**PROJECT PROPOSAL**

**Proposal :**  Visual Captioning and Neural machine translation

**Project Proposal Description:**

The idea is to develop an application that can generate a brief textual description of the content of an image and translate it to user specified language among the available language options provided to them. Visual captioning would require the combination of computer vision and natural language processing. Neural Machine Transalation can be achieved using Neural Networks and Natural Language Processing. The Challenge here is to preserve the meaning of the image description while translating it from English to other languages. The other challenge is to choose a diverse image dataset to build the application to cover multiple disciplines.

**Project Title and Team Members:**

| Name | Email Id | Student Id |
| --- | --- | --- |
| Aniv Chakravarty | anivchakravarty@my.unt.edu | 11424113 |
| Priya Enuganti | priyaenuganti@my.unt.edu | 11508112 |
| Shridhar Kshirsagar | ShridharKshirsagar@my.unt.edu | 11541383 |
| Pankaj Yadav | pankajyadav@my.unt.edu | 11486346 |

**Motivation :**

* Differently abled people who are visually challenged with incapacities such as color blindness, night blindness, etc. have a hard time understanding visuals.
* Most of the image captioning applications today conform to a certain standard language. English is preferred the most due to its standard and globally accepted grammar and vocabulary. Parts of the world, where English isn’t a primary language, people may find it difficult to decipher the summary generated by such image captioning applications which are trained using English corpus.
* Human summarization of visuals falls into a subjective territory as it may add up a potential bias depending on the perception of every individual. This is also the reason manual captioning becomes obsolete.

**Significance:**

* Summarizing the visual scenes that visually challenged people come across in everyday life and later translating the summary into audio later might help them lead a better life.
* The application will naturally exceed human capacity of captioning as a machine can caption a large number of images in seconds depending on the available processing power.
* By removing the human bias, we attain a certain standard for summarizing texts. This makes the final product less ambiguous to the user.

**Objectives :**

* Our overall objective is to build a model capable of generating a text description of an image first into english and then translate it into another language. We split up our main objective into multiple sub objectives to complete in order to achieve it. These sub objectives are as follows:
* **Increment 1 Objectives**:
* Identifying a proper image captioning and machine translation datasets for training the object description model.
* Identify any pre-existing models and methods along with their documented performance.
* Build basic working model of somewhat coherent text description
* **Increment 2 Objectives**:
* Achieve relevant machine translation into the target language.
* Improve performance and accuracy of the built methods if possible.

**Features :**

Will be exploring the COCO dataset and OID dataset shared by google for Object detection/Captioning of Images. Will have following features

* Object identification with instance annotations
* Context recognition
* Dense Pixel to identify vectors
* High number of labeled images
* 1.5 Mio object instances
* Multiple categories of object
* Many stuff categories which has boundary to build context
* pre-trained key points

**References :** [**https://ieeexplore.ieee.org/document/9197977**](https://ieeexplore.ieee.org/document/9197977)

[**https://viso.ai/computer-vision/coco-dataset/**](https://viso.ai/computer-vision/coco-dataset/)

[**https://datagen.tech/guides/image-datasets/ms-coco-dataset-using-it-in-your-computer-vision-projects/#:~**](https://datagen.tech/guides/image-datasets/ms-coco-dataset-using-it-in-your-computer-vision-projects/#:~)**:**

[**https://www.researchgate.net/publication/317420233\_Image\_Captioning\_with\_Object\_Detection\_and\_Localization**](https://www.researchgate.net/publication/317420233_Image_Captioning_with_Object_Detection_and_Localization)

[**https://arxiv.org/abs/1810.04101**](https://arxiv.org/abs/1810.04101)