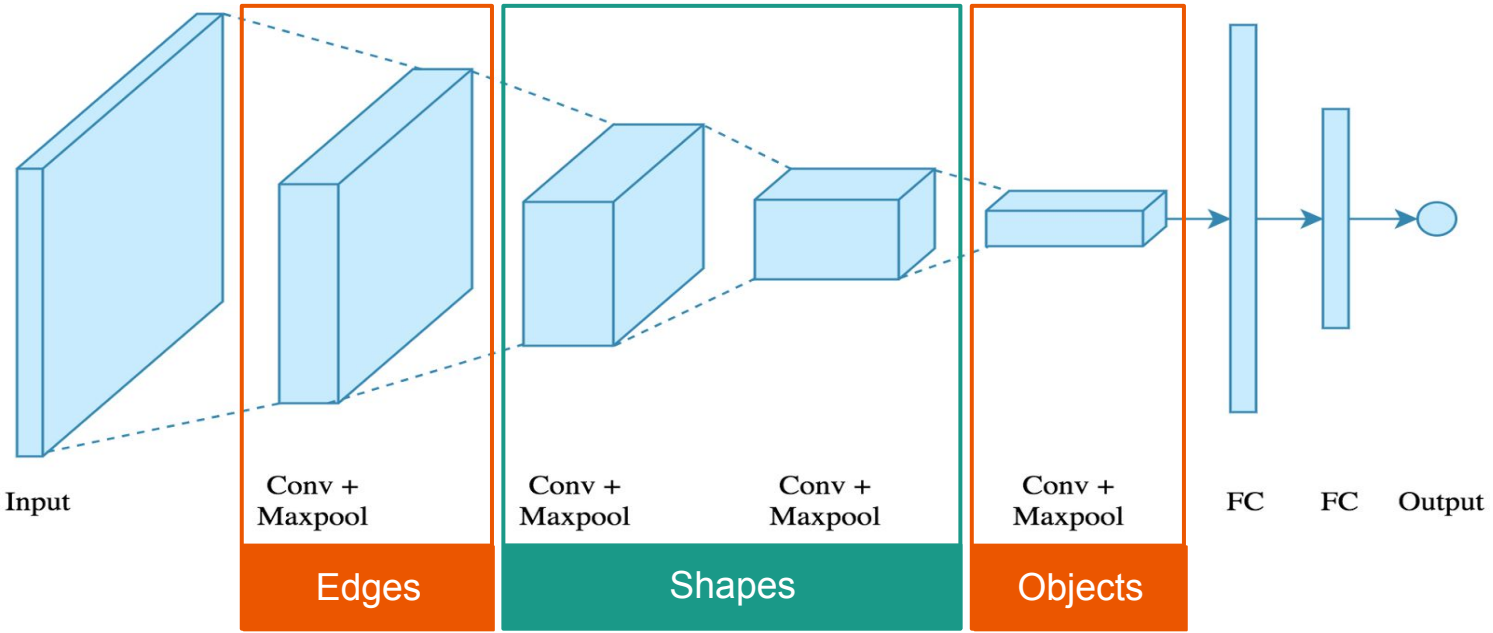
- "Slide" a filter over image
- Output is a convolved image that's **smaller** than the original



ConvNet layers

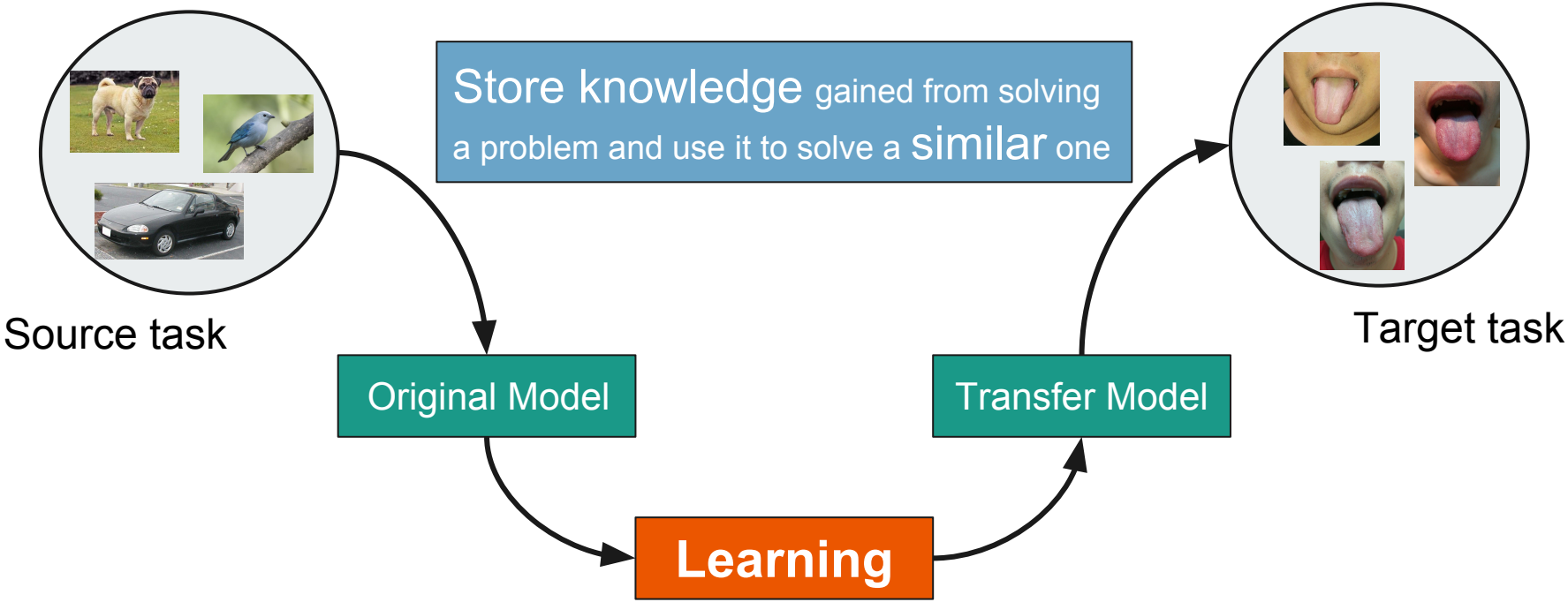
INPUT	raw pixel values of image
CONV	compute dot product between weights and small connected portion in input volume
POOL	downsampling operation along spatial dimensions (width and height)
RELU	applies element-wise activation function
FC	(i.e. fully-connected) computes probability of being in a class

ConvNet architecture



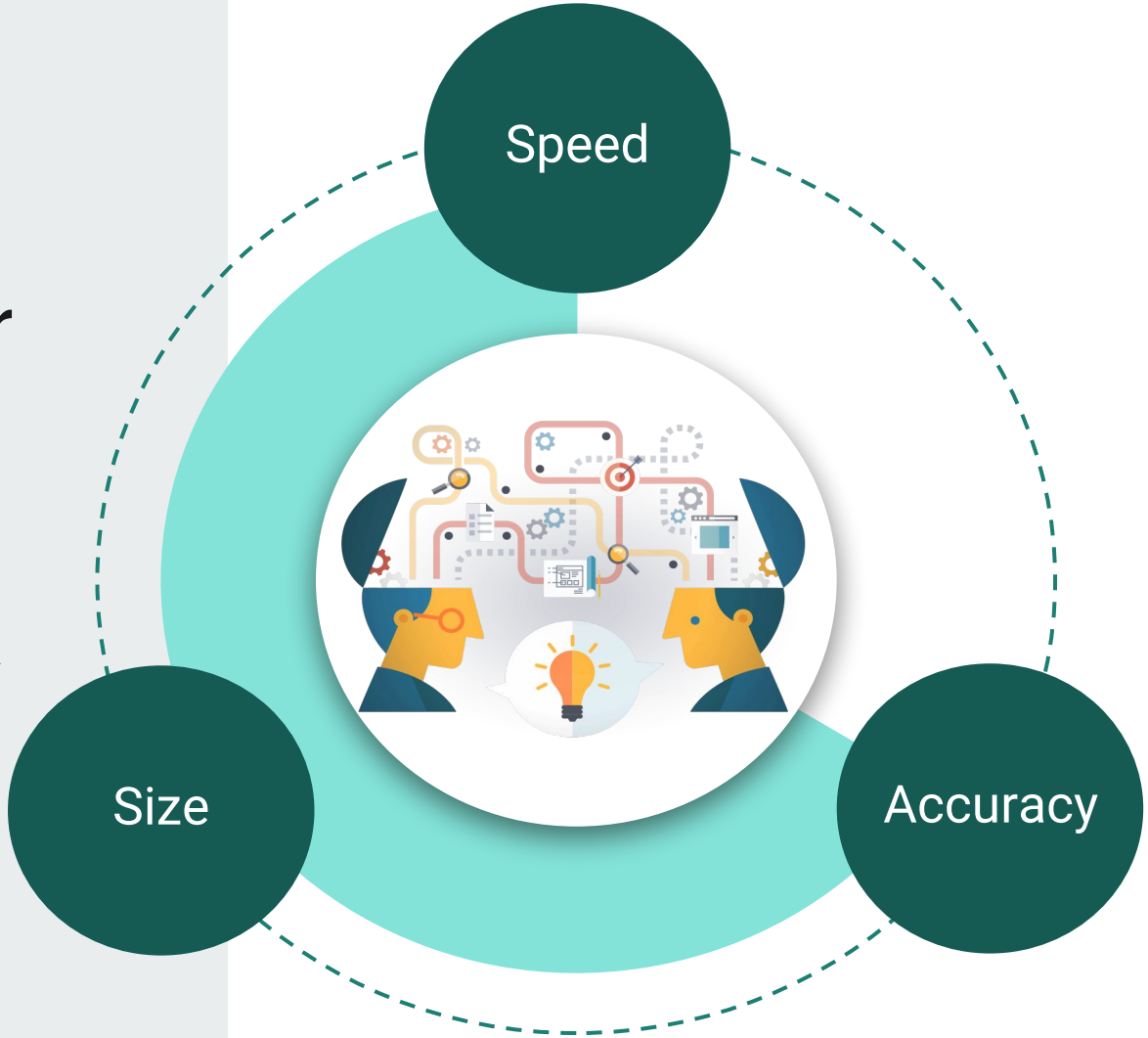
Transfer Learning

What is transfer learning



Why use transfer learning?

- Small dataset
- Similar to ImageNet

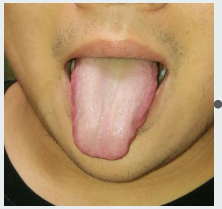


Problem 1

Can we diagnose diabetes with the picture of a tongue?

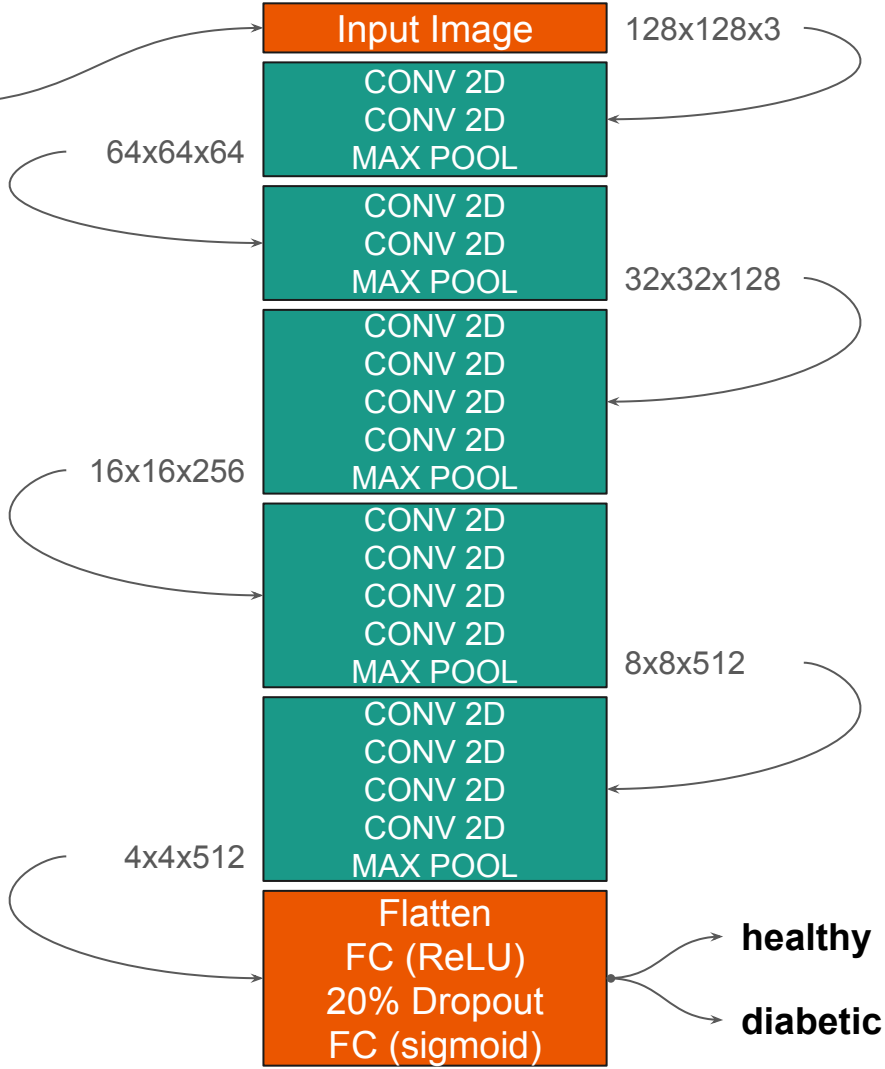
Data preprocessing

Label Images	{ healthy = 0 : diabetic = 1 }
Train Set	497 healthy, 204 diabetic (pull extra samples to create balanced dataset)
Test Set	20 images of each class



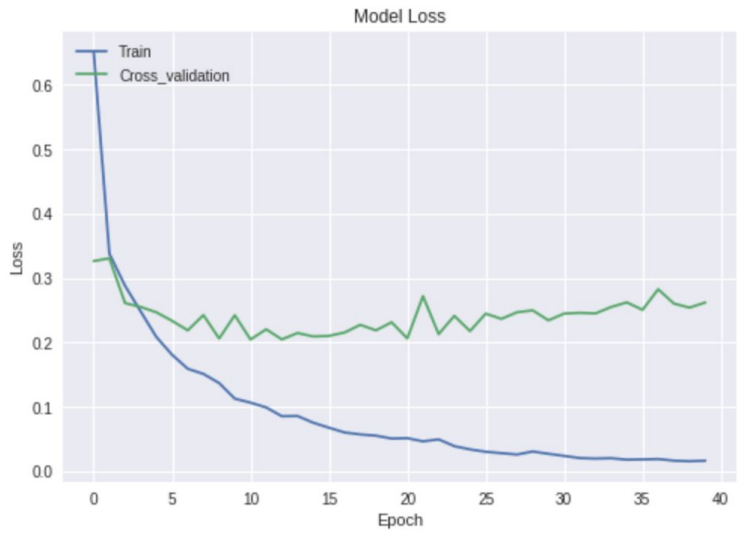
Model architecture

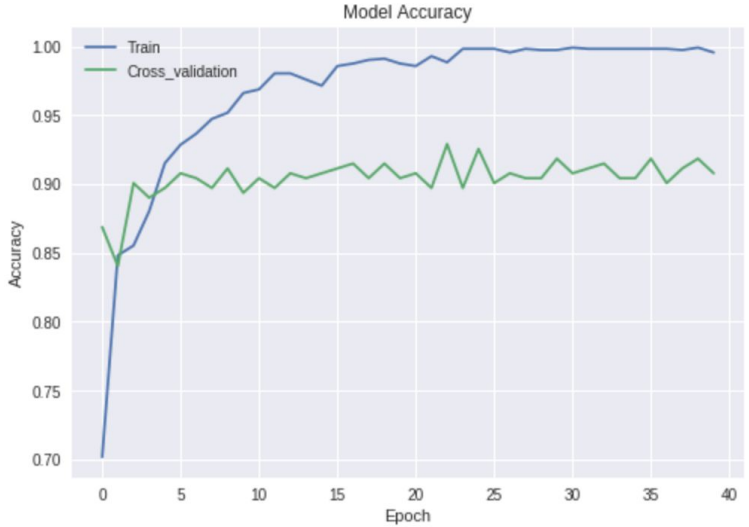
- Input image 128x128x3
- VGG-19 “ImageNet” base model
- Fine-tune top model
  - Flatten
  - 64-unit F.C. & ReLU activation
  - Dropout 20%
  - 2-unit F.C. & sigmoid activation



Training results

- 40 epoch
- 64 mini-batch
- Test accuracy: 87.5%





Hyperparameter tuning

Input Size	Epoch	Mini-Batch	FC 1	Dropout	Accuracy
256x256x3	40	64	64	20%	82.5%
128x128x3	60	64	64	20%	82.5%
128x128x3	40	32	64	20%	87.5%
128x128x3	40	64	32	20%	86.25%
128x128x3	40	64	64	10%	85%
128x128x3	40	64	64	20%	87.5%

Model comparisons

Model	Image Size	Layers	Parameters	Epoch	Mini-Batch	Train Time	Accuracy
Scratch	128x128x3	21	14,731,074	30	64	300 sec.	57.5%
CapsuleNet	128x128x3	9	62,256,096	10	x	224 sec.	62.5%
VGG16 Transfer	128x128x3	21	131,122	30	64	80 sec.	82.5%
VGG19 Transfer	128x128x3	25	524,482	40	64	105 sec.	87.5%

Problem 2

Can we diagnose the *stage* of diabetes?

Multi-class classification

- 5 unique stages of diabetes
  - Healthy
  - Pre-diabetes
  - Mild
  - Moderate
  - Severe

Multi-class classification

Model	Image Size	Layers	Parameters	Epoch	Mini-Batch	Train Time	Accuracy
Random Guess	--	--	--	--	--	--	20%
Multi-Class Transfer	128x128x3	21	125,353	20	64	72 sec.	37%

Problem 3

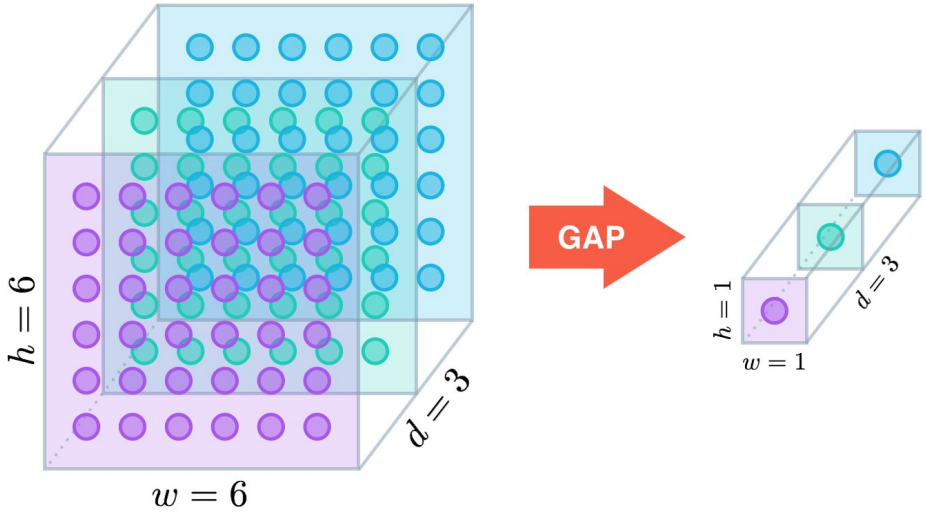
Can we make our results more *interpretable*?

Unboxing the “black box”

Question 1	Which *layers* collect specific feature information?
Question 2	What *parts of the tongue* are contributing to diabetes classifications?
Question 3	Can we find a more interpretable model?

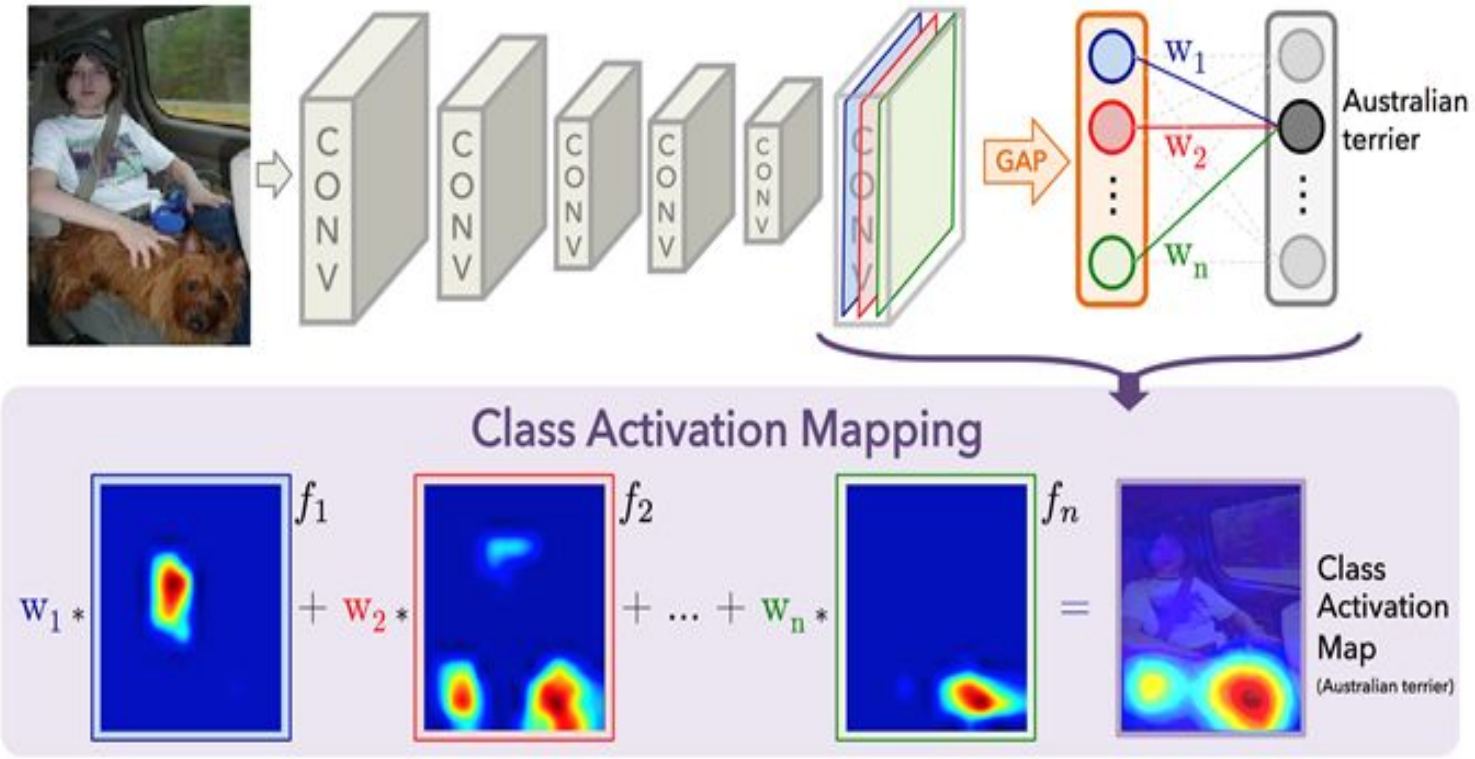
Global average pooling (GAP)

- Map to one prediction per color channel



Grad-CAM (Gradient-weighted Class Activation Mapping)

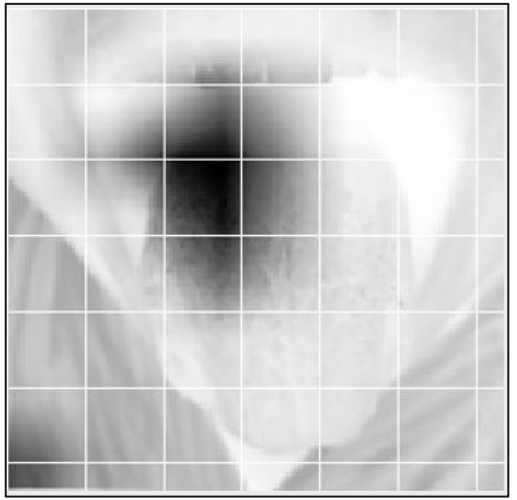
Step 1	Train **CNN** model
Step 2	Extract **class probabilities** from final convolution layer
Step 3	Multiply **feature map** by pooled gradients → 8x8x512
Step 4	**Average** the weighted feature map along channel dimension → 1x512


CONV
CONV
CONV
CONV
CONV
CONV
GAP
w_1
w_2
w_n
Australian terrier
Class Activation Mapping
$w_1 *$
f_1
$+ w_2 *$
f_2
$+ \ldots + w_n *$
f_n
$=$
Class Activation Map
(Australian terrier)

Grad-CAM

Results?

- Activations effectively localize “hotspots” for distinguishing diabetes
- Allows us to present distinguishable features to health experts



Conclusion

Conclusion

I. Binary accuracy: 87.5%

II. Multi-class accuracy: 37%

III. Identified localized areas of tongue images that distinguish diabetes

Future work

Future work

- Filter survey results such that we retain a subset of most important questions
- Extend algorithm to include classification based off survey results
- Apply computer vision techniques to other areas of healthcare