Isack Odero

ML Eng
LogAI

Innovation Week Tanzania 2025

Future Ready Summit 2025

Log.ai

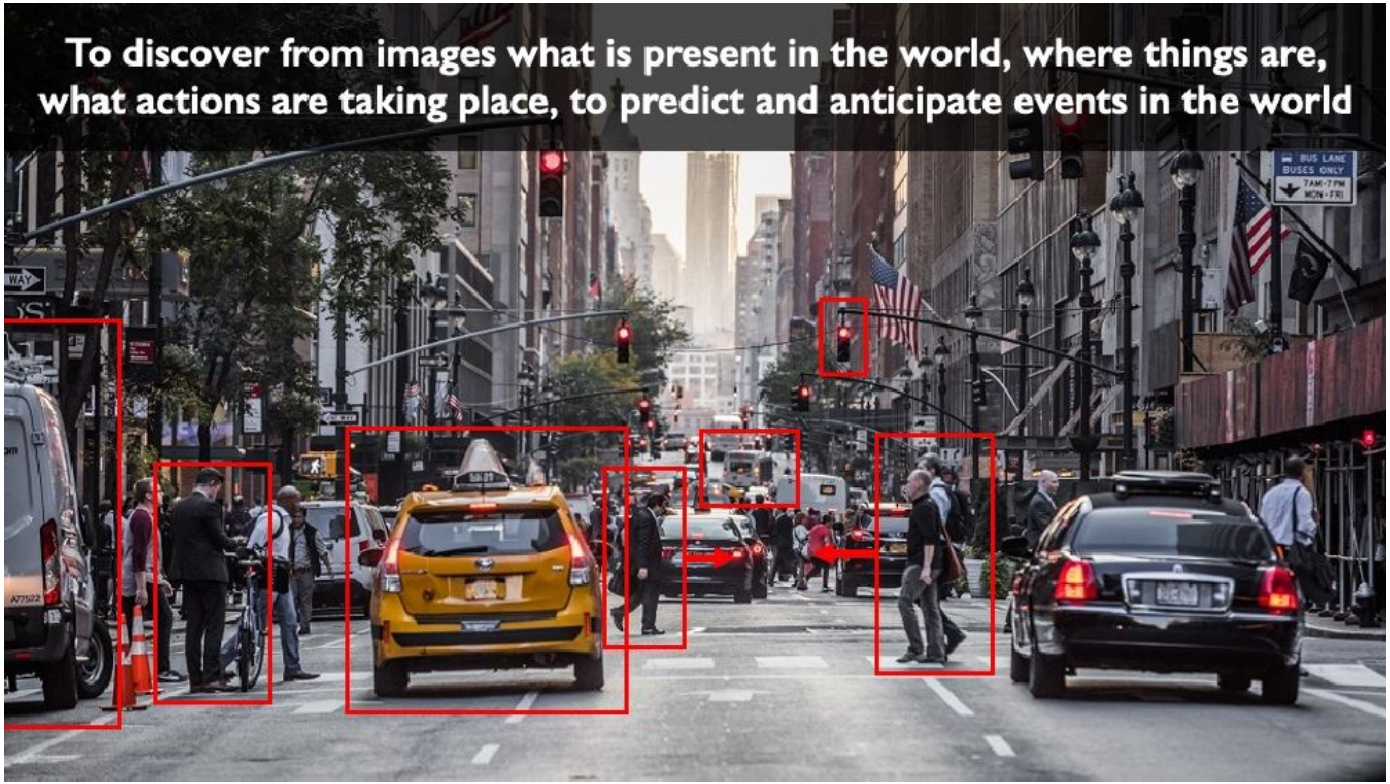
How Computer See and Talk

➔ Computer Vision

➔ Sequence Modeling

Computer Vision

Computer Vision



Innovation Week Tanzania 2025

Future Ready Summit 2025

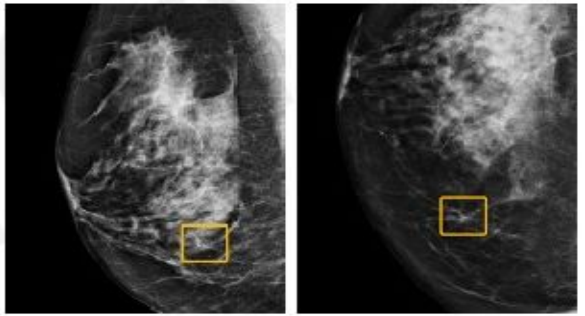
Log.ai

The rise and impact of computer vision

Robotics

Accessibility

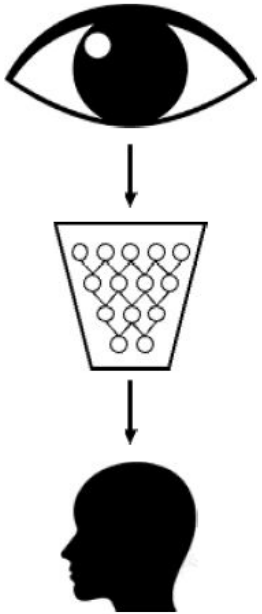
Biology & Medicine

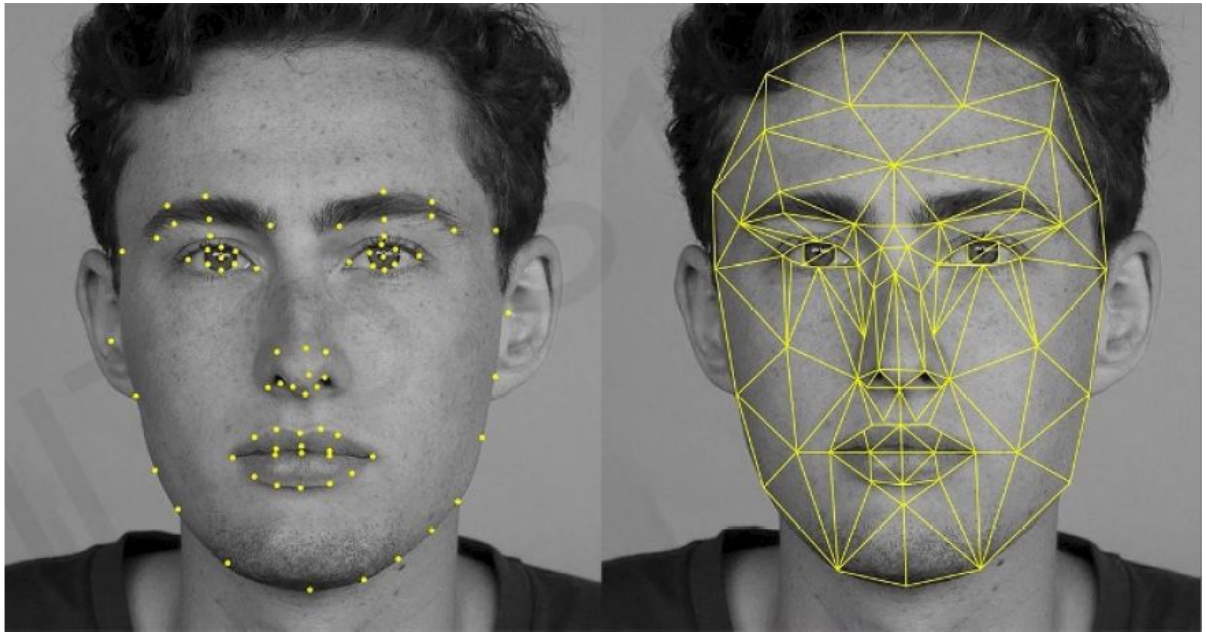


Autonomous driving

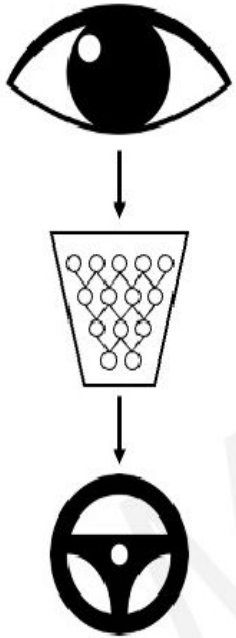
Mobile computing

Impact: Facial Recognition

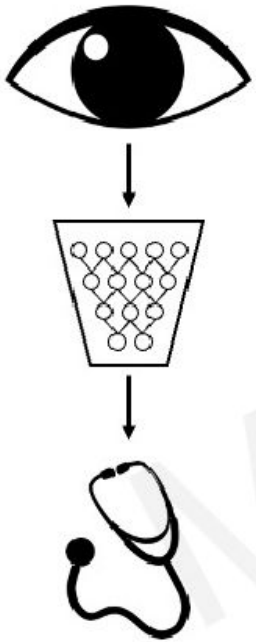




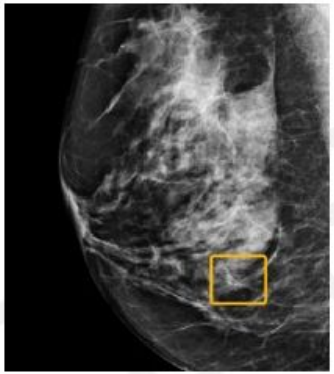
Impact: Autonomous Driving



Impact: Medicine, Biology, Healthcare

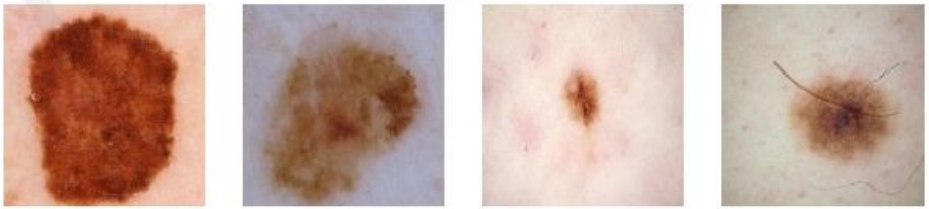


Breast cancer

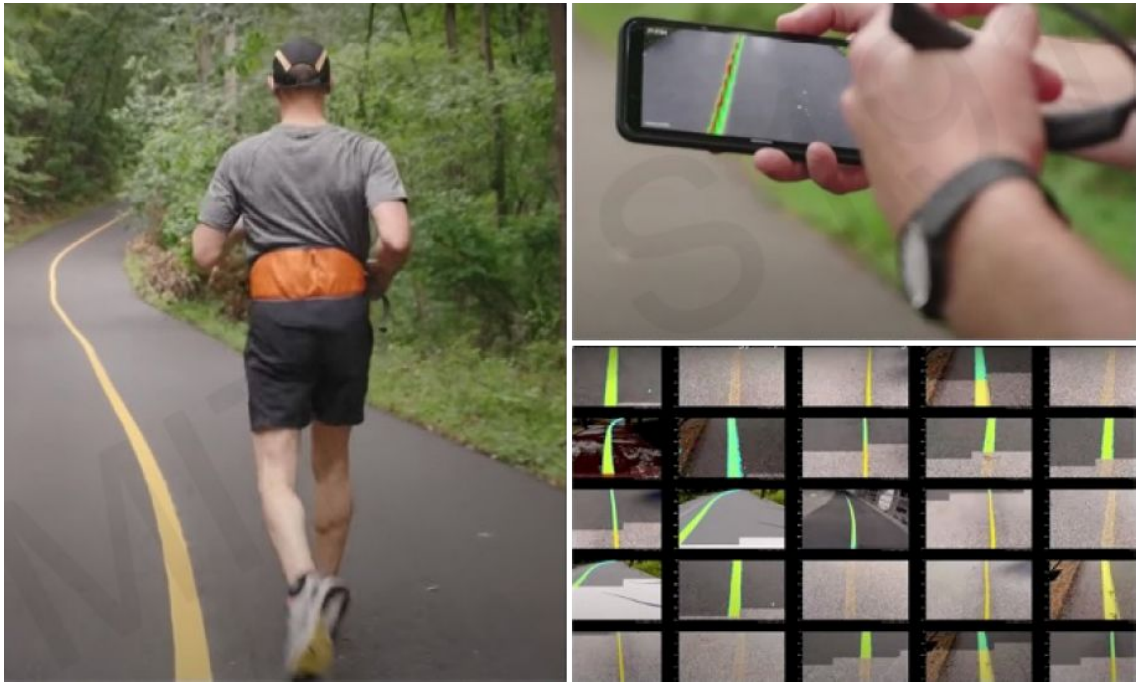


COVID-19

Skin cancer



Impact: Medicine, Biology, Healthcare



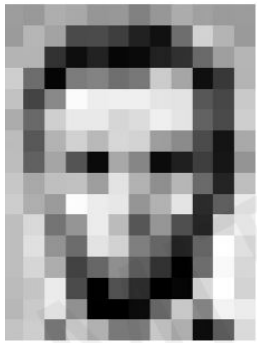
Innovation Week Tanzania 2025

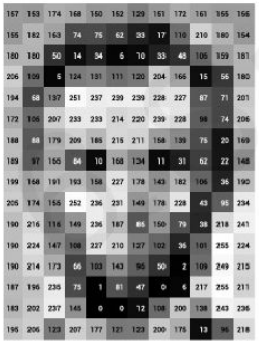
Future Ready Summit 2025

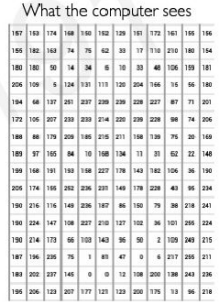
Log.ai

What Computers “See”

Images are Numbers

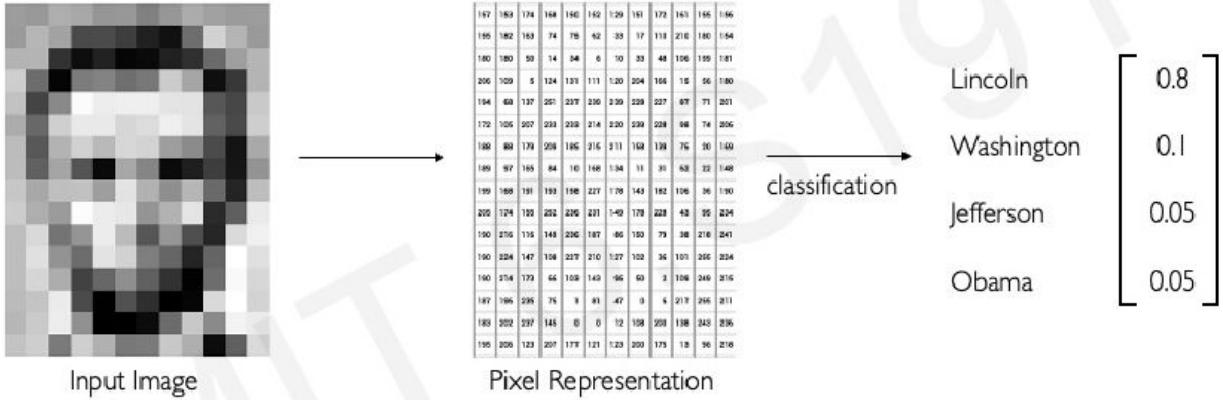






- ➔ An image is just a matrix of numbers [0, 255]
- ➔ I.e. 1080x1080x3 an RGB image

Tasks in Computer Vision

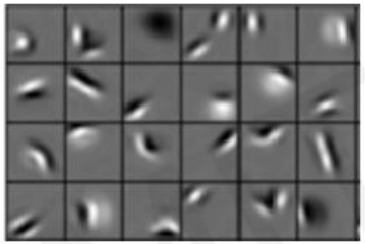


- **Regression**: output variable takes continuous value
- **Classification**: output variable takes class label. Can produce probability of belonging to a particular class

Learning Feature Representations

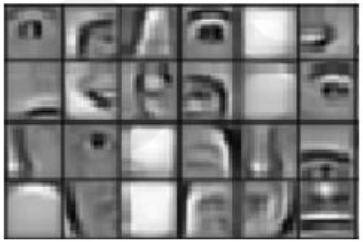
Can we learn a **hierarchy of features** directly from the data instead of hand engineering?

Low level features



Edges, dark spots

Mid level features

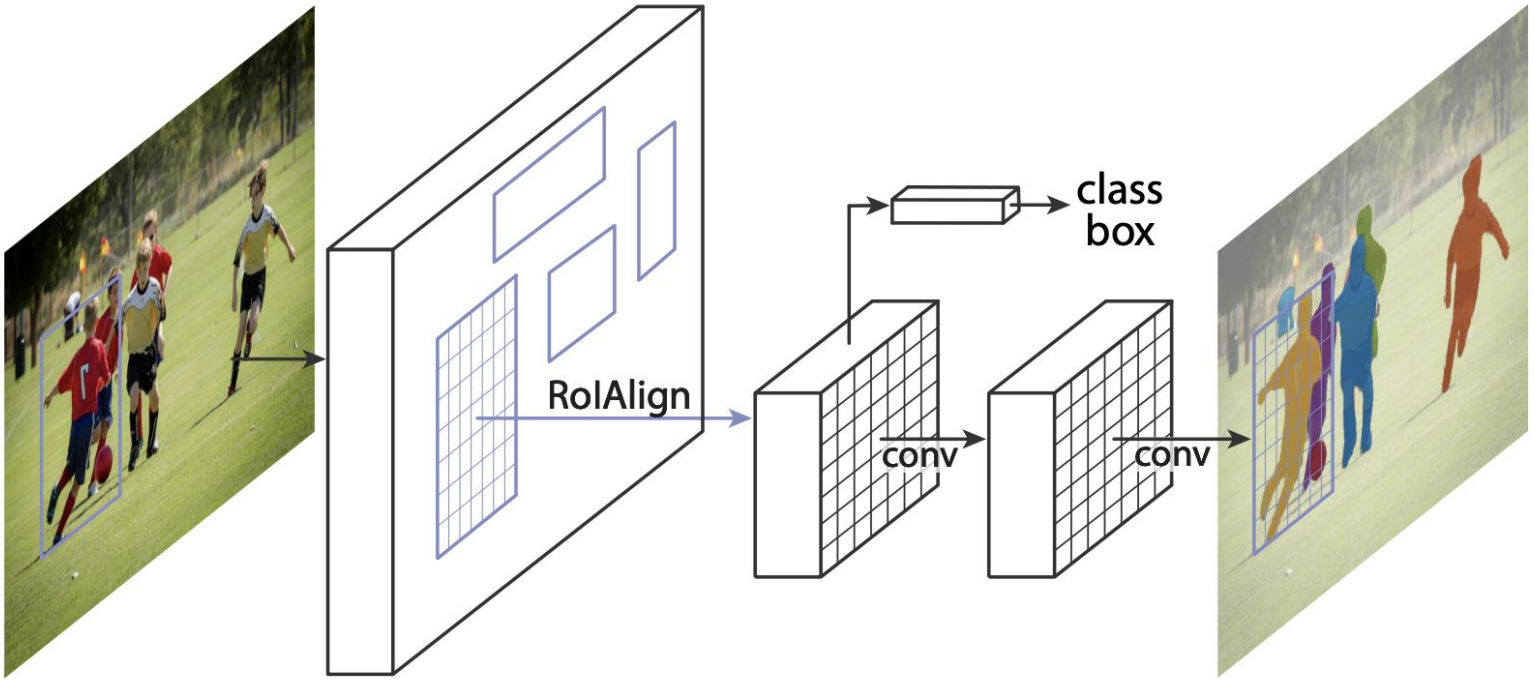


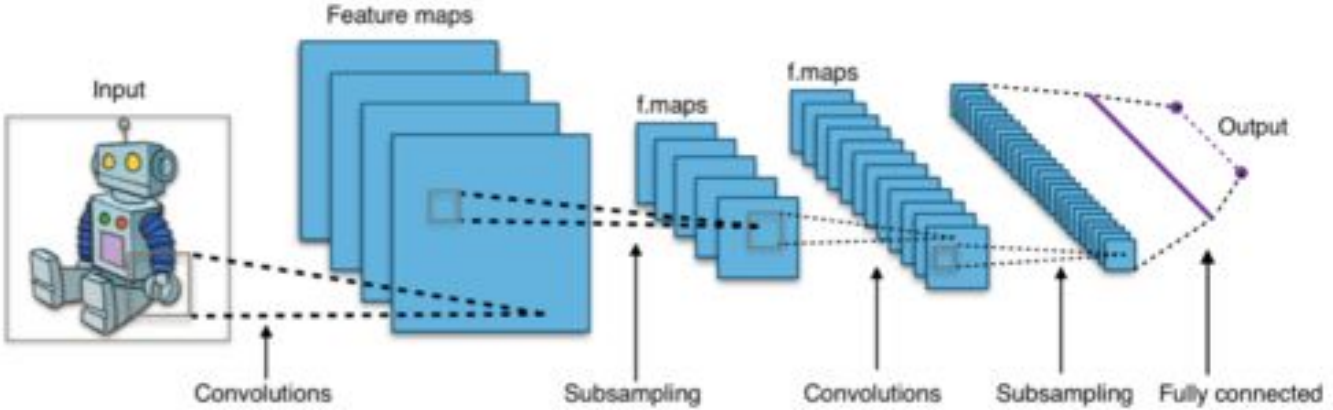
Eyes, ears, nose

High level features

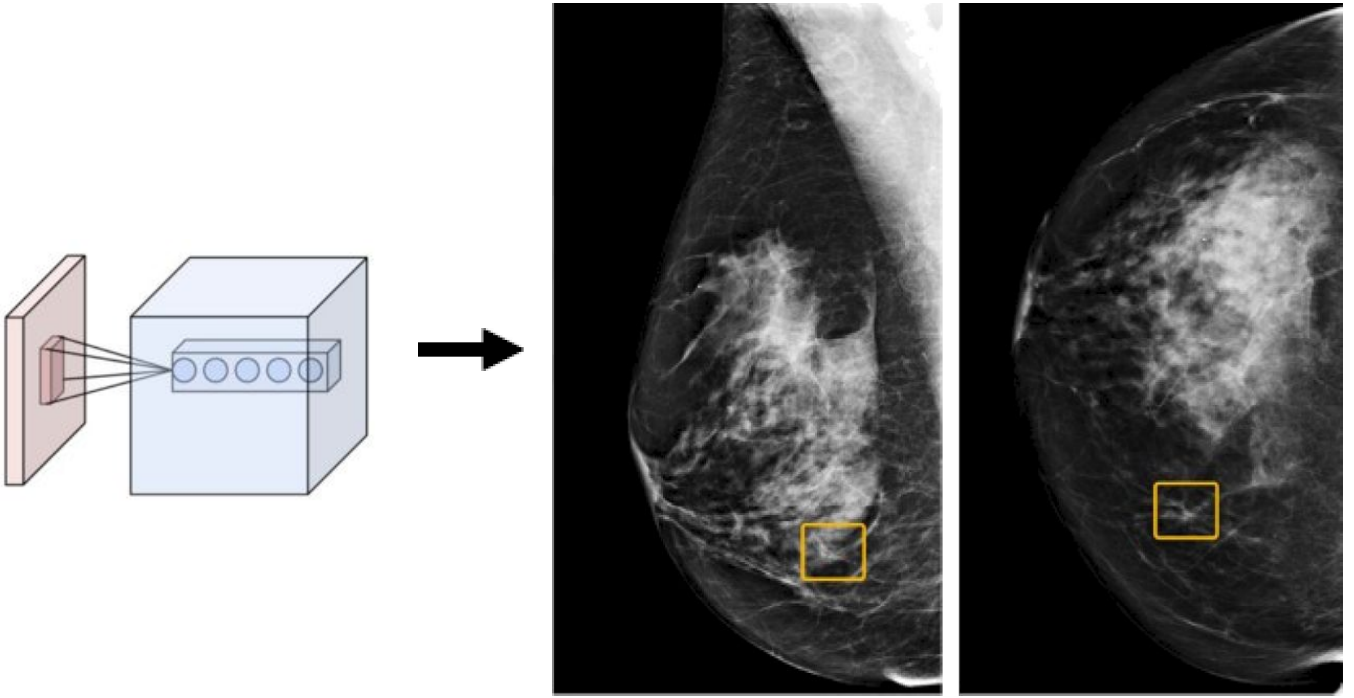
Facial structure

An Architecture for Many Applications

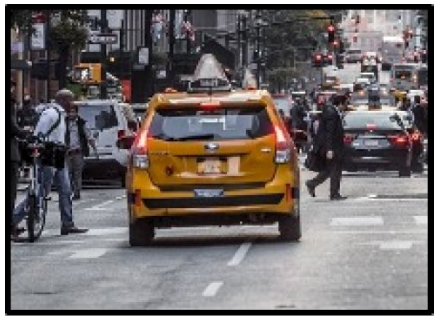

RoIAlign
class
box
conv
conv



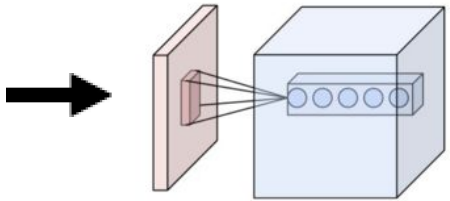
Classification: Breast Cancer Screen



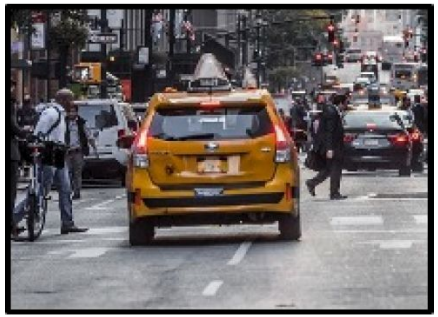
Object Detection



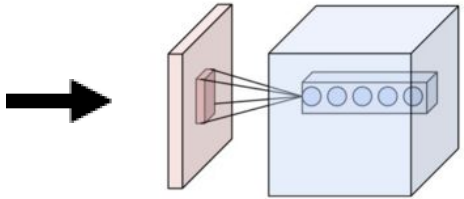
Image



Taxi

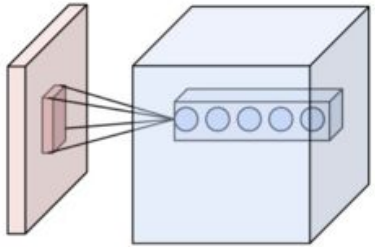


Image



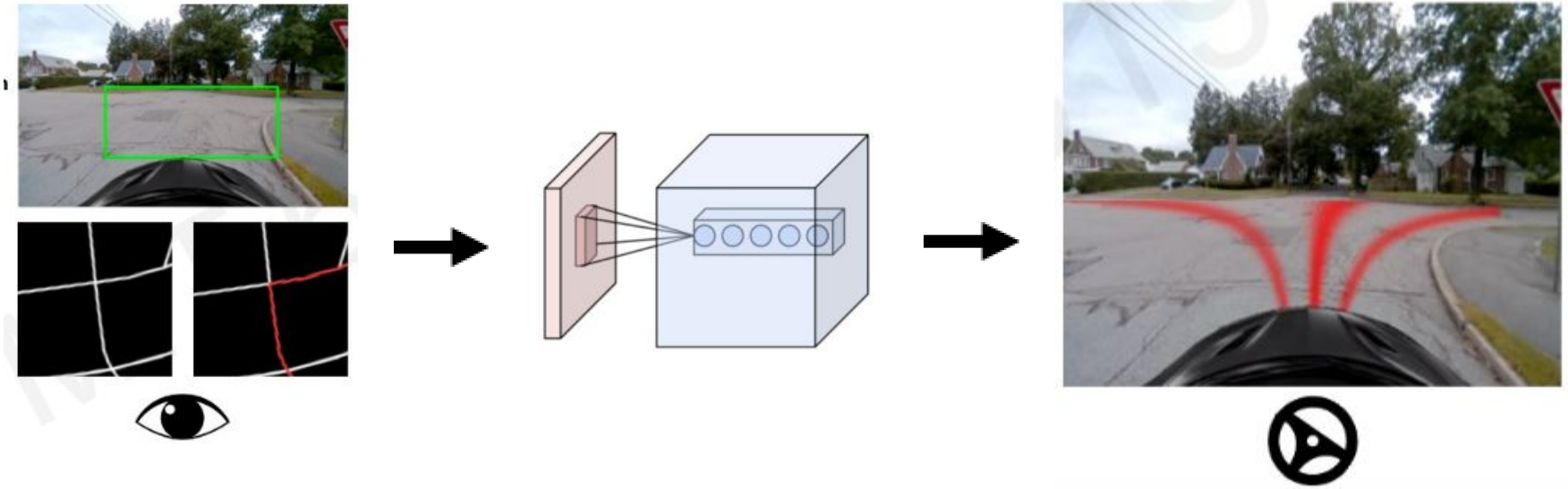
Semantic Segmentation

Input:
3 x H x W



Predictions:
H x W

Continuous Control Navigation



Deep Learning for Computer Vision

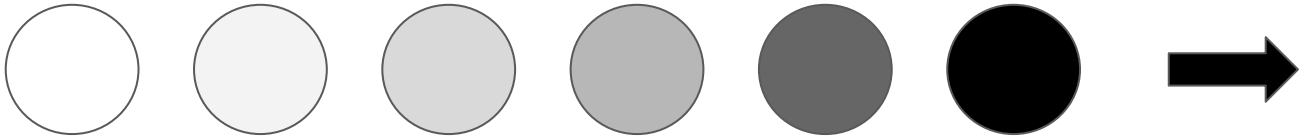
Innovation Week Tanzania 2025

Future Ready Summit 2025

Log.ai

Sequence Modeling

Given an Image of a ball, can you predict where it will go next?

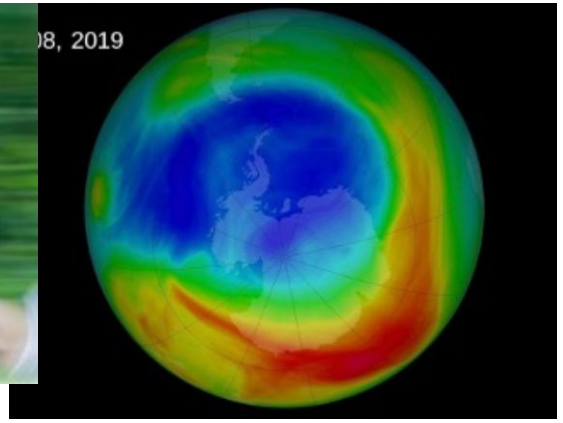


Innovation Week Tanzania 2025

Future Ready Summit 2025

Log.ai

Sequence data



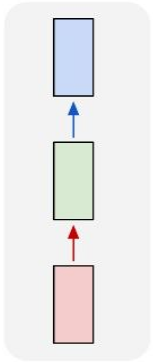
Innovation Week Tanzania 2025

Future Ready Summit 2025

Log.ai

Sequence Modeling Application

one to one



Binary Classification

"Will I pass this class?"
Student → Pass?

one to many

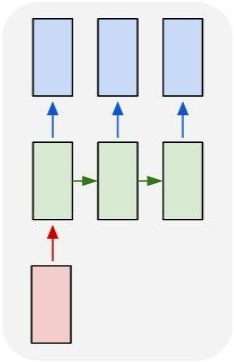
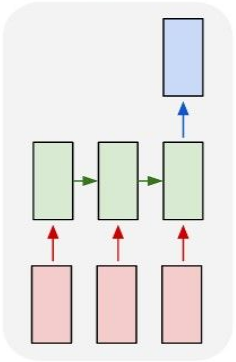


Image Captioning

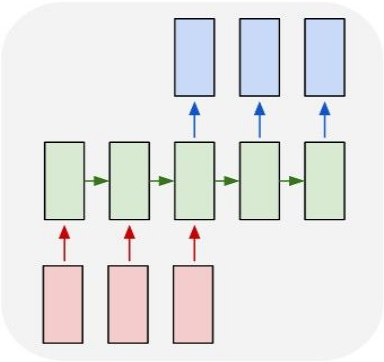
"A baseball player throws a ball."

many to one



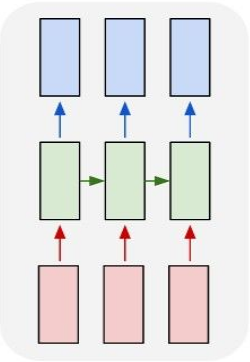
Sentiment Classification

many to many



Machine Translation

many to many

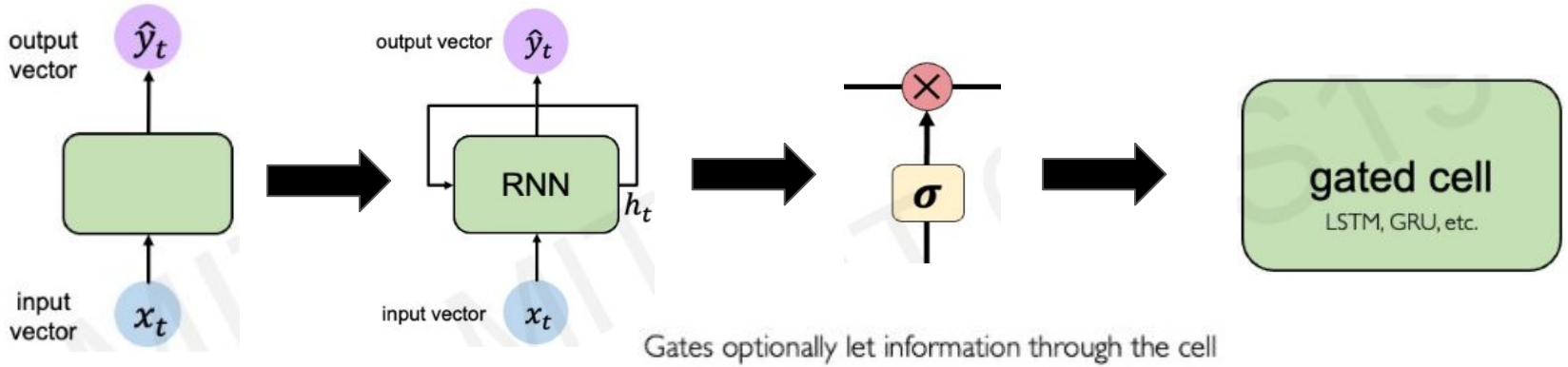


Machine Translation

Sequence Modeling Type

1. Sequence Modeling with Recurrence

Models: RNN, LSTM, GRU
Core Idea: Process sequences **step-by-step**, maintaining a hidden state that captures past information.



Sequence Modeling Type

1. Sequence Modeling with Recurrence

Models: RNN, LSTM, GRU
Core Idea: Process sequences **step-by-step**, maintaining a hidden state that captures past information.

Pros	Cons
✔ Handles variable-length sequences well	✖ Struggles with **long-range dependencies**
✔ Low memory usage (sequential processing)	✖ Slow training (no parallelization)
✔ Simple architecture	✖ Prone to vanishing/exploding gradients (RNNs)

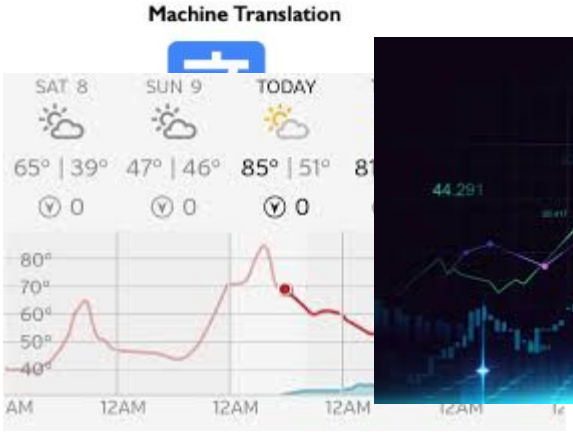
Sequence Modeling Type

1. Sequence Modeling with Recurrence

Models: RNN, LSTM, GRU
Core Idea: Process sequences **step-by-step**, maintaining a hidden state that captures past information.

Applications

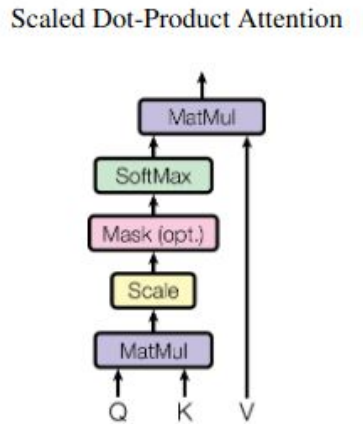
- Early NLP (machine translation, sentiment analysis)
- Time-series forecasting (stock prices, weather)

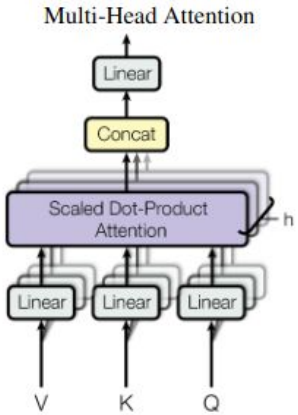


Sequence Modeling Type

2. Sequence Modeling with Attention

Models: Transformers (BERT, GPT)
Core Idea: Weigh the importance of all past inputs dynamically using **self-attention**.





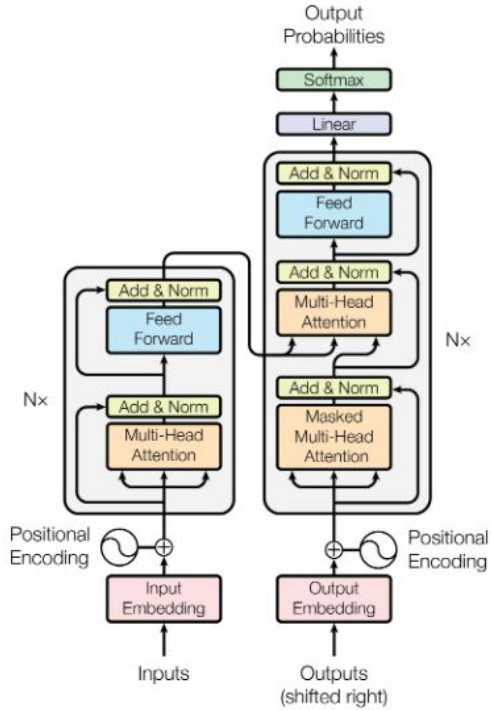


Figure 1: The Transformer - model architecture.

Innovation Week Tanzania 2025

Future Ready Summit 2025

Log.ai

Sequence Modeling Type

2. Sequence Modeling with Attention

Models: Transformers (BERT, GPT)
Core Idea: Weigh the importance of all past inputs dynamically using **self-attention**.

Pros	Cons
✔ Captures **long-range dependencies** better	✖ High memory usage (stores all tokens)
✔ Parallel processing (faster training)	✖ Computationally expensive
✔ State-of-the-art performance (e.g., GPT-4)	✖ Requires large datasets

Sequence Modeling Type

2. Sequence Modeling with Attention

Models: Transformers (BERT, GPT)
Core Idea: Weigh the importance of all past inputs dynamically using **self-a**

Applications

- Modern NLP (ChatGPT, translation, summarization)
- Vision tasks (ViT - Vision Transformers)