Study Materials for REU Students in Natural Language Processing and Computer-Vision Applications

**Research Objective:** The objective of this study document is to help the REU students to understand and familiarize themselves with multi-modal computer vision and NLP-based research applications. The document enlists the resources to learn the basic commands of GitHub, Python, ML open-source codebases. It also enlists a few interesting high-level research problems such as:

- Visual and Audio Navigation - Rescue Mission, Self-driving, etc.
- Scene Understanding from the video and associate a descriptive text. This can be applied across various fields such as sports analytics, smart home and smart health.
- VQA (Visual Questioning Answering) system for physically challenged people, smart flood systems, etc.

For better understanding of the research area and various state-of-the-art algorithms, please refer to this GitHub repository: [Link](#)

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 1</strong></td>
<td>Understand the basic commands of GitHub and ML open-source libraries.</td>
</tr>
</tbody>
</table>
| **Background** | 1.  [GitHub](#)  
2.  [PyTorch](#)  
3.  [TensorFlow](#)  
Learn how to use basic GitHub, ML open sources using any programming language. |
| **Week 2 - 3** | Study in-depth about various deep learning architectures and its functions (loss, hyperparameters, training and testing setup, etc).  
1.  [Basic image processing techniques](#)  
2.  [Basic understanding of Transformers](#)  
3.  [Basic Architecture of Seq-Seq translation](#)  
4.  [Basic understanding of LSTM](#)  
5.  [Other Popular Deep Learning Architectures](#)  
Gain hands-on experience by implementing various deep learning architectures and pre-processing techniques to design research frameworks or algorithms. |
| Week 4 - 6 | Read through state-of-the-art literature in CV and NLP application areas.  
1. [End-to-End Video Captioning](#)  
2. [End-to-End Video Captioning with Multitask Reinforcement Learning](#)  
3. [Deep Learning Contextual Models for Prediction of Sport Events](#)  
4. [Knowledge-Based Video Question Answering with Unsupervised Scene Descriptions](#)  
5. [VideoBERT: A Joint Model for Video and Language Representation Learning](#)  
6. [Attention-Based Multimodal Fusion for Video Description](#)  
7. [Long-term Recurrent Convolutional Networks for Visual Recognition and Description](#)  
8. [Fine-grained Video Captioning for Sports Narrative](#)  
9. [Sports Video Captioning via Attentive Motion Representation and Group Relationship Modeling](#) |
|---|---|
| Week 7 - 9 | Read through the following tutorial links, GitHub codes, and papers.  
Open-source references for NLP and CV applications  
1. [Deep Learning Contextual Models for Prediction of Sport Events (Code)](#)  
2. [Video and Text Captioning](#)  
3. [PyTorch Implementation Guide](#)  
4. [Implementation of Attention Mechanism](#)  
5. [Evaluation Metrics for NLP based algorithm](#) |
| Week 10 | Prepare the final report and presentation slides. |