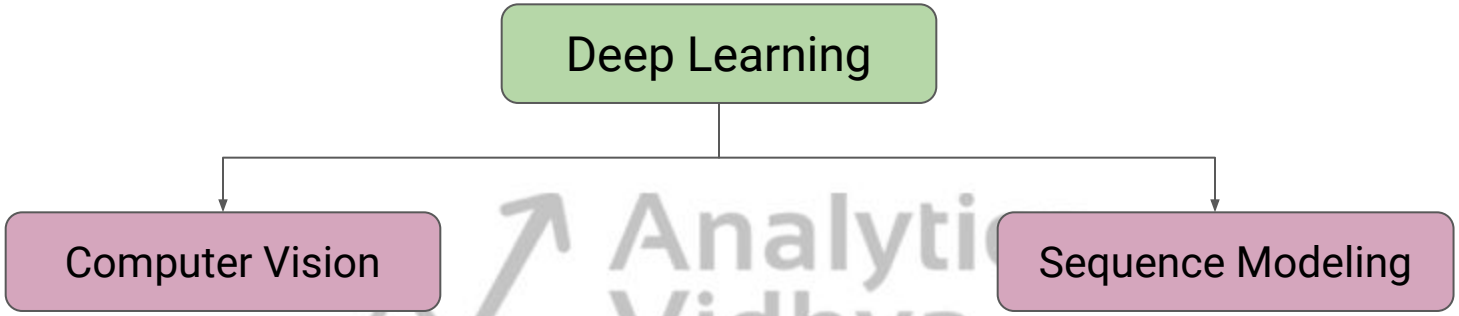
Where to go from here?

Overview of the Course

Basics of Neural Network	Handling Image Data	Keras	Improving Model Performance	Convolutional Neural Network
Working with Text Data	Recurrent Neural Network	Handling Audio Data	Unsupervised Deep Learning	PyTorch

Realm of Deep Learning



- Image Classification
- Object Detection
- Image Segmentation
- Video Analytics
- Image Denoising
- Image Colorization

- Text Classification
- Neural Machine Translation
- Sequence Tagging
- Time Series Forecasting
- Audio Classification
- Speech-to-Text

Computer Vision

Image Classification

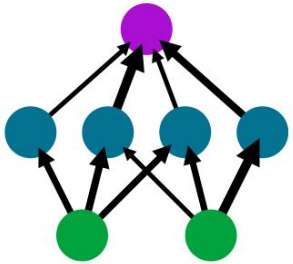
- Most popular use case of Computer Vision

Emergency Vehicle

Non Emergency Vehicle

Image Classification

- Most popular use case of Computer Vision
- Can be solved using MLP and CNN



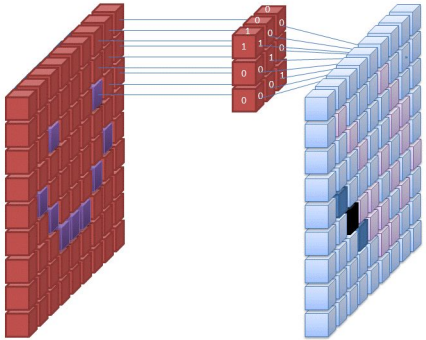
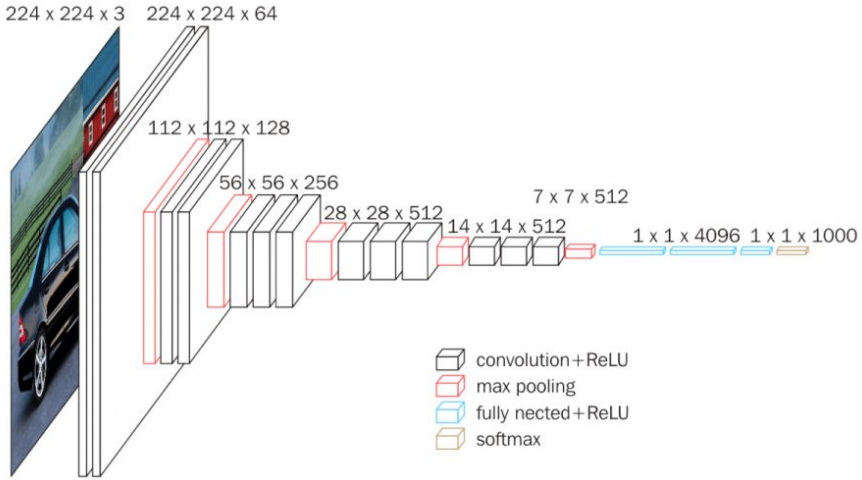


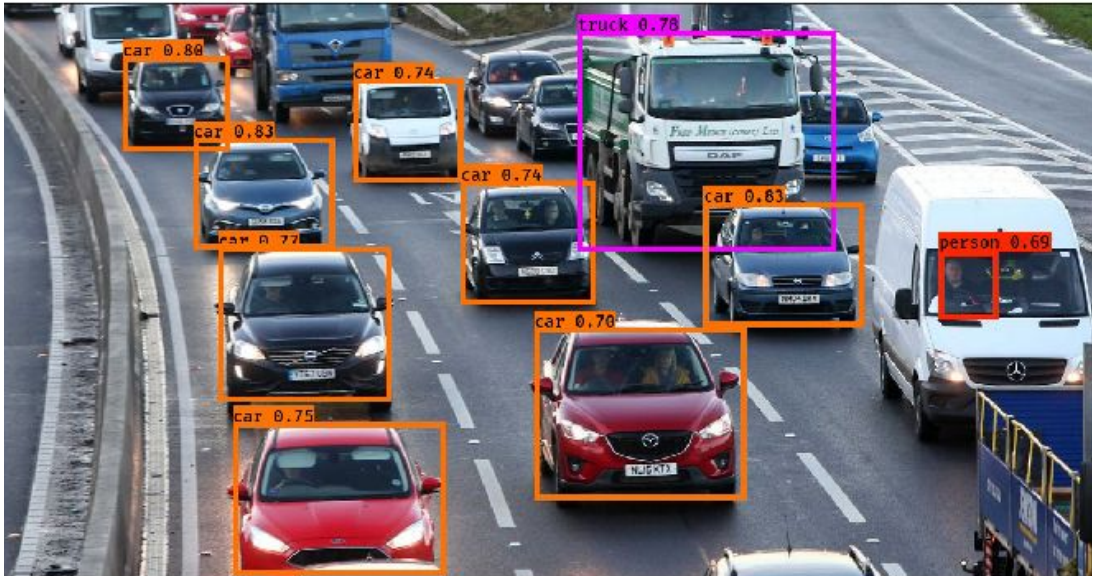
Image Classification

- Most popular use case of Computer Vision
- Can be solved using MLP and CNN
- Transfer Learning
  - VGG16
  - Inception
  - ResNet



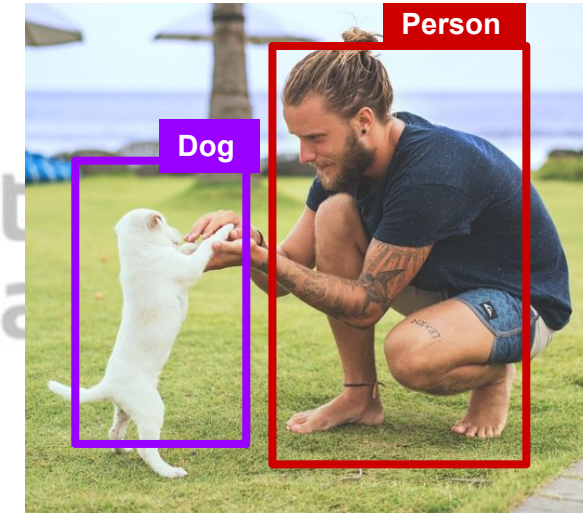
Source: https://neurohive.io/en/popular-networks/vgg16/

Object Detection



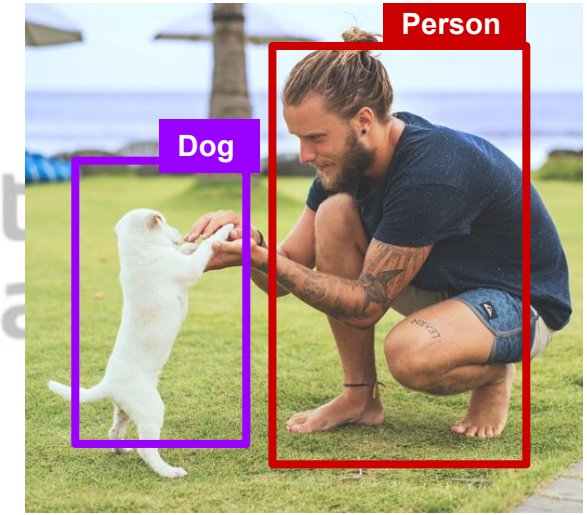
Object Detection

- Identifying and Locating objects
- **Goal:** Find a bounding box around the object of interest



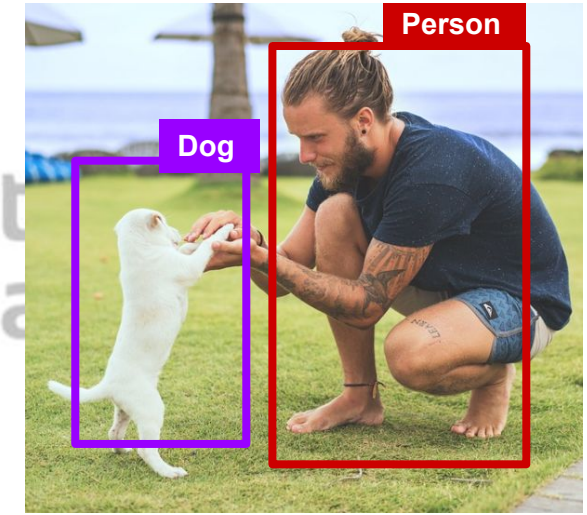
Object Detection

- Identifying and Locating objects
- **Goal:** Find a bounding box around the object of interest
- Single-class object detection and Multi-class object detection

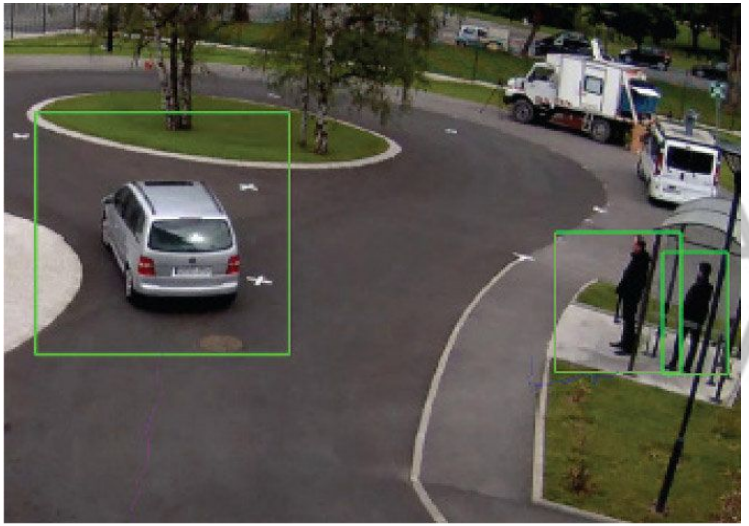


Object Detection

- Identifying and Locating objects
- **Goal:** Find a bounding box around the object of interest
- Single-class object detection and Multi-class object detection
- Pre-trained Models
  - Single Shot Detector (SSD)
  - RCNN and Faster RCNN
  - YOLO



Object Detection: Use Cases



Source: researchgate.net

Vehicle Detection

Source: arxiv.org/pdf/1808.01050.pdf

Crowd Counting

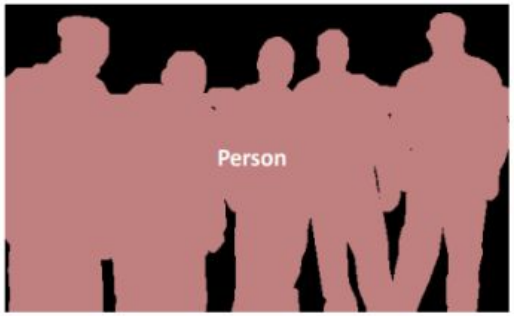
Image Segmentation

Image Segmentation

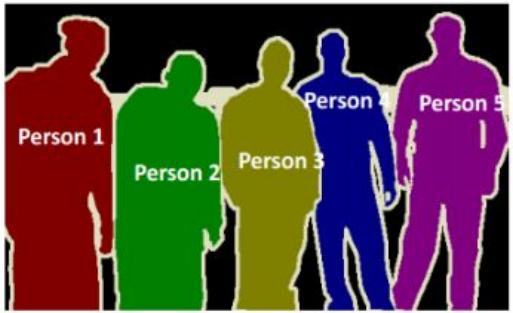
- Group together image pixels that have similar attributes
- Image segmentation creates a pixel-wise mask for the objects in an image.

Image Segmentation

- Group together image pixels that have similar attributes
- Image segmentation creates a pixel-wise mask for the objects in an image.
- Types of Image Segmentation
  - Semantic Segmentation
  - Instance Segmentation

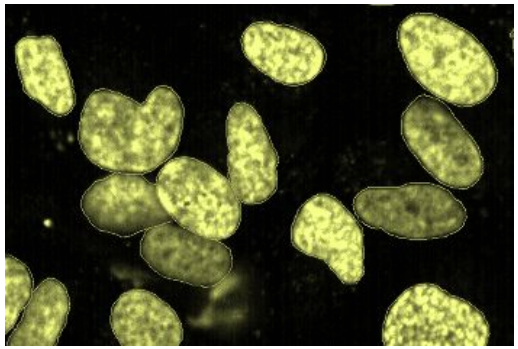


Semantic Segmentation



Instance Segmentation

Image Segmentation - Use Cases



Source: Wikipedia

Finding Shape of Cancerous Cells

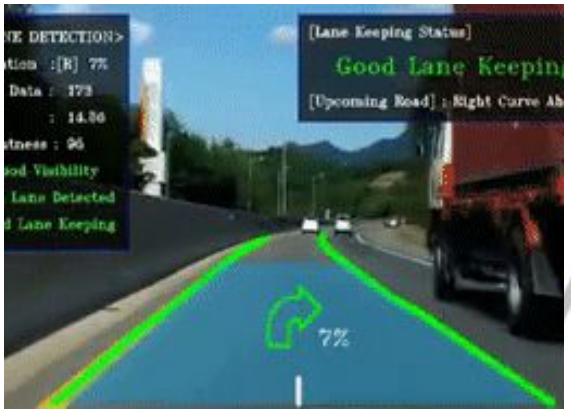
Source: mathworks.com

Lane Segmentation for Self-driving cars

Source: researchgate.net

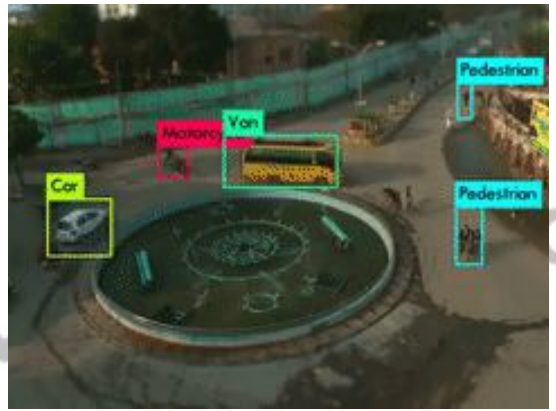
Semantic Segmentation of Satellite Imagery

Video Analytics



Source: https://youtu.be/VIH3OEhZnow

Real Time Lane Detection and Segmentation



Source: https://youtu.be/DeCFxPQlOVk

Traffic Monitoring System

Source: https://youtu.be/iRlWw8GD0xc

Sports Analytics: Ball Tracking System

Deep Generative Models

Which one is fake?

Source: https://www.thispersondoesnotexist.com/

Deep Generative Models

- Generative modeling is a subset of unsupervised learning

Source: blogs.nvidia.com

Deep Generative Models

- Generative modeling is a subset of unsupervised learning
- Learn distribution of the given data and use it to generate new data

Source: blogs.nvidia.com

Deep Generative Models

- Generative modeling is a subset of unsupervised learning
- Learn distribution of the given data and use it to generate new data
- Deep Generative Models
  - Pixel RNN
  - Pixel CNN
  - Variational Autoencoders
  - Generative Adversarial Networks (GANs)

Source: blogs.nvidia.com

Sequence Modeling

Language Modeling

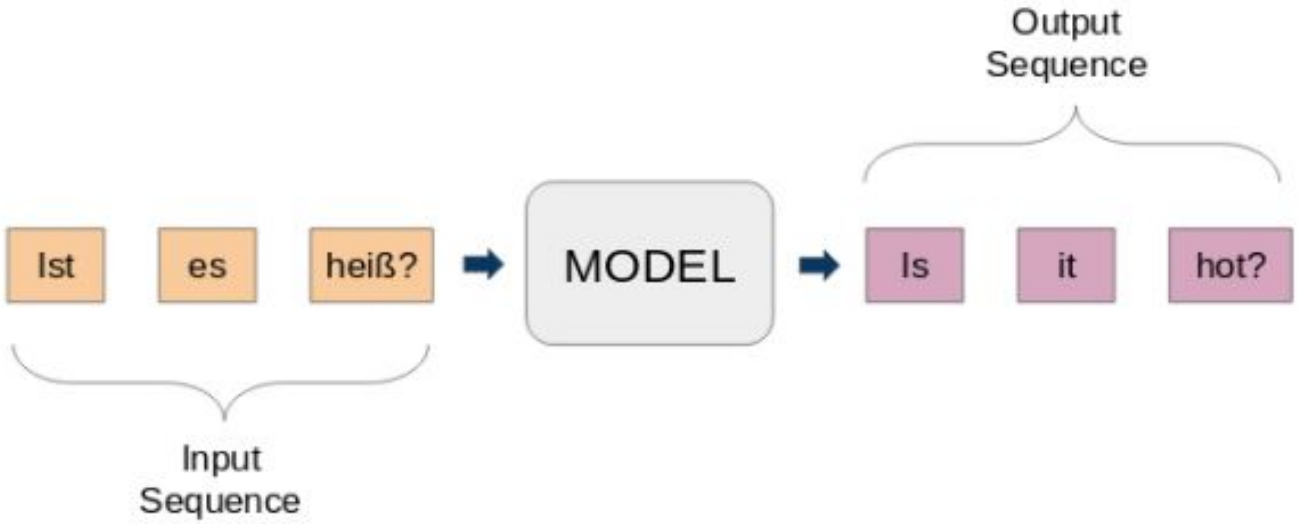
Language Modeling

- A language model essentially predicts a sequence of words given a text
- **Examples:** Auto-complete in messenger apps, email, search engines

Language Modeling

- A language model essentially predicts a sequence of words given a text
- **Examples:** Auto-complete in messenger apps, email, search engines
- Model Architectures:
  - RNN based
  - LSTM or GRU based

Sequence-to-Sequence Modeling



Sequence-to-Sequence Modeling

- Both the input and output are sequences

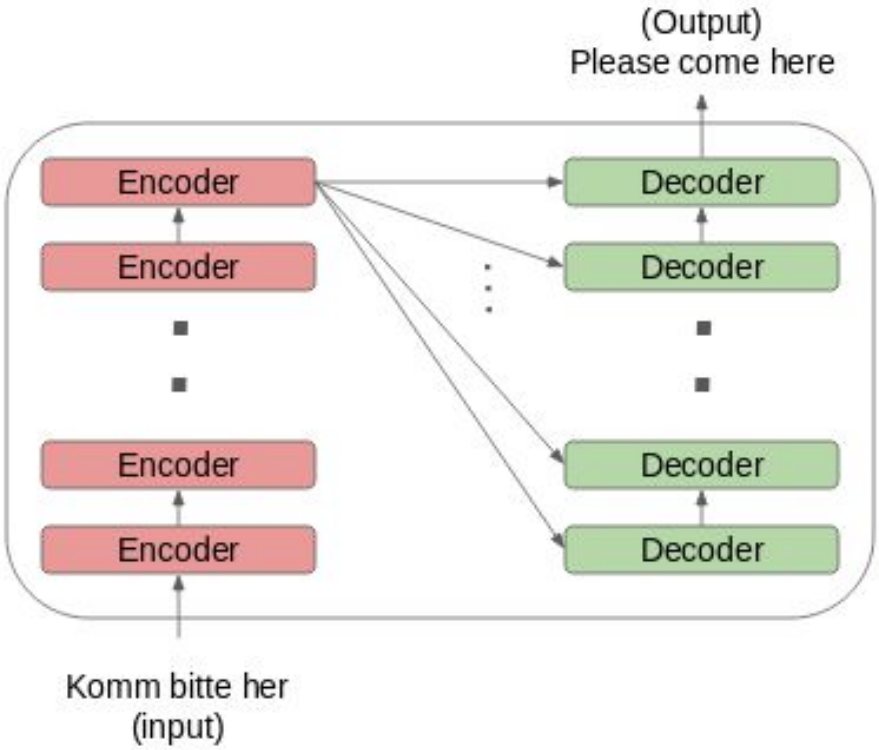
Sequence-to-Sequence Modeling

- Both the input and output are sequences
- Input and Output sequences can have different lengths

Sequence-to-Sequence Modeling

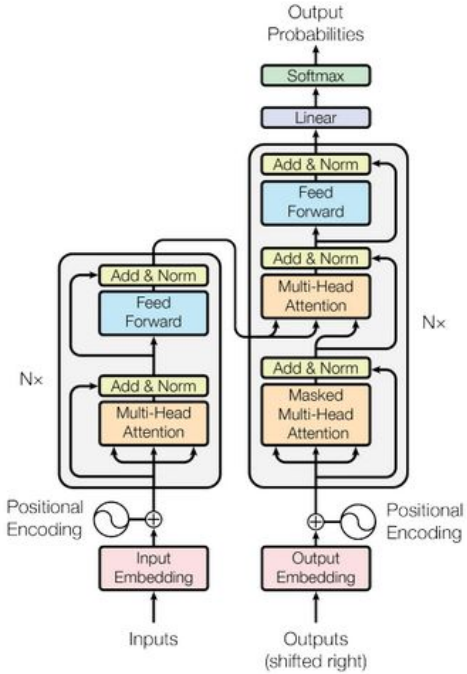
- Both the input and output are sequences
- Input and Output sequences can have different lengths
- Real-world use cases:
  - Machine Translation
  - News Headline Generation
  - Text Summarization

Transformer Models



Transformer Models

- Perform Language modeling and sequence-to-sequence tasks

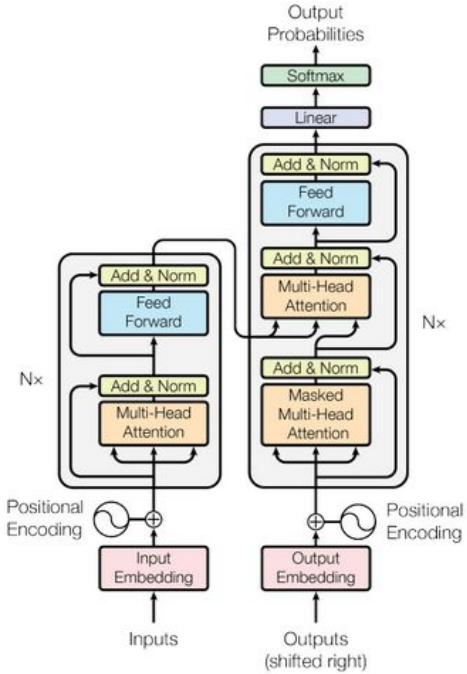


The Transformer Model

(Source: arxiv.org/abs/1706.03762)

Transformer Models

- Perform Language modeling and sequence-to-sequence tasks
- Does not use RNN or LSTM

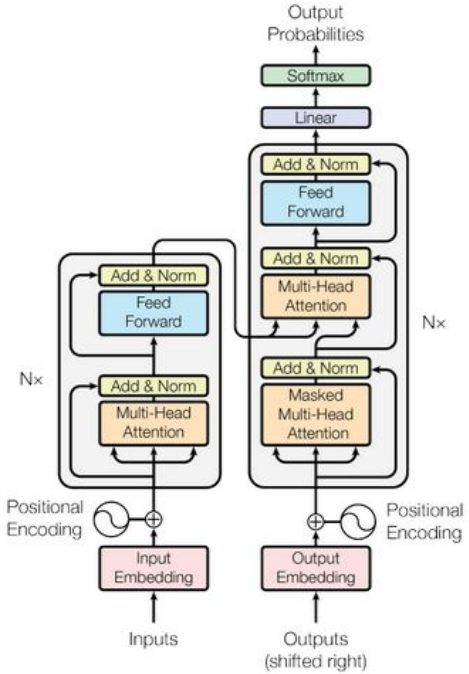


The Transformer Model

(Source: arxiv.org/abs/1706.03762)

Transformer Models

- Perform Language modeling and sequence-to-sequence tasks
- Does not use RNN or LSTM
- State-of-the-Art models in NLP
  - BERT
  - GPT-2
  - XLNet
  - RoBERTa



The Transformer Model

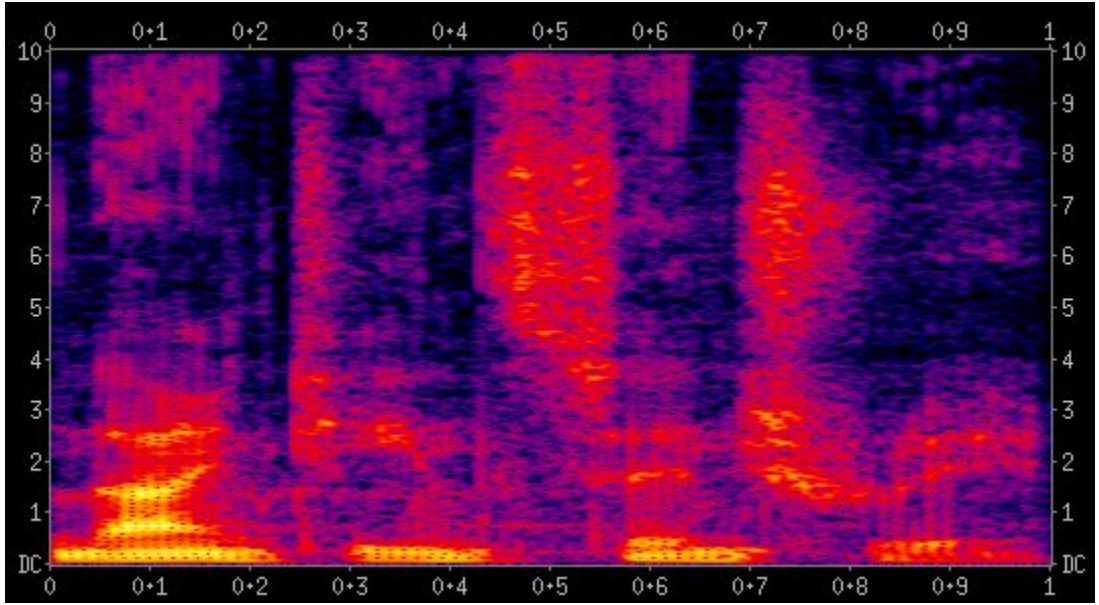
(Source: arxiv.org/abs/1706.03762)

Transformer Models

- Perform Language modeling and sequence-to-sequence tasks
- Does not use RNN or LSTM
- State-of-the-Art models in NLP
  - BERT
  - GPT-2
  - XLNet
  - RoBERTa
- Transformers library by **HuggingFace**
  - Pre-trained models
  - Model fine-tuning

huggingface.co

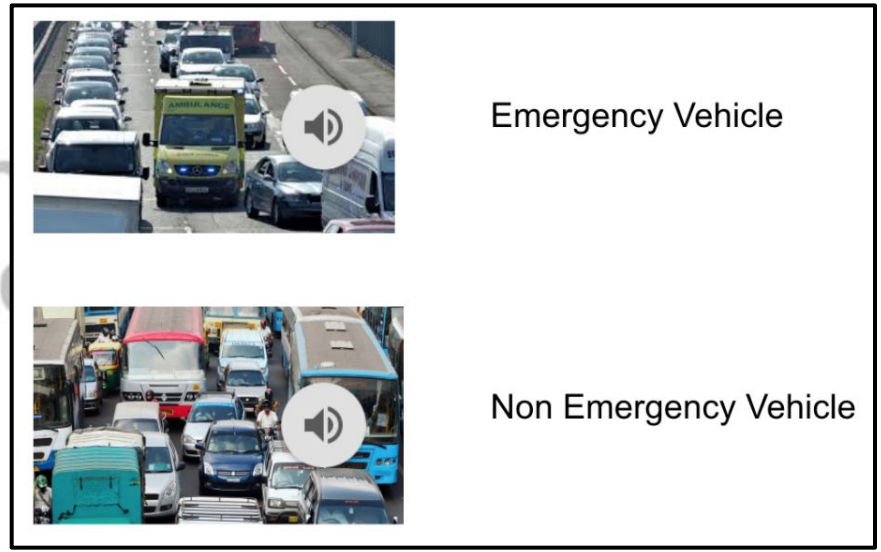
Audio Processing and Modeling



Audio Processing and Modeling

- Emergency vs Non Emergency Audio Classification

- Audio Features:
  - Time domain features
  - Spectrogram Features



Audio Processing and Modeling

- Automatic Speech Recognition (ASR)
    - Extract information from speech
    - Convert speech to text
- Siri, Alexa, Google Assistant

Source: cloud.google.com

Audio Processing and Modeling

- Automatic Speech Recognition (ASR)
    - Extract information from speech
    - Convert speech to text
- Siri, Alexa, Google Assistant
- State-of-the-art Speech Recognition Models
    - Deep Speech 2
    - Wave2Vec
    - Wave2Letter

Source: cloud.google.com

Image Captioning

A young boy is playing basketball. 	Two dogs play in the grass. 	A dog swims in the water. 
A group of people walking down a street. 	A group of women dressed in formal attire. 	Two children play in the water. 

Image Captioning

- Combination of Computer Vision and Natural Language Processing

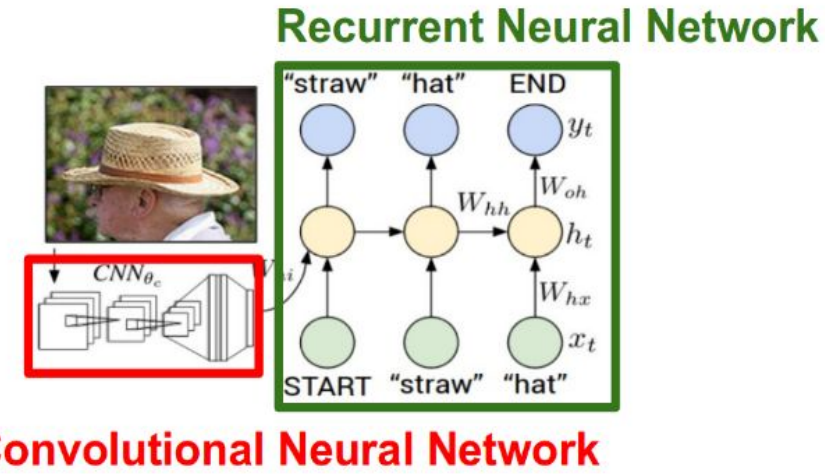


Image Captioning

- Combination of Computer Vision and Natural Language Processing
- Features are extracted from the input image.

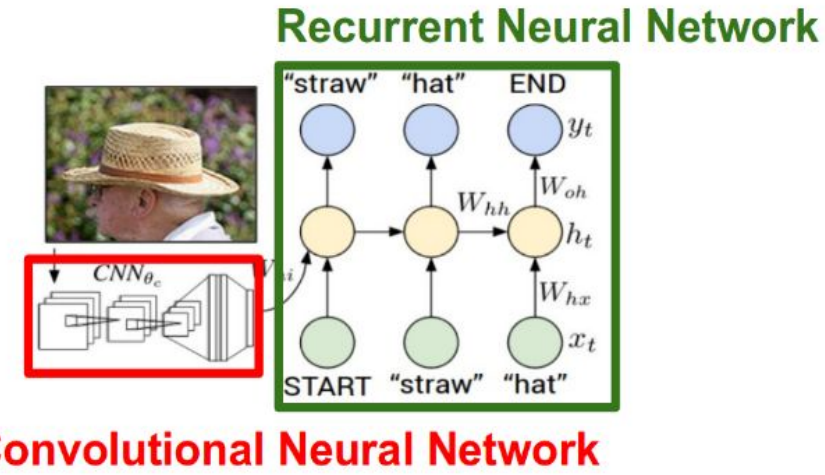
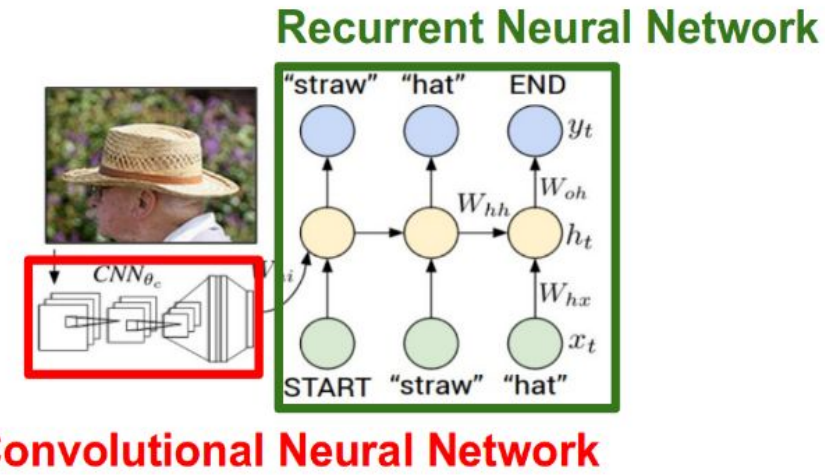


Image Captioning

- Combination of Computer Vision and Natural Language Processing
- Features are extracted from the input image.
- Extracted features are mapped to a natural language sentence or phrase



Learning Resources

arXiv.org

Thank You