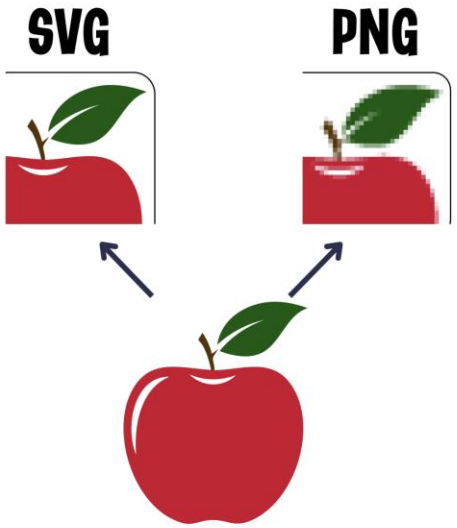
Week 3
Structured Vector Graphics Editing With VLMs

Matthew Perlman

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

Background

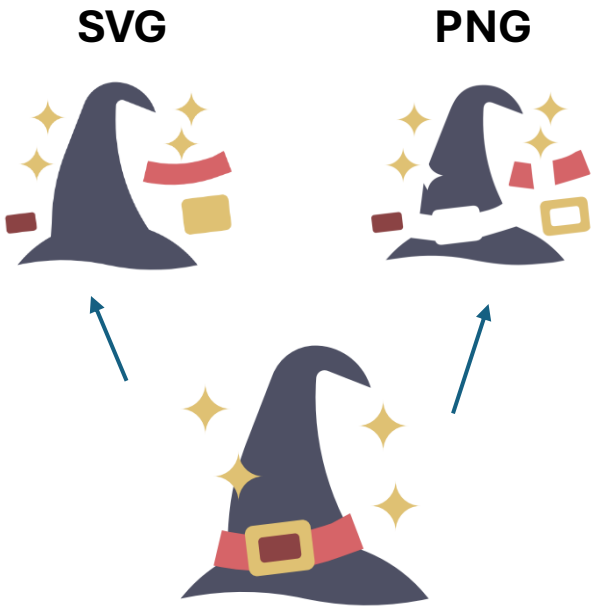
Why do we want to use SVG images instead of raster images?



Lossless Upscaling



Reduced Storage



Human Editable

Existing Work

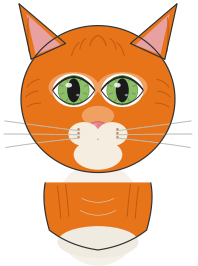
Combining the expressive power of frontier VLMs with diffusion models has "solved" generating rasters. For SVGs… not so much.



"Convert this cat image to a cartoon style svg"



GPT-5.5



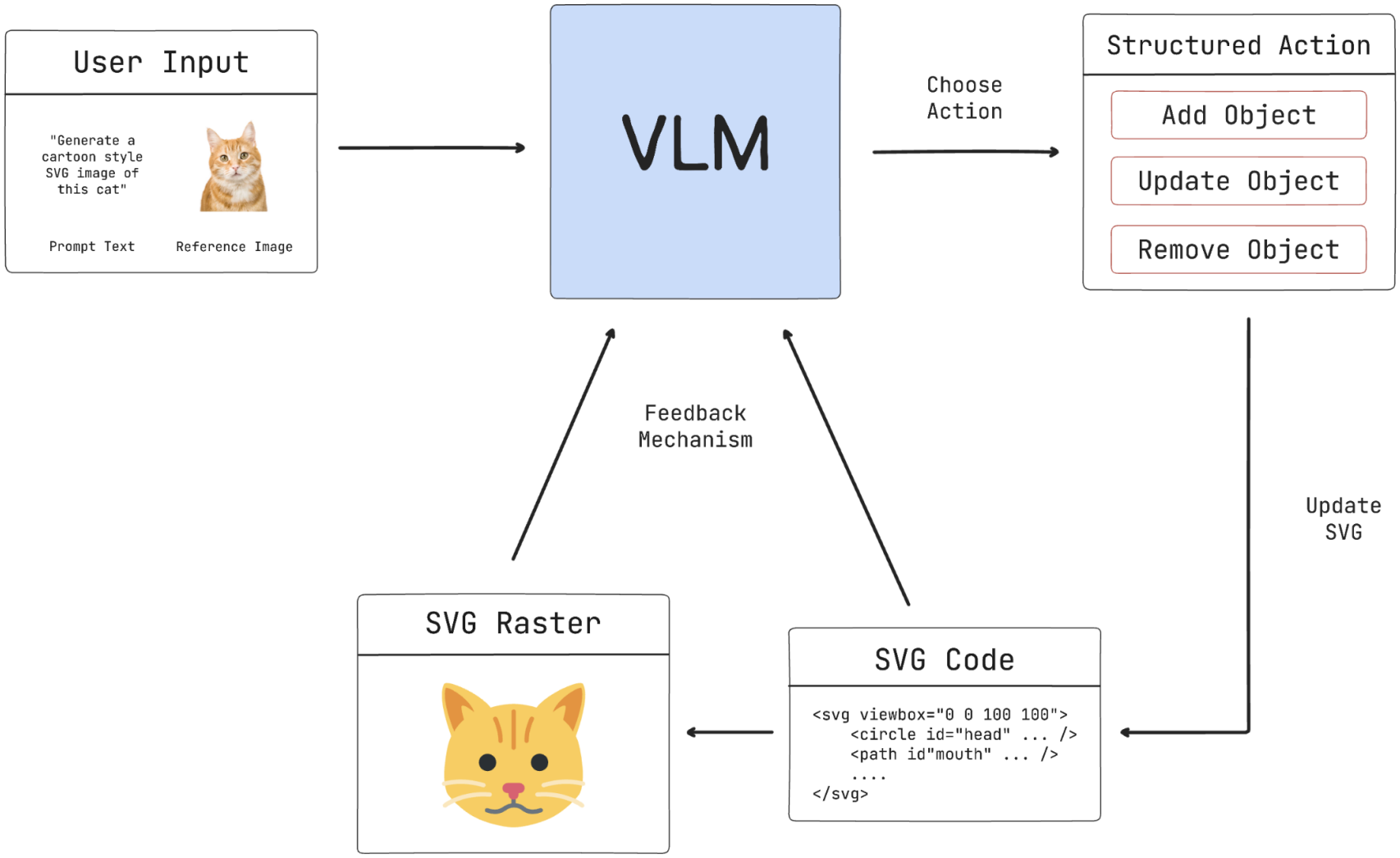
Sonnet-4.6

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

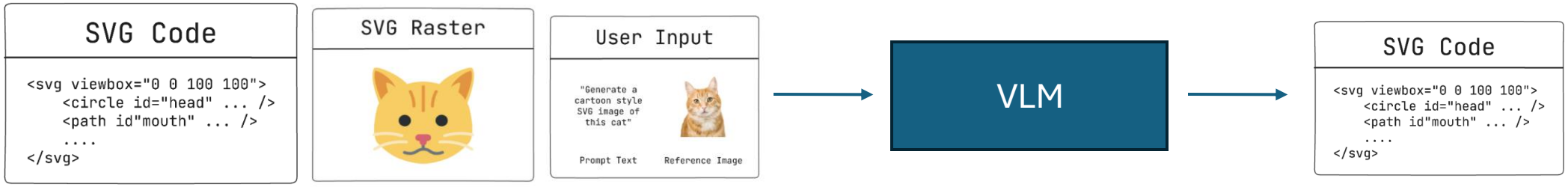


Progress This Week

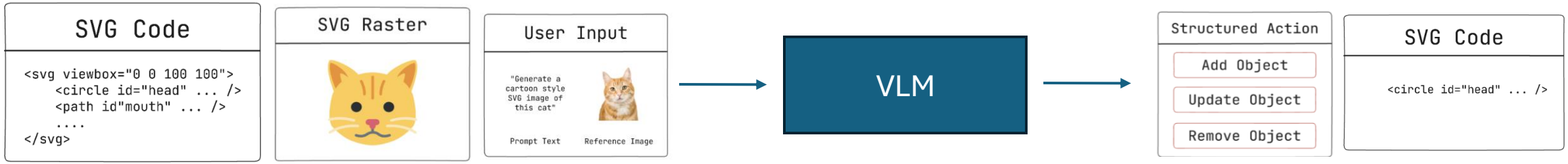
- Setup SVG generation on Newton cluster with **Qwen 72B Instruct**
- Setup SVG generation using cloud APIs (**ChatGPT models**)
- Implemented initial **pipeline variants**:
  - Unstructured – the VLM outputs code for the complete SVG every iteration
  - Semi-Structured – the VLM outputs code for individual SVG components
  - Structured – the VLM outputs an action with defined parameters
- Started implementing **performance metrics**:
  - Successful compilation
  - CLIP visual alignment
  - DINO visual alignment
- Reviewed **papers**: StarVector, OmniSVG, IntroSVG

Pipeline Variants

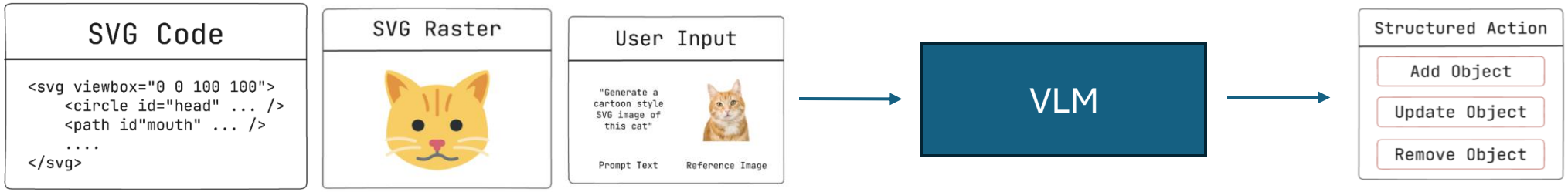
Unstructured Actions



Semi-Structured Actions



Unstructured Actions



Results Using GPT 5.4



Reference Image

"Generate a picture of a cat"

Unstructured



Step 1



Step 3



Step 5

Semi-Structured

Step 3

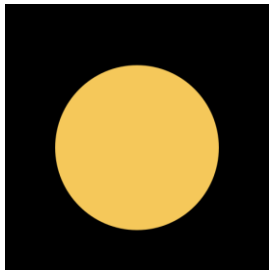


Step 7

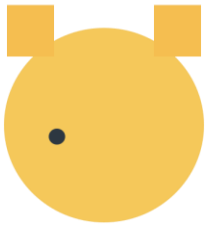


Step 5

Structured



Step 2



Step 4



Step 5

Next Steps

- Collect **quantitative results** on Stack-SVG Dataset
- Finish implementing performance metrics from **StarVector Benchmark**
- **Compare results** across different system prompts and model sizes
- **Paper direction**: Visual Feedback Unlocks SVG Manipulation in Off-the-Shelf VLMs