Silver Medal Algorithm Overview for

Drawing with LLMs

**Competition Overview:**

This competition is a text-to-image generation challenge in which participants generate SVG (Scalable Vector Graphics) code from English descriptions so that the rendered image matches the original description as closely as possible. The task emphasizes a model’s ability to grasp abstract semantics, deliver precise visual rendering, and keep the output file compact. The official scoring system centers on the SVG Image Fidelity Score, which combines VQA accuracy, an OCR penalty, and a CLIP-based aesthetic score.

**Algorithm Descriptions:**

The proposed solution for this competition implements a complete automated pipeline from text to SVG with the following key stages:

1. Prompt optimization stage:
   1. Expand the original prompt semantically and constrain the generation to a cartoon style.
   2. Enhance the prompt with supplemental keywords while suppressing background elements that could introduce noise.
   3. Apply a structured template: “[subject description], cartoon style, minimal background, high-definition details”.
2. Image generation stage:
   1. Use the Stable Diffusion 3.5 Large Turbo model to produce multiple candidate images.
   2. Employ dynamic seed scheduling to encourage diverse outputs.
   3. Retain three to six candidate images per iteration to form a candidate pool.
3. Vector conversion stage:
   1. Apply an improved feature-extraction algorithm to convert bitmaps to SVG.
   2. Use an adaptive color-quantization strategy (down to 8-bit color depth).
   3. Employ a path-simplification algorithm to satisfy the file size constraint (<100 KB)..
4. Multimodal evaluation stage:
   1. Deploy the paligemma2-10b-mix-448 multimodal model to build the evaluation framework.
   2. Compute the VQA (visual-question-answering) score as the primary metric.
   3. Add an aesthetic score (normalized 0-1) as a secondary metric.

This approach creatively combines prompt engineering, image-compression optimization, and multimodal evaluation, achieving a balance between semantic fidelity and visual quality while keeping SVG files compact. Future work will explore fine-tuning the diffusion model to further enhance stylistic consistency.