

Final Year Project Proposal

Muhammad iftikhar ,Aimen Wadood, Hamza Majeed.

{p180054, p180002, p180128}@nu.edu.pk

Suggested Supervisor:

Faculty Member's Name: Muhammad Amin

Signature: _____

Date (17 September 2021)

Project Details

Project Title	<i>Image and Video Caption Generator</i>		
Project Area of Specialization	Deep learning, Computer Vision, Natural language processing		
Project Start Date	2021-9-17	Project End Date	2022-09-17
Project Summary (less than 2500 characters)	Our project automatically Generate the Caption of the content observed on the image and video in the form of a Natural language process. Image and video Caption Generator has very Applications for example enriching and created new image Datasets, human-reboot interaction, enhancing the power of google search engine and enhance the functionality of the system similar to google photos.		
Project Objectives (less than 2500 characters)	<ol style="list-style-type: none">1. This model will be made in order to help visually impaired people to have knowledge of pictures so it is necessary that the model generate captions that is close to human language, the captions should be grammatically accurate and this accuracy will be measured through BLEU score.2.The model will be made in order to help the separation of content from multiple videos.3.The model will be made in order to enhance the power google search engine inside the content of the image and video.4: The model will be made in order to increase human-reboot interaction.		
Project Implementation	Deep Learning: CNN Model, LSTM		

Method (less than 2500 characters)	Object/Detection: Computer Vision Language Transformation: NLP Rest API: Python, Flask/Django, Azure
Benefits of the Project (less than 2500 characters)	The application of Image and video Caption Generator has many applications like enhancing the power of google search engine, and helping impaired people, human-robot interaction, enriching and created new images dataset, enhancing the functionality of system similar to google photos.
Technical Details of Final Deliverable (less than 2500 characters)	Image and Video caption Generator recognizes image Frame by frame. It will have two components: Encoder and Decoder. Encoder utilizes Convolutional Neural Network (CNN), which extracts objects and features from image or video frame. Decoder utilizes a neural to generate a natural sentence based on available information. CNN will be used to recognize image. Recurrent Neural Network (RNN) will be used in speech recognition, video activity recognition. Long Short-Term Memory (LSTM) networks will be used for learning order dependence in sequence prediction problems. Video caption generator will also use these models to generate caption.
Final Deliverable of the Project	Developing a Web application Where user can upload their images and Our System will generate the image caption.
Type of Industry	Artificial intelligence, Data Science
Technologies	Python, OpenCV. TensorFlow, GitHub, Keras, Django, VM, GPU
Sustainable Development Goals	Developing a system that will help the impaired people and blind people and also the power existing system like google search engine and enhance the human-robot interaction.

Project Key Milestones

Elapsed time in (days or weeks or month or quarter) since start of the project	Milestone	Deliverable
FYP-1	1st 4 Months	Model Trained on Image Caption
FYP-2	2nd 4 Months	Deployed model on both image and Video caption

Project Equipment Details

Item Name	Type	No. of Units	Per Unit Cost (in Rs)	Total (in Rs)
Azure Credits	Web service	2	500	1000
			Total in (Rs)	1000