DEV DISTRUCT

DATA SCIENCE ROLE

Assignment: Image-to-HTML Model Development

Submitted to: Dev Distruct
Deadline: 10th February 2025
Submission Requirements: Model file, GitHub repository, Implementation demonstration video in Google Colab, and this document.

Job Post: https://www.linkedin.com/jobs/view/4142984672

Objective

The goal of this assignment is to develop a deep learning model that takes an image as input and generates the corresponding HTML code. The dataset used for this task is **WebSight** from Hugging Face, which consists of images and their corresponding HTML representations.

Dataset

- **Source:** HuggingFace WebSight Dataset
- **Structure:** Each entry consists of an image and its corresponding HTML code.
- **Preprocessing:** HTML needs to be tokenized before training.

Approach

Model Selection

You are free to use any technique to generate HTML from images. Some possible approaches include:

1. **Vision-Language Models:**
   - **CLIP + Transformer:** CLIP converts images to embeddings, followed by a fine-tuned GPT-2 or T5 model to generate HTML.
   - **Image Captioning Models:** Using models like BLIP or OFA, treating HTML as a caption.

2. **Deep Learning Architectures:**
   - CNN, RNN, or Transformer-based architectures to process image inputs and generate HTML sequences.
   - Any hybrid or ensemble methods to improve accuracy.
3. **Other Techniques:**
   - Any custom pipeline that effectively maps images to HTML code is encouraged.

Implementation Steps

1. **Data Preprocessing**
   - Tokenize HTML using a transformer-based tokenizer.
2. **Model Training**
   - Use pre-trained models for feature extraction and sequence generation.
   - Train on a subset of the dataset for efficiency.
   - Use **Amazon SageMaker** or **Google Cloud Console** for training.
3. **Evaluation & Metrics**
   - Compare generated HTML with ground truth using BLEU score and token-level accuracy.
   - Check structural validity of the generated HTML.
4. **Deployment & Testing**
   - Save the trained model and provide a **Google Colab notebook** for demonstration.
   - Fine-tune the model for better accuracy.

Results & Findings

After training, we will compare different approaches to determine the most effective model for accurate HTML generation.

Submission Details

- **Model File:** Trained model checkpoint.
- **GitHub Repository:** Contains all source code, including preprocessing, training, and evaluation scripts.
- **Google Colab Demonstration:** Jupyter notebook for testing the model.
- **Implementation Video:** Explanation and demo of the model in action.
- **Final Report:** This document.

Conclusion

This project aims to develop an accurate image-to-HTML model using deep learning techniques. The final submission will include all necessary components to demonstrate the model's effectiveness and practical applications.

Prepared by: Nishkarsh
Email: nishkarsh1215@gmail.com