



ભારતીય માહિતી તંત્રજ્ઞાન સંસ્કૃત રાયચૂર  
ભારતીય સૂચના પ્રૌદ્યોગિકી સંસ્થાન રાયચૂર  
Indian Institute of Information Technology Raichur

# Image Captioning Using LLM

**a group of young boys playing soccer on a field.**



Instructor: Dr. Neha Agarwal

**What Do You See in The Picture??**



When we look at an image, we effortlessly come up with descriptions like

- "A boy in a brown jacket holds a large chunk of snow"
- "A boy wearing blue mittens gets ready to throw snow"
- "A child preparing to throw a snowball"
- "A young boy grins as he holds up a clump of snow" or
- "A young boy in a brown jacket playing with snow" .

These captions, though different, all accurately depict the image. It's a natural skill for humans, even a child can do it effortlessly. However, can we replicate this ability in a computer program? Can we create Model that takes an image as input and generates a relevant caption? Do You See in The Picture?



# Exploring Diverse Image Datasets

## **Dataset Selection: Flickr 8k**

Utilizing the Flickr 8k dataset provides access to a rich collection of 8,000 images, offering a wide range of scenarios and contexts essential for image captioning projects.

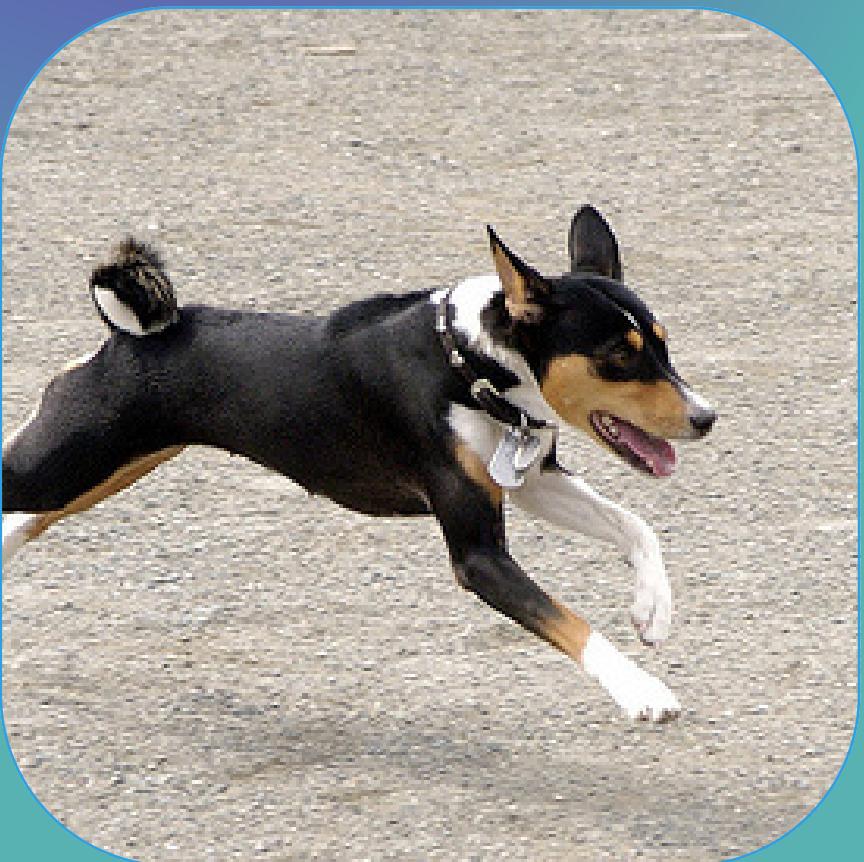
**Flickr 8k (8,000 images)**

**Flickr 30k (30,000 images)**

**MS COCO (180,000 images)**

# Example

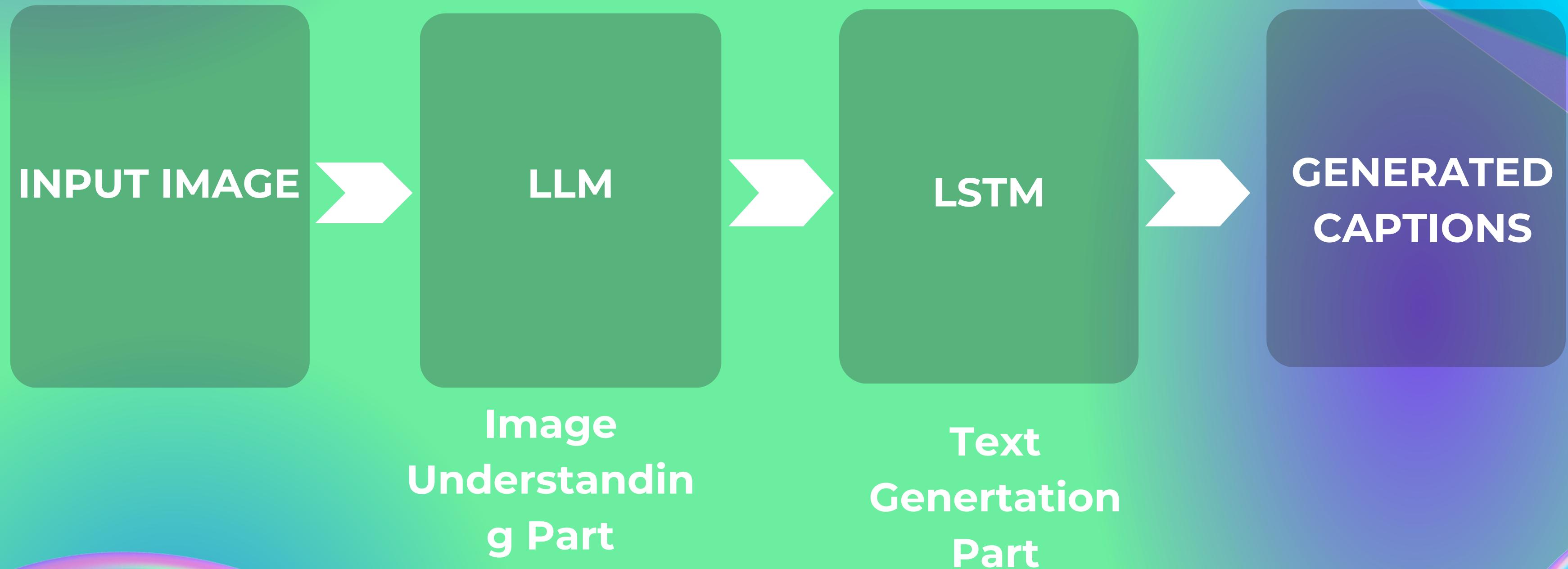
## Image



## Caption.txt

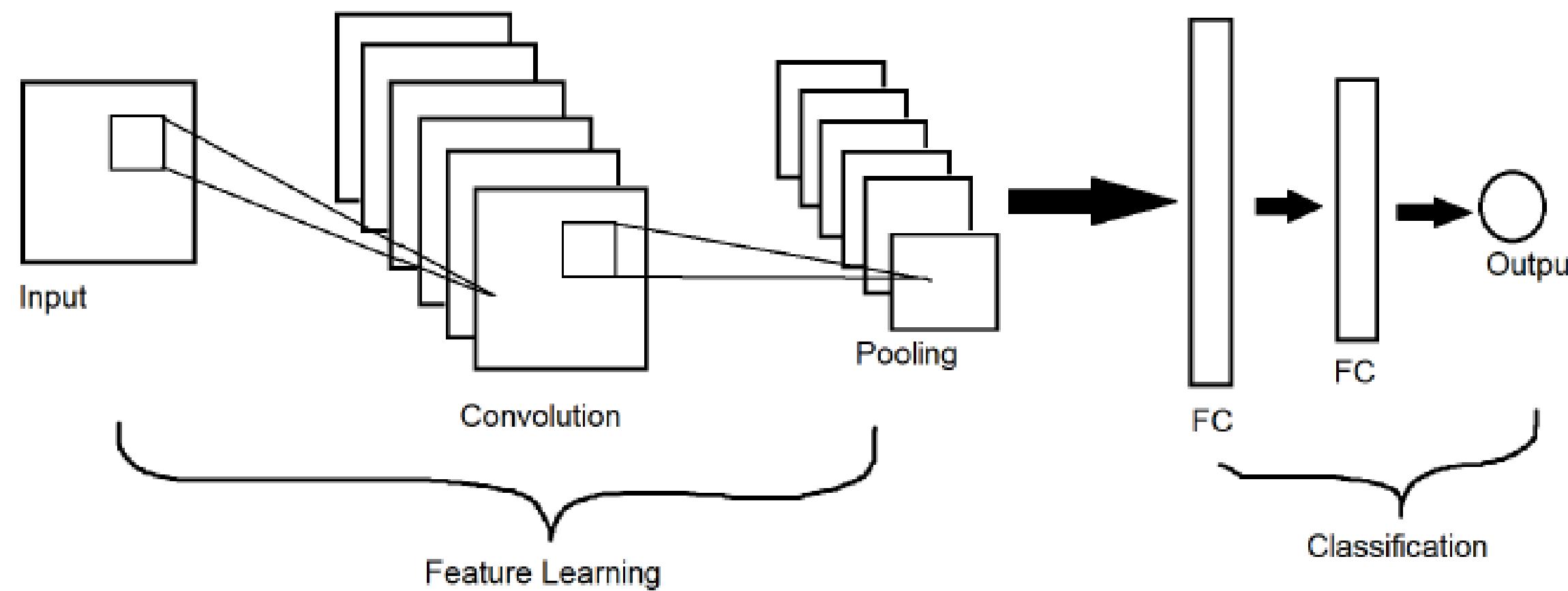
```
image,caption
1000268201_693b08cb0e.jpg,A child in a pink dress is climbing up a set of stairs in an entry way .
1000268201_693b08cb0e.jpg,A girl going into a wooden building .
1000268201_693b08cb0e.jpg,A little girl climbing into a wooden playhouse .
1000268201_693b08cb0e.jpg,A little girl climbing the stairs to her playhouse .
1000268201_693b08cb0e.jpg,A little girl in a pink dress going into a wooden cabin .
1001773457_577c3a7d70.jpg,A black dog and a spotted dog are fighting
1001773457_577c3a7d70.jpg,A black dog and a tri-colored dog playing with each other on the road .
1001773457_577c3a7d70.jpg,A black dog and a white dog with brown spots are staring at each other in the street .
1001773457_577c3a7d70.jpg,Two dogs of different breeds looking at each other on the road .
1001773457_577c3a7d70.jpg,Two dogs on pavement moving toward each other .
1002674143_1b742ab4b8.jpg,A little girl covered in paint sits in front of a painted rainbow with her hands in a bowl .
1002674143_1b742ab4b8.jpg,A little girl is sitting in front of a large painted rainbow .
1002674143_1b742ab4b8.jpg,A small girl in the grass plays with fingerpaints in front of a white canvas with a rainbow on it .
1002674143_1b742ab4b8.jpg,There is a girl with pigtails sitting in front of a rainbow painting .
1002674143_1b742ab4b8.jpg,Young girl with pigtails painting outside in the grass .
1003163366_44323f5815.jpg,A man lays on a bench while his dog sits by him .
1003163366_44323f5815.jpg,A man lays on the bench to which a white dog is also tied .
1003163366_44323f5815.jpg,a man sleeping on a bench outside with a white and black dog sitting next to him .
1003163366_44323f5815.jpg,A shirtless man lies on a park bench with his dog .
1003163366_44323f5815.jpg,man laying on bench holding leash of dog sitting on ground
1007129816_e794419615.jpg,A man in an orange hat staring at something .
1007129816_e794419615.jpg,A man wears an orange hat and glasses .
1007129816_e794419615.jpg,A man with gauges and glasses is wearing a Blitz hat .
1007129816_e794419615.jpg,A man with glasses is wearing a beer can crocheted hat .
1007129816_e794419615.jpg,The man with pierced ears is wearing glasses and an orange hat .
1007320043_627395c3d8.jpg,A child playing on a rope net .
1007320043_627395c3d8.jpg,A little girl climbing on red roping .
1007320043_627395c3d8.jpg,A little girl in pink climbs a rope bridge at the park .
1007320043_627395c3d8.jpg,A small child grips onto the red ropes at the playground .
1007320043_627395c3d8.jpg,The small child climbs on a red ropes on a playground .
1009434119_febe49276a.jpg,A black and white dog is running in a grassy garden surrounded by a white fence .
```

# Flowchart For Model



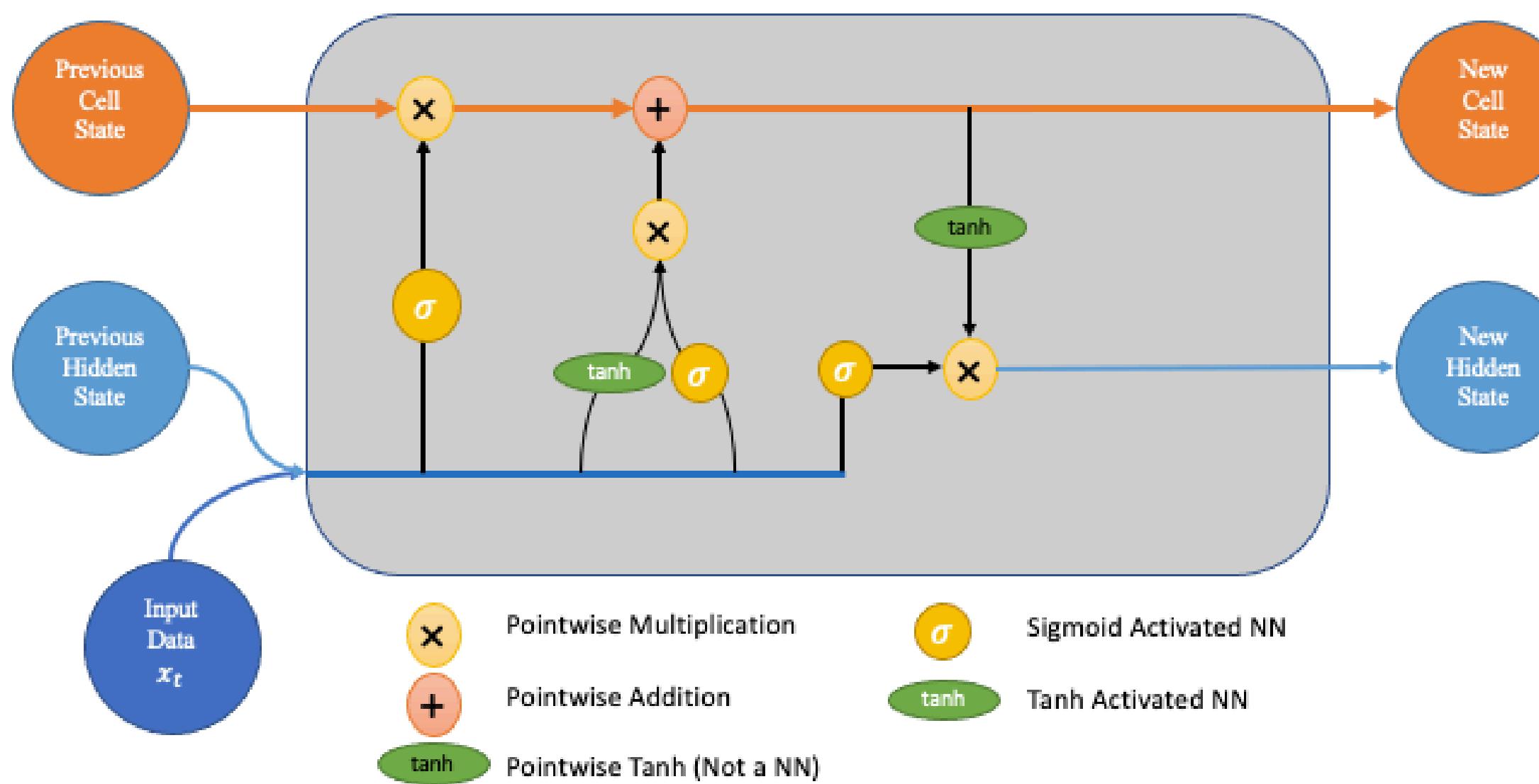
# CNN

**Convolutional Neural Network (CNN) is a type of deep learning model for processing data that has a grid pattern, such as images.**

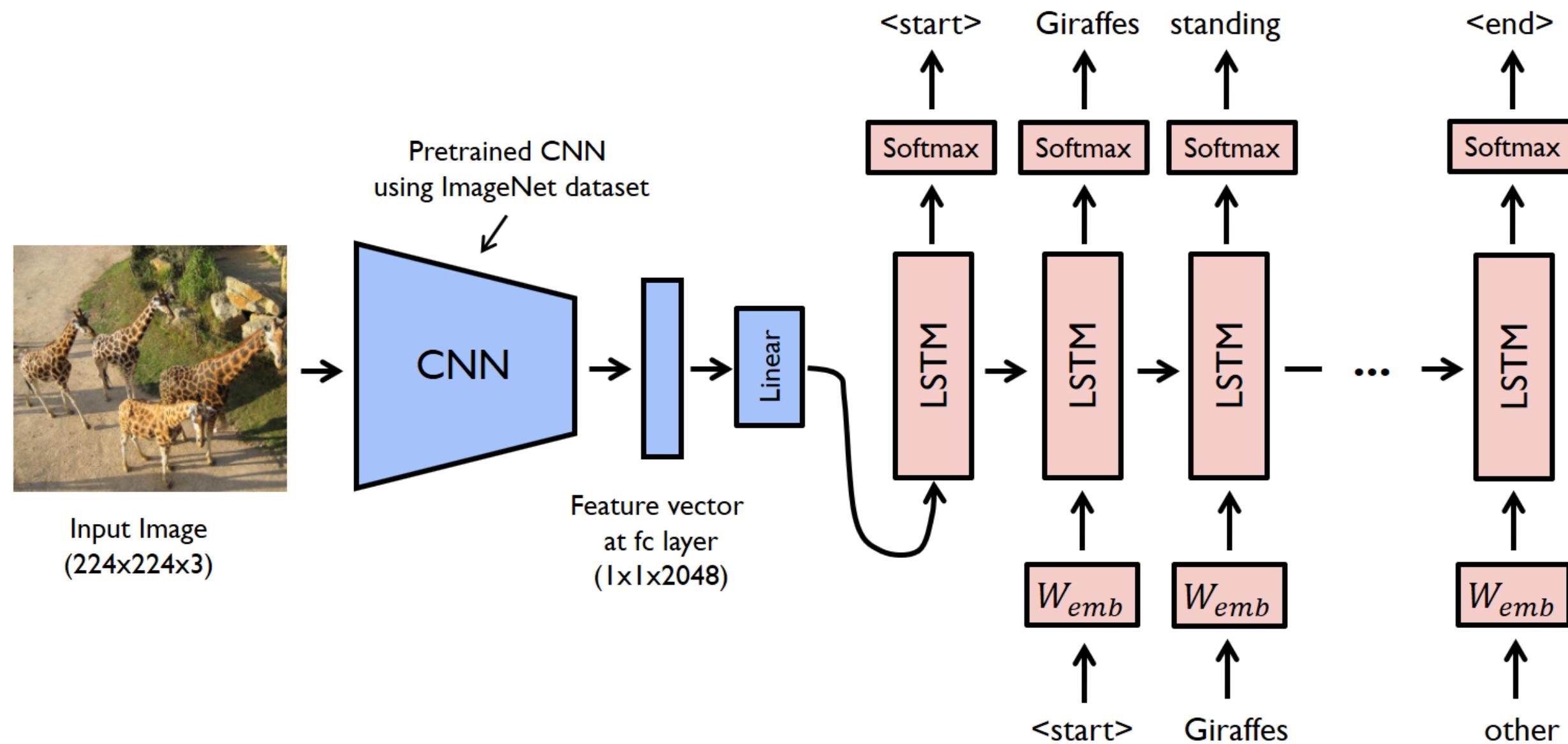


# LSTM

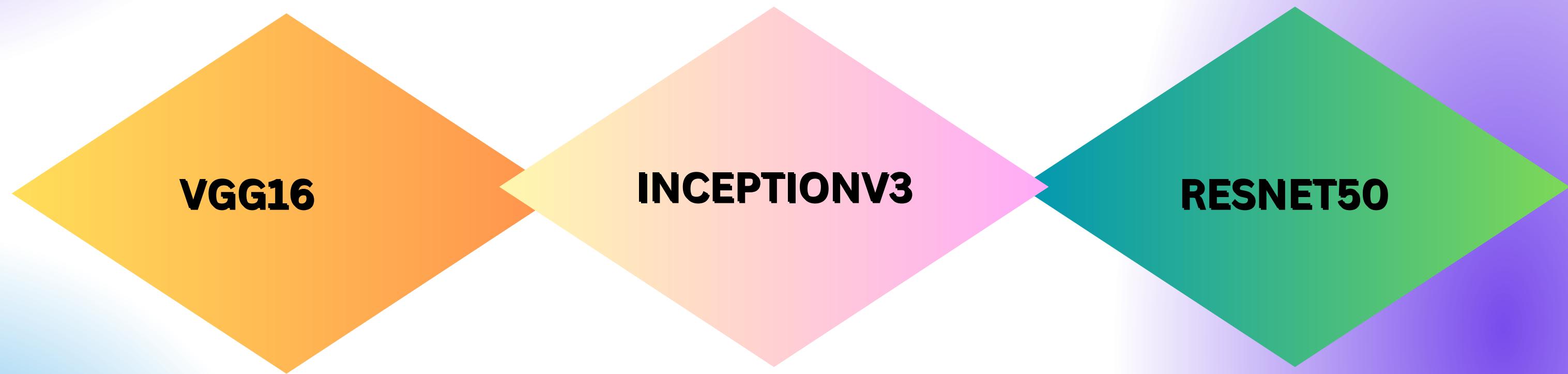
**LSTM networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more.**



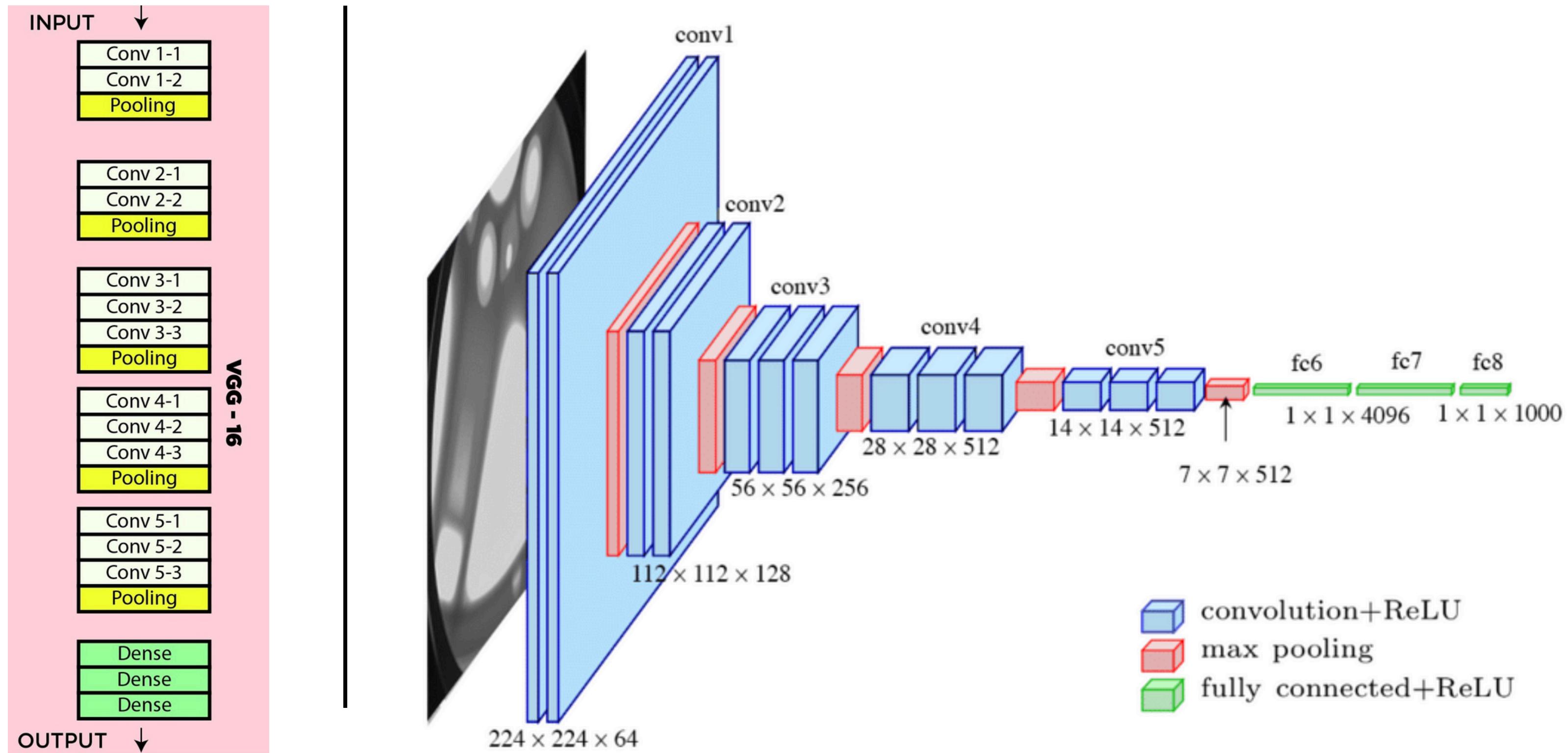
# MODEL OF IMAGE CAPTIONING



# PRETRAINED CNN

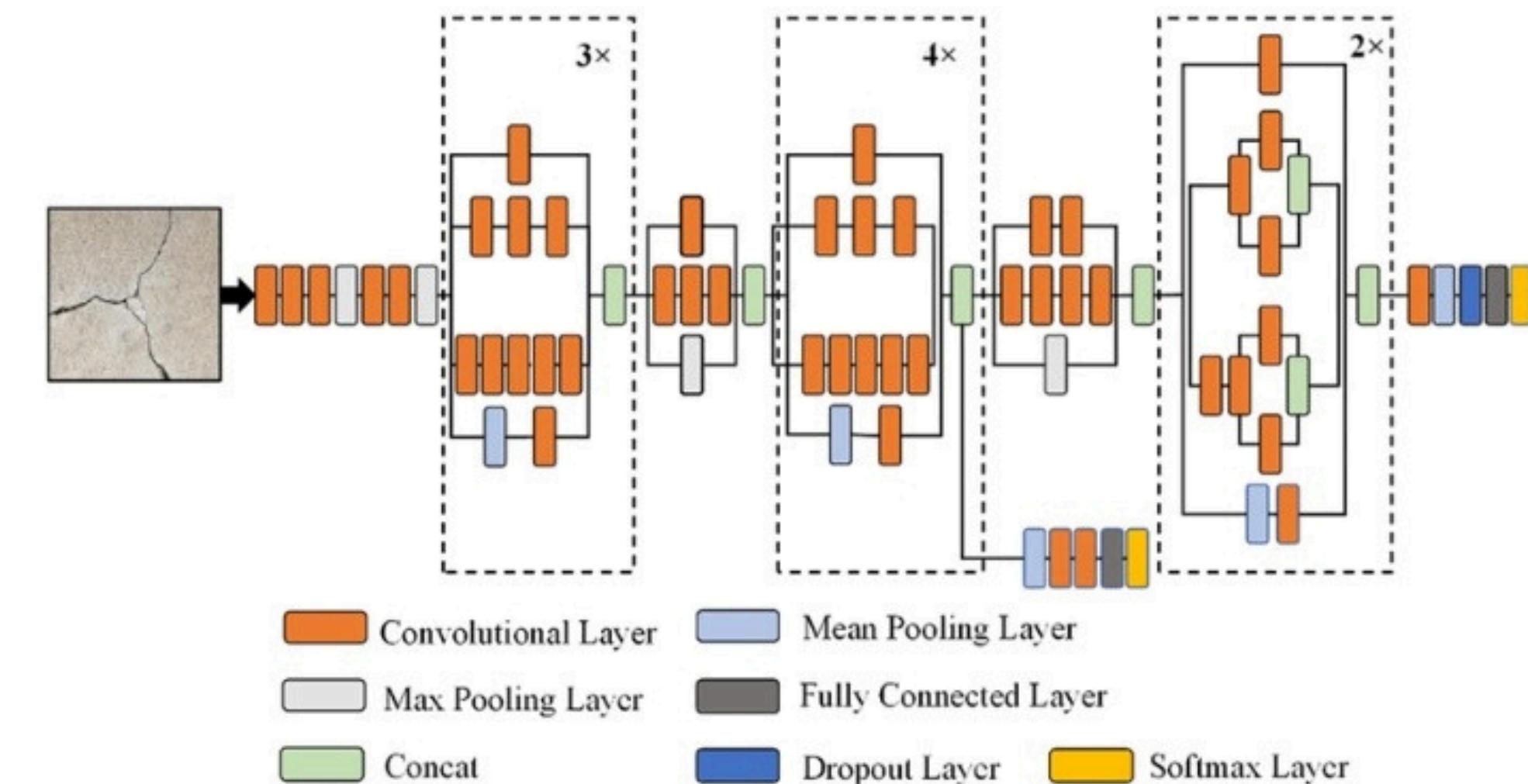
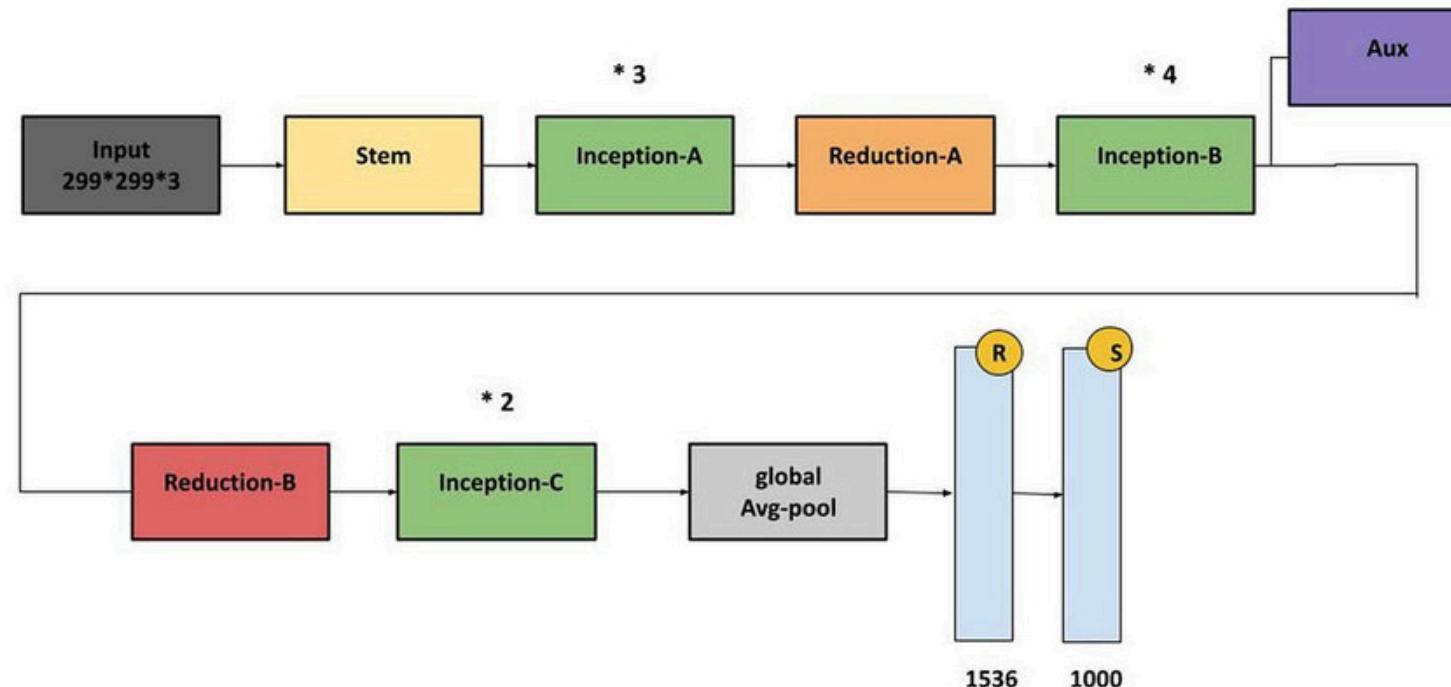


# VGG16

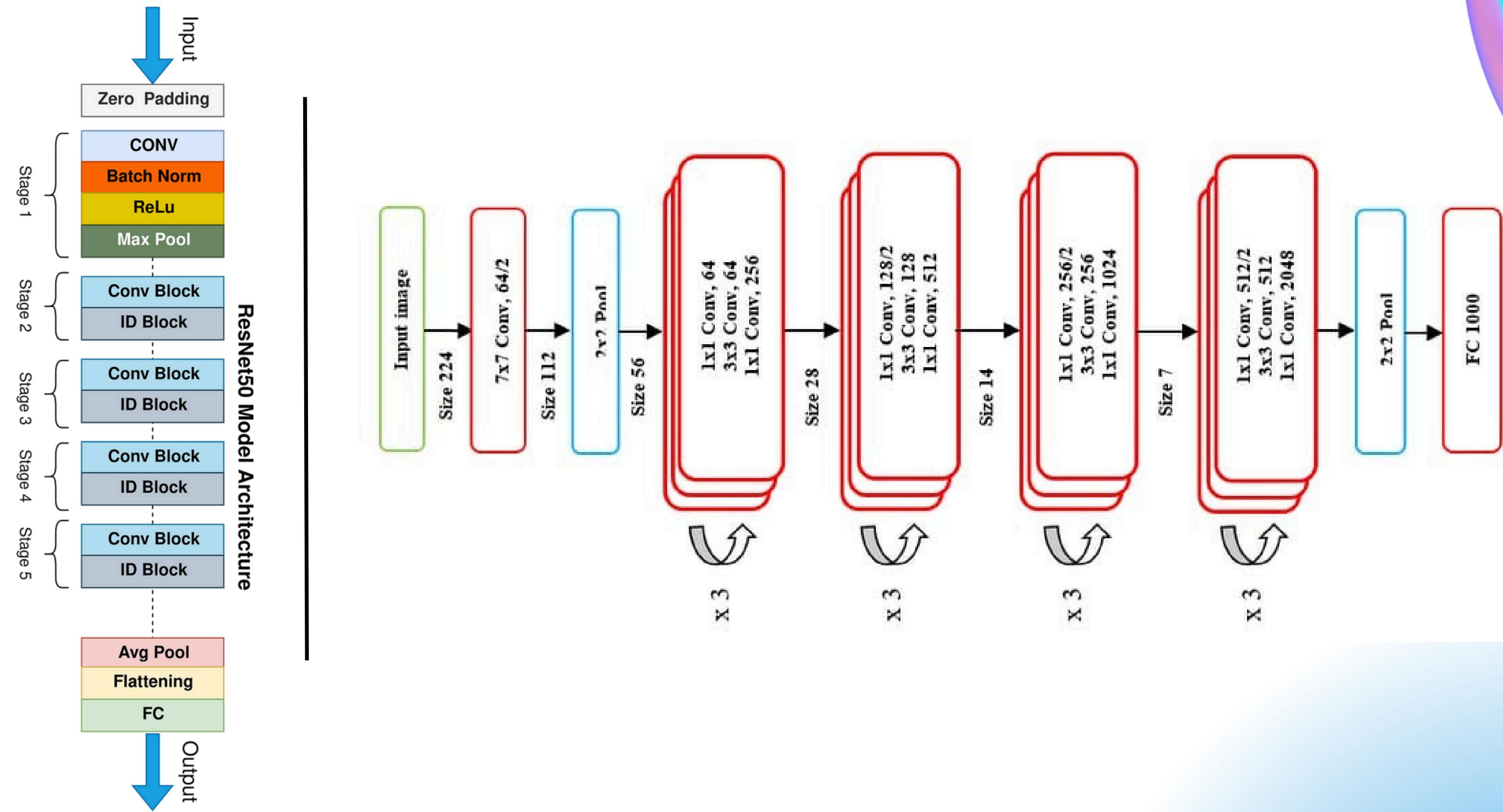


# INCEPTION-V3

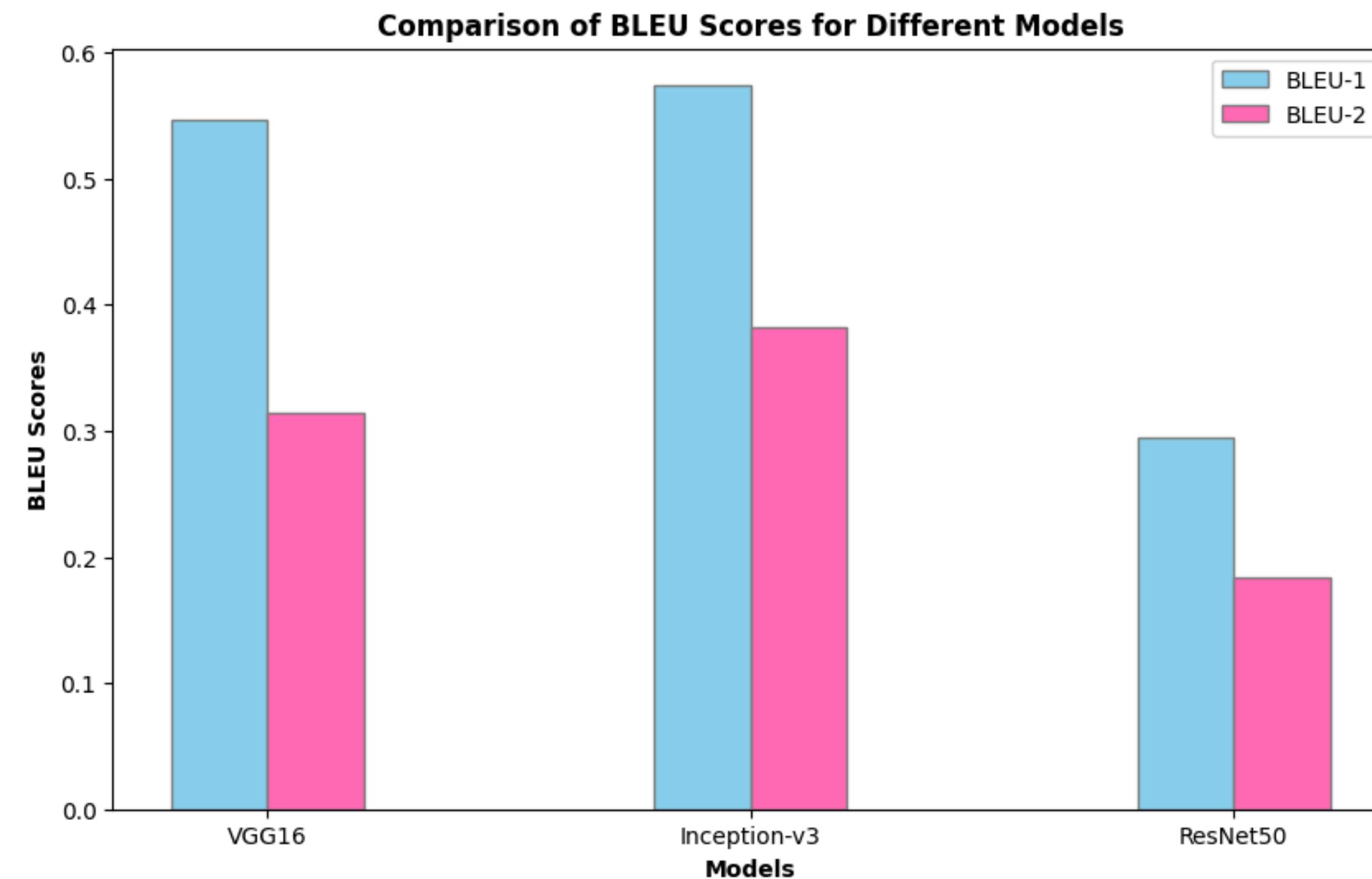
Inception V3



# RESNET50



# VGG16, INCEPTION-V3, RESNET50



# STREAMLIT UI INTERFRANCE

Image Caption Generator with CNN & LSTM

Choose an image...

Drag and drop file here  
Limit 200MB per file • JPG, JPEG, PNG

Browse files

10815824\_2997e03d76.jpg 60.9KB

X

Image uploaded successfully.



Generated Caption:

two horses are looking at barrel

# THANK YOU