# CSCE 5290: Natural Language Processing

# Project Proposal

## **GitHub repository Link:**

<https://github.com/nagasai-gummadi/CSCE_5290_Project>

## **Project Title:**

## Image Captioning by Using Auto encoders

## **Project Proposal Description:**

The Project idea is to generate Textual description of an image based on user stated language from the list of available languages. For image captioning we need computer vision and natural language processing techniques. The main challenge in this project is translating textual image description from one language to another language which means that doing translation from English to other languages. For captioning and summary of the image we are going to use auto-encoders.

## **Team Members:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Email\_Id** | **Student\_Id** |
| Nagasai Gummadi | [nagasaigummadi@my.unt.edu](mailto:nagasaigummadi@my.unt.edu) | 11540905 |
| Divya Sri Vakkala | [divyasrivakkala@my.unt.edu](mailto:divyasrivakkala@my.unt.edu) | 11590101 |
| Arun Kumar Gangireddy | <arunkumargangireddy@my.unt.edu> | 11656859 |
| Mohammed Abdul Moin | [abdulmoinmohammed@my.unt.edu](mailto:abdulmoinmohammed@my.unt.edu) | 11508271 |

## **Motivation:**

* People who are facing vision issues find it difficult to visualize the images.
* There are several image captioning applications available but every application translates text data into English but in some places where English is not native language people may face understanding issues of data description.
* Every individual person is having their own perception of image captioning after seeing the visual and that translation may result in incorrect conclusion.

## **Significance:**

* Translating the text summary into audio will help people with vision issues.
* If we have large number of images we have to caption and translate it into text as it is complicate for humans which may take so much of time but if you process those images through machine it will process those in quick time.
* By avoiding manual processing of captioning and text translation we will get more accurate product.

## **Objectives:**

* The Main Object is to build an application that should be able to create image description into English after that we have to translate it into user specified language. Here we separated this project into 2 parts.
* **Increment-1:**
* In this we are going to identify and render appropriate image captioning and translation datasets for training the model.
* Identify the previous available model and their performance.
* Finally, we are going to build consistent text description model.
* **Increment-2:**
* Translate text description into user specified language.
* Based on final accuracy going to improve accuracy, performance of model.

## **Features:**

We are analysing the COCO and Open image dataset. The dataset contains fallowing features

* Labeled images
* Around 1 million images
* Image identification with annotations
* Different types of images
* For identifying the vectors that has dense pixels are available.

## **References:**

<https://ieeexplore.ieee.org/abstract/document/9071372>

<https://proceedings.neurips.cc/paper/2016/hash/eb86d510361fc23b59f18c1bc9802cc6-Abstract.html>

<https://ieeexplore.ieee.org/abstract/document/10054038>

<https://github.com/ankushjain2001/Image-Captioning-with-Autoencoders>