Few-Shot Fine-Tuning of Stable Diffusion XL for Defect Image Generation

Babak Badnava

August 12, 2025

Project Overview

- Goal: Generate high-quality synthetic images of defective pills (e.g., **color** or **scratch** defects) from few-shot examples.
- Approach:
  1. Sample small subsets from real dataset.
  2. Fine-tune **Stable Diffusion XL** with LoRA and/or Textual Inversion.
  3. Generate synthetic images conditioned on defect type.
  4. Evaluate quality using standard image generation metrics.
- Supports **multiple defect types** and different sample sizes.

Model Selection

Chosen Models:

- **SDXL Base 1.0**:
  - High-resolution generation (1024×1024).
  - Well-suited for fine detail (important for defect representation).
- **SDXL Refiner 1.0**:
  - Refines base model outputs for sharper textures.
  - Especially helpful for subtle defects.

Why good for this task?

- Pretrained on large, diverse datasets $\rightarrow$ strong visual priors.
- LoRA fine-tuning $\rightarrow$ efficient adaptation to new visual concepts.

Dataset and Preprocessing

- Dataset contains multiple defect types: *color*, *scratch*, etc.
- **Few-Shot Sampler** selects N samples for a given defect.
- Data is split for:
  - LoRA fine-tuning
  - Textual Inversion
- Masks and captions:
  - Masks localize defects for inpainting.
  - Captions describe object and defect (*"a macro photo of a pill with a color defect"*).

Pipeline Workflow

1. Few-shot sampling. Distributed and scalable.
2. Fine-tuning (LoRA + TI).
3. Inference: Generate synthetic images.
4. Evaluation: Compare to real dataset.

Evaluation Metrics

- **ISC** (Inception Score): Measures image quality and diversity.
- **FID** (Frechet Inception Distance): Measures realism by comparing distribution of features.
- **KID** (Kernel Inception Distance): Similar to FID but unbiased.
- **PRC** (Precision Recall for Distributions): Measures coverage and fidelity of generated samples.

Using Segmentation Masks in Fine-Tuning

- **Purpose:** Highlight the defective region so the model focuses learning on relevant areas.
- In **inpainting fine-tuning**, the mask tells the model:
  - Which pixels to **replace** (defect region).
  - Which pixels to **keep** (background / non-defective areas).
- This improves:
  - **Defect localization** (model learns where defects occur).
  - **Sample efficiency** (few-shot training focuses on critical areas).

base image

mask image

generated image

Using Masks in Inference and Generation

- During inference, masks allow:
  - **Defect injection:** Generate a defect only in a given masked area.
  - **Defect removal or replacement:** Inpaint over defective area.
- Workflow:
  1. Provide *initial image* + *segmentation mask*.
  2. Specify prompt (e.g., "a pill with a scratch defect").
  3. Model edits only masked region, preserving the rest.

Benefit

Mask-guided generation increases **control**, **accuracy**, and **consistency** of defect placement.

Results and Next Steps

- **Results:**
  - Generated images closely match defect type and preserve pill structure.
  - LoRA + TI improves fine detail and defect accuracy.
- **Next Steps:**
  - Extend to additional defect types.
  - Experiment with Refiner model for all outputs.
  - Experiment with Inpainting model to incorporate masks for defect generation.
  - Automate hyperparameter search for better few-shot adaptation.

References and Acknowledgments

- Hugging Face Diffusers Library.
- Stability AI Stable Diffusion XL.
- ISC, FID, KID, PRC evaluation methods from literature.