



Image Captioning Using NLP and Deep Learning

Done by :
Renad Albishri
Shahad Alkaltham

TABLE OF CONTENTS

Problem Statement

01

Tools

02

Data Story

03

04

Transfer Learning

05

07

Model
Architecture

Greedy Search

06

Final Results

01



02

03

04

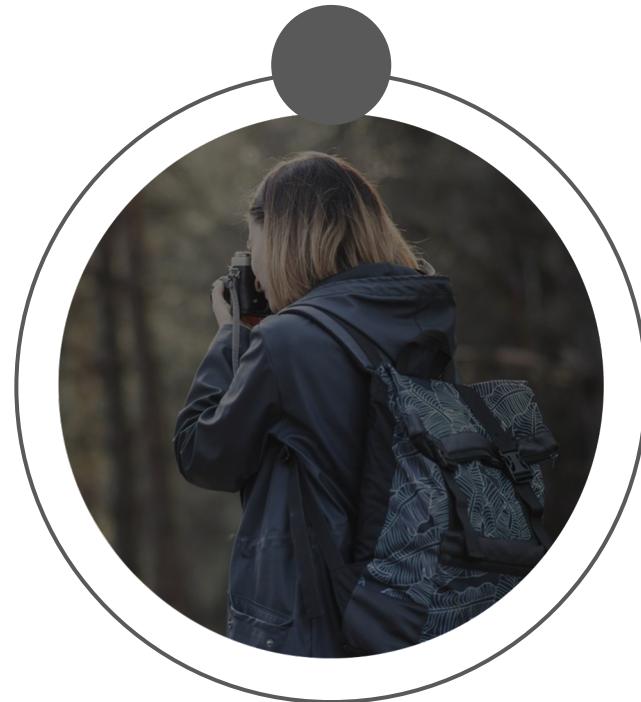
05

06

07

Problem Statement

This project aims to generate a textual description for images through Neural Networks and NLP, for reducing the time consumed on generated manually.



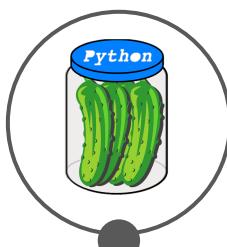
Tools



Sklearn



Keras



Pickle



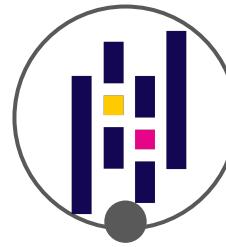
TensorFlow



Flask



NumPy



Pandas

Data story



39,874
Rows

- Combine Flicker8k ,
Flicker30k from
Kaggle.
- Each image with 5
Captions (199,370)



Captions Cleaning

- Remove Digits,
Punctuations, tabs.
- Convert it to lower case.
- Adding ‘startseq’ ,
‘endseq’.

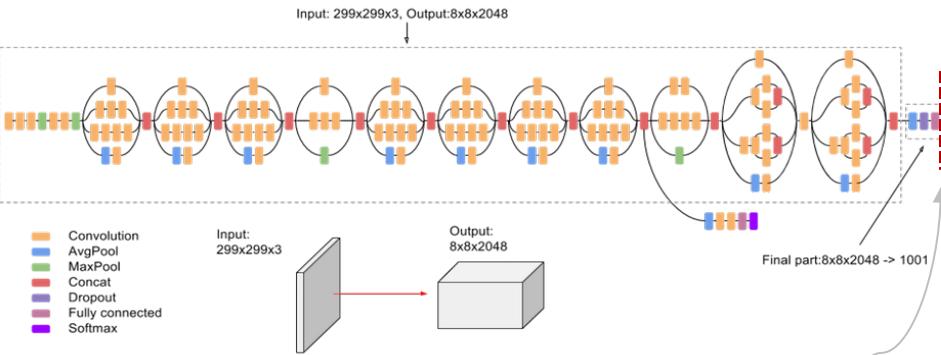


Dataset Problem

▪ | → ,

Transfer Learning

InceptionV3

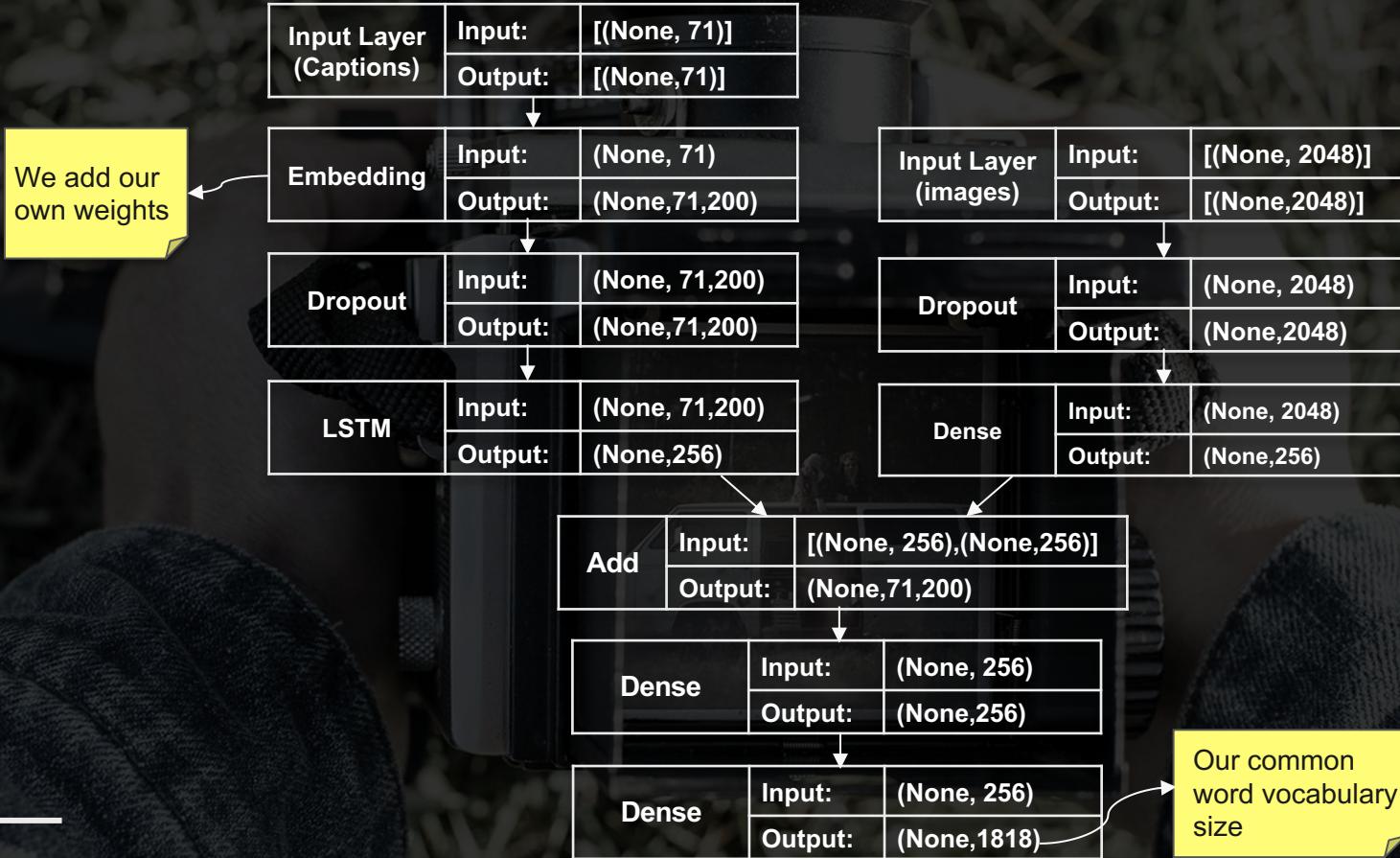


- Image Recognition Model
- Trained on **ImageNet** dataset
- Has Accuracy **78.1%**
- We take **Bottleneck Features (2048,1)**
- We Freeze the layers
- We convert our image to (299,299)
- We encode our image using it

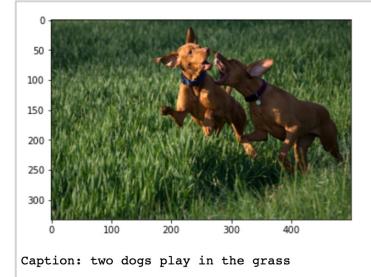
GloVe

- Unsupervised Learning Algorithm.
- Obtaining vector representations for words.
- Map it with our common word vocabulary.
- **200D Dense Vector** obtained for each word.

Model Architecture



Greedy Search



Caption: two dogs play in the grass

Iteration	Inputs	\hat{y}
1	 startseq	two
2		dogs
3		play
4		in
5		the
6		grass
7		endseq

Cosine Similarity

```
evaluate_model()
```

```
text_to_vector(txt)
```

```
get_avg_cosine(actuals, generated)
```

```
get_cosine(vec1, vec2)
```

Final Model Scores

Cosine Similarity: 0.3429 ← ON sample 500 only

Training Accuracy: 50.59% ← ON sample of 5,000

Final Results



Caption: a boy swims in a pool of water

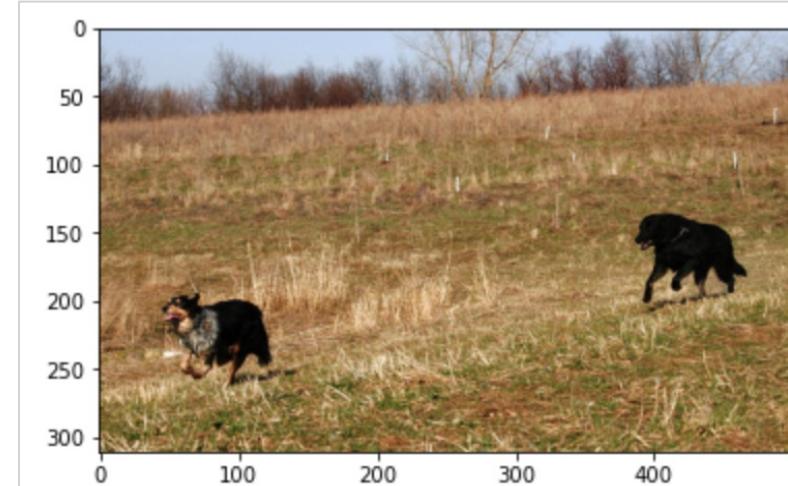


Caption: a black and white dog swimming in the ocean

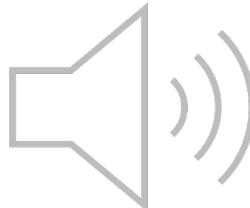
Final Results



الوصف: كلبان يركضان عبر العشب



Caption: two dogs run through the grass



Challenges

- The output of dataset is **sequence** not **classes**.
- The long time of fitting the model.
- Trying to work on **Google Cloud Platform** and **AWS**.
- Trying to implement android application.

Future Work

We would like to work on smart app to help blindness and low vision people to feel like others by using a camera of it.



Thank you!

