

Image Captioning

Team - The Scalars

Team members :-

Ishita Gupta
Raina Dsouza
Saeed Bachute
Sania Valiyani
Urja Tendolkar

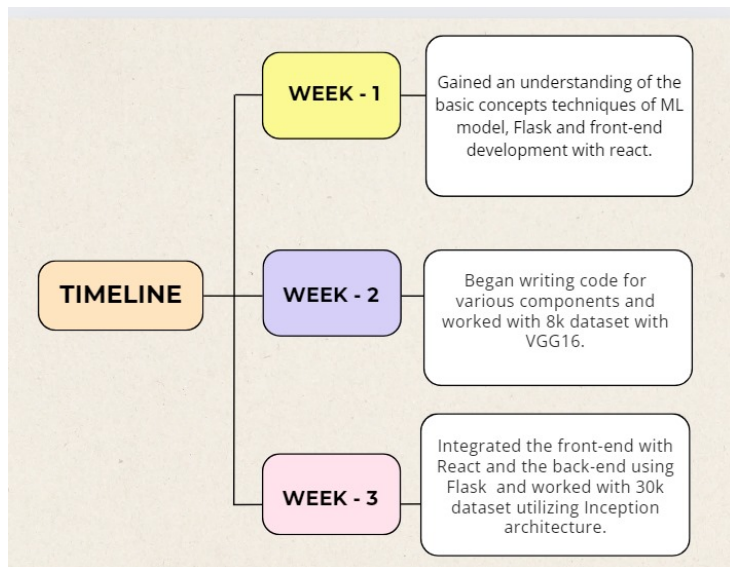
July 26, 2024

Objective

The objective of our project was to develop an image captioning website that generates a one-line description when a picture is uploaded.

- Python
 - TensorFlow
 - VGG16 model
 - Inception model (GoogLeNet)
- Frontend
 - React
 - Tailwind
 - Email JS
- Backend
 - Flask

Timeline Chart

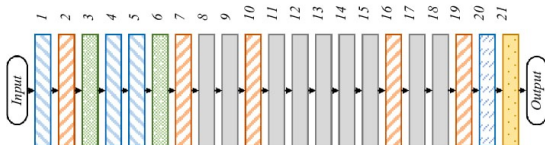


VGG16 (Visual Geometry Group 16)

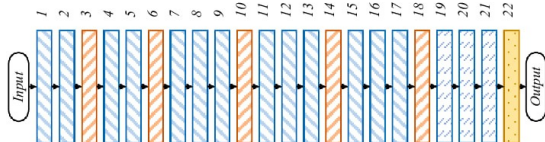
- 1) VGG-16 achieved 92.7% top-5 test accuracy on the ImageNet dataset
- 2) Depth: VGG-16 has 16 layers, including 13 convolutional layers and 3 fully connected layers.
- 3) Uniformity: All convolutional layers use a small filter size (3x3) with a stride of 1 and same-padding.

Model Explanation

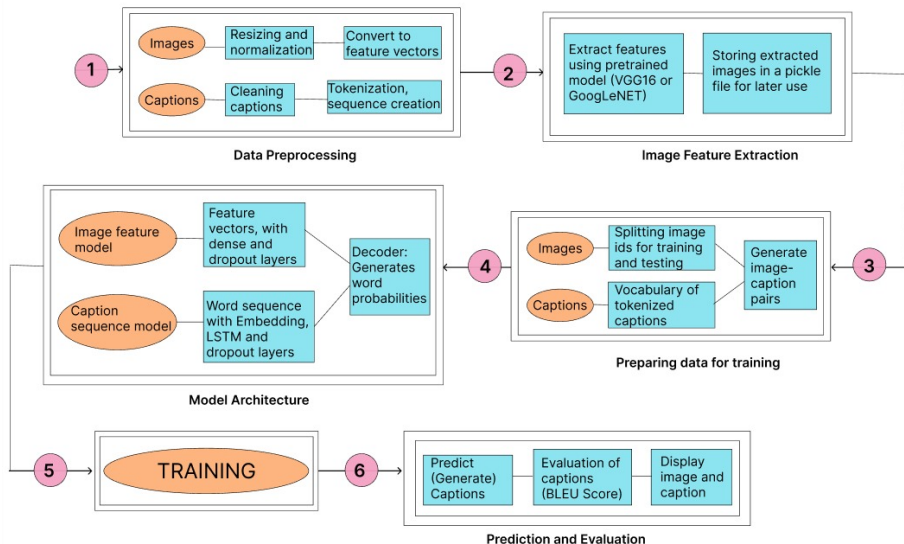
GoogLeNet



VGG16



Block diagram of model



- Explored topics in machine learning, deep learning, and natural language processing.
- Learnt about different pre-trained models and their differences.
- Learned how to integrate frontend and backend.
- Cooperation with teammates, used Git, and learned how to debug Git errors.

Challenges

Challenges faced by us were :-

- Faced several Git issues
- Improper format of CSV files
- Memory issues and speed constraints (absence of GPU)
- Integration of frontend and backend
- Overfitting and inaccurate BLEU scores
- Implementation of model with Scala

- ① Model Enhancement
 - Attention Mechanism
 - Data Augmentation
 - Beam Search
- ② Documentation

Thank You !