[Show and Tell: A Neural Image Caption Generator](https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Vinyals_Show_and_Tell_2015_CVPR_paper.pdf)

<https://www.researchgate.net/publication/329037107_Image_Captioning_Based_on_Deep_Neural_Networks>

<https://ieeexplore.ieee.org/document/8697360/references#references>

<https://ieeexplore.ieee.org/document/8308186> (Understanding a convolution neural network)

[https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43905.pdf](https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43905.pdf(LSTM)) (LSTM)

Image captioning is the process of automatically describing the content of an image and is a fundamental problem in artificial intelligence that connects computer vision and natural language processing[1].

**Introduction**

Being able to automatically describe the content of an image using properly formed English sentences is a very challenging task, but it could have great impact, for instance by helping visually impaired people better understand the content of images on the web.[1]

This task of automatically generating

captions and describing the image is significantly harder than image classification and object recognition. The description of an image must involve not only the objects in the image, but also relation between the objects with their attributes.[3]

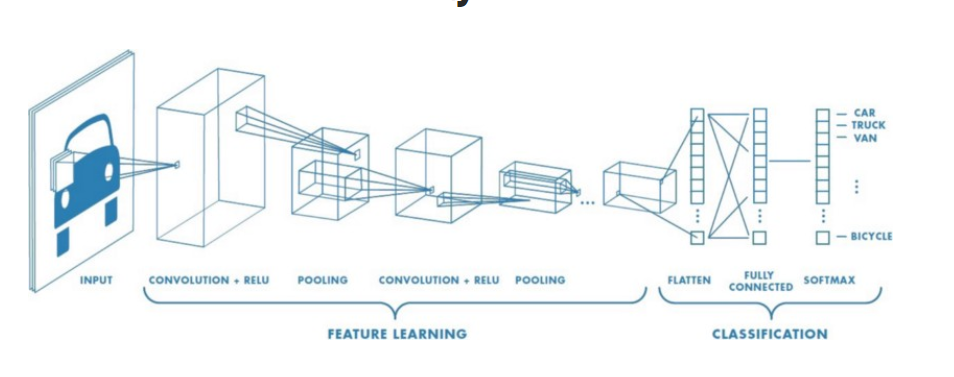
With extensive research the task of image captioning is now done using Artificial Neural Networks[4]. Artificial Neural Networks is one of the strongest tools in Machine Learning which can handle large amounts of data.recent years most image captioning methods employ a ConvolutionalNeural Network (CNN)[4] as the encoder and a Recurrent

Neural Network(RNN) as the decoder, especially Long Short-Term Memory(LSTM)[5] to generate captions, with the objective to maximize the likelihood of a sentence given the visual features of an image.

**Approach**

Since the task of image captioning can be divided into two parts, that is, Image processing and generating text based on the features identified. We use a combination of Convolution Neural Networks and Long Short Term Memory to carry out this given task.

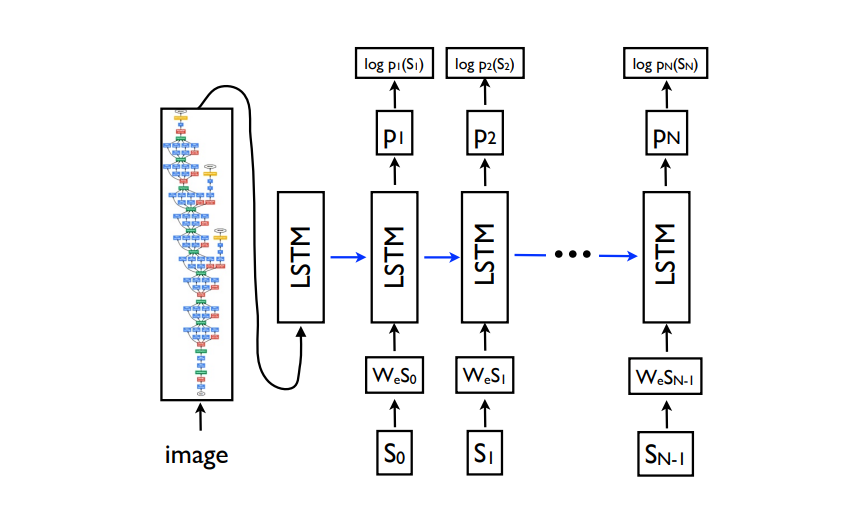
1. **Convolution Neural Network(CNN)**

****

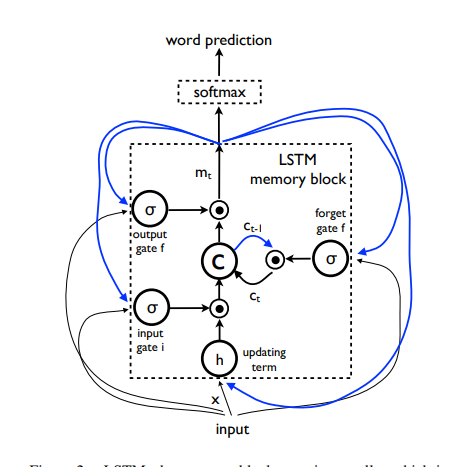
Convolutional Neural Network has had groundbreaking results over the past decade in a variety of fields related to pattern recognition; from image processing to voice recognition.

It has been convincingly shown that CNNs can produce a rich representation of the input image by embedding it into a fixed-length vector.

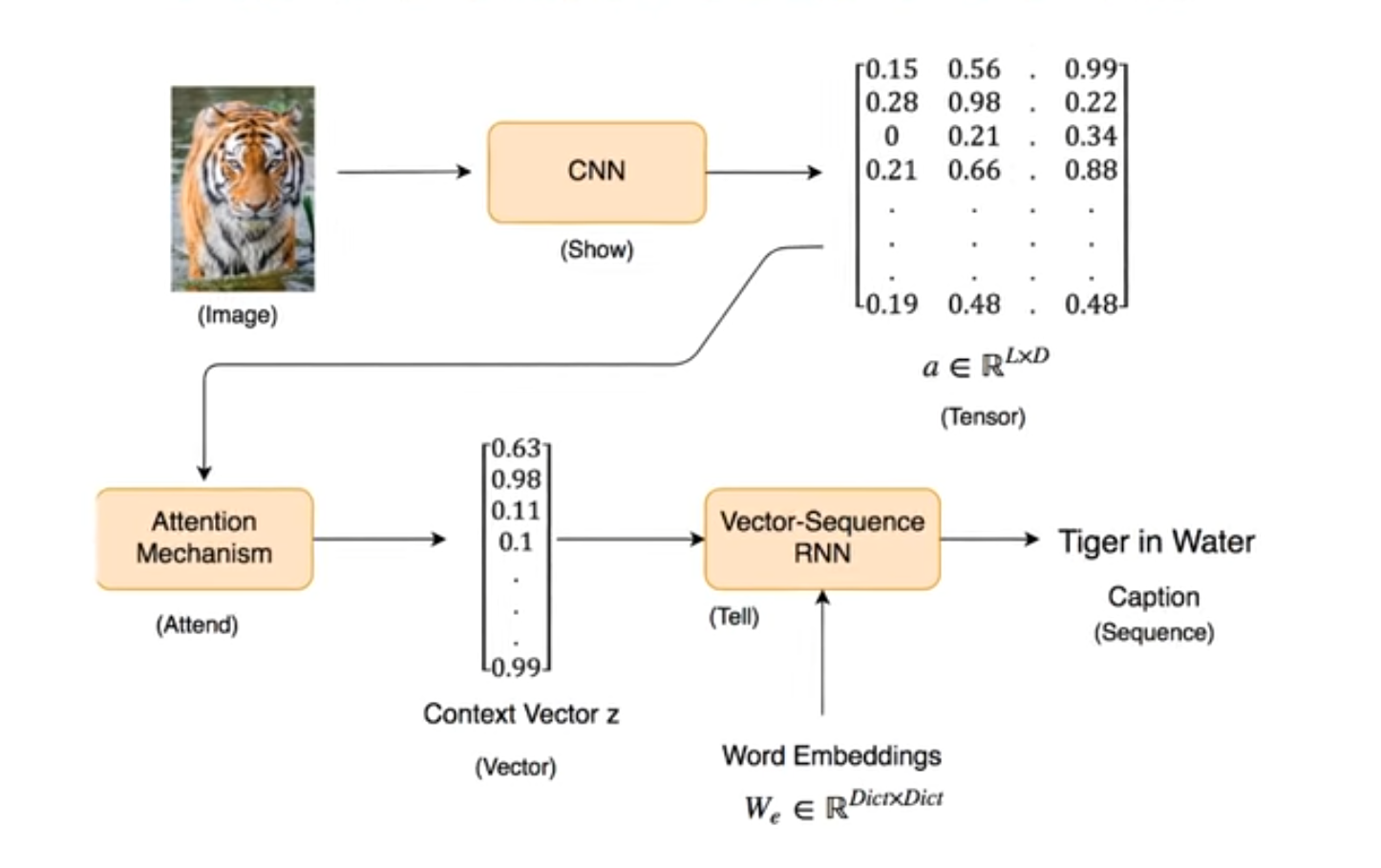
1. **Long Short Term Memory (LSTM)**

****

Long Short-Term Memory (LSTM) is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. The core of the LSTM model is a memory cell c encoding knowledge at every time step of what inputs have been observed up to this step. The behaviour of the cell is controlled by “gates” – layers which are applied multiplicatively and thus can either keep a value from the gated layer if the gate is 1 or zero this value if the gate is 0



**Final Architecture**



**Dataset Used**

The dataset used is MSCO

**Conclusion**

Image captioning has made significant advances in recent

years. Recent work based on deep learning techniques has

resulted in a breakthrough in the accuracy of image

captioning. The text description of the image can improve

the content-based image retrieval efficiency, the

expanding application scope of visual understanding in the

fields of medicine, security, military and other fields,

which has a broad application prospect. At the same time,

the theoretical framework and research methods of image

captioning can promote the development of the theory and

application of image annotation and visual question

answering (VQA), cross media retrieval, video captioning

and video dialog, which has important academic and

practical application value.