

Supplementary Material

This supplementary section provides additional details, visualizations, and implementation specifics that support the main claims and methodology of our work. It is organized as follows:

A. Additional Qualitative Results

Figure 1b showcases additional samples from our synthetic dataset. Each row depicts the clean background, expert-annotated mask, and the final inpainted output. These examples illustrate the diversity of spill locations and materials (oil, water, rust) our pipeline can handle.

B. Prompt Engineering Details

Stable Diffusion Prompts (Scene Generation):

- **Positive:** "A factory interior, close-up of industrial equipment, image captured via a colored high-quality high-end inspection camera, clear mechanical details, realistic metallic textures, authentic lighting, natural industrial setting, accurate machinery components, subtle equipment variations, realistic wear and tear."
- **Negative:** "Text, watermark, low quality, jitter, nsfw, stickers, labels, blurred details, distorted equipment, cartoonish or unrealistic textures, unnatural colors, overly bright lighting, irrelevant objects, human presence, animals, plants, visible text, duplicated or repeated elements, unrealistic proportions, overly polished surfaces, plastic-like or artificial appearance."

Inpainting Prompts (Hazard Synthesis):

- **Positive** "Realistic oil spill in factory with brown or black stains, industrial scene with dark oil leakage stains, brown-black oily patch on factory floor, factory oil spill with realistic black sludge, realistic factory environment with oil smears, black or brown oil leakage on industrial surface, dirty oil-stained floor in realistic factory, blackened spill area in a manufacturing plant, authentic oil spill marks on brown concrete, industrial realism with black or brown oil spill."
- **Negative** "Cartoon, anime, illustration, painting, drawing, lowres, blurry, pixelated, overexposed, unrealistic,

stylized, clipart, animated, text, watermark, signature, frame, border, extra limbs, distorted hands, shiny, plastic, toy-like, glossy, yellow tint, white overlay, newspaper texture, poster art, human figures, fingers, deformed body parts, 3D render, CGI, artifact, sketch."

System Prompt for VLM: As detailed in the main text, the system prompt emphasized verifiability, physical cues, and the avoidance of speculation. Minor adjustments to wording (e.g., stressing "clearly visible hazard") impacted precision marginally.

C. Adaptation and LoRA Configuration

All LoRA experiments used the following setup:

- Rank: 8
- Scaling factor α : 1/8
- Optimizer: AdamW, learning rate 5e-5
- Hardware: NVIDIA A100 80GB

Annotation Sample (COCO-style):

```
{
  "image_id": 134,
  "category_id": 3,
  "bbox": [256, 411, 142, 95],
  "score": 0.97
}
```

D. Extended Evaluation Trends

In Table ??, we provide hit-rates at stricter IoU thresholds for Qwen-7B across adaptation strategies. These reinforce that models adapted using LoRA (especially on vision pathways) maintain more accurate localization under tighter criteria.

Table 1. Hit-rate of Qwen-7B under varying IoU thresholds

Method	0.5	0.6	0.7	0.8	0.9
Zero-Shot	0.35	0.22	0.15	0.08	0.02
ICL (10-shot)	0.53	0.41	0.29	0.17	0.05
LoRA (V)	0.58	0.49	0.38	0.26	0.10
LoRA (V+L)	0.63	0.54	0.42	0.29	0.13

E. Dataset License and Reuse

We release SynthSpill dataset under the **CC BY-NC 4.0 license**. Researchers may:

- Use the data for non-commercial research purposes
- Modify or build upon the dataset
- Share derived work with attribution

F. Visuals

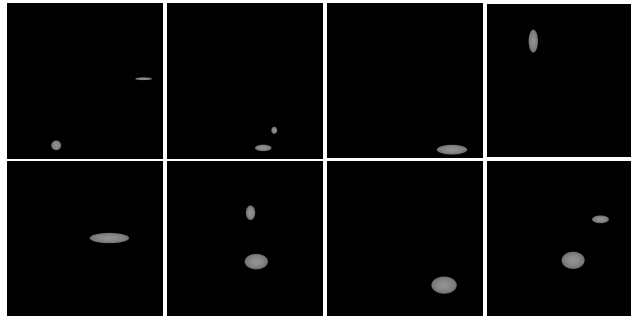


(a) Public (Web Scraped) Dataset Images

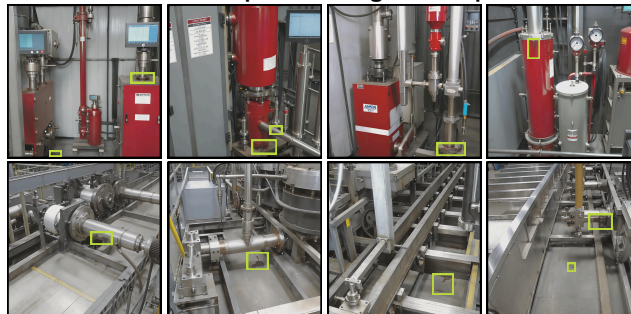
SDXL Generated Synthetic Images



Manually Annotated Masks → Gaussian Filter + Feathering



SDXL Turbo Inpainted Images with Spills



(b) Synthetic Data Images

Figure 1. Dataset Samples