[Samsung PRISM] Preliminary Discussion

SAMSUNG PRISM
PREPARING AND INSPIRING STUDENT MINDS

Deep Cache Implementation

Team

1. College Professor(s):
    1. Dr. G.K.Sandhia / sandhiag@srmist.edu.in
    2. Dr.D.Viji / vijid@srmist.edu.in
2. Students:
    1. Abhishek Soni / as8819@srmist.edu.in
    2. Arnav Agarwal / aa5579@srmist.edu.in
    3. Devanand K / dk7180@srmist.edu.in
    4. Kushagra Saxena / ks3780@srmist.edu.in
3. Department: CTECH

Date: 7th March 2024

DeepCache Implementation

SAMSUNG

Problem Statement

Context

DeepCache helps improve inference speed of Stable Diffusion model. It does this by skipping few steps in U-Net during inference. The results are faster with very little degradation in quality. We can have this worklet to compare this method as well as combining this with other methods to improve inference speed of SD.

Statement

Implement DeepCache on latest Stable Diffusion model.

Worklet Details

6

Duration (Months)

4

Members Count

Mentors

Pranal Prasad Dongare
+91-7022250561
pranal.p@samsung.com

Tushar Madaan
+91-9205301569
Tushar.m2@samsung.com

Pre-Requisite

- 2312.00858.pdf (arxiv.org)

Expectations

Undertaken Tasks

- Conduct Literature survey.
- Implement DeepCache.
- Identify whether LoRA technique post DeepCache implementation is feasible and helps in getting more inference time.

KPI

- DeepCache implementation
- Inference speed should have at least 25% improvement for similar prompt on vanilla Stable Diffusion model.
- LoRA implementation.

Timeline

Kick Off
< 1st Month >

- Problem Briefing
- Check Feasibility
- Literature Survey
- LLM Setup

Milestone 1
< 3rd Month >

- Initial implementation of architecture.
- Initial benchmarking of performance.

Milestone 2
< 6th Month>

- Optimization and enhancement of implemented model.

Complexity

Work-let Name: DeepCache Implementation

Worklet Details

1. Worklet ID: 23GAI30
2. College Name: SRM Institute of Science and Technology

Last Meet's Progress

- Last month, our team focused on understanding the core concepts outlined in the project worklet.
- We delved into implementing DeepCache in diffusion models and exploring LoRA technology integration.
- Key milestones achieved include:
    - Gaining familiarity with project terminology.
    - Establishing a foundational understanding of the project scope.
- These milestones signify significant progress and set the stage for upcoming tasks.

Date: 7th March 2024

Work-let Name: DeepCache Implementation

Worklet Details

1. Worklet ID: 23GAI30
2. College Name: SRM Institute of Science and Technology

KPIs achieved till now

- While specific KPIs were not outlined in the worklet, our progress aligns with our overarching goal of mastering the concepts essential for project implementation.
- We have diligently pursued learning objectives and have made substantial progress in understanding machine learning fundamentals.

Any Challenges/ Issues faced

- One of the primary challenges we encountered was the need to bridge the gap between our existing knowledge and the specialized requirements of the project.
- To address this, the team initiated a comprehensive learning program focused on machine learning basics.
- We also embarked on studying end-to-end ML projects to gain insights into project structure and best practices for maintaining a clean development environment.

Next Steps

- Moving forward, our focus will be on applying our newfound knowledge to practical project tasks.
- We plan to transition from theory to practice by initiating hands-on experimentation with DeepCache implementation and LoRA integration.
- Additionally, we aim to continue our learning journey, reinforcing and expanding our understanding of machine learning principles.

Key Achievements/ Outcome till now

- As of now, our primary achievement lies in the collective upskilling of the team in machine learning fundamentals.
- While tangible project results are yet to be realized, our enhanced understanding lays a solid foundation for future implementation efforts.

Date: 7th March 2024

Thank you