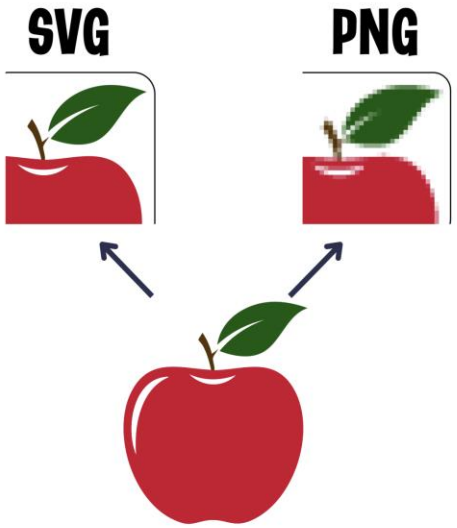
Week 5
Structured Vector Graphics Editing With VLMs

Matthew Perlman

James Beetham, Dr. Amrit Bedi, Dr. Mubarak Shah

Background

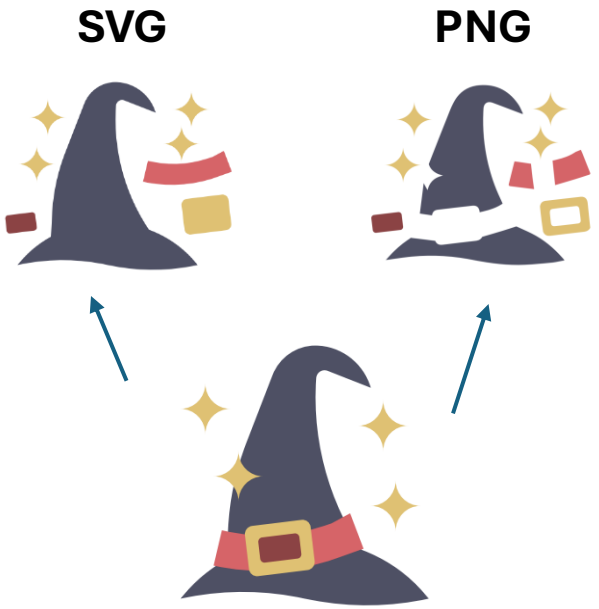
Why do we want to use SVG images instead of raster images?



Lossless Upscaling



Reduced Storage



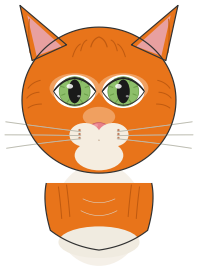
Human Editable

Existing Work

Combining the expressive power of frontier VLMs with diffusion models has "solved" generating rasters. For SVGs… not so much.



GPT-5.5



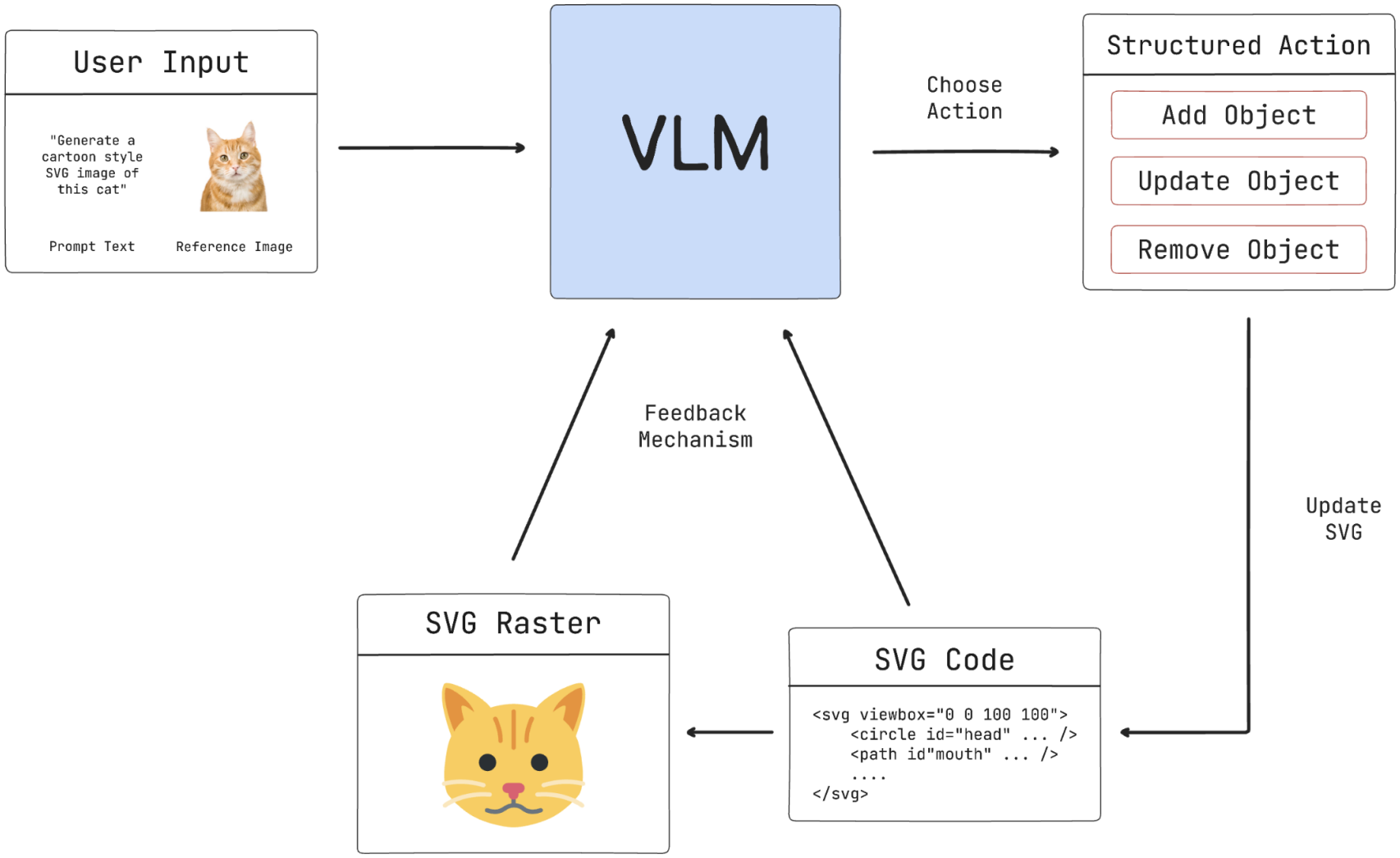
Sonnet-4.6

Dedicated raster-to-vector models perform better, but are **limited** to mostly icons and logos, often produce **overly complex** code, and require **separate training** pipelines.

Why Off-the-Shelf VLMs Fail

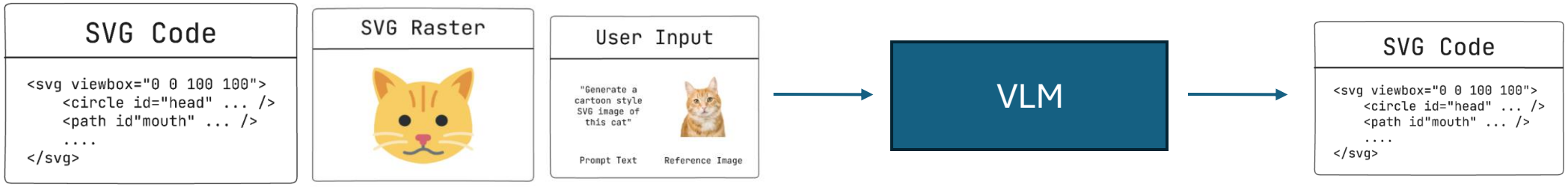
- Smaller models **lack reasoning skill** needed to produce full SVG code
- VLMs **never see** the result, they guess how code looks
- They produce code autoregressively, making choices **irreversible**

Our Solution – Iterative Refinement

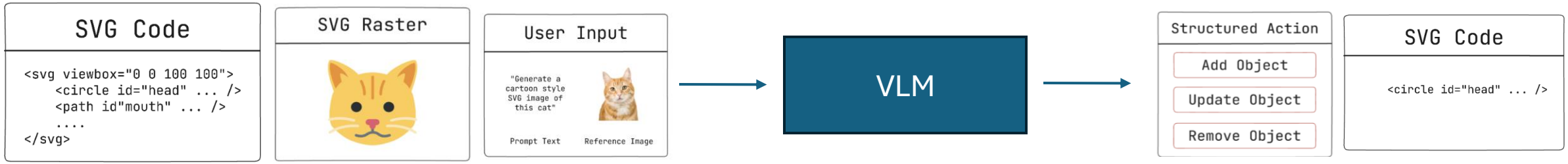


Pipeline Variants

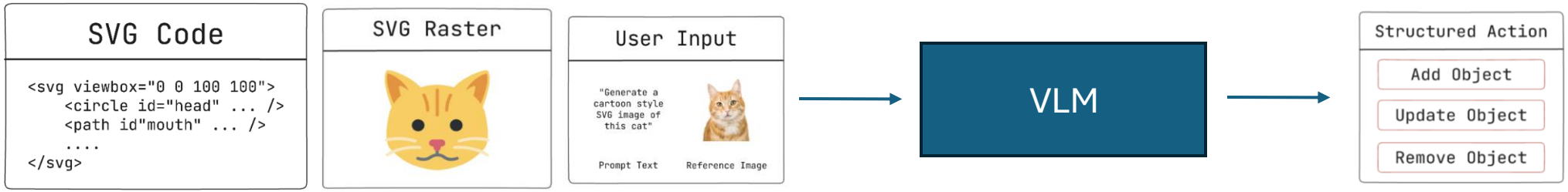
Unstructured Actions



Semi-Structured Actions



Structured Actions



Progress This Week

- **Set up vLLM**, a library for optimal runtime batching and KV caching
  - Roughly 15x speedup
- Evaluated multiple models with iterative refinement on **MMSVG-Bench**
  - Qwen 3B, 7B, 72B
- Finished implementing all image **metrics** (Added LPIPS and SSIM)
- Began **optimizing structured actions** for higher compilation rate

Text-to-SVG Results: Model × Pipeline

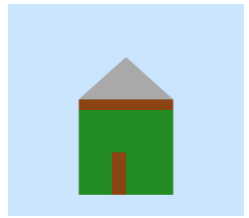
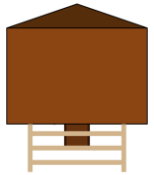
Model Input

<SystemPrompt />

"A stylized tree house with a ladder."

Generation Pipeline Variant

Qwen Model Parameter Count

	3B	7B	72B
Single-Turn Generation			
Iterative Refinement: Unstructured Actions			
Iterative Refinement: Semi-Structured Actions			
Iterative Refinement: Structured Actions			

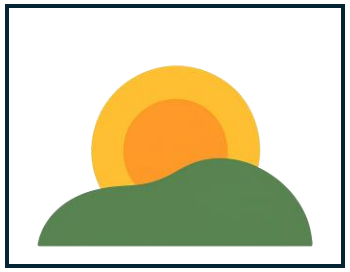
Text-to-SVG Results: Model × Pipeline

Model	Generation Method	Compilation Rate	CLIP Alignment
Qwen2.5-VL-3B	Single-Turn	0.18	0.216
	Structured	**0.25**	0.188
	Semi-Structured	0.01	0.168
	Unstructured	0.03	**0.221**
Qwen2.5-VL-7B	Single-Turn	**0.82**	0.202
	Structured	0.11	0.168
	Semi-Structured	0.18	0.190
	Unstructured	0.57	**0.206**
Qwen2.5-VL-72B	Single-Turn	0.12	0.246
	Structured	0.33	0.201
	Semi-Structured	0.92	0.237
	Unstructured	**0.97**	**0.247**

Image-to-SVG Results: Model × Pipeline

Model Input

<SystemPrompt />



Generation Pipeline Variant

Qwen Model Parameter Count

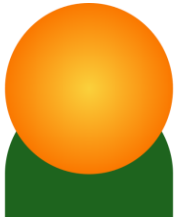
	3B	7B	72B
Single-Turn Generation			
Iterative Refinement: Unstructured Actions			
Iterative Refinement: Semi-Structured Actions			
Iterative Refinement: Structured Actions			

Image-to-SVG Results: Model × Pipeline

Model	Generation Method	Compilation Rate	CLIP Alignment
Qwen2.5-VL-3B	Single-Turn	0.05	0.553
	Structured	**0.07**	0.492
	Semi-Structured	0.003	**0.566**
	Unstructured	0.03	0.559
Qwen2.5-VL-7B	Single-Turn	**0.46**	0.558
	Structured	0.01	**0.636**
	Semi-Structured	0.34	0.260
	Unstructured	0.45	0.573
Qwen2.5-VL-72B	Single-Turn	0.48	0.604
	Structured	0.03	0.536
	Semi-Structured	0.88	0.611
	Unstructured	**0.89**	**0.633**

Next Steps

- Set up **prompt optimization** loop
  - Optimize system prompt using frontier LLM
- Continue **optimizing structured actions** and recompute benchmark
- Collect **qualitative examples** and **finalized evaluation** results
- Continue drafting **paper**
  - SIGGRAPH Asia due July 27
  - AAAI Submission due July 21
  - Eurographics CGF rolling submissions
  - Pacific Graphics poster due July 15