COMPUTER VISION:
A Deep Learning Approach

by

Dr. Andrew Katumba, Makerere University

8th May 2023

MARCONI
LAB @MAK

Outline

Part I

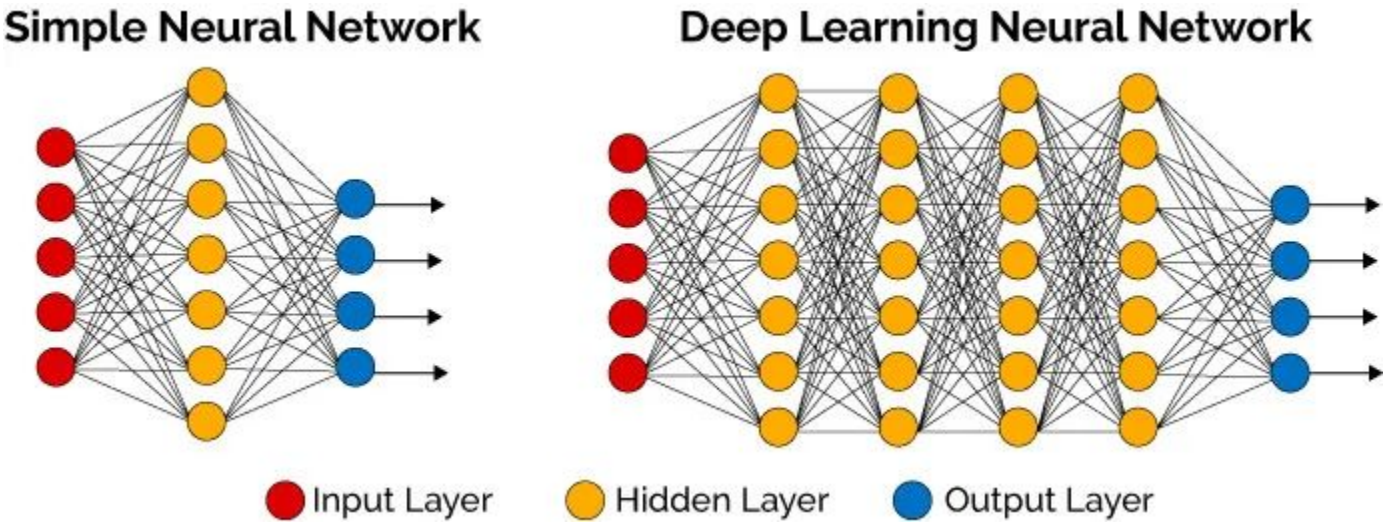
- Overview
- Why DL?
- Development of Deep
- Intro to Computer Vision with Deep Learning
- Computer Vision tasks (with applications)
  - Classification
  - Detection
  - Segmentation

Part II

Hands-on preparation

Overview: What is Deep Learning?

A machine learning technique where computers learn by example.



Consist of multiple layers of interconnected nodes.

Source: https://www.pipionshop.com/

Difference from Machine Learning? (Overview cont'd)

Representation Learning: Automatic discovery of the features needed for classification or identification.

- Automatic feature extraction.

Why does it matter? (Overview cont’d)

Higher accuracy than ever before on a broad range of tasks.

.

Why deep learning now?

- Better algorithms & understanding
- Computing power (GPUs, TPUs, ...)
- Data with labels
- Open source tools and models

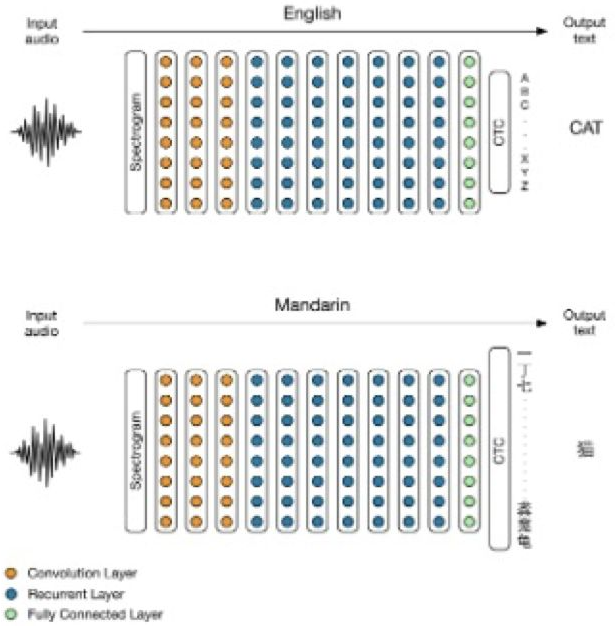
Source: https://www.nvidia.com/

Source: https://cloud.google.com/

Applications (Overview)

- Self driving cars.
- Medical research, e.g detection of cancerous cells.
- Speech processing.
- Industrial automation.
- etc.

DL Today: Speech-to-Text



[Baidu 2014]

DL Today: Vision

[Krizhevsky 2012]

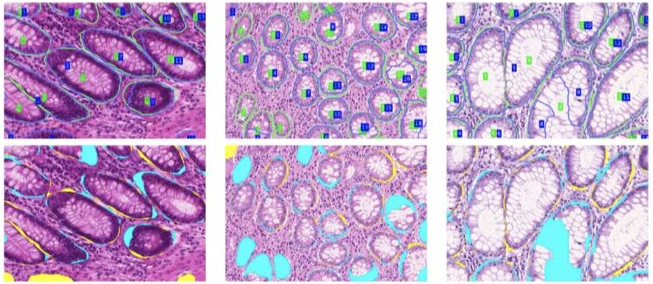
[Ciresan et al. 2013]

[Faster R-CNN - Ren 2015]

[NVIDIA dev blog]

DL Today: Vision

[Stanford 2017]



[Nvidia Dev Blog 2017]

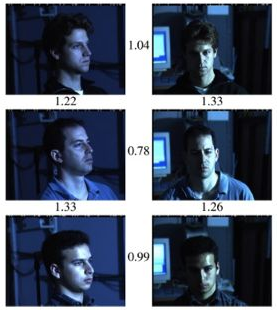
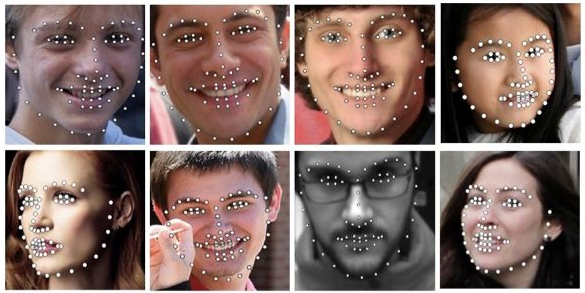


Figure 1. **Illumination and Pose invariance.**

[FaceNet - Google 2015]



[Facial landmark detection CUHK 2014]

DL Today: Generative models

Sampled celebrities [Nvidia 2017]

StackGAN v2 [Zhang 2017]

How does it work?

- Supervised Learning*: labelled data
- Methodology
  - Large Dataset
  - Model
  - Train and test

Deep learning models are trained on large datasets from which which they automatically learn features

Dataset

A **dataset** is always required to train a model. A model learns from the features in the dataset.

- Split into train, (val), test.
- Train dataset is labelled (Supervised learning).
- Test dataset, labels are removed.

Model

An **algorithm** that learns from training data. Choice of deep learning model depends on nature of the dataset e.g

- For images, use Convolutional Neural Networks (transformers too??)

Neural Networks

Key Concepts

- Neuron
- Activation Functions

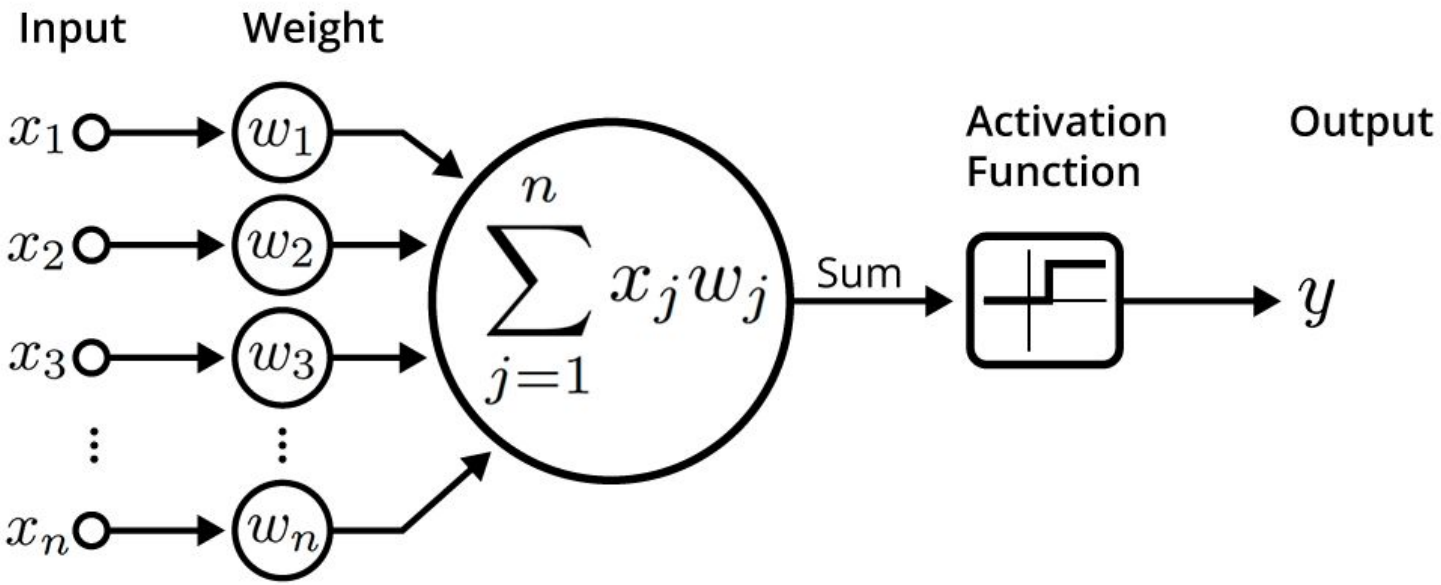
Training a neural network

- Gradient Descent
- Backpropagation
- Overfitting and dropout
- Transfer Learning

Convolutional Neural Networks and Recurrent Neural Networks

The Neuron

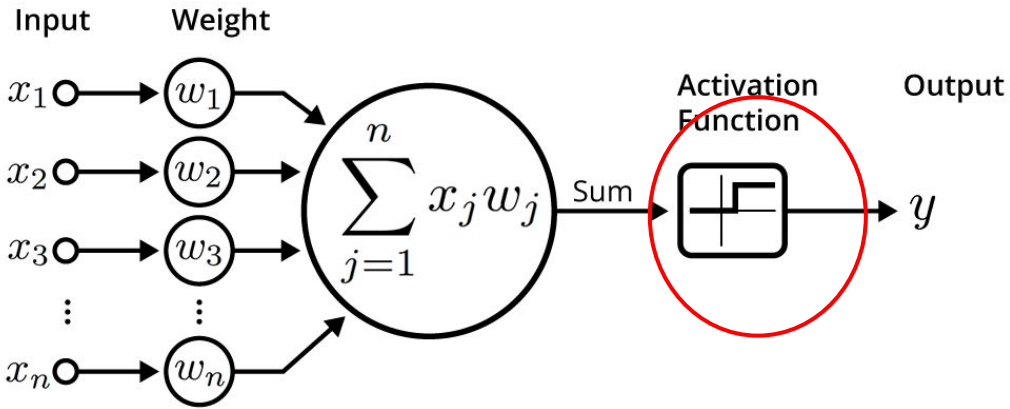
- The basic unit of a Neural Network.



An illustration of an artificial neuron. Source: Becoming Human.

Activation Function

- Perceptron/Threshold Function
- Sigmoid Function
- Rectifier Function
- Hyperbolic Function

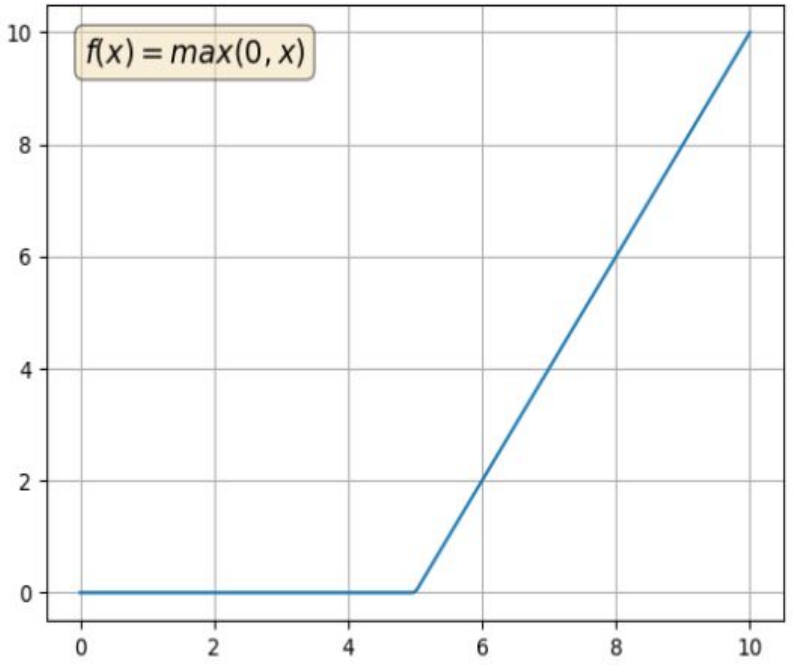


An illustration of an artificial neuron. Source: Becoming Human.

Rectifier Function (RELU)

Thresholded at zero.

Most popular non-linear function at the moment.



Training a Neural Network

A neural network model is trained using supervised learning.

- **Forward pass:** data is provided to the model to obtain an output.
- **Backward pass**: adjust the weights of the nodes through backpropagation.

Learning of a neural network simply means **adjusting the weights of the neurons until the error is minimized**

Gradient Descent

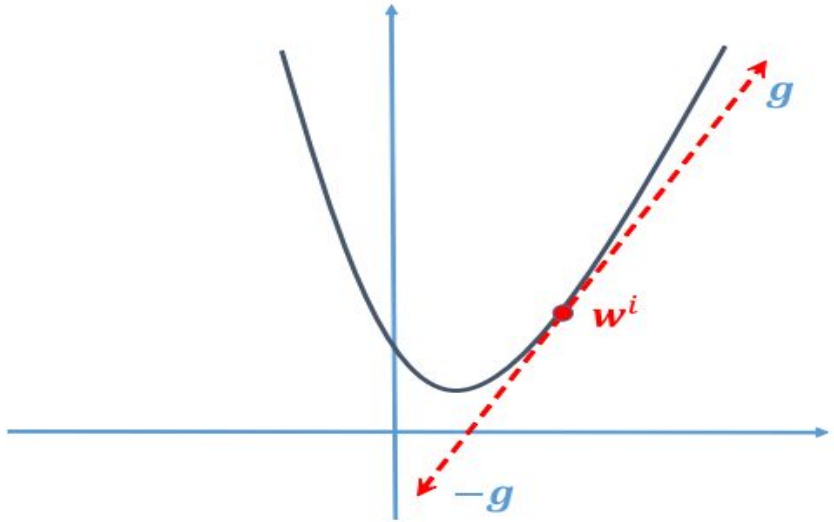
Update parameters in the d**irection opposite to the sign of the gradient**.

$$\theta = \theta - \eta \nabla J(\theta)$$

η- learning rate - controls how fast or slow the model learns

$J(\theta)$- loss function w.r.t θ.

Stochastic Gradient Descent computes gradient and updates values using a single example.



Source: Introduction to Convolutional Neural Networks
https://cs.nju.edu.cn/wujx/paper/CNN.pdf

Issues

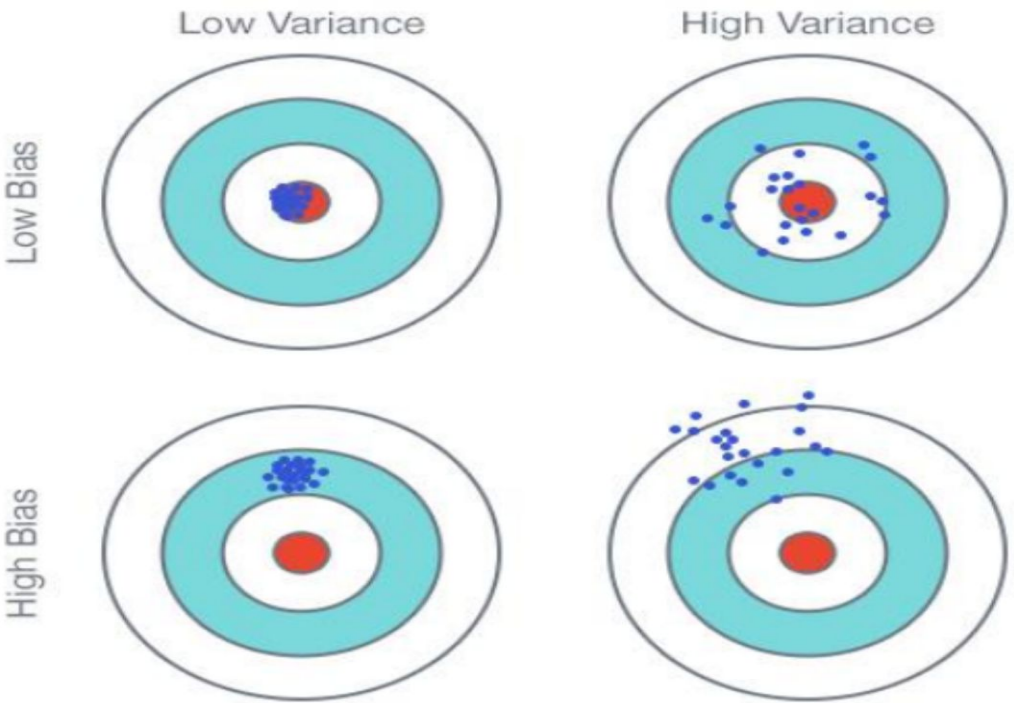
- Overfitting

- Underfitting

Issues

Overfitting - Low bias, high variance

Underfitting - High bias, Low variance



Early Stopping, Dropout

Early Stopping

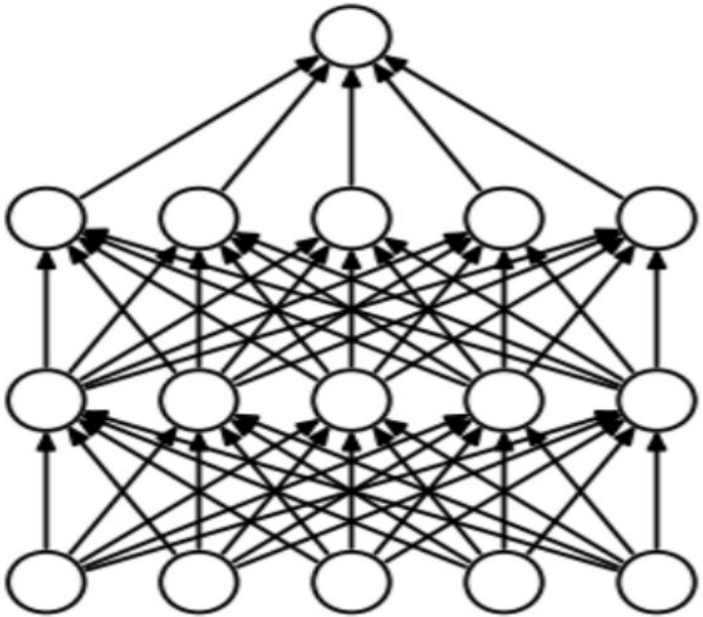
- Monitor a model’s validation error and stop as soon as it’s constant.

Randomly drop some neurons along with their connections from the network.

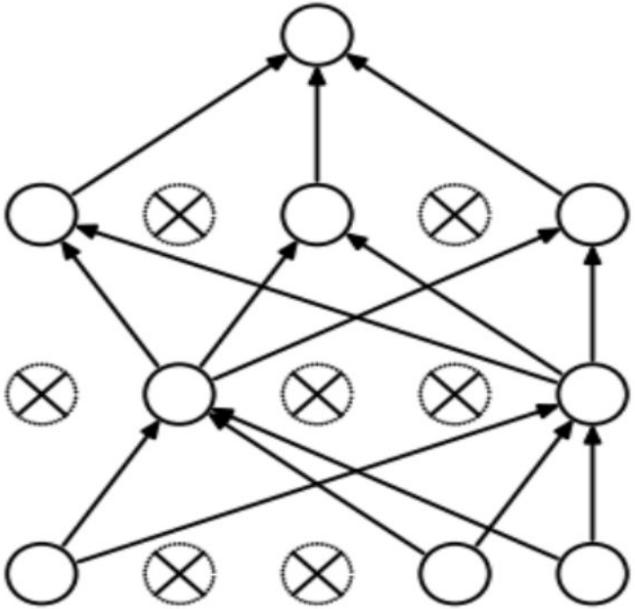
- To prevent over-fitting

Dropout Illustration

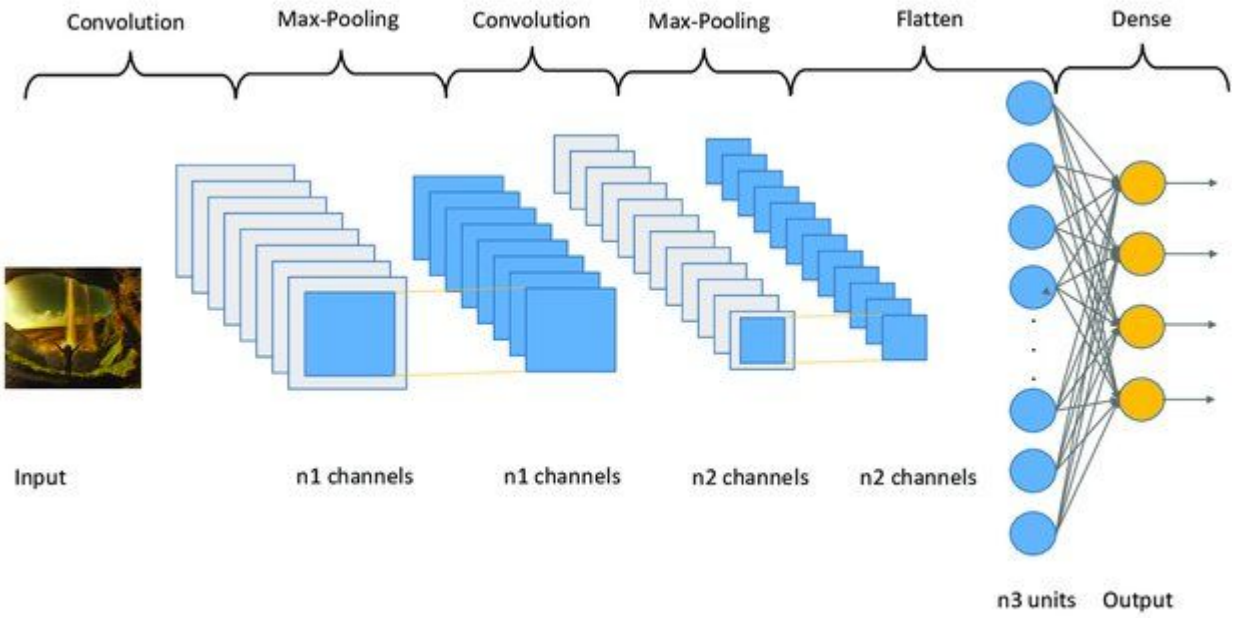
Before dropout



After dropout

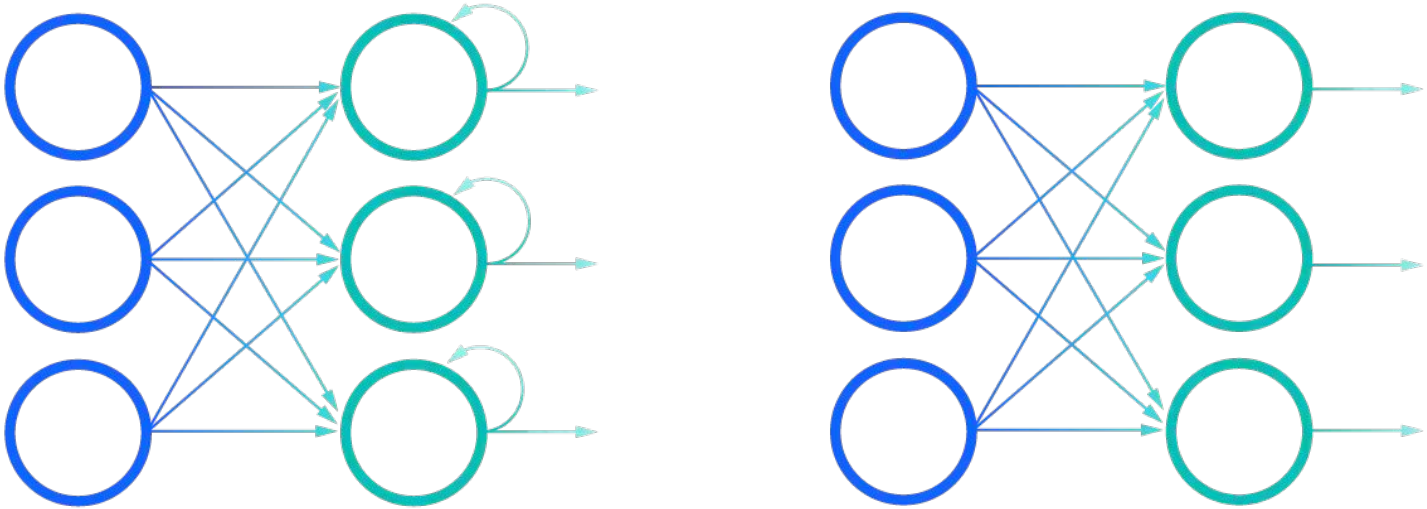


Convolutional Neural Networks



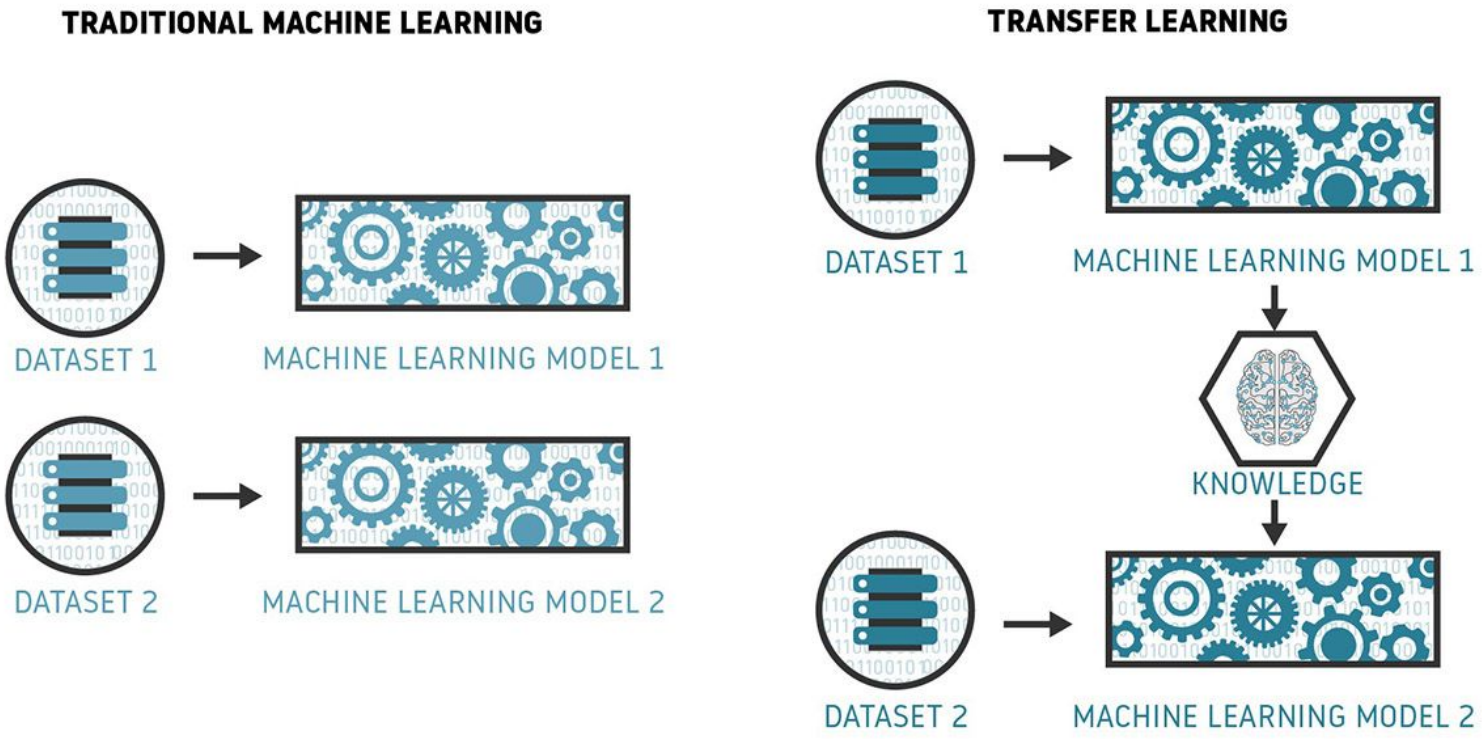
Recurrent Neural Networks

Work on sequential data to solve common temporal problems seen in language translation and speech recognition.



Source: https://www.ibm.com/

Transfer Learning



Source: pinecone.io

Computer Vision tasks

In this section, we will explore the following concepts:

- Image Classification
- Object detection
- Image Segmentation

The majority of computer vision tasks employ these in some form

Image Classification

The task of assigning an input image to one of several predefined categories

Example: Identifying a healthy vs a diseased cassava plant leaf

cmd

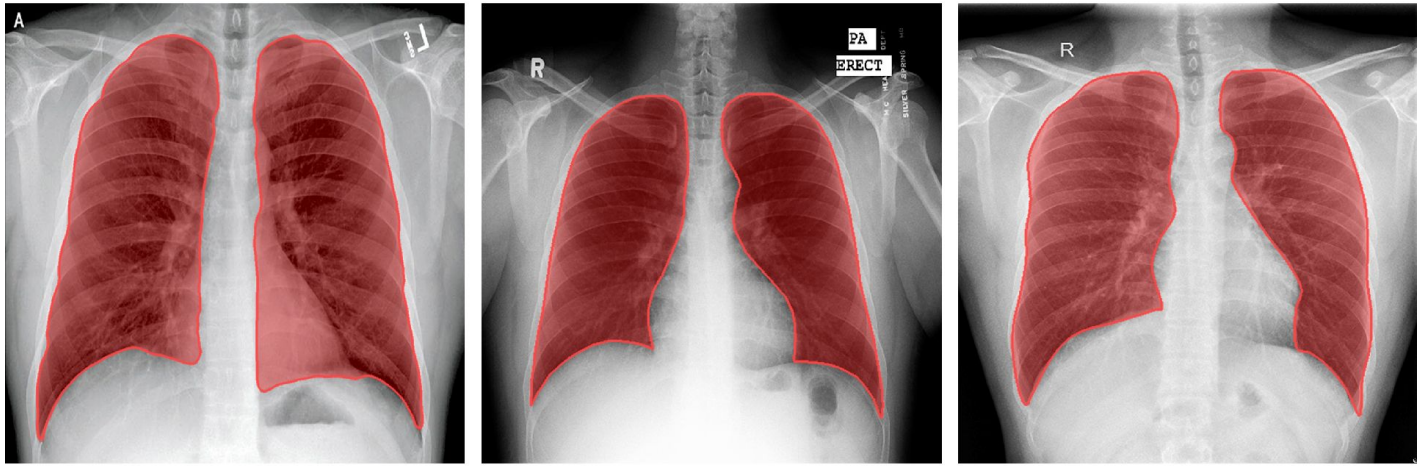
healthy

cbsd

Semantic Segmentation

The process of dividing an image into regions, each corresponding to a specific object or background.

Example: Disease Segmentation in Chest X-rays



(a) Darwin (b) Montgomery (c) Shenzhen

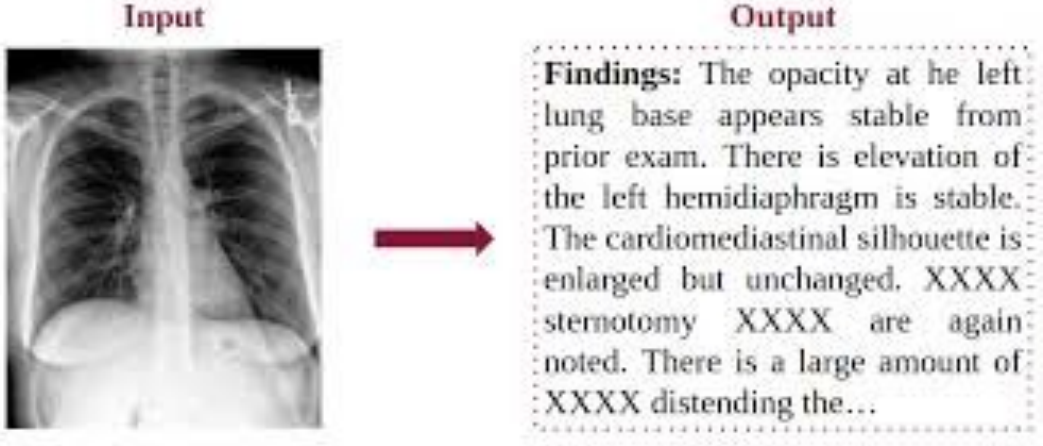
Object Detection

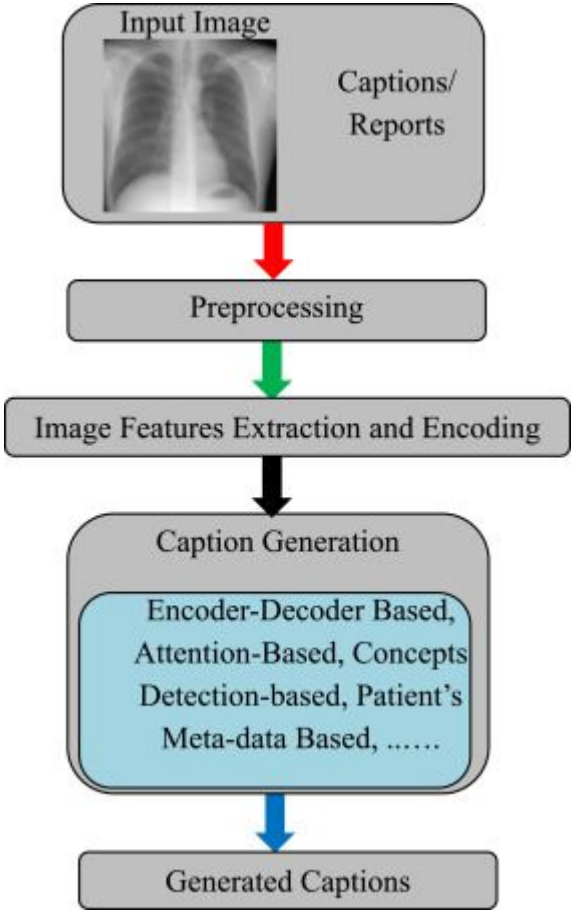
The task of identifying and locating objects within an image, often represented by bounding boxes.

Example: Passion Fruit Leaf and Fruit detection for disease detection and precision Agriculture

Bonus task (Caption Generation)

- Combine CV and NLP aspects
- Great for explainability for healthcare applications





Part II: Practical Considerations

Practical modeling strategy

- Problem definition
- Dataset creation (collecting, labelling, sanitization)
- EDA - know your data
- Data splitting (train/val/test)
- For CV - Typically start from a state-of-the-art pre-trained model.
- Fine-tune depending on amount of available data and compute.
- Leverage all general best practices e.g. data pre-processing, augmentation, learning rate decay, regularization, etc.

Practical modeling strategy …

- Use domain knowledge any other close domain data (could be unlabeled)
- Where applicable add “unknown” class for robustness of use (when applicable).
- Follow **responsible AI** practices - PII anonymization, Gender and Inclusiveness, consenting e.g. for health data, model explainability
- Experiment tracking
- Deployment (Covered in another track)
- Model and Dataset cards

Practical modeling: Dataset Creation

- Use public available datasets if available
  - Repositories with available open datasets
  - Hugging face datasets
  - Zenodo, Dataverse, Mendeley etc (searchable)
- Create a dataset from scratch
  - Expensive
  - Quality assurance is a must (and by domain experts) e.g annotator agreement
- If possible package in well-known formats e.g. COCO etc
- **Course**: Artificial Intelligence for All: How to Create and Sustain Open AI Image Training Data (atingi.org)

Practical modeling: Dataset Labelling

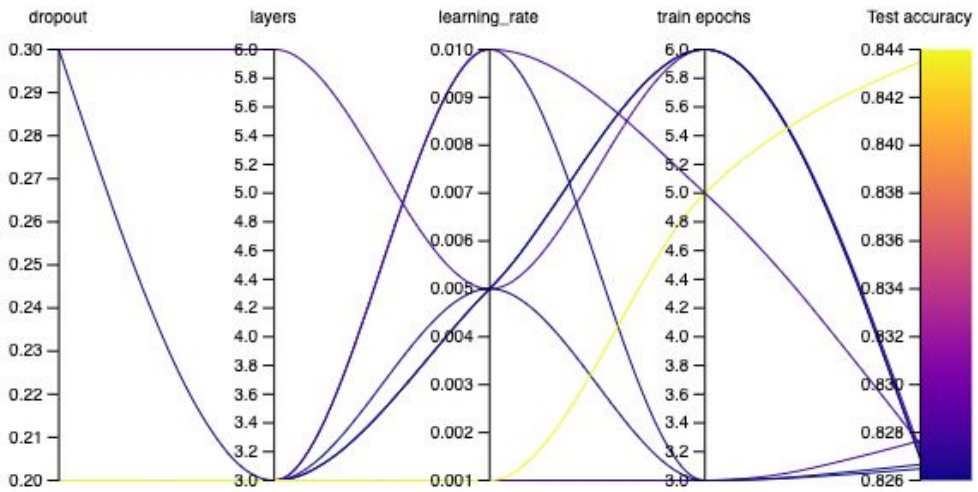
- Both open and closed-source options for annotation
  - Labelbox (commercial but educational licenses available)
  - RectLabel
  - VGG Image Annotator (VIA)
  - makesense.ai (can self-host and customize)
  - Roboflow (commercial with free tier, data solutions platform)

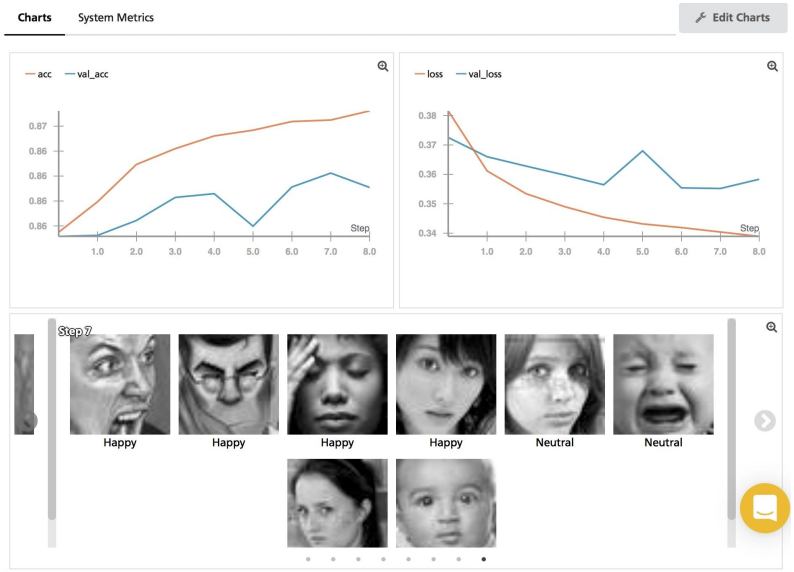
- Support various tasks (classification, segmentation, annotation) and export and import features and formats
- **Pro-tip:** Some of these support some form of model assisted labelling to support an active learning workflow

Practical modeling: Experiment tracking

- Track different runs as you try out various concepts
- Tools
  - **Weights & biases (wandb)** - commercial but free for individual or academic research teams
  - MLFlow
- Tools usually support hyperparameter optimization as well (random, BO, grid search, …)
- Support artifact tracking - datasets, models
- Sample validation set during training and display intermediate inference results
- Important for your sanity***

Practical modeling: Experiment tracking



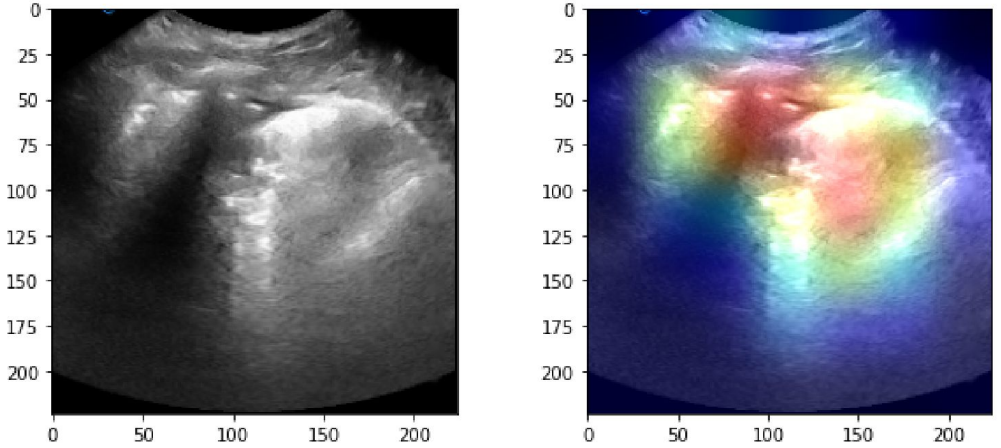


- Easily identify parameter combinations that result in the best performance
- Visual CV model training insights/debugging

Practical modeling: Model explainability

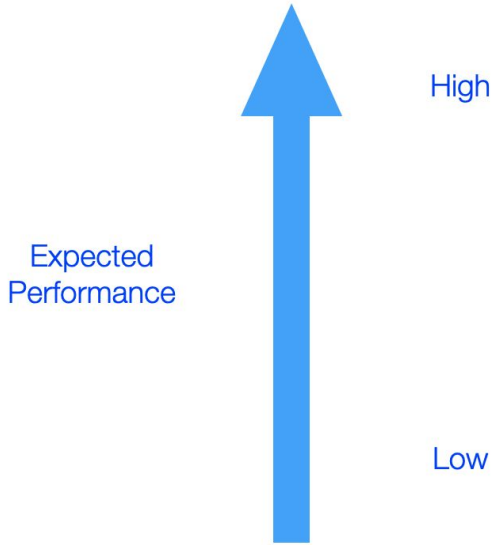
- Increases trust in models
- A must for healthcare applications
- Techniques: Gradcam, …, etc
- Careful: doesn't always work

Example: Consolidation detection in COVID-19 Lung Ultrasound →



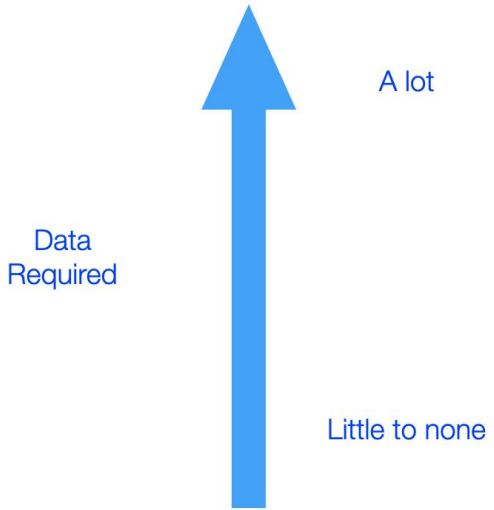
Pre-trained vs from scratch

- Train a model from scratch with all data.
- Fine-tune a pre-trained model.
- Utilize representations learned from a pre-trained model.



Pre-trained vs from scratch

- Train a model from scratch with all data.
- Fine-tune a pre-trained model.
- Utilize representations learned from a pre-trained model.



Where next?

Tools/resources

- Tensorflow Model Maker - quickly fine-tune a model targeting mobile devices.
- Tensorflow Hub - leverage state-of-the-art pre-trained models and classifiers.
- Tensorflow datasets - access curated datasets, particularly for “unknown class”
- Several similar resources with Pytorch

Tools/resources

- 🌋 LLaVA: Large Language and Vision Assistant - https://llava.hliu.cc/ (chat with your images)

- HuggingFace - https://huggingface.co/ :
  a. Datasets (pip install datasets) - can create your own
  b. HuggingFace - Quickly deploy models for inference. See e.g. https://huggingface.co/andrewkatumba/cassava-leaf-diseases-1

- Transformers for vision - https://huggingface.co/docs/transformers/model_doc/vit

Courses on Open and Inclusive AI datasets for the African Context

Artificial Intelligence for All: How to prepare open training data and share it for re-use by others following international standards (https://online.atingi.org/enrol/index.php?id=653)

Artificial Intelligence for All: How to eliminate or mitigate biases in Artificial Intelligence training data (https://online.atingi.org/enrol/index.php?id=689)

Artificial Intelligence for All: How to Create and Sustain Open AI Image Training Data (https://online.atingi.org/enrol/index.php?id=690) ***

FAIR FORWARD
Artificial Intelligence for all.

Crop Pest and Disease Diagnosis Datasets

Crop/Location	Dataset	Digital Object Identifier
Cassava (UG)	Healthy - 5000, Cassava Mosaic Disease - 5000, Cassava Brown Streak Disease - 5000	https://doi.org/10.7910/DVN/T4RB0B
Maize (UG)	Healthy - 5326, Maize Streak Virus - 5216, Maize Leaf Blight - 5279	https://doi.org/10.7910/DVN/LPGHKK
Beans (UG)	Healthy - 5284, Angular Leaf Spot -5031, Bean Rust - 5020	https://doi.org/10.7910/DVN/TCKVEW
Cassava (UG) Spectrometry	Screenhouse - 20,400 readings Open field - 18,900 readings	https://doi.org/10.7910/DVN/R0KL7R
Maize (TZ)	Healthy - 5542, Maize Lethal Necrosis - 5068, Maize Streak Virus - 6667	https://doi.org/10.7910/DVN/GDON8Q
Bananas (TZ)	Healthy - 5883 , 6147 - Black Sigatoka, 5038 - Fusarium Wilt Race 1 class.	https://doi.org/10.7910/DVN/LQUWXW
Maize (GN)	Healthy - 5,392, Fall Armyworm - 5,110, Maize Streak Virus - 5,063	https://doi.org/10.7910/DVN/CXUMDS
Cocoa (GN)	Healthy - 5,056, Anthracnose - 5,133, Cocoa Swollen Shoot Virus - 6,757	https://doi.org/10.7910/DVN/BBGQSP
Maize (NB)	Healthy - 2611, Maize Streak Virus -3009, Fall Armyworm - 3414	https://doi.org/10.7910/DVN/6R78HR

Part III: Hands-on preparation

https://github.com/andrewkatumba/dsa2023_cv

END