

Isack Odera

ML Eng
LogAI

How Computer See and Talk

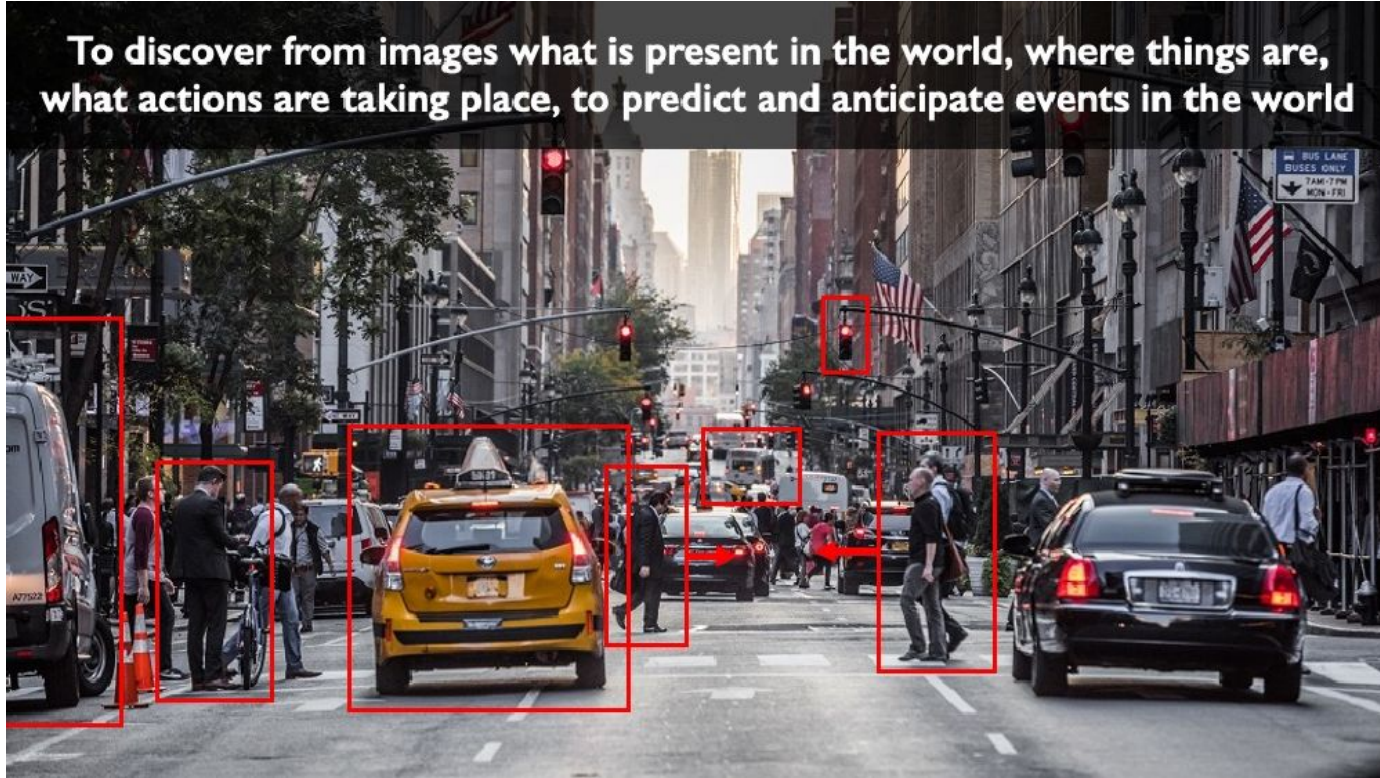
- Computer Vision
- Sequence Modeling

Computer Vision



Computer Vision

To discover from images what is present in the world, where things are, what actions are taking place, to predict and anticipate events in the world



The rise and impact of computer vision

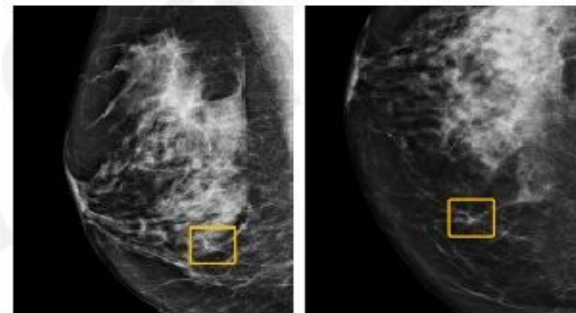
Robotics



Accessibility



Biology & Medicine

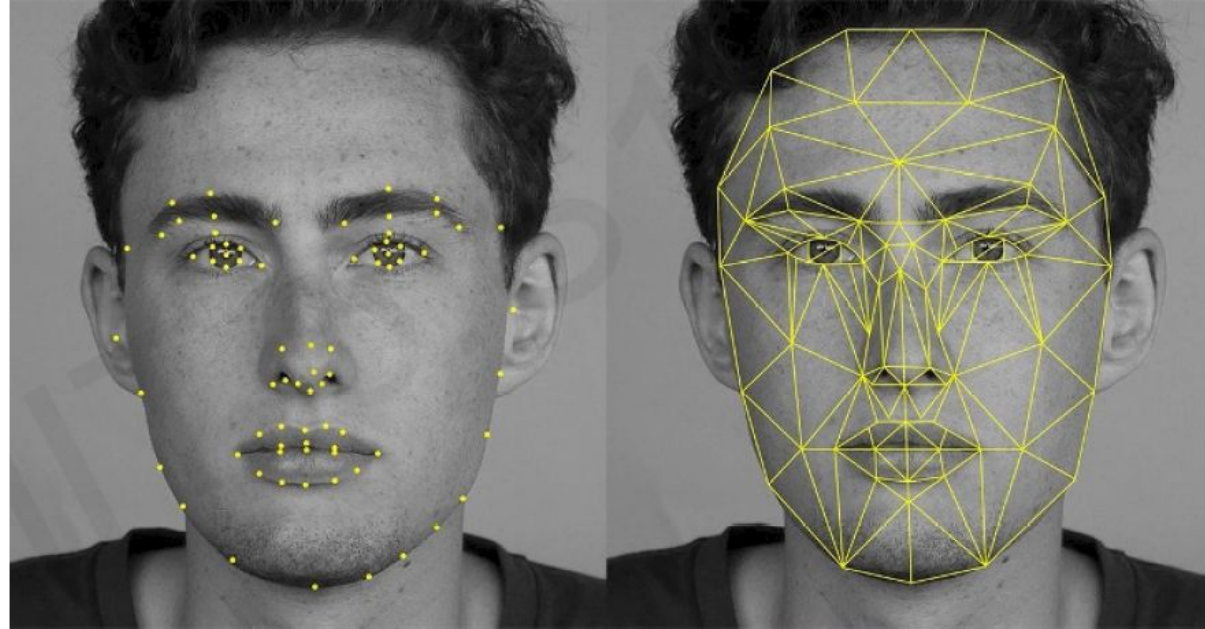
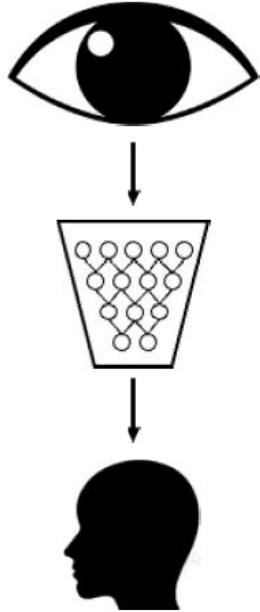


Autonomous driving

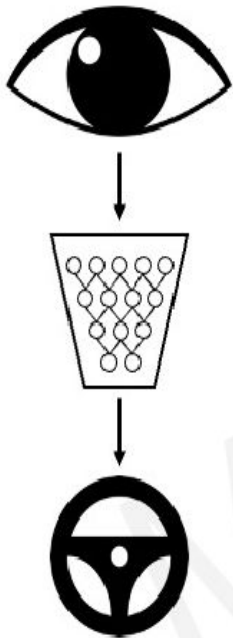


Mobile computing

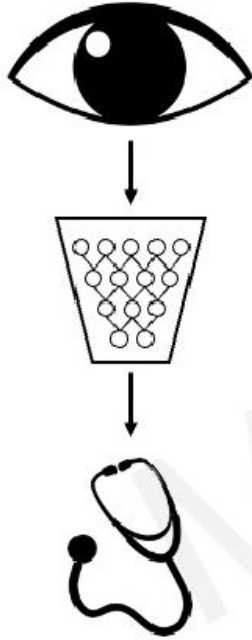
Impact: Facial Recognition



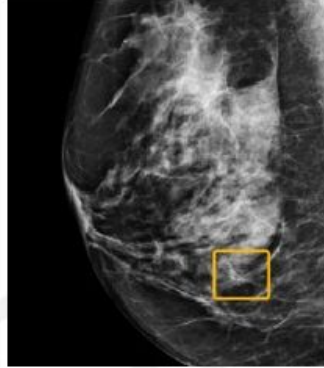
Impact: Autonomous Driving



Impact: Medicine, Biology, Healthcare



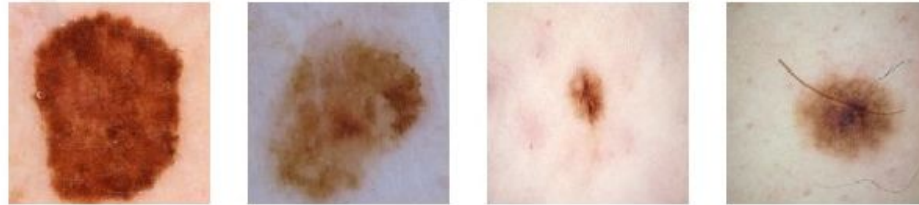
Breast cancer



COVID-19



Skin cancer

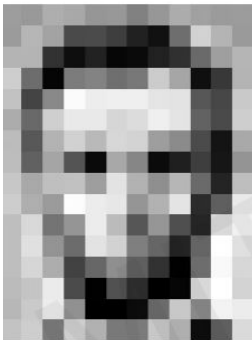


Impact: Medicine, Biology, Healthcare



What Computers “See”

Images are Numbers



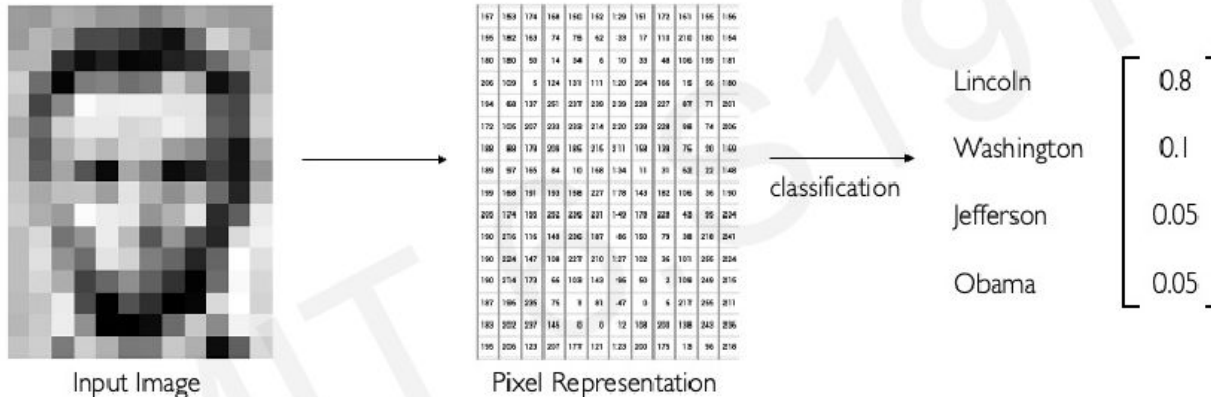
187	163	174	168	160	163	129	161	172	161	165	166
186	182	163	74	75	62	33	17	110	230	180	164
180	180	50	14	34	6	10	33	48	106	159	181
206	109	6	124	131	131	120	204	166	16	66	180
194	68	197	251	297	299	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	309	185	215	211	164	139	76	80	169
189	81	166	66	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	140	178	238	43	65	234
190	216	116	149	236	187	86	150	79	38	318	241
190	224	147	106	227	210	127	102	36	101	253	224
190	214	173	66	183	143	50	50	2	109	249	215
187	196	236	75	1	81	47	0	6	217	256	211
183	202	237	145	0	0	12	138	200	138	243	236
196	206	123	207	177	171	123	200	178	13	76	218

What the computer sees

187	163	174	168	160	163	129	161	172	161	165	166
186	182	163	74	75	62	33	17	110	230	180	164
180	180	50	14	34	6	10	33	48	106	159	181
206	109	6	124	131	131	120	204	166	16	66	180
194	68	197	251	297	299	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	309	185	215	211	164	139	76	80	169
189	81	166	66	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	140	178	238	43	65	234
190	216	116	149	236	187	86	150	79	38	318	241
190	224	147	106	227	210	127	102	36	101	253	224
190	214	173	66	183	143	50	50	2	109	249	215
187	196	236	75	1	81	47	0	6	217	256	211
183	202	237	145	0	0	12	138	200	138	243	236
196	206	123	207	177	171	123	200	178	13	76	218

- An image is just a matrix of numbers $[0, 255]$
- I.e. $1080 \times 1080 \times 3$ an RGB image

Tasks in Computer Vision

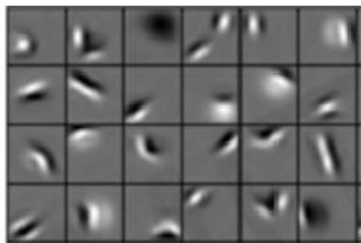


- **Regression:** output variable takes continuous value
- **Classification:** output variable takes class label. Can produce probability of belonging to a particular class

Learning Feature Representations

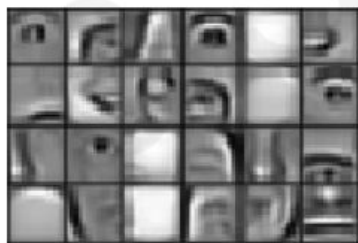
Can we learn a **hierarchy of features** directly from the data instead of hand engineering?

Low level features



Edges, dark spots

Mid level features



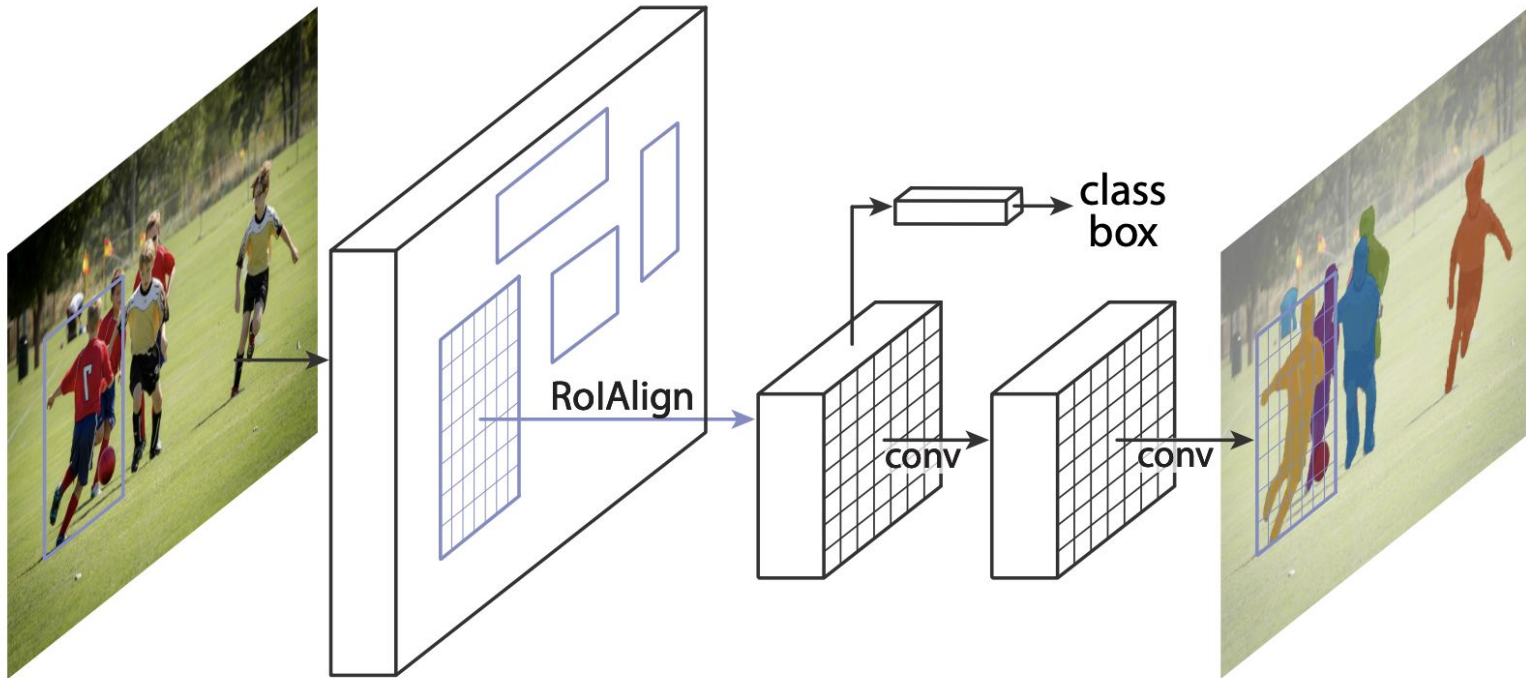
Eyes, ears, nose

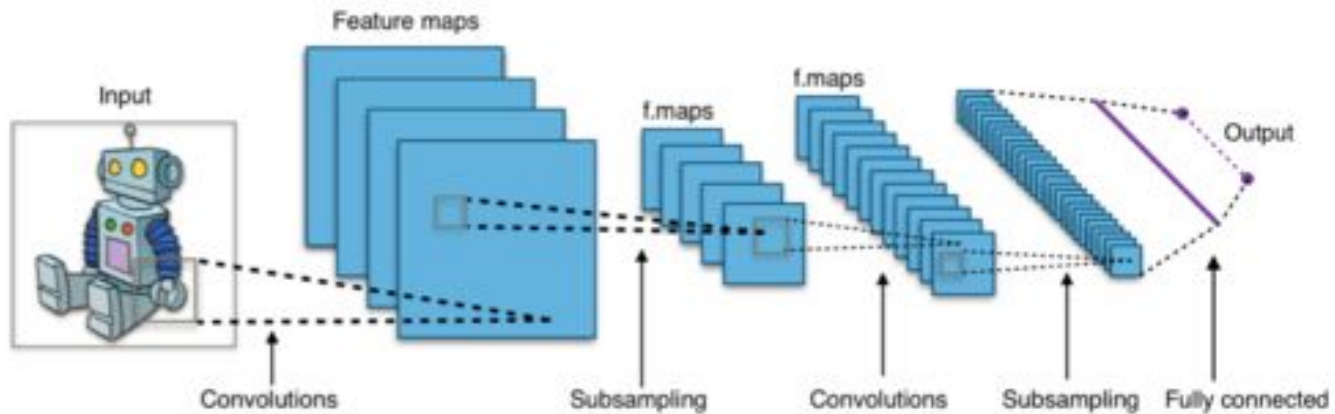
High level features



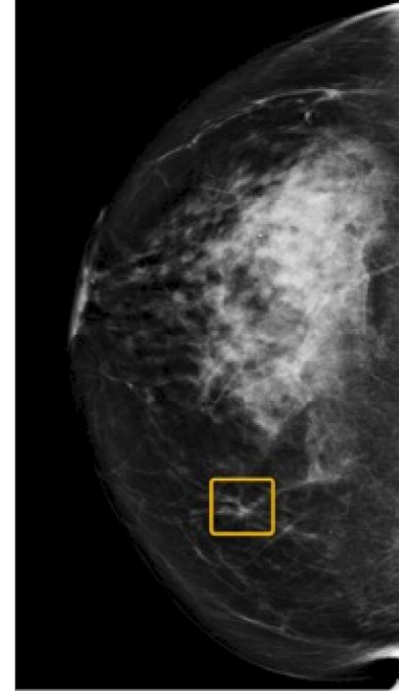
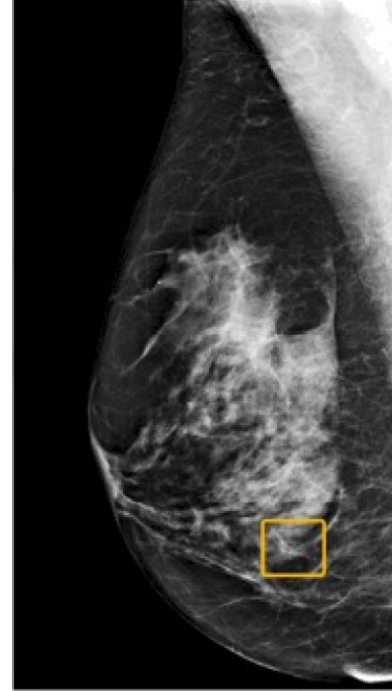
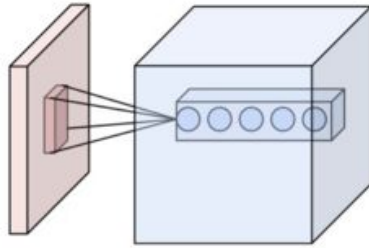
Facial structure

An Architecture for Many Applications





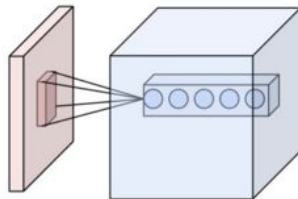
Classification: Breast Cancer Screen



Object Detection



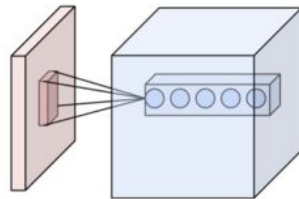
Image



Taxi



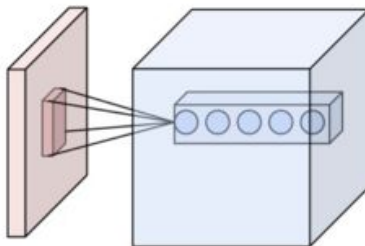
Image



Semantic Segmentation

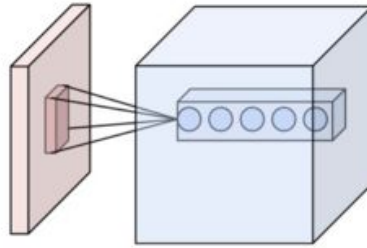
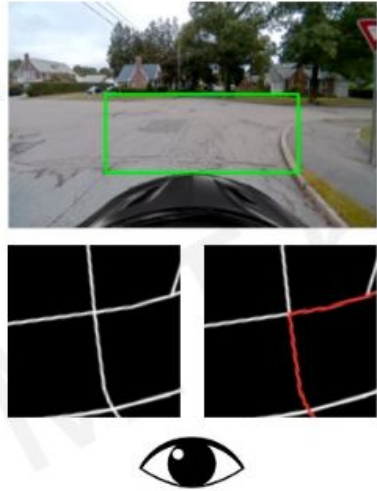


Input:
 $3 \times H \times W$



Predictions:
 $H \times W$

Continuous Control Navigation



Possible Control Commands

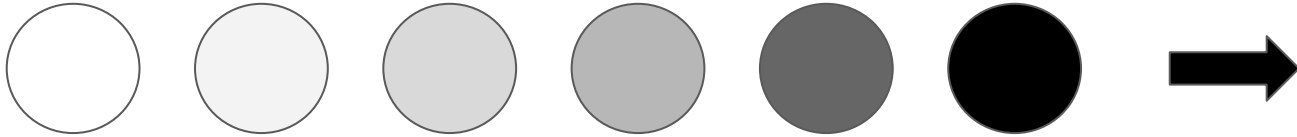


Deep Learning for Computer Vision

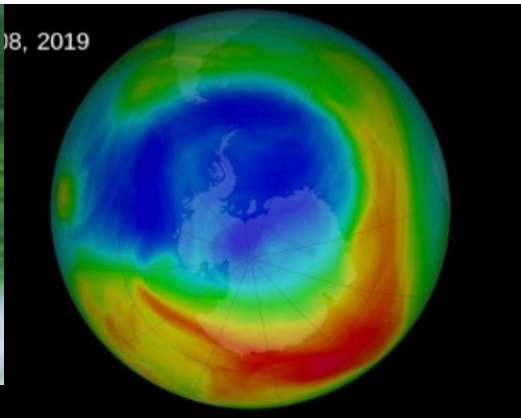


Sequence Modeling

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

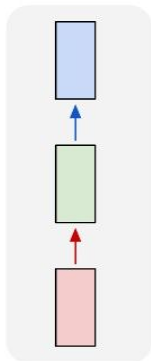


Sequence data

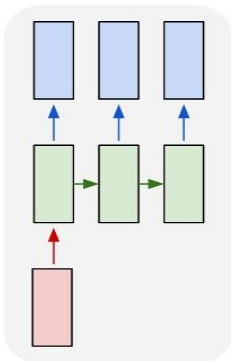


Sequence Modeling Application

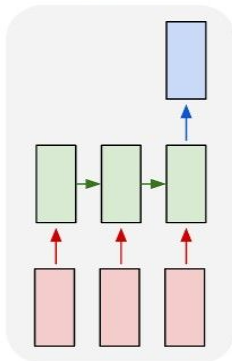
one to one



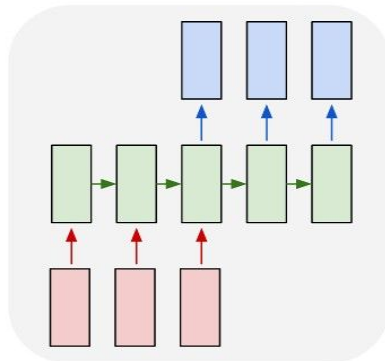
one to many



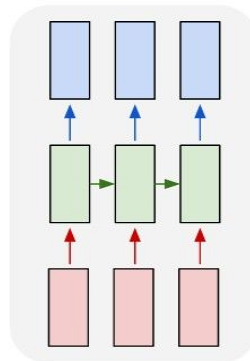
many to one



many to many



many to many



Binary Classification



Image Captioning



"A baseball player throws a ball."

Sentiment Classification



Machine Translation



Machine Translation

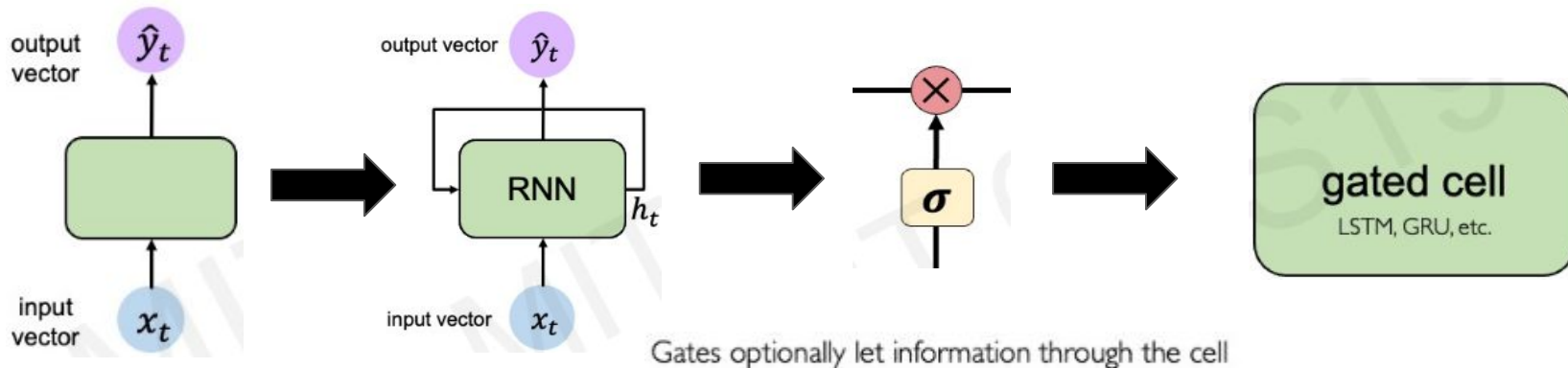


Sequence Modeling Type

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

- ✓ Handles variable-length sequences well
- ✓ Low memory usage (sequential processing)
- ✓ Simple architecture

Cons

- ✗ Struggles with **long-range dependencies**
- ✗ Slow training (no parallelization)
- ✗ Prone to vanishing/exploding gradients (RNNs)

Sequence Modeling Type

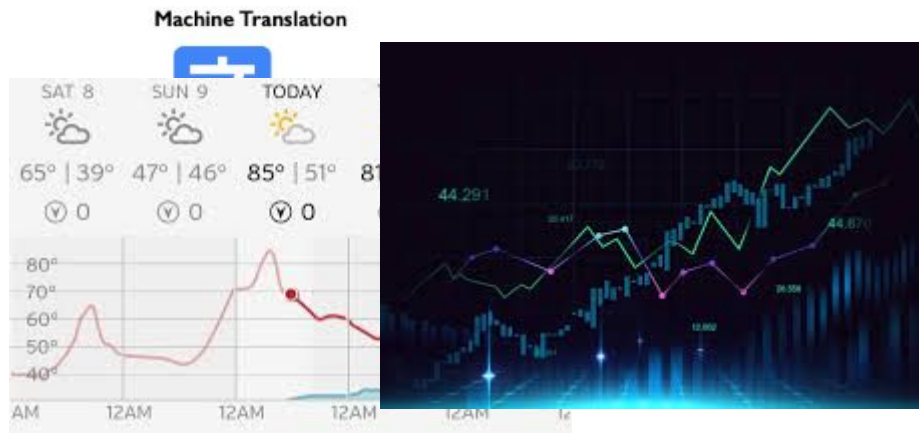
1. Sequence Modeling with Recurrence

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)



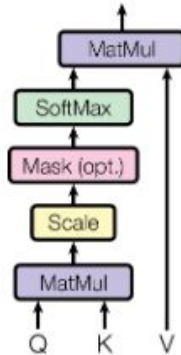
Sequence Modeling Type

2. Sequence Modeling with Attention

Models: Transformers (BERT, GPT)

Core Idea: Weigh the importance of all past inputs dynamically using **self-attention**.

Scaled Dot-Product Attention



Multi-Head Attention

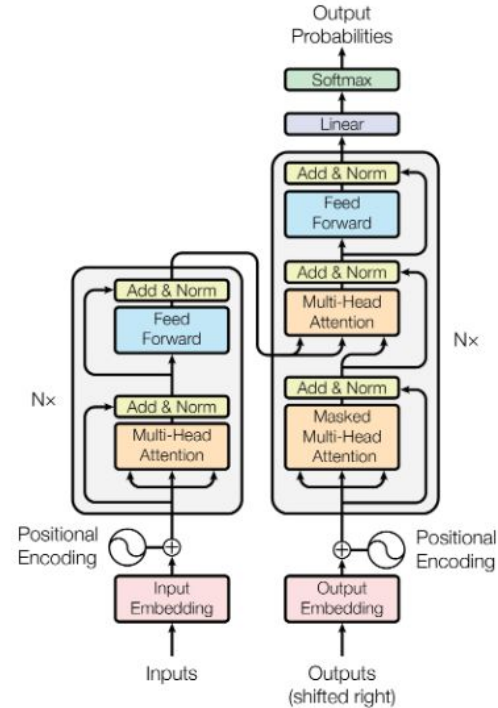
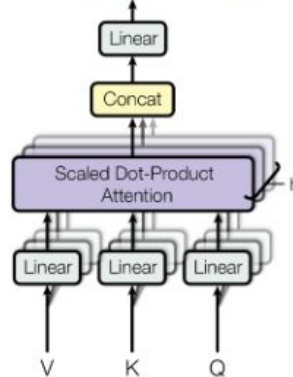


Figure 1: The Transformer - model architecture.

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

- ✓ Captures **long-range dependencies** better
- ✓ Parallel processing (faster training)
- ✓ State-of-the-art performance (e.g., GPT-4)

Cons

- ✗ High memory usage (stores all tokens)
- ✗ Computationally expensive
- ✗ 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)

