CS550/DSL501: Machine Learning (2024–25–M)
Project-II: Zero-Shot Text-to-Image Generation Using TurboSD and Text Ensembling
Progress Report

1 Project Title

Zero-Shot Text-to-Image Generation Using TurboSD and Text Ensembling

2 Group Name

Zeroshot Dreamers

3 Contributors

- **Nitin Mane (M24MT004)** - Project Lead
- **Himanshu Rana (12240700)** - Data Processing and Feature Engineering
- **Deepak Kumar (12240510)** - Feature Transformation
- **Mahesh Kesgire (M24MT001)** - Model Pipeline Development
- **Mohd abdul Mukheet (M24MT002)** - Model Integration and Inference

4 Progress Summary

4.1 Nitin Mane (Project Lead)

Contribution:
Nitin has coordinated the project phases, ensuring alignment between data collection, preprocessing, model development, and inference. He has structured the GitHub repository and facilitated the ongoing integration of various modules. The project is currently halfway complete, and further steps are being planned for fine-tuning and final model adjustments.

Current Progress:

- Organized and managed the GitHub repository.
- Coordinated team meetings and set deadlines for different components.
- Collaborated with Mahesh to prepare for the integration of the LoRA-based fine-tuning process into Stable Diffusion.

4.2 Himanshu Rana (Data Processing and Feature Engineering)

Contribution:
Himanshu has made significant progress in data collection and preprocessing. He has successfully completed the initial dataset loading and text preprocessing. His work is 50% completed, and he will proceed with further data augmentation and optimization in the next phase.

Current Progress:

- Loaded image-text pairs from Hugging Face (LAION-400M).
- Processed textual descriptions using the CLIPTokenizer for initial feature extraction.
- Next phase will involve more comprehensive data augmentation and final optimizations.

4.3 Deepak Kumar (Feature Transformation)

Contribution:
Deepak has made progress in transforming the text and image data into feature embeddings. He is currently at 50% completion, with feature extraction using CLIP successfully implemented. In the next phase, he will refine the transformation process and ensure optimal alignment between text and image embeddings.

Current Progress:

- Implemented CLIP-based text and image feature extraction.
- Collaborated with Himanshu to ensure the proper format of feature vectors.
- Will refine feature alignment and optimization in the next phase.

4.4 Mahesh Kesgire (Model Pipeline Development)

Contribution:
Mahesh has been working on setting up the model pipeline using Stable Diffusion with LoRA integration. While the base model architecture is ready, the next phase will focus on fine-tuning the model, applying text-ensembling techniques, and incorporating diffusion regularization methods. The model pipeline is approximately 50% complete.

Current Progress:

- Set up the basic Stable Diffusion model with LoRA integration.
- Implemented preliminary text ensembling techniques.
- Next steps include fine-tuning, diffusion style regularization, and content regularization.

4.5 Mohd Mukheet (Model Integration and Inference)

Contribution:
Mohd has partially integrated the model and set up initial inference tests. The current focus is on running test cases with preliminary results, which show promise but require further refinement. In the next phase, he will focus on completing the inference pipeline and optimizing the output for quality.

Current Progress:

- Integrated early versions of the model and ran initial inference tests.
- Preliminary visualizations are completed but require further adjustments.
- Will finalize the inference and output evaluation in the next phase.

5 Next Steps

1. Complete the integration of the TurboSD model with LoRA fine-tuning and text ensembling techniques.
2. Fine-tune the model and apply advanced diffusion regularization methods to improve image quality.
3. Perform detailed testing and evaluation using quantitative and qualitative metrics such as Fréchet Inception Distance (FID) and human evaluations.
4. Prepare the final report and presentation summarizing the project results and future directions.

6 GitHub Repository

For detailed code implementation and version control, please refer to the project's GitHub repository:
`https://github.com/Nitin-Mane/Zeroshot-Text-to-ImageAI`