Computational Biology Project

Day 1

Welcome to day 1!

Overview:

- Overview of the computational biology project
- Introduction to computer vision and medical image analysis
- Brief review of neural networks and classification
- Coding activities we will cover this week

Feel free to ask questions whenever they come up!

Icebreaker!

Before we get started, let's first meet one another!

In the icebreaker,

- Introduce yourself
- What's your favourite subject?
- If you could have any superpower, what would it be?

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare

https://www.healtheuropa.eu/ai-technology-identifies-160-possible-new-drug-treatments-for-covid-19/109609/

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare

https://www.news-medical.net/news/20210628/Transfer-learning-technique-achieves-992425-success-when-detecting-COVID-19-in-chest-x-rays.aspx

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare

https://www.theguardian.com/society/2021/jul/01/ai-software-may-help-spot-early-signs-of-oesophageal-cancer

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare, especially in cancer research, **although how is AI doing this**?

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare, especially in cancer research, although how is AI doing this?

The goal of this project is to demystify the field of AI in medicine by exploring how medical image analysis is performed on a cancer dataset

Computational biology project:
Classification and segmentation of medical images

Everyday we hear more about how artificial intelligence is being used to solve problems in healthcare, especially in cancer research, although how is AI doing this?

The goal of this project is to demystify the field of AI in medicine by exploring how medical image analysis is performed on a cancer dataset

For this project, we will be exploring medical image analysis on a dataset of liver tumours [dataset link, original challenge]

- We will be using a modified version of the 3D dataset; our dataset only has 2D images from the 3D CT volumes that had the liver in them

Introduction to computer vision

What is computer vision?

Introduction to computer vision

What is computer vision?

Area of computer science that creates algorithms to automate what our visual system does

Introduction to computer vision

What is computer vision?

Area of computer science that creates algorithms to automate what our visual system does

What are some applications of computer vision?

Introduction to computer vision

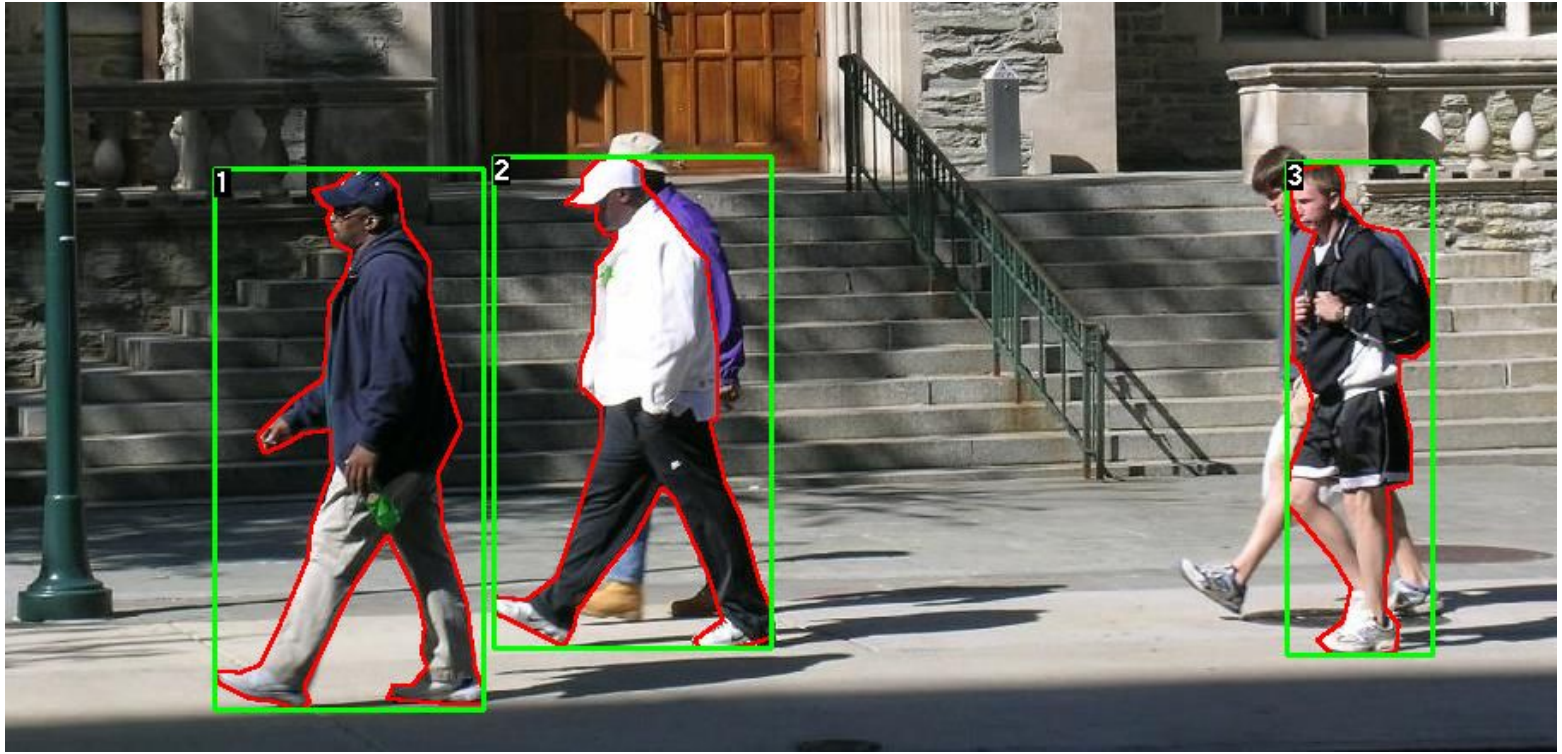
What is computer vision?

Area of computer science that creates algorithms to automate what our visual system does

What are some applications of computer vision?

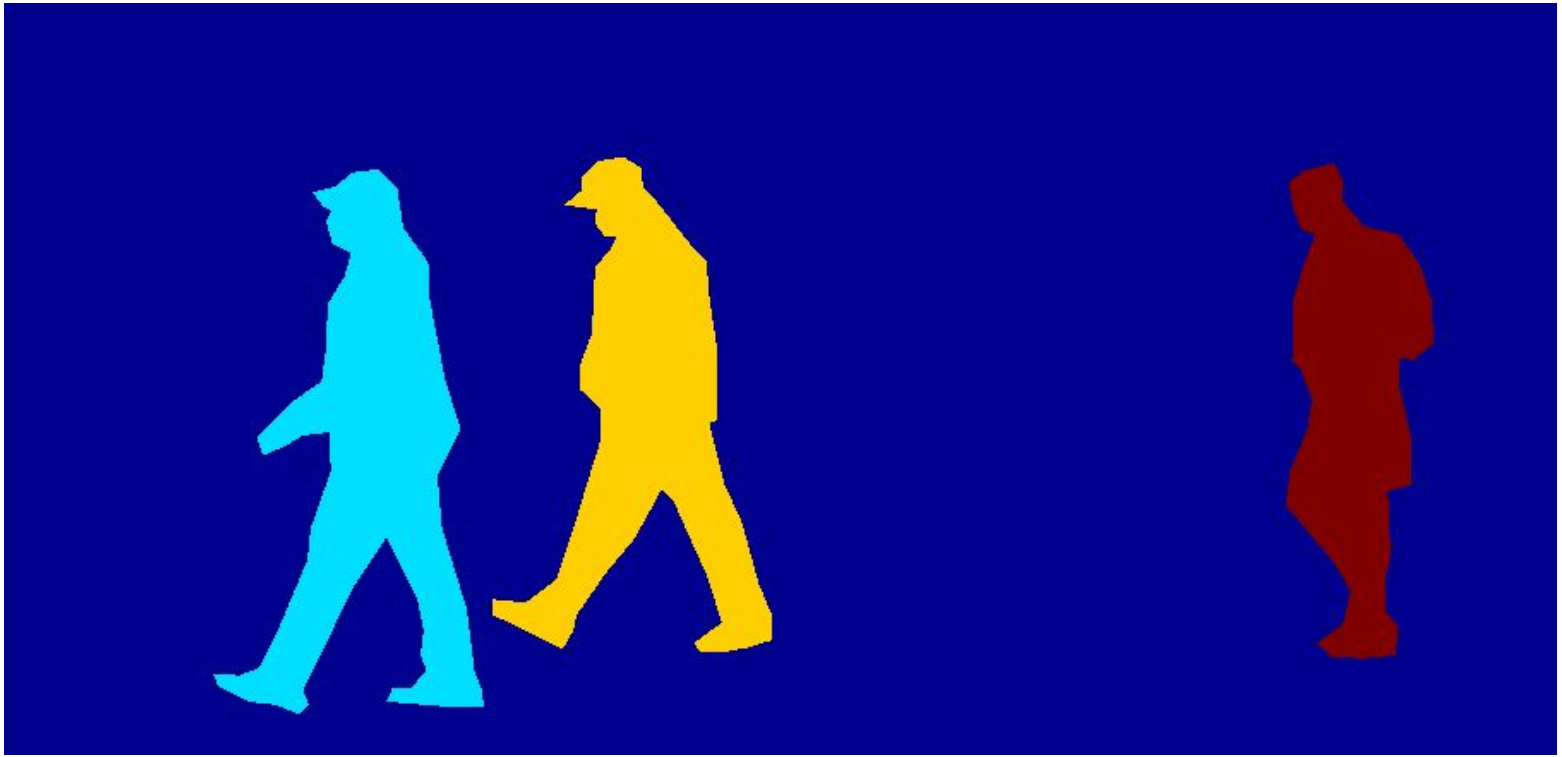
- Object detection
- Object tracking
- 3D scene generation
- Edge detection
- Image interpretation
- Object segmentation

Introduction to computer vision: object detection



https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html

Introduction to computer vision: object segmentation



https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html

Introduction to computer vision

What are some applications of computer vision?

- Object detection
- Object tracking
- 3D scene generation
- Edge detection
- Image interpretation
- Object segmentation

Wait a minute.... what are images?

Introduction to computer vision: defining images

An image is defined by a grid of numbers, where each point in the grid is a pixel

The number of pixels in the image is given by the number of rows and columns in the grid

Introduction to computer vision: defining images

An image is defined by a grid of numbers, where each point in the grid is a pixel

The number of pixels in the image is given by the number of rows and columns in the grid

The value at each pixel represents the light intensity at that point in the grid

Intensity values are in the range 0-255, where 255 represents the maximum intensity and 0 represents no light intensity

255 = white pixel, 0=black pixel, values in between are different shades of gray

Introduction to computer vision: defining images

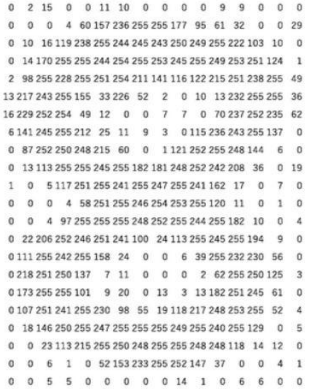
An image is defined by a grid of numbers, where each point in the grid is a pixel

The number of pixels in the image is given by the number of rows and columns in the grid

What we see

vs

What the computer stores



Introduction to computer vision: defining images

An image can be defined by a grid of numbers, where each point in the grid is a pixel

The number of pixels in the image is given by the number of rows and columns in the grid

Black and white (ie. grayscale) images are made up of one grid

Coloured images are made of three grids stacked together (one for the red, blue and green colour “channels”)

```
  0   2  15   0   0  11  10   0   0   0   0   9   9   0   0   0
  0   0   0   4  60 157 236 255 255 177  95  61  32   0   0  29
  0  10  16 119 238 255 244 245 243 250 249 255 222 103  10   0
  0  14 170 255 255 244 254 255 253 245 255 249 253 251 124   1
  2  98 255 228 255 251 254 211 141 116 122 215 251 238 255  49
 13 217 243 255 155  33 226  52   2   0  10  13 232 255 255  36
 16 229 252 254  49  12   0   0   7   7   0  70 237 252 235  62
  6 141 245 255 212  25  11   9   3   0 115 236 243 255 137   0
  0  87 252 250 248 215  60   0   1 121 252 255 248 144   6   0
  0  13 113 255 255 245 255 182 181 248 252 242 208  36   0  19
  1   0   5 117 251 255 241 255 247 255 241 162  17   0   7   0
  0   0   0   4  58 251 255 246 254 253 255 120  11   0   1   0
  0   0   4  97 255 255 255 248 252 255 244 255 182  10   0   4
  0  22 206 252 246 251 241 100  24 113 255 245 255 194   9   0
  0 111 255 242 255 158  24   0   0   6  39 255 232 230  56   0
  0 218 251 250 137   7  11   0   0   0   2  62 255 250 125   3
  0 173 255 255 101   9  20   0  13   3  13 182 251 245  61   0
  0 107 251 241 255 230  98  55  19 118 217 248 253 255  52   4
  0  18 146 250 255 247 255 255 255 249 255 240 255 129   0   5
  0   0  23 113 215 255 250 248 255 255 248 248 118  14  12   0
  0   0   6   1   0  52 153 233 255 252 147  37   0   0   4   1
  0   0   5   5   0   0   0   0   0  14   1   0   6   6   0   0
```

Introduction to computer vision: defining images

Black and white (ie. grayscale) images are made up of one grid

Coloured images are made of three grids stacked together (one for the red, blue and green colour “channels”)

=

Blue

+

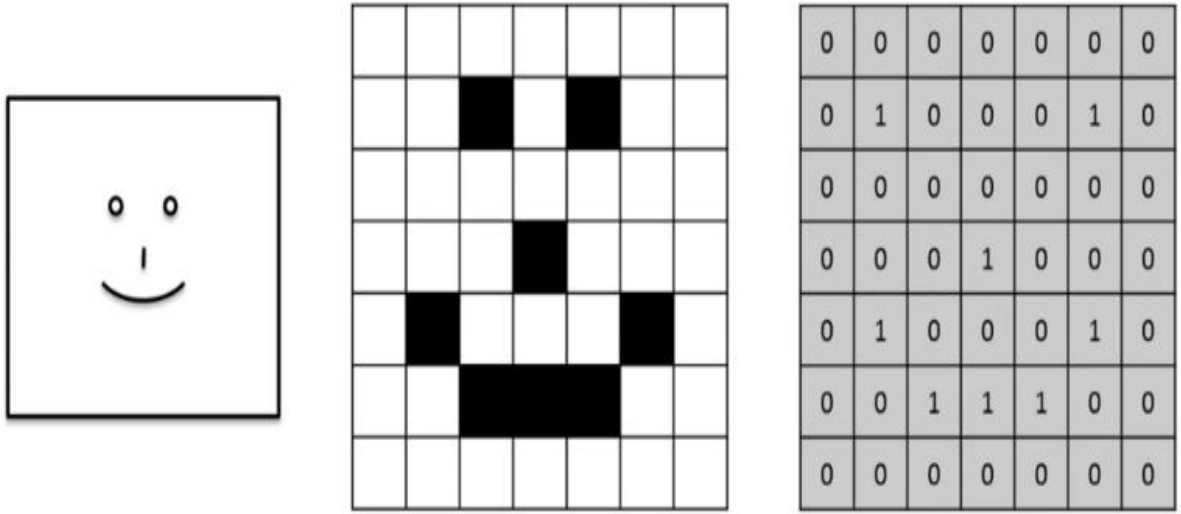
Green

+

Red

Introduction to computer vision: defining images

Binary images are represented by one grid of numbers where the values are either 0 or 1. This representation is very useful in segmentation tasks when trying to define background from foreground.



0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Introduction to computer vision: binary image applications

https://docs.opencv.org/master/d1/dc5/tutorial_background_subtraction.html

Introduction to medical image analysis

What is medical image analysis?

Introduction to medical image analysis

What is medical image analysis?

Area of computer science and computer vision to automatically analyse medical images

Introduction to medical image analysis

What is medical image analysis?

Area of computer science and computer vision to automatically analyse medical images

Many topics within medical image analysis:

Introduction to medical image analysis

What is medical image analysis?

Area of computer science and computer vision to automatically analyse medical images

Many topics within medical image analysis:

- Image classification (ie. image contains healthy anatomy or anatomy with disease?)
- Disease detection
- Disease segmentation
- Image reconstruction
- Explainable AI (ie. field of study to show why and how AI made the decisions it did)

Introduction to medical image analysis: medical images

Different types of medical images obtained from different imaging techniques are called modalities

Different modalities can capture different anatomical information depending on how that imaging technique works

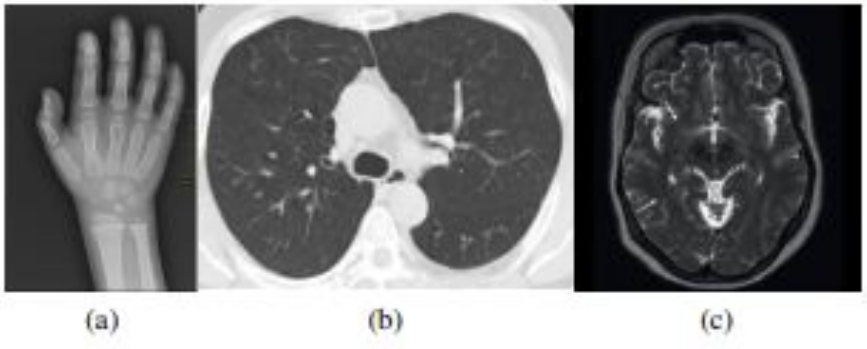


Figure 3 from https://arxiv.org/pdf/2012.02364.pdf

Introduction to medical image analysis: medical images

Different types of medical images obtained from different imaging techniques are called modalities

Different modalities can capture different anatomical information depending on how that imaging technique works

The values at the pixels of the image can represent anatomical information

- ie. in CT scans, the values at the pixels are in Hounsfield units (HU) and different ranges of HU values represent different types of anatomy such as soft-tissue, bone, air, etc

Day 1 project: exploring classification of liver images

Using our subset of data of liver images, we are going to implement a Monai classification workflow

In this workflow, we will train a model to classify images as healthy or unhealthy by learning patterns among the images of both healthy and unhealthy liver images

Day 1 project: exploring classification of liver images

Using our subset of data of liver images, we are going to implement a Monai classification workflow

In this workflow, we will train a model to classify images as healthy or unhealthy by learning patterns among the images of both healthy and unhealthy liver images

The classification will be performed using neural networks

Review of neural networks

What are neural networks?

Review of neural networks

What are neural networks?

Layers of neurons that build a network, trained to learn patterns in data

Review of neural networks

What are neural networks?

Layers of neurons that build a network, trained to learn patterns in data

Training a neural network (in a supervised fashion) consists of:

- Feeding in example inputs to the model

Review of neural networks

What are neural networks?

Layers of neurons that build a network, trained to learn patterns in data

Training a neural network (in a supervised fashion) consists of:

- Feeding in example inputs to the model
- Calculating the difference (ie. the loss) between the model's prediction and the expected output corresponding to the input

Review of neural networks

What are neural networks?

Layers of neurons that build a network, trained to learn patterns in data

Training a neural network (in a supervised fashion) consists of:

- Feeding in example inputs to the model
- Calculating the difference (ie. the loss) between the model's prediction and the expected output corresponding to the input
- Backpropagate the loss throughout the network, to determine how to adjust the weights of the network to minimize this error

Review of neural networks

What are neural networks?

Layers of neurons that build a network, trained to learn patterns in data

Training a neural network (in a supervised fashion) consists of:

- Feeding in example inputs to the model
- Calculating the difference (ie. the loss) between the model's prediction and the expected output corresponding to the input
- Backpropagate the loss throughout the network, to determine how to adjust the weights of the network to minimize this error
- Update the weights of the network and continue learning

Training a network for a classification task

The network learns to encode key features (ie key information) from the image by a sequence of operations that reduce it’s size to a vector of features

Training a network for a classification task

The network learns to encode key features (ie key information) from the image by a sequence of operations that reduce it’s size to a vector of features

From the training procedure, this learned feature vector should have key local and whole image information

Training a network for a classification task

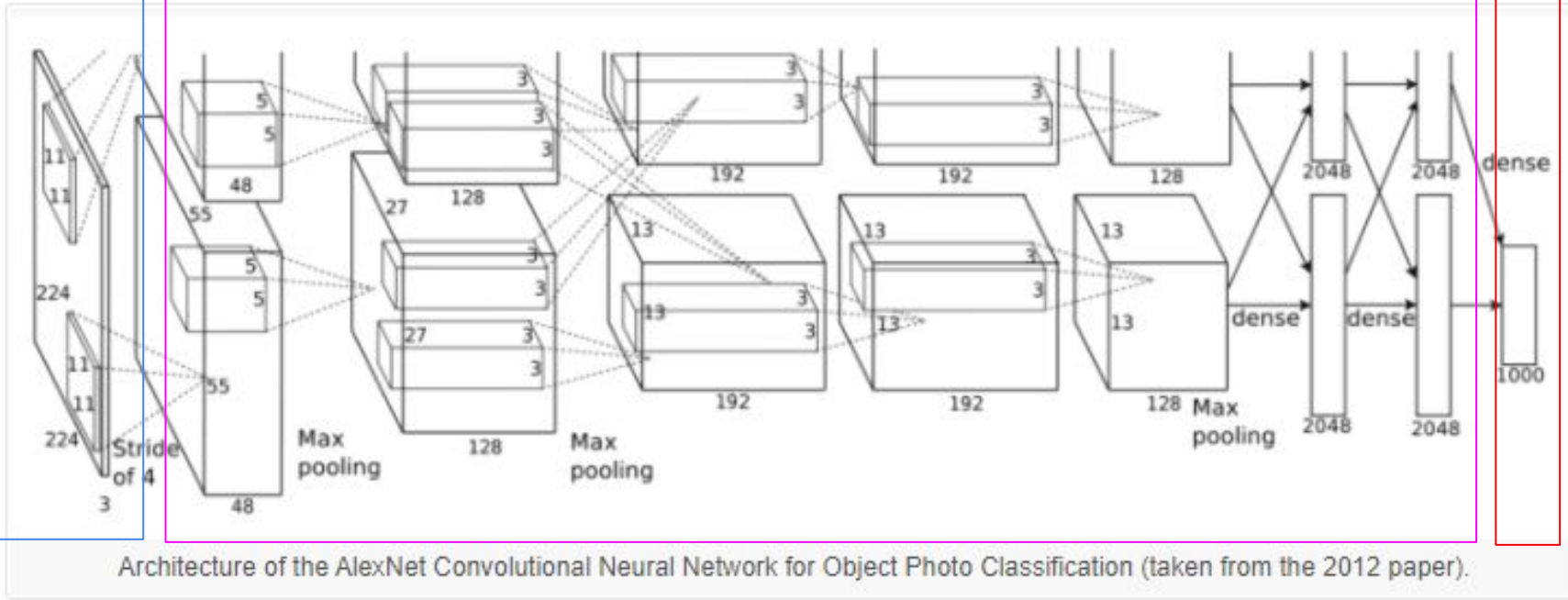
The network learns to encode key features (ie key information) from the image by a sequence of operations that reduce it's size to a vector of features

From the training procedure, this learned feature vector should have key local and whole image information

The model then uses this feature vector to predict whether the image is healthy or not

AlexNet classification example from Machine learning mastery

Input Image Series of operations to extract key features from the input image Feature vector



Architecture of the AlexNet Convolutional Neural Network for Object Photo Classification (taken from the 2012 paper).

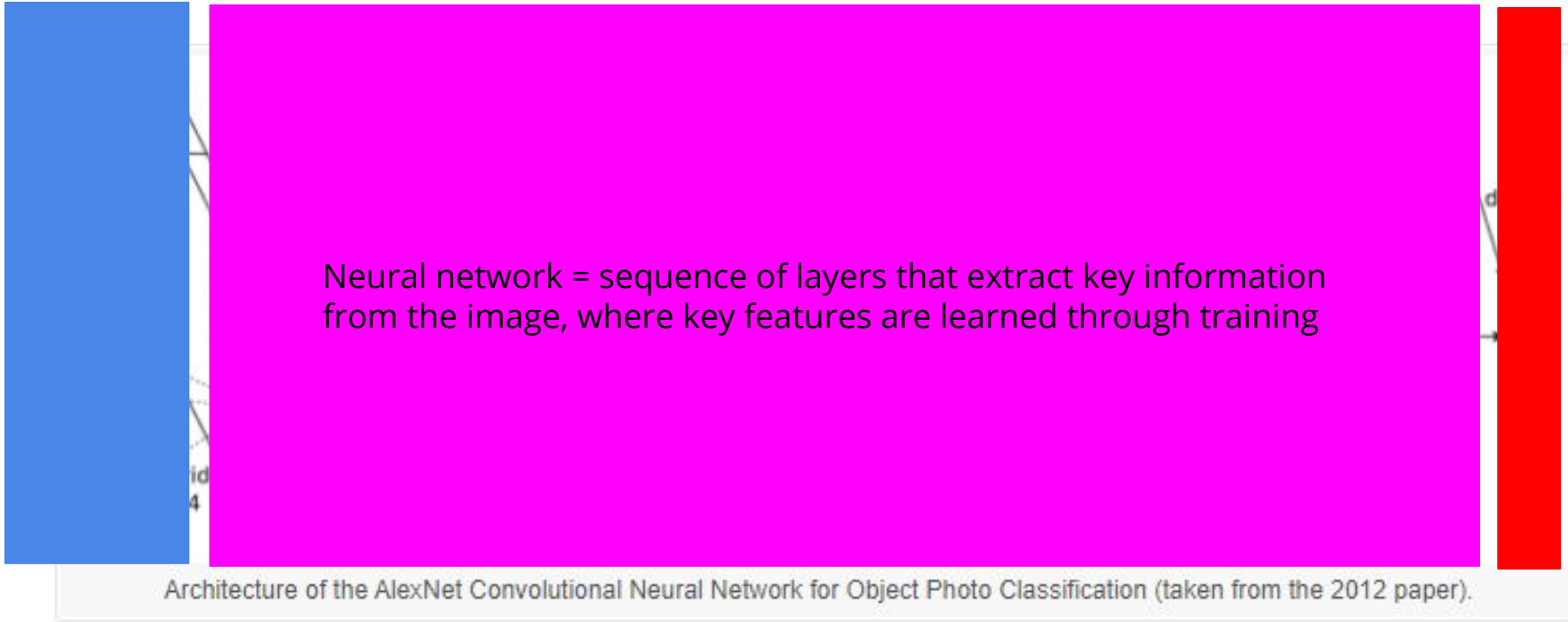
https://machinelearningmastery.com/review-of-architectural-innovations-for-convolutional-neural-networks-for-image-classification/

AlexNet classification example from Machine learning mastery

Input Image

Series of operations to extract key features from the input image

Feature vector



Architecture of the AlexNet Convolutional Neural Network for Object Photo Classification (taken from the 2012 paper).

https://machinelearningmastery.com/review-of-architectural-innovations-for-convolutional-neural-networks-for-image-classification/

Evaluating a classification network's performance

Can evaluate the network based on accuracy:

$$accuracy = \frac{correct\ classifications}{total\ classifications}$$

Or using a confusion matrix of the following values:

	Predicted yes	Predicted no
Actually yes	True positives	False negatives
Actually no	False positives	True negatives

Evaluating classification network's performance

Can evaluate the network based on accuracy:

$$accuracy = \frac{correct\ classifications}{total\ classifications}$$

Or using a confusion matrix of the following values:

	Predicted yes	Predicted no
Actually yes	True positives	False negatives
Actually no	False positives	True negatives

MONAI: Medical Open Network for AI

New Python library that offers many different workflows for medical image analysis tasks.

They offer excellent tutorials and use public datasets so that anyone can reproduce their results in Google colab.

For their documentation: https://docs.monai.io/en/latest/

For their wide variety of tutorials: https://github.com/Project-MONAI/tutorials

We will leverage this library today and tomorrow to explore classification and segmentation on our liver dataset.

Week overview:

- To demystify AI in medical imaging by exploring classification and segmentation tasks on a dataset of liver cancer
- Dataset: A 2D modified version of a 3D Liver tumour dataset (original dataset)
- Data visualization, train/test classification and segmentation workflow using Monai library, write UNet from scratch, experiment with improving models through data augmentation and apply evaluation techniques.

Day 1 - Intro and classification workflow

- Introduction to computer vision and it's application to medical images, medical image analysis (MIA), and overview of common tasks in MIA (image classification, object detection, object segmentation)
- Visualization of the liver cancer dataset
  - Activity 1: create a function to display examples of healthy and tumour images
- Experiment with the Monai classification workflow
  - Activity 2: fill in the blanks of the classification model training loop
  - Advanced activity 3: Experiment with improving the results through different hyperparameters, data augmentation

Day 1 - Intro and classification workflow

- Introduction to computer vision and it's application to medical images, medical image analysis (MIA), and overview of common tasks in MIA (image classification, object detection, object segmentation)
- Visualization of the liver cancer dataset
  - Activity 1: create a function to display examples of healthy and tumour images
- Experiment with the Monai classification workflow
  - Activity 2: fill in the blanks of the classification model training loop
  - Advanced activity 3: Experiment with improving the results through different hyperparameters, data augmentation

Day 1 - Intro and classification workflow

- Introduction to computer vision and it's application to medical images, medical image analysis (MIA), and overview of common tasks in MIA (image classification, object detection, object segmentation)
- Visualization of the liver cancer dataset
  - Activity 1: create a function to display examples of healthy and tumour images
- Experiment with the Monai classification workflow
  - Activity 2: fill in the blanks of the classification model training loop
  - Advanced activity 3: Experiment with improving the results through different hyperparameters, data augmentation

Day 2 - Segmentation workflow

- Visualization of the liver cancer dataset
  - Activity 1: Display examples of tumors images and their segmentations from the dataset
- Experiment with the Monai segmentation workflow:
  - Activity 2: fill in the blanks of the training loop
  - Activity 3: add a validation step to the training loop to save the best model
- Evaluate the segmentation model:
  - Activity 4: Display the test results ie. test images, ground truth segmentation, model output)
  - Activity 5: Calculate dice score

Day 2 - Segmentation workflow

- Visualization of the liver cancer dataset
  - Activity 1: Display examples of tumors images and their segmentations from the dataset
- Experiment with the Monai segmentation workflow:
  - Activity 2: fill in the blanks of the training loop
  - Activity 3: add a validation step to the training loop to save the best model
- Evaluate the segmentation model:
  - Activity 4: Display the test results ie. test images, ground truth segmentation, model output)
  - Activity 5: Calculate dice score

Day 2 - Segmentation workflow

- Visualization of the liver cancer dataset
  - Activity 1: Display examples of tumors images and their segmentations from the dataset
- Experiment with the Monai segmentation workflow:
  - Activity 2: fill in the blanks of the training loop
  - Activity 3: add a validation step to the training loop to save the best model
- Evaluate the segmentation model:
  - Activity 4: Display the test results ie. test images, ground truth segmentation, model output)
  - Activity 5: Calculate dice score

Day 3 - Writing UNet from scratch

- Understand the architecture of UNet and its layers
  - Activity 1: Build the model, train the first iteration
- Build variations of the UNet model
  - Activity 2: Experiment with convolution blocks, re-train, and plot learning curves
- Evaluate the models to get preliminary results
  - Activity 3: Use the model weights to predict segmentation maps

Day 3 - Writing UNet from scratch

- Understand the architecture of UNet and its layers
  - Activity 1: Build the model, train the first iteration
- Build variations of the UNet model
  - Activity 2: Experiment with convolution blocks, re-train, and plot learning curves
- Evaluate the models to get preliminary results
  - Activity 3: Use the model weights to predict segmentation maps

Day 3 - Writing UNet from scratch

- Understand the architecture of UNet and its layers
  - Activity 1: Build the model, train the first iteration
- Build variations of the UNet model
  - Activity 2: Experiment with convolution blocks, re-train, and plot learning curves
- Evaluate the models to get preliminary results
  - Activity 3: Use the model weights to predict segmentation maps

Day 4 - Improve performance of previous models

- Techniques to improve the performance of the model
  - Activity 1: Hyperparameters
  - Activity 2: Data Augmentation
- Compare validation scores to choose the best model
- Evaluate on test dataset to secure final results
  - Activity 3 & 4: Dice score and Binary Cross Entropy loss

Day 4 - Improve performance of previous models

- Techniques to improve the performance of the model
  - Activity 1: Hyperparameters
  - Activity 2: Data Augmentation
- Compare validation scores to choose the best model
- Evaluate on test dataset to secure final results
  - Activity 3 & 4: Dice score and Binary Cross Entropy loss

Day 4 - Improve performance of previous models

- Techniques to improve the performance of the model
  - Activity 1: Hyperparameters
  - Activity 2: Data Augmentation
- Compare validation scores to choose the best model
- Evaluate on test dataset to secure final results
  - Activity 3 & 4: Dice score and Binary Cross Entropy loss

Today's plan

- During mentor led project time, we will open up the day 1 notebook together and go through downloading the data, the Monai library, and a few visualization activities for the dataset as a group.

- During the self-directed project time, you will work with your team (communicating over slack or zoom link) to complete the activities for day 1.

- Work through the activities at your own pace. The last activity will be tuning the network to improve results.