

Outlines



The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Advanced Scene Understanding

Technical Super-Resolution

The Recommendation Engine

Outlines

The Generative Leap

Throwback – The SD1.5 & ControlNet Era

Introducing Flux2Klein 4B

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Advanced Scene Understanding

Technical Super-Resolution

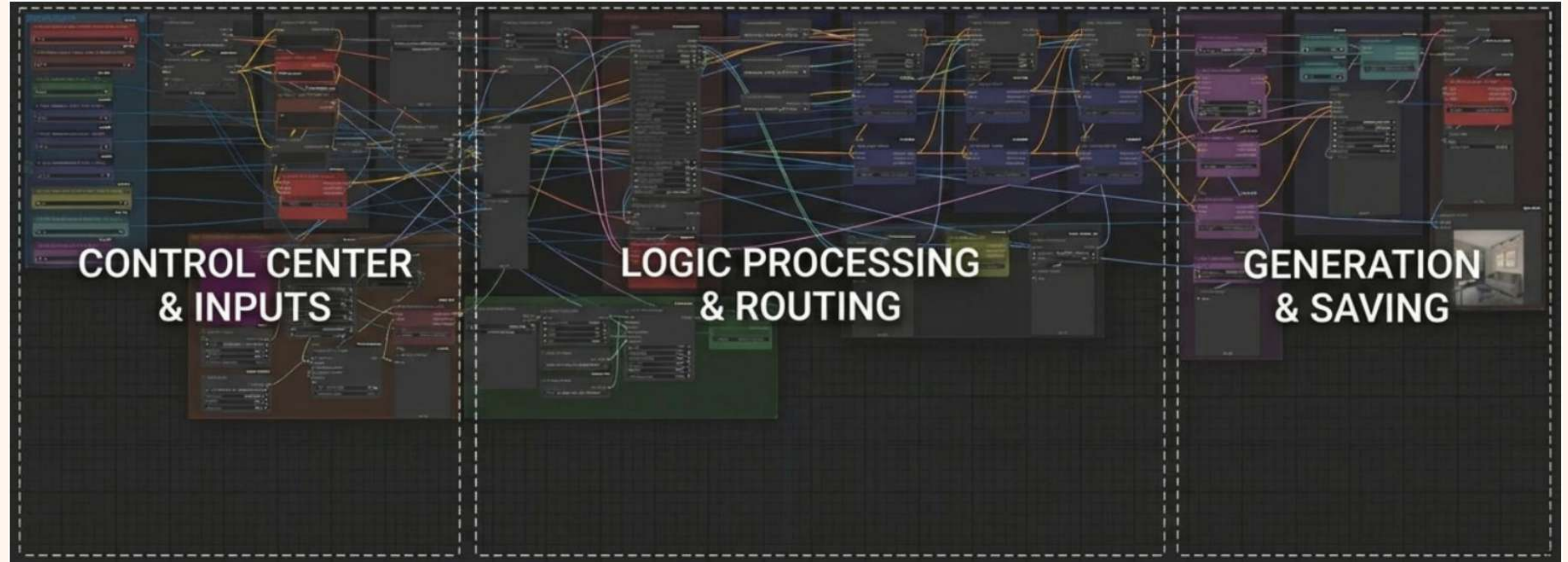
The Recommendation Engine

The Generative Leap: Throwback – The SD1.5 & ControlNet Era

The Approach: Stable Diffusion 1.5 + Multi-ControlNet Stacks (Canny, M-LSD, Depth).

The Bottleneck: High debugging complexity and multiple points of failure.

The Flaw: Complex routing required to prevent the AI from "hallucinating" or breaking room geometry.



The Goal ?

The Goal ?

“We needed a smarter, more native architectural brain.”

Introducing Flux2Klein 4B

Introducing Flux2Klein 4B

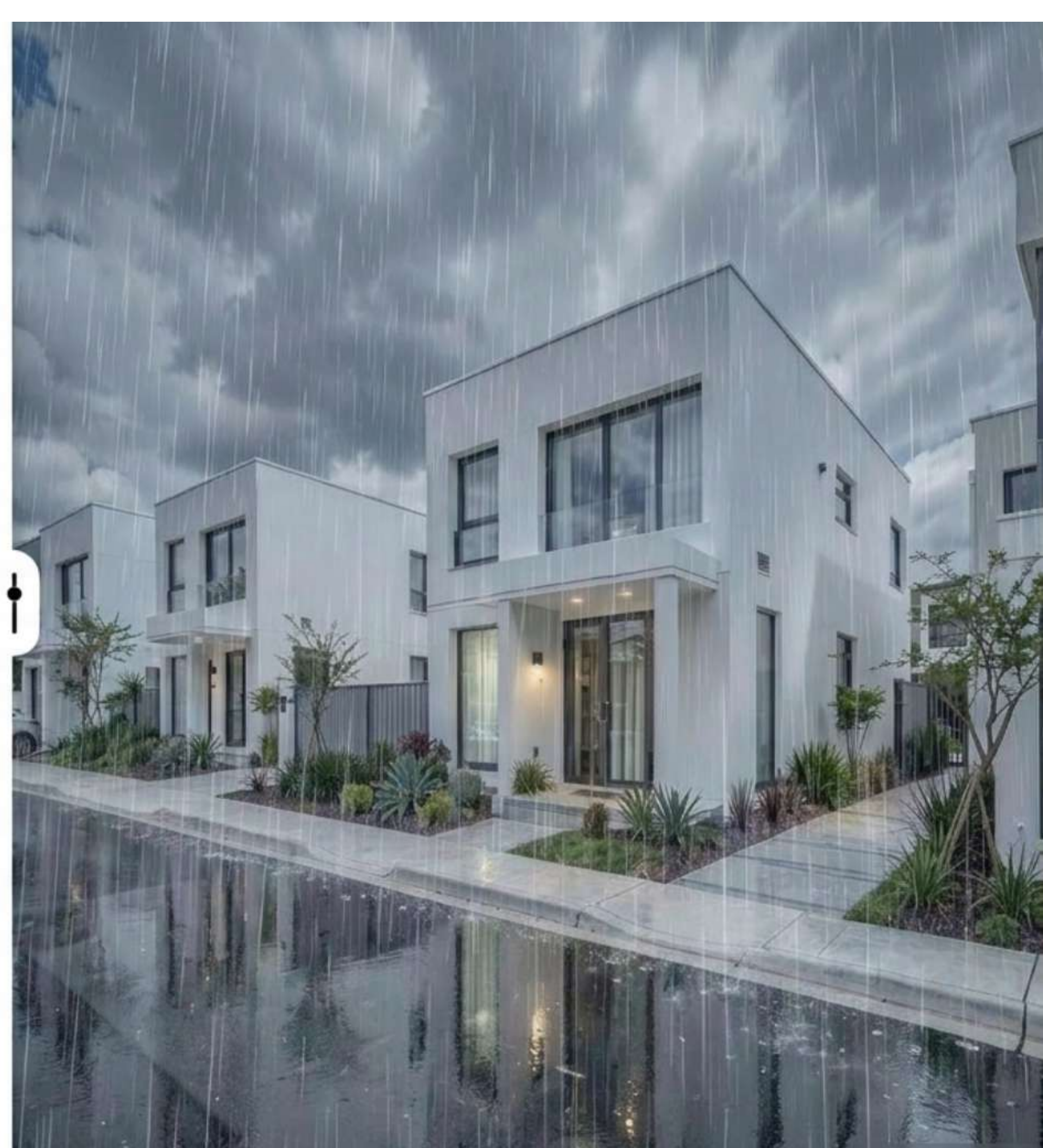
The Commercial-Viable Beast

Here's a little taste of what's to come.



Exterior Environmental Controls

Varying the atmospheric, lighting, and weather profile of a neighborhood photograph. Notice the dynamically generated water reflections on the asphalt and the atmospheric diffusion of the sky in the rainy render.

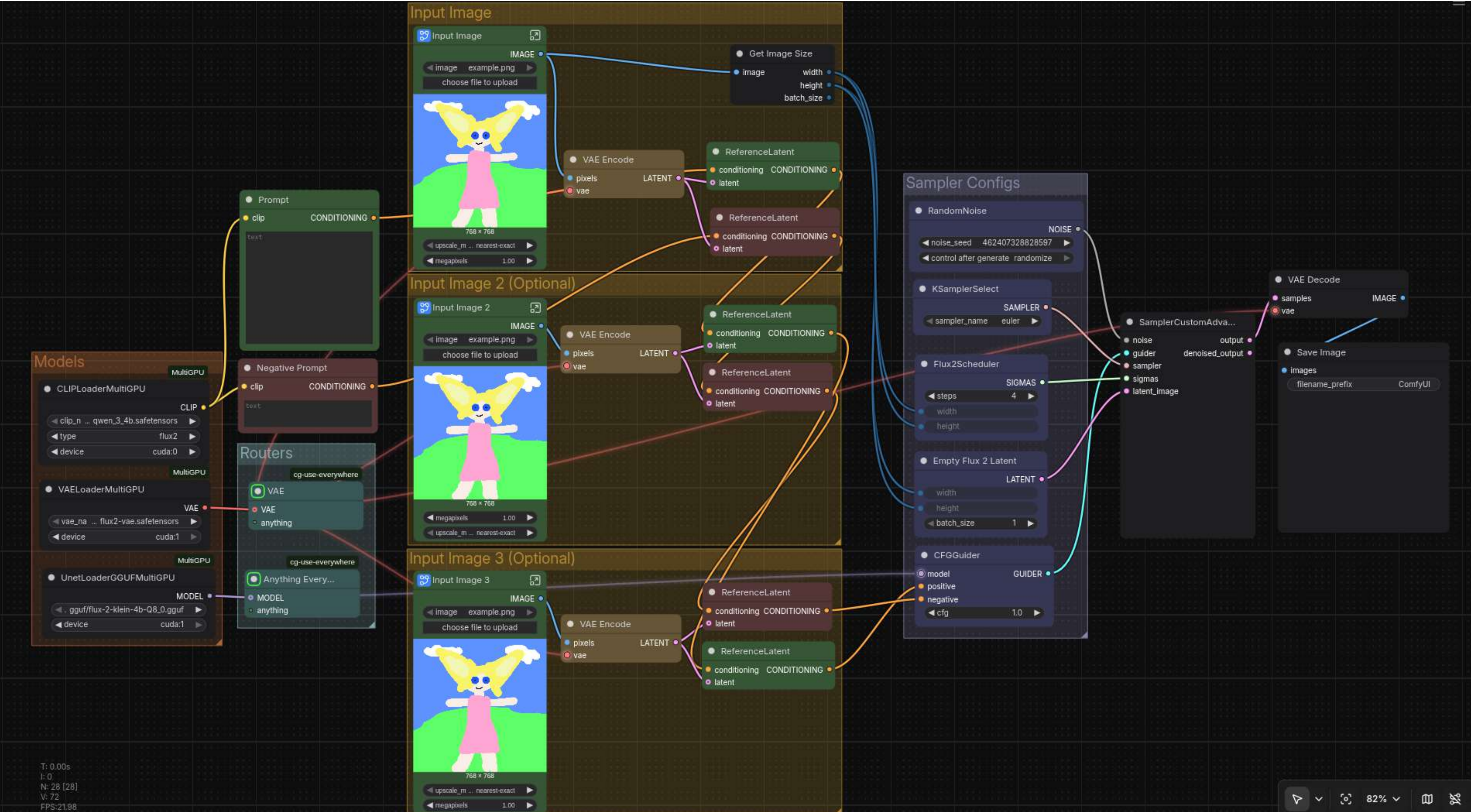


Introducing Flux2Klein 4B

The Commercial-Viable Beast

x2

Flux2Klein Workflow: Featuring Flux2klein 4B Q8 GGUF & Qwen-3 4b text encoder



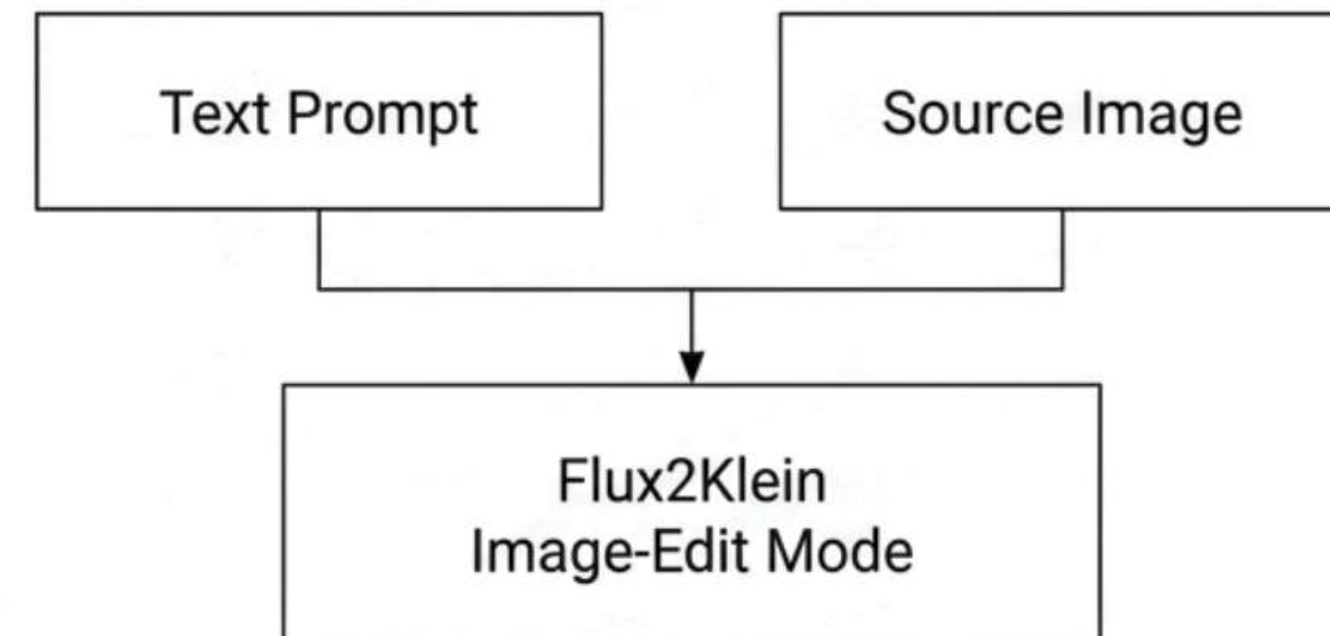
The Generative Leap: Introducing Flux2Klein 4B

- **Diffusion Transformer (DiT):** Replaces legacy UNet for superior global scene coherence.
- **Native Image-Edit Mode:** Locks walls, windows, and structural lines without needing 3 separate ControlNets.
- **High Semantic Density:** Understands complex lighting, physics, and styles naturally.
- **Commercial Viability:** Apache 2.0 license with an optimized footprint.

UNet Iterative Denoising



Diffusion Transformer (DiT) & Flow Matching



Outlines

The Generative Leap

Core Staging & Furnishing

Throwback – Empty Room Staging

The New Era – Flux2Klein Standard Staging

Complete Redecorating

Sketch-to-Render Upgrade

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

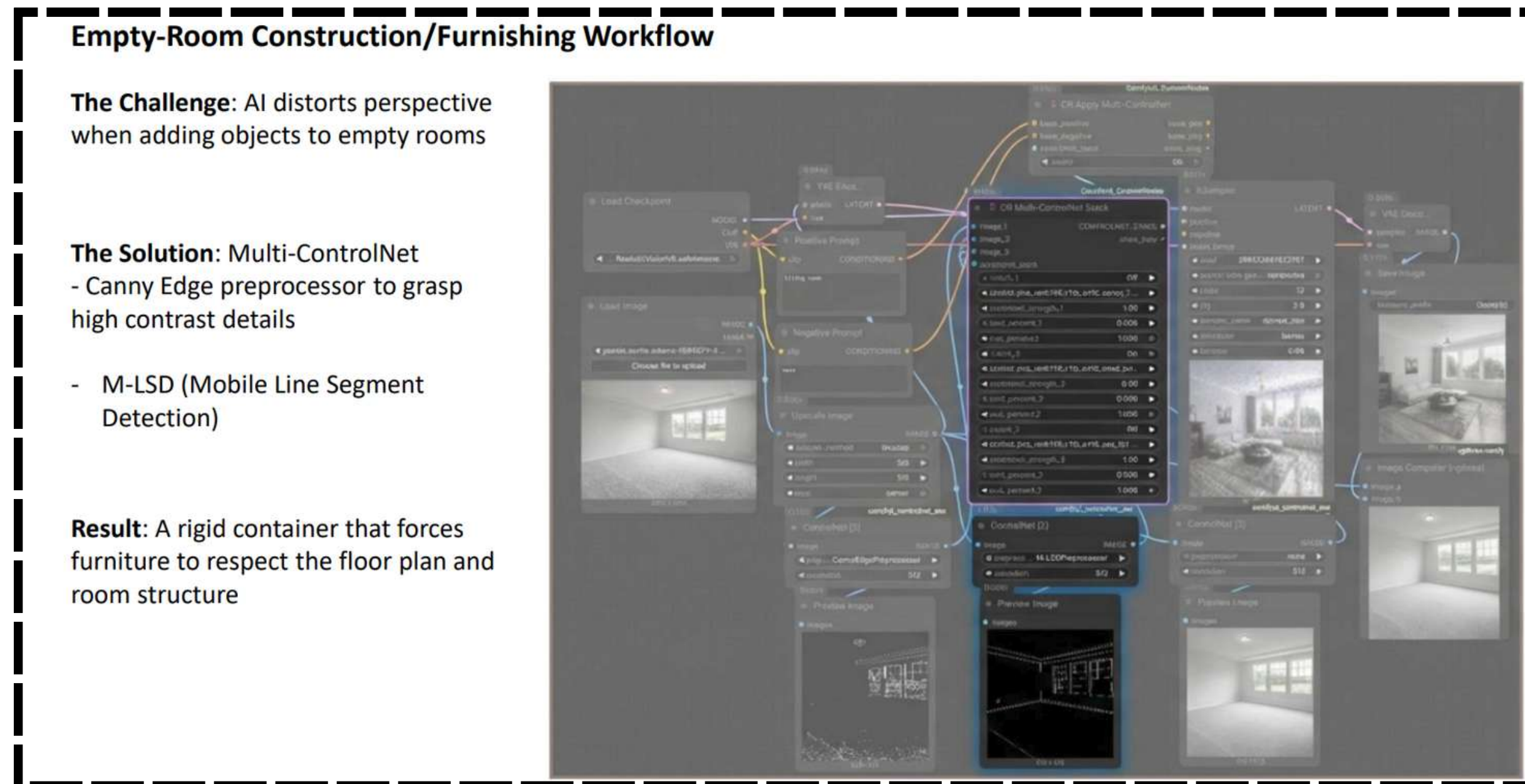
Advanced Scene Understanding

Technical Super-Resolution

The Recommendation Engine

Core Staging & Furnishing: Throwback – Empty Room Staging

Throwback to when Furnishing an Empty Room was a separate workflow and stacked some controlnets...



Old Method: Forced the model to respect room bounds using Canny and M-LSD.

Result: Rigid, but prone to lighting inconsistencies and required preprocessing.

Core Staging & Furnishing: The New Era – Flux2Klein Standard Staging

Next-Gen Staging: Native Spatial Understanding

- Zero controlnet preprocessing required.
- Perfect preservation of floor dimensions, doors, and windows.
- Highly accurate natural light integration.



Mathematical
Locking Intact.



Core Staging & Furnishing: Complete Redecorating

Input images generated by Flux2Klein **Text-to-Image** Workflow

Output images generated by Flux2Klein **Image-Edit** Workflow

Finish Replacement:

Substituting legacy finishes while preserving exact spatial volume.



Void Population:

Populating a completely empty architectural void with coherent, to-scale furnishings.



Core Staging & Furnishing: Sketch-to-Render Upgrade

- **Legacy Workflow:** Required complex Lineart + M-LSD ControlNets.
- **New Workflow:** Flux2Klein translates raw sketches directly into photorealistic renders natively, eliminating the need for ControlNet stacks.



Outlines

The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Throwback – Object Replacement

The New Era – Multi-Reference Integration

Multi-Reference Results

Aesthetic Flexibility & Style Presets

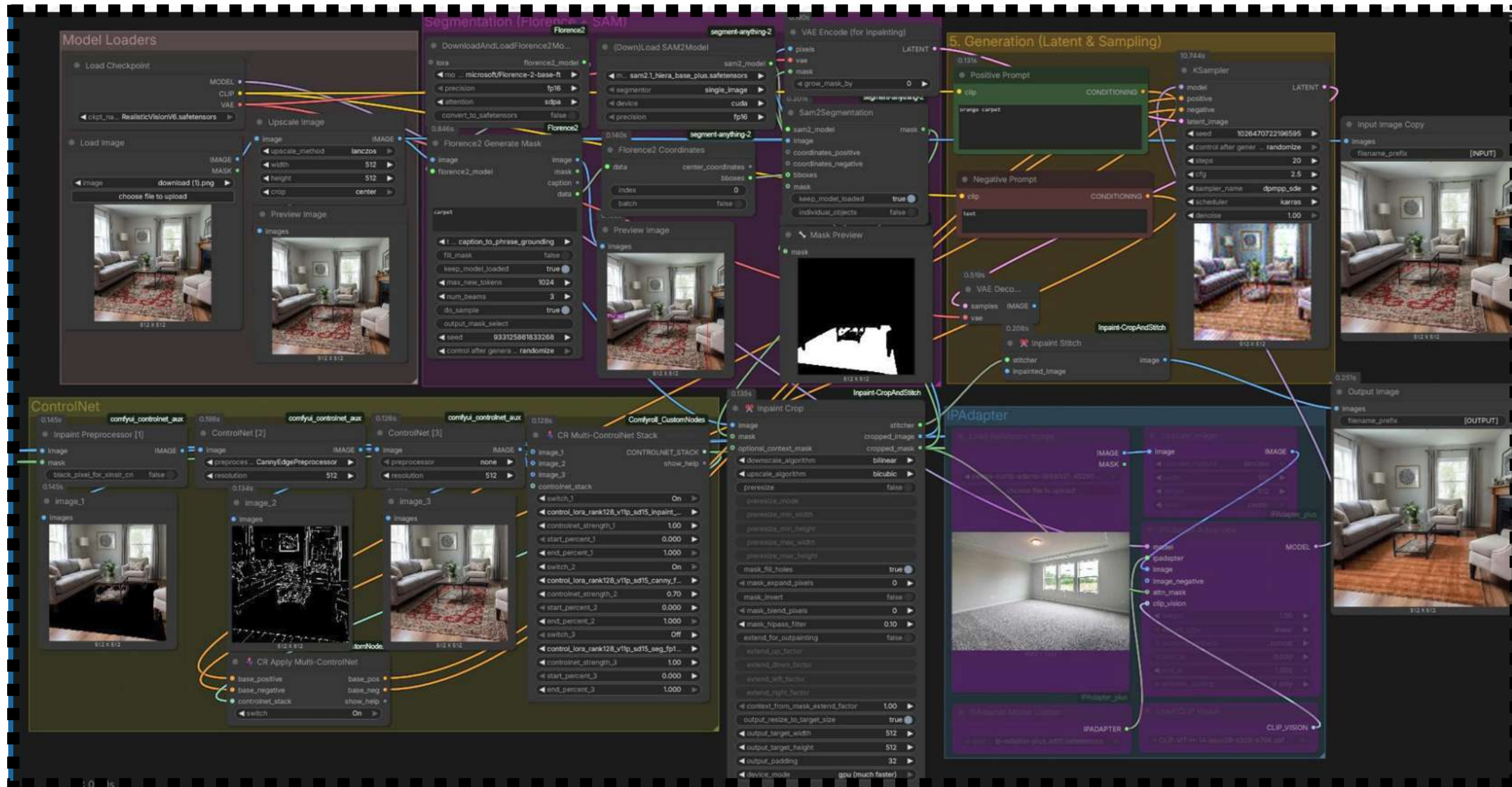
Advanced Scene Understanding

Technical Super-Resolution

The Recommendation Engine

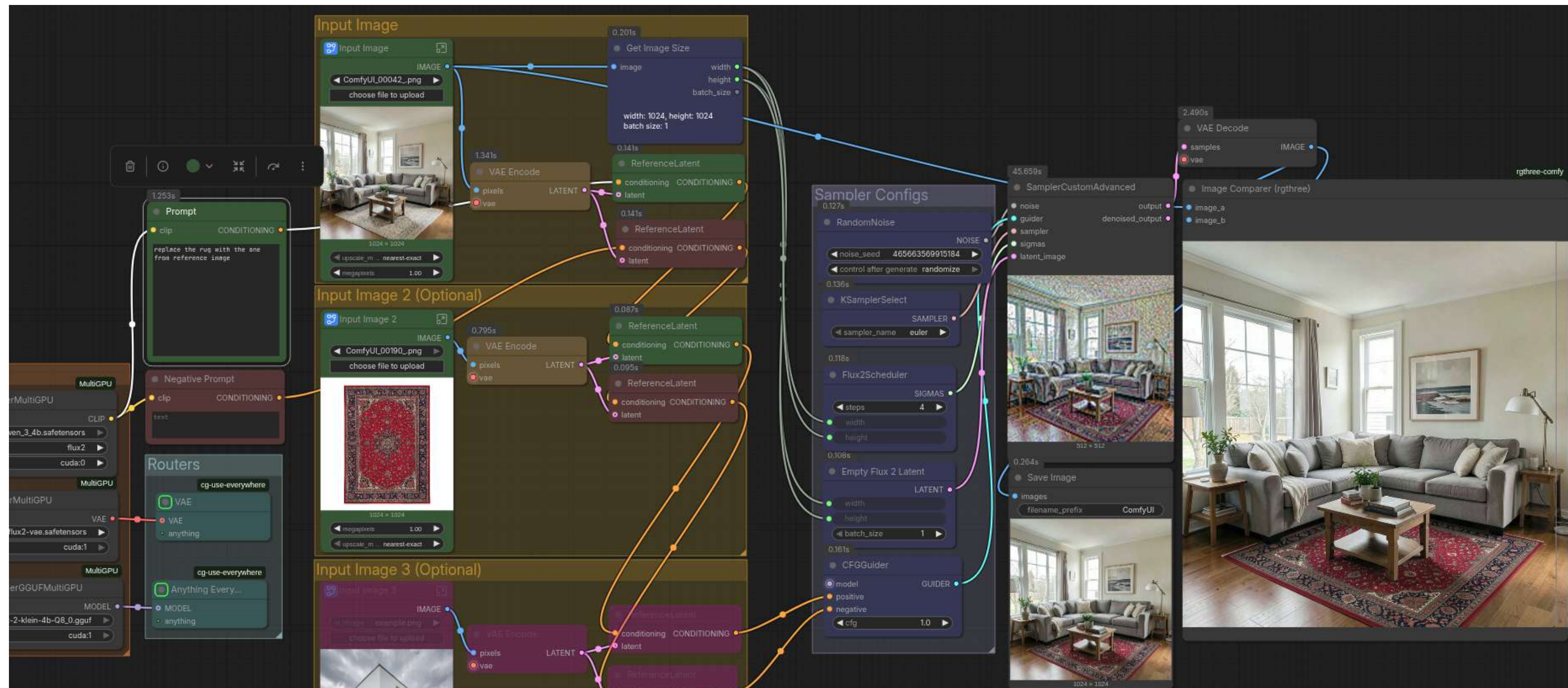
Object Replacement & "Shop the Look": Throwback – Object Replacement

- **Old Method:** Florence2 (Detection) + SAM2 (Masking) + IP-Adapter (Texture injection).
- **Challenge:** Masking was precise, but integrating products or matching exact global lighting was difficult.



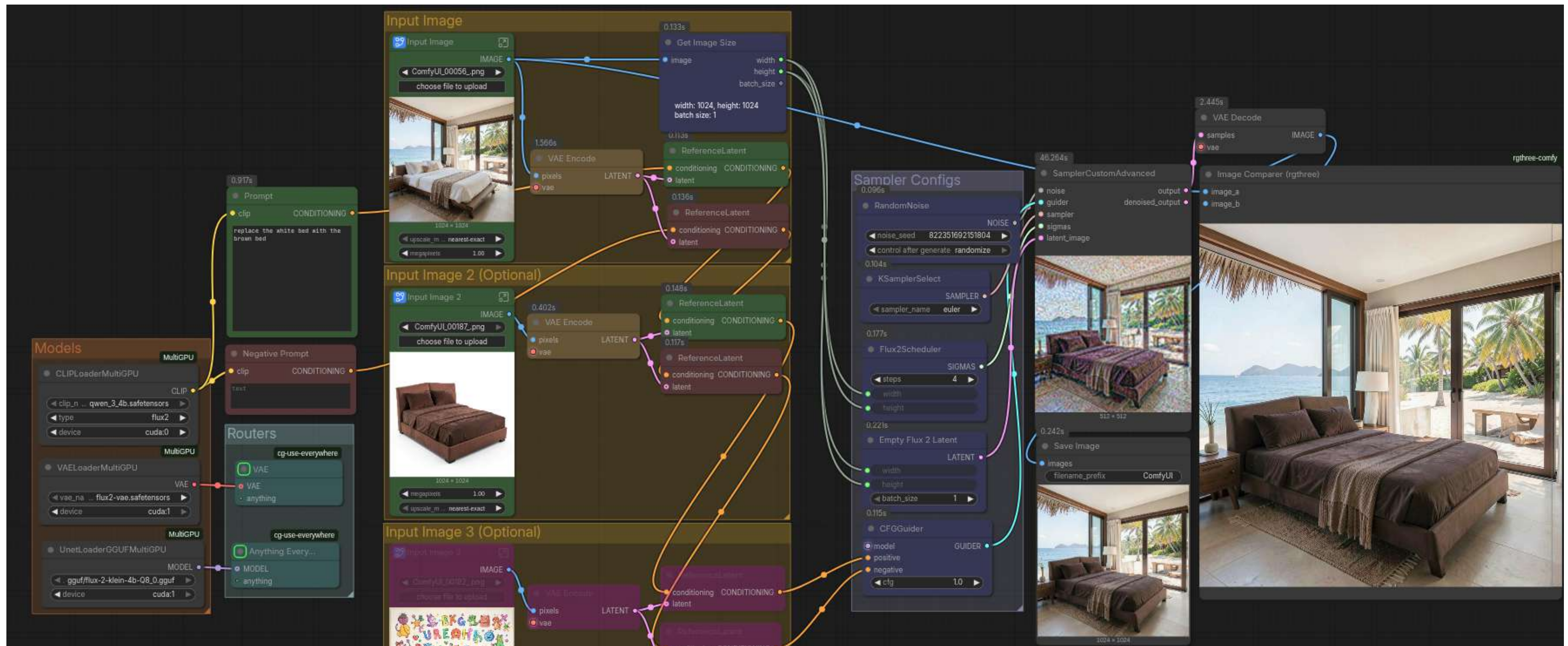
Object Replacement & "Shop the Look": The New Era – Multi-Reference Integration

- "Shop The Look" at scale.
- Injecting specific commercial products directly into empty room blockouts.
- Testing individual and combined reference product injections simultaneously.



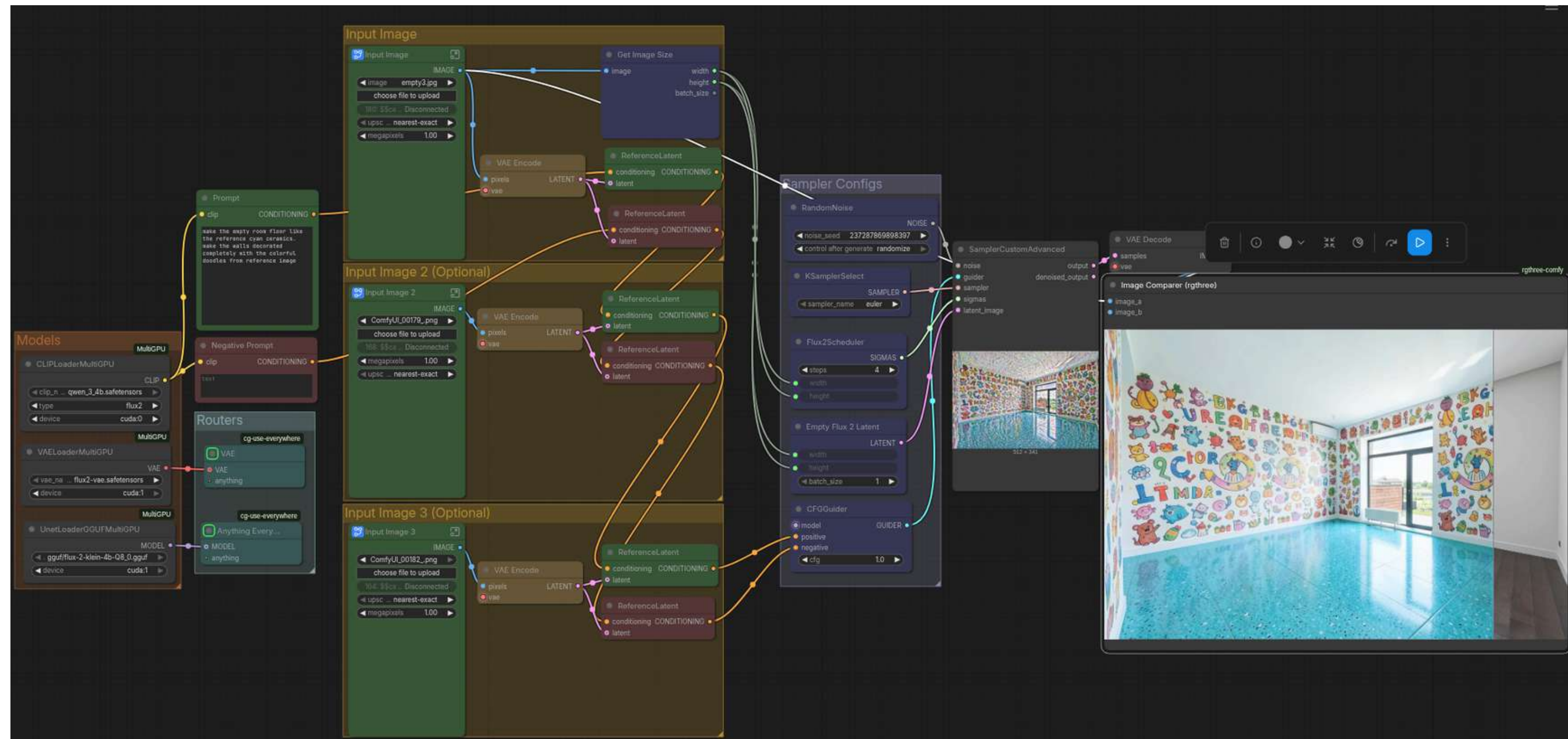
Object Replacement & "Shop the Look": The New Era – Multi-Reference Integration

- "Shop The Look" at scale.
- Injecting specific commercial products directly into empty room blockouts.
- Testing individual and combined reference product injections simultaneously.



Object Replacement & "Shop the Look": The New Era – Multi-Reference Integration

- "Shop The Look" at scale.
- Injecting specific commercial products directly into empty room blockouts.
- Testing individual and combined reference product injections simultaneously.



Object Replacement & "Shop the Look": The New Era – Multi-Reference Integration

*The Difference between **Hallucination** and **Integration***



Base Target Image



Reference Product



Option A (Standard): AI hallucinates a generic, random bed.



Option B (Product Guided):
AI perfectly injects the exact reference product for e-commerce visualization.

Object Replacement & "Shop the Look": Multi-Reference Results

Base Target Canvas



Reference Injection

+



→

The Solution: Commercial Integration



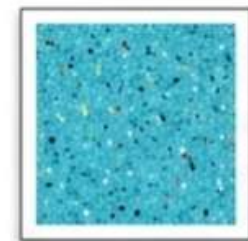
Injects the exact commercial reference product with physically accurate lighting and shadows.

Object Replacement & "Shop the Look": Multi-Reference Results

Multi-Step Object replacement from reference images



Input image



Flooring SKU
Swatch



Base Space + Flooring SKU



Wall Texture
SKU



Wall Textures + Flooring SKU

Outlines

The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Throwback – Style Transfer

The New Era – High Semantic Density

Extreme Stylization: Minecraft

Advanced Scene Understanding

Technical Super-Resolution

The Recommendation Engine

Aesthetic Flexibility & Style Presets: Throwback – Style Transfer

- **Old Method:** Relied heavily on IP-Adapter.
- **Issue:** IP-Adapters were great, but getting them to apply the exact texture we wanted sometimes felt like playing the lottery.

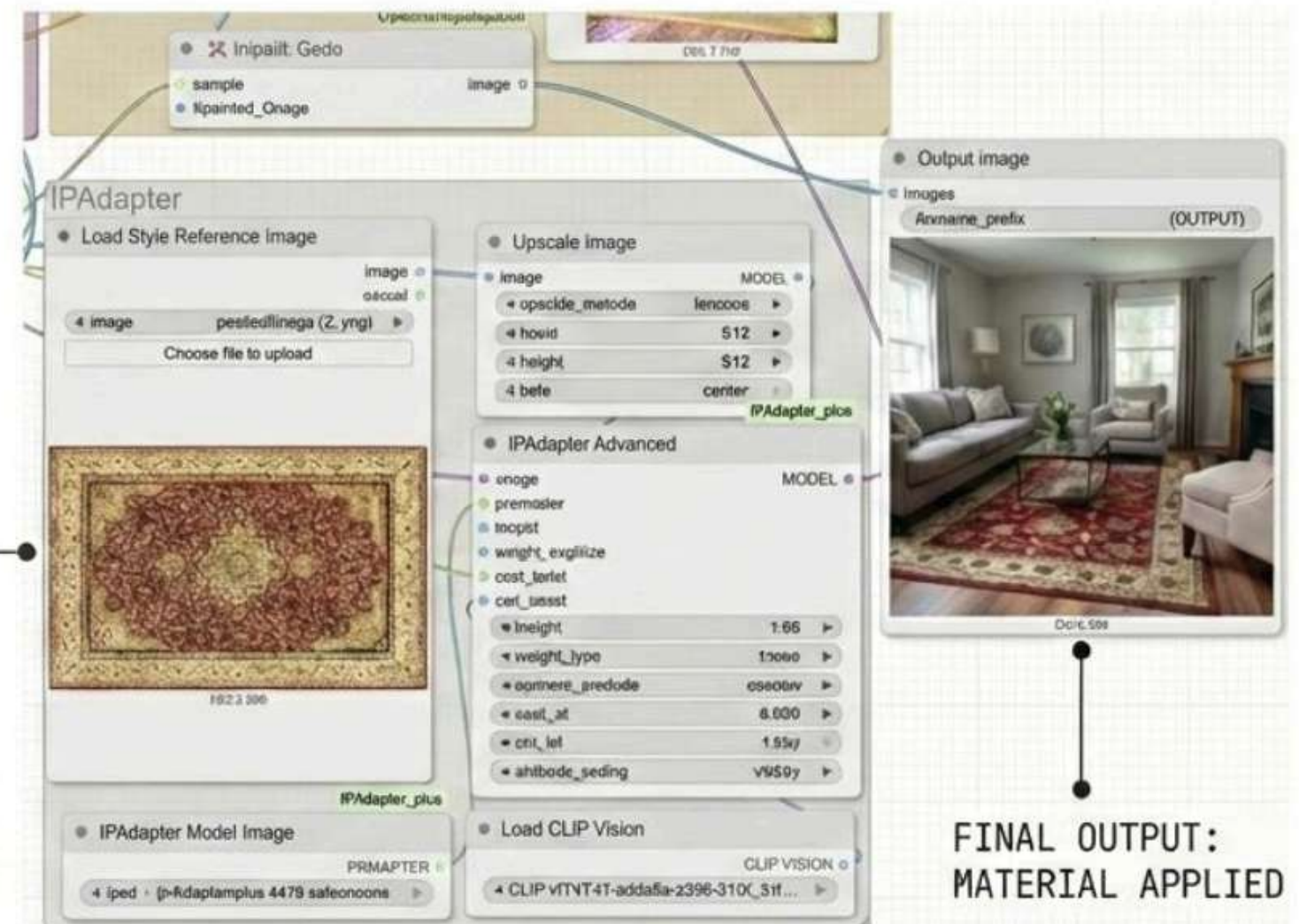
Object Replacement Workflow: Introducing Style Transfer

Style Injection via IP-Adapter “Material Swapping”

Instead of relying on vague text descriptions like ‘red persian rug’, we use IP-Adapters.

This allows us to **inject** exact visual **embeddings** from a **reference image** directly into the masked area, guaranteeing material accuracy.

INPUT:
VISUAL EMBEDDING



FINAL OUTPUT:
MATERIAL APPLIED

What about a hard test ?

Aesthetic Flexibility & Style Presets: The New Era – High Semantic Density

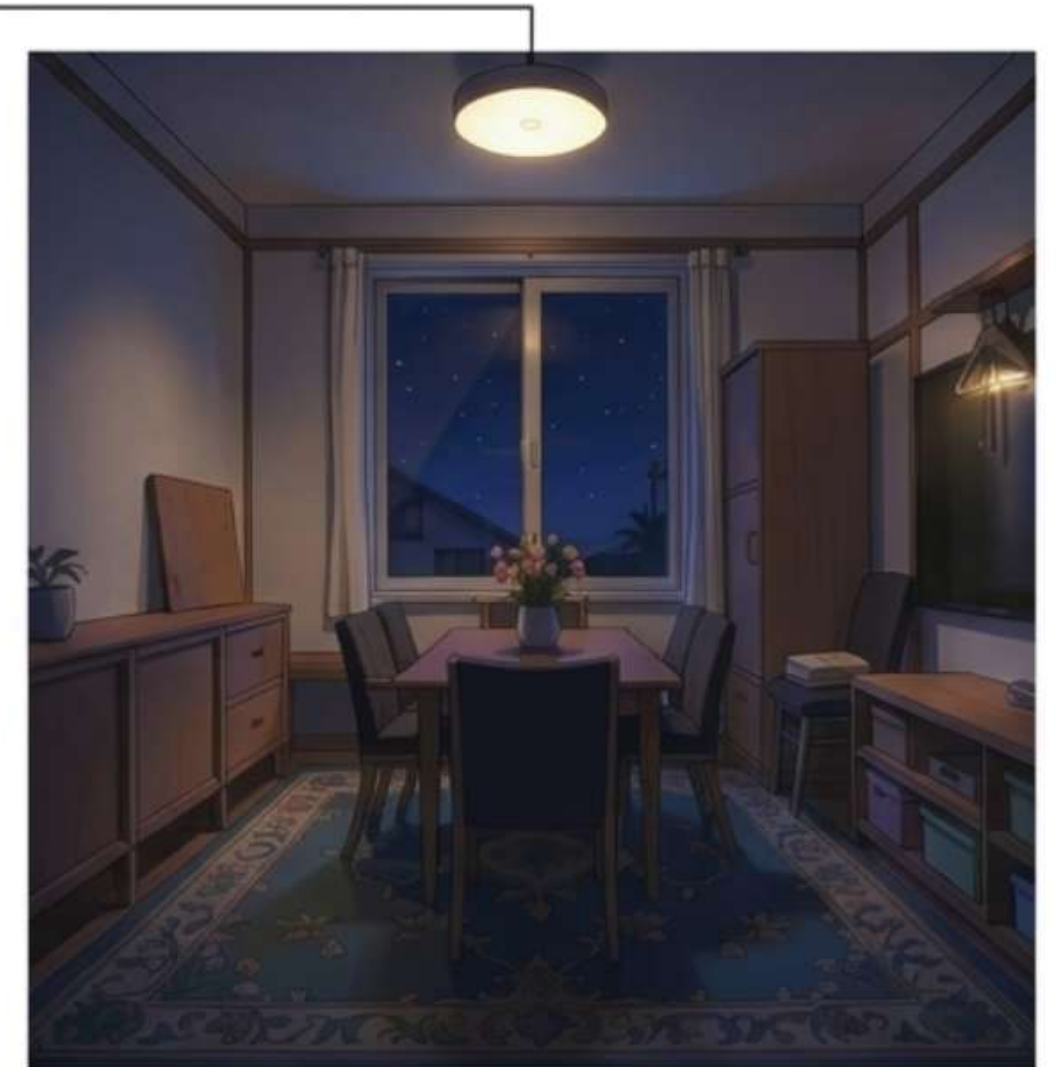
- Taking highly stylized layouts (yes, even anime rooms) and forcing them into reality.
- Zero loss of structural integrity (furniture shapes, window frames remain identical).



Anime Photographic Input



Anime Daytime Render



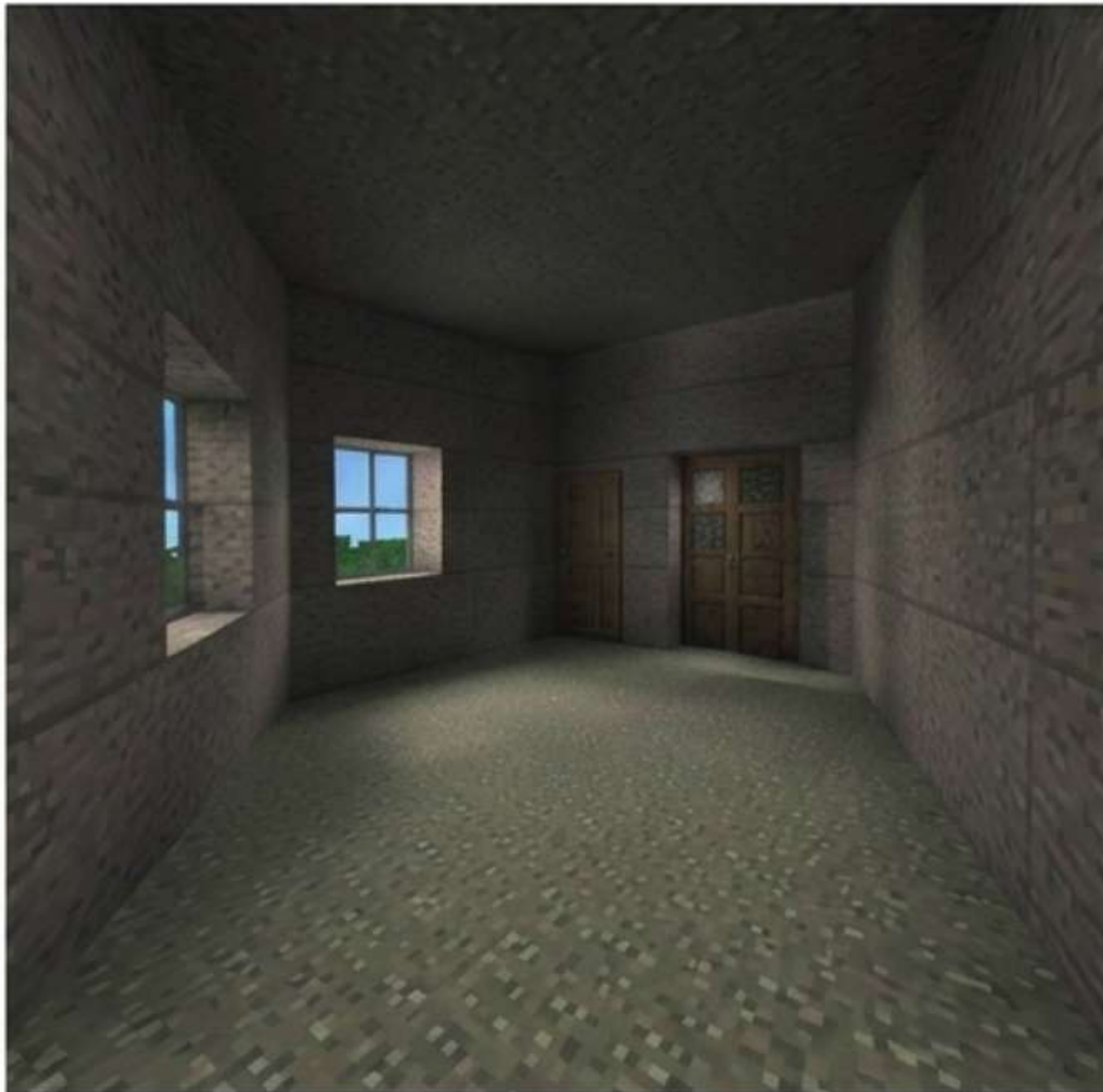
Anime Nighttime Render

What about a harder test ?

Aesthetic Flexibility & Style Presets: Extreme Stylization: Minecraft

- A creative stress-test of the workflow.
- The AI understands "blocks" as a texture, applying them accurately to a same-context architectural layout.

Before



After



Outlines

The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Advanced Scene Understanding

The New Capability – Scene Decluttering

Lighting & Time-of-Day Transitions

Intelligent Coloring

Technical Super-Resolution

The Recommendation Engine

Ok, Enough Furnishing Tests

What about Decluttering a messy room ?

Advanced Scene Understanding: The New Capability – Scene Decluttering

Deletes scattered laundry, trash, or whatever else you want gone, while the AI logically guesses what the hidden floorboards underneath actually look like.



FIG 1A: CLUTTERED INPUT



FIG 1B: DECLUTTERED OUTPUT
(INFERRED STRUCTURE)

Beyond Aesthetics: This capability demonstrates true physical scene understanding, unlocking entirely new architectural workflows.

What about Changing Lighting ?

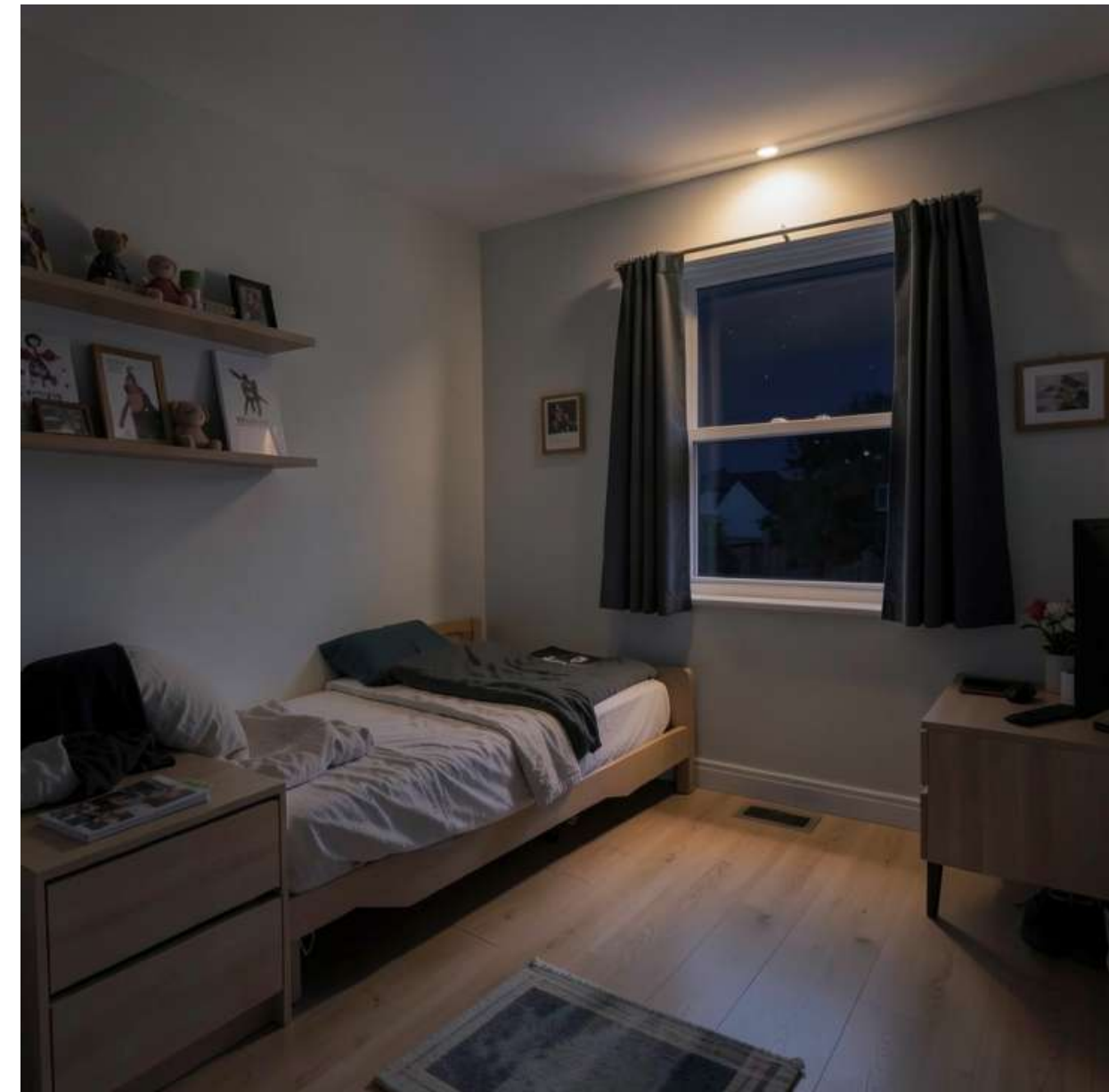
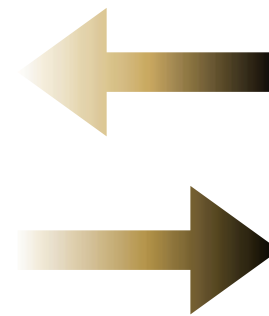
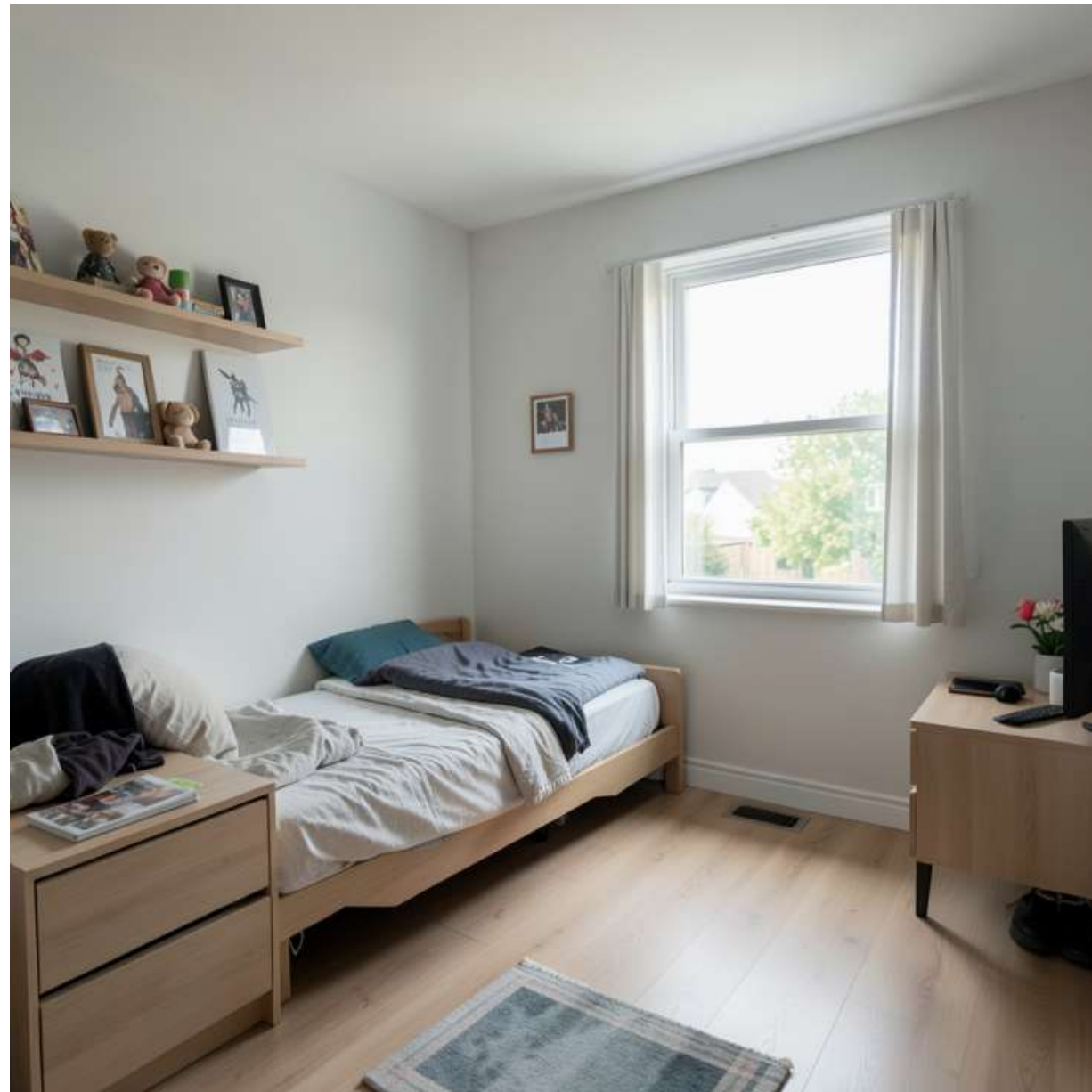
Advanced Scene Understanding: Lighting & Time-of-Day Transitions

- Altering a scene's light sources (daylight to artificial warm lamps).
- Generates accurate cast-shadows.
- Keeps every piece of furniture mathematically locked in place.



Advanced Scene Understanding: Lighting & Time-of-Day Transitions

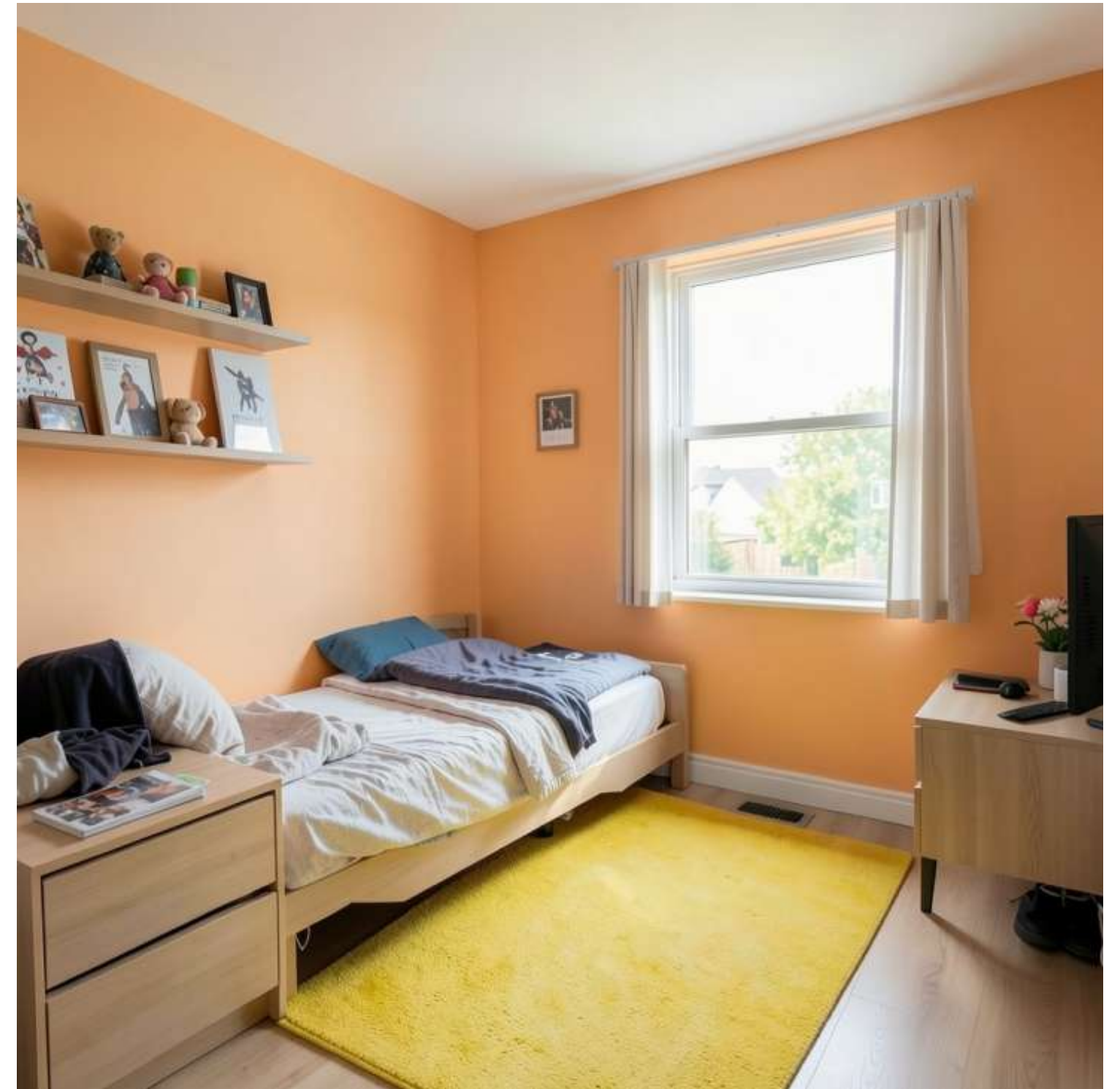
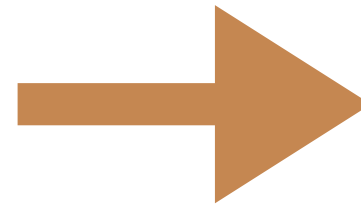
- Altering a scene's light sources (daylight to artificial warm lamps).
- Generates accurate cast-shadows.
- Keeps every piece of furniture mathematically locked in place.



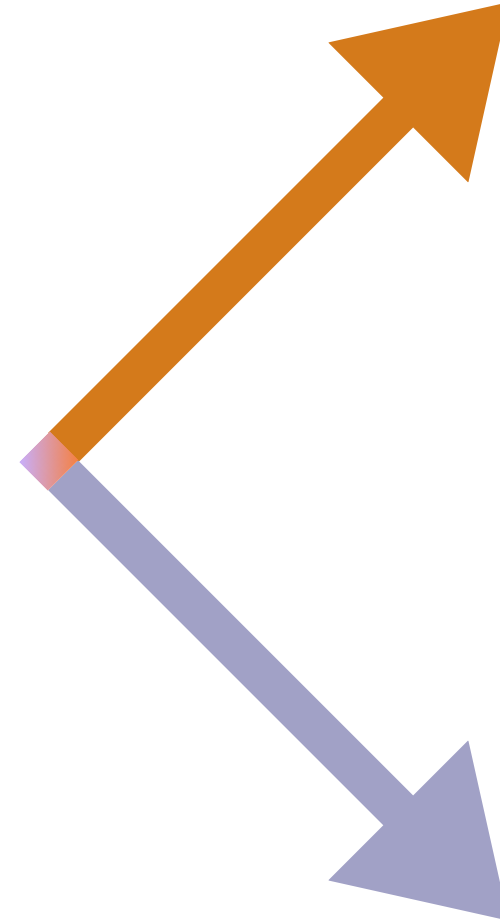
What about Changing Colors ?

Advanced Scene Understanding: Intelligent Coloring

Users can virtually paint their rooms and see the results before making a terrible, irreversible mistake with real paint...



Advanced Scene Understanding: Intelligent Coloring



Advanced Scene Understanding: Intelligent Coloring

We can even bring lifeless, grey-boxed 3D blockouts to life.



FIG 1A: VOLUMETRIC MONOCHROME INPUT



FIG 1B: FINALIZED COMMERCIAL OUTPUT

Outlines

The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Advanced Scene Understanding

Technical Super-Resolution

Revisiting The Upscaling Problem

The New Era – SeedVR2 Upscaling

The Recommendation Engine

Technical Super-Resolution: Revisiting The Upscaling Problem

Technical Super-Resolution: Revisiting The Upscaling Problem

Last Semester, **Ultimate Tiled Upscale** actually achieved very good results

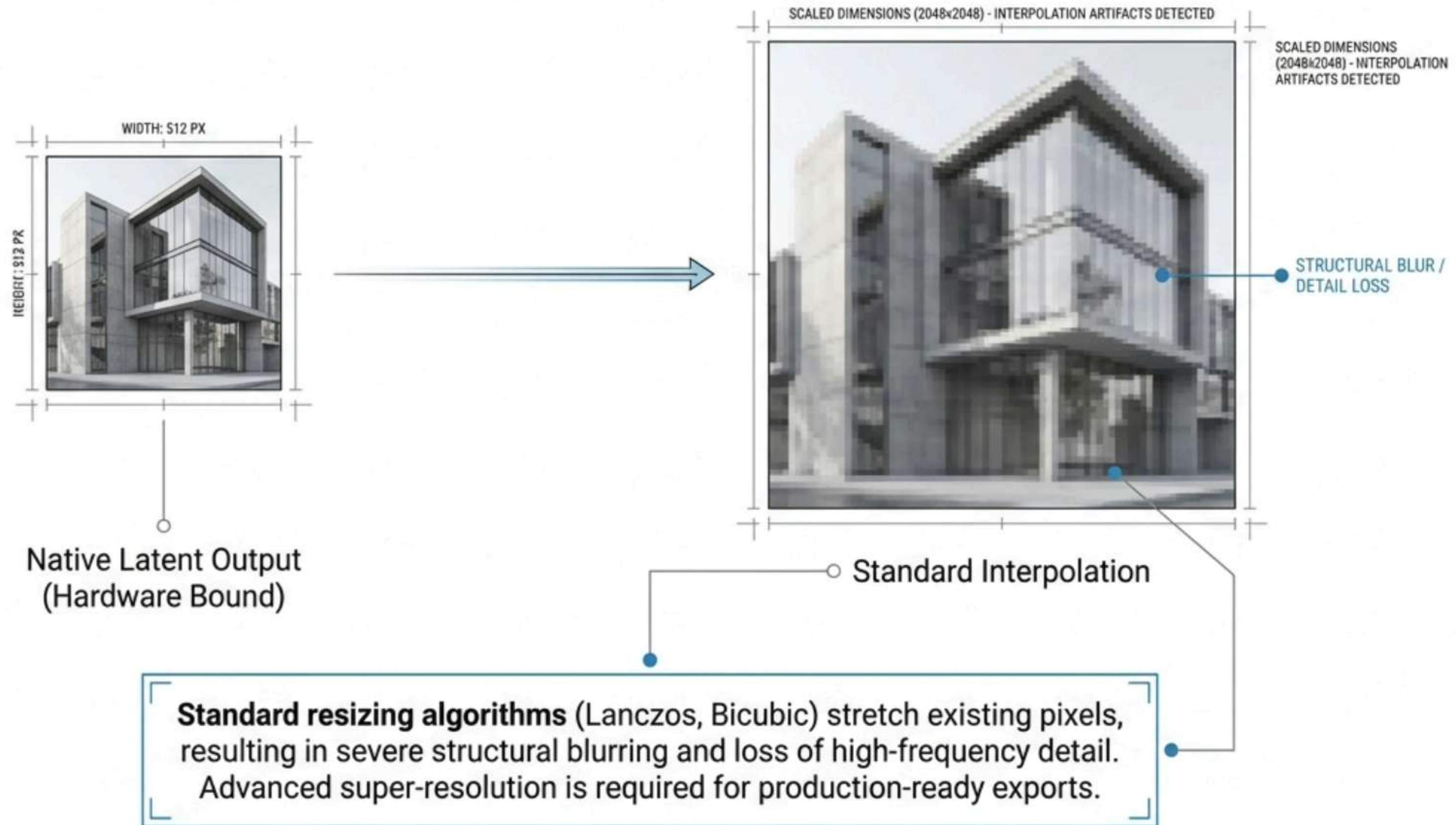
Technical Super-Resolution: Revisiting The Upscaling Problem

Last Semester, Ultimate Tiled Upscale actually achieved very good results

But let's analyze the general upscaling problem from its roots

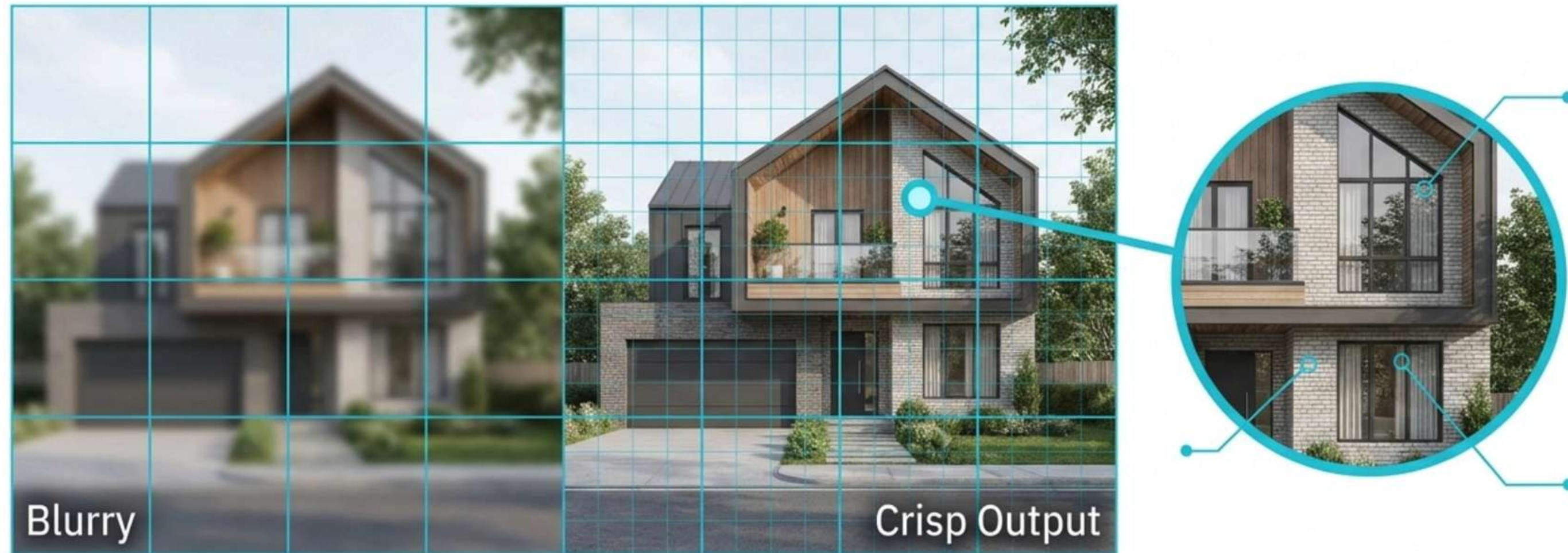
Technical Super-Resolution: Revisiting The Upscaling Problem

Standard Resizing is unreliable for high-quality upscaling



Technical Super-Resolution: Revisiting The Upscaling Problem

Paradigm A: The Detail Generator “Ultimate SD Upscaling”



Ultimate SD Upscale bypasses hardware limits by **slicing** the image into independent 512x512 **tiles**. It runs standard image-to-image generation on each tile using a **base model** (e.g., Realistic Vision) and a **ControlNet Tile** model to invent new micro-details.

Technical Super-Resolution: Revisiting The Upscaling Problem

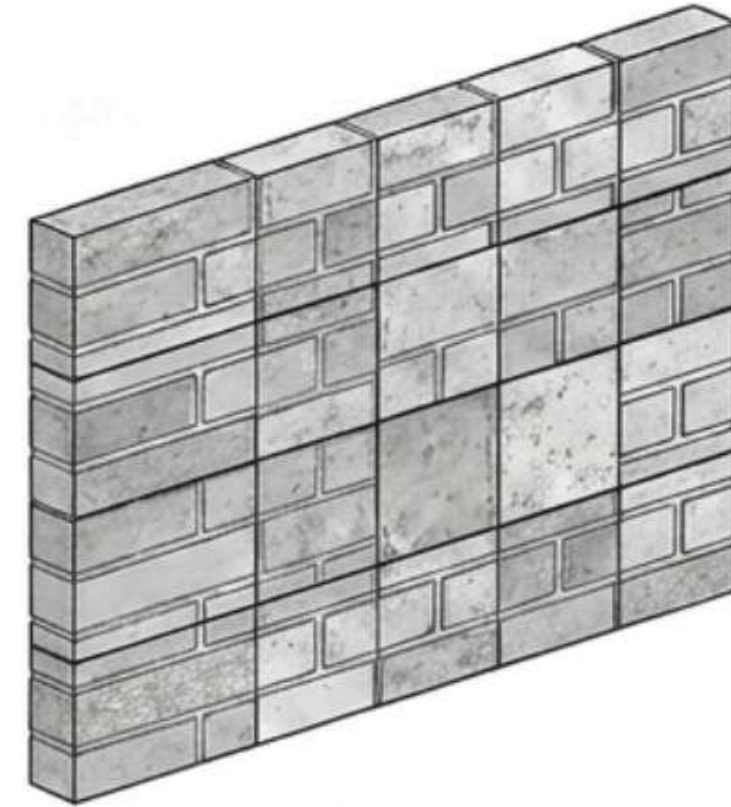
Paradigm A: The Detail Generator “Ultimate SD Upscaling”

Benefit: Detail Injection via Micro-Segmentation



Invents incredible **micro-details** (fabric threads, realistic lighting) absent from the source.

Drawback: Global Context Loss & Structural Hallucination



Because processing is localized, the model **lacks global context**. This makes it prone to **visible tile seams** and other **inconsistencies** between the independent tiles, as illustrated by the brick wall.

Technical Super-Resolution: Revisiting The Upscaling Problem

Paradigm B: The Faithful Restorer “SeedVR2”

Global Spatial Awareness: Enhances existing pixels, removes noise, and sharpens edges without inventing random artifacts.

SeedVR2 abandons the tiled img2img paradigm entirely.

Built on a massive Diffusion Transformer (DiT) architecture, it analyzes the **entire image frame globally**.

Low-Resolution Raw Generation:



Cleaned FHD Upscale Output:



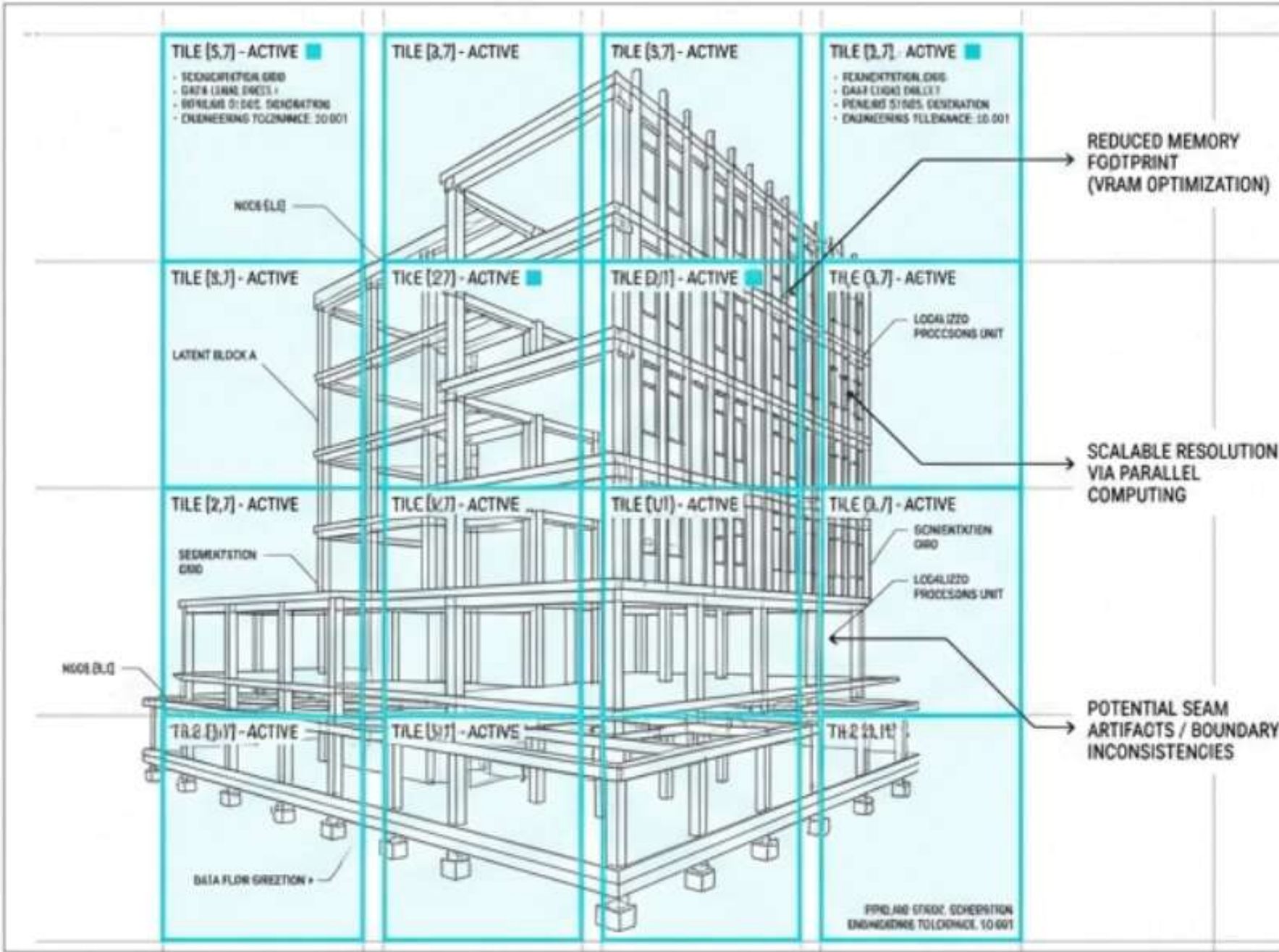
Technical Super-Resolution: Revisiting The Upscaling Problem

Overview:

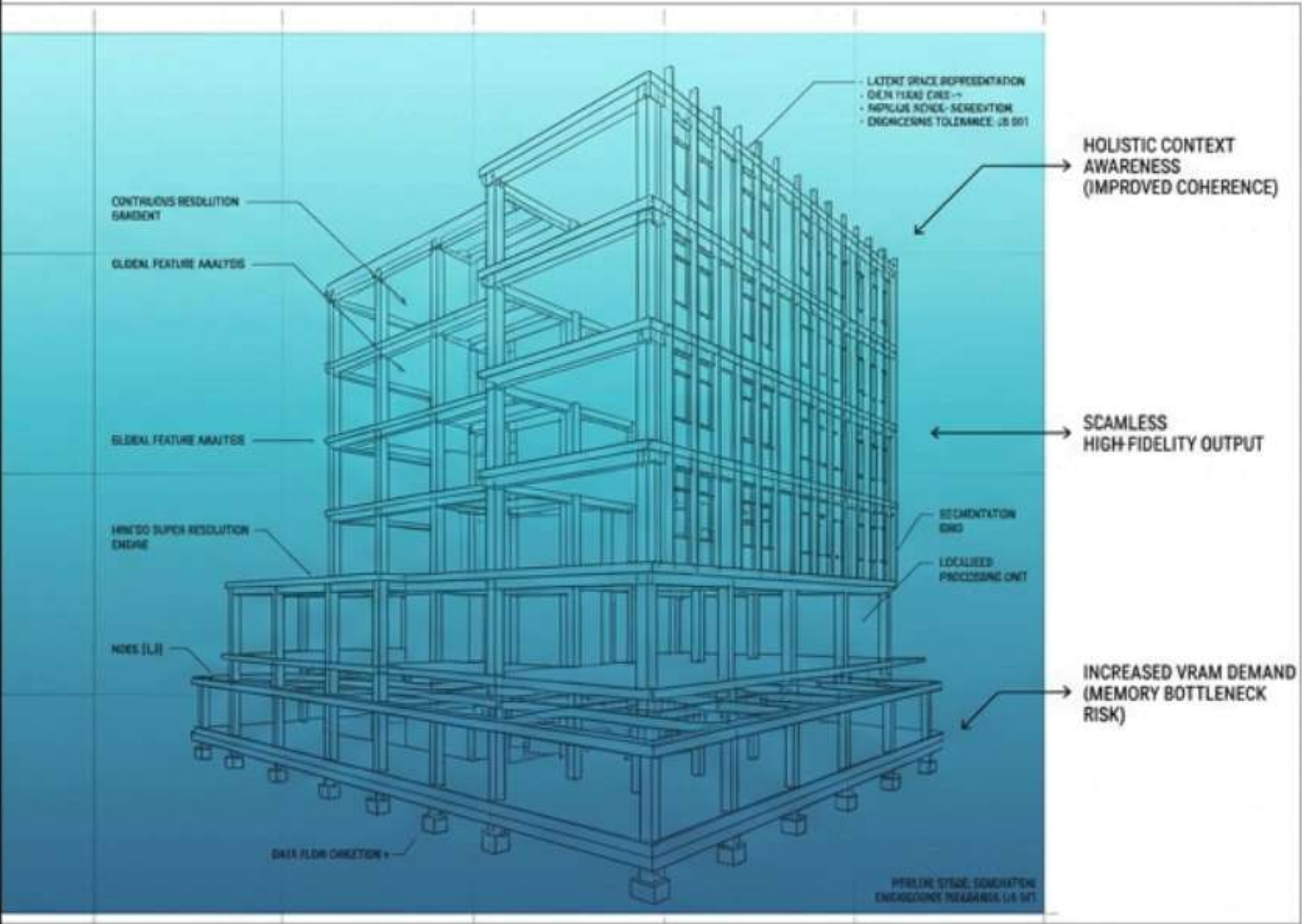
“Like Ultimate SD Upscaling”

“Like SeedVR2”

TILED GENERATION ARCHITECTURE



GLOBAL SUPER-RESOLUTION ARCHITECTURE



Outlines

The Generative Leap

Core Staging & Furnishing

Object Replacement & "Shop the Look"

Aesthetic Flexibility & Style Presets

Advanced Scene Understanding

Technical Super-Resolution

The Recommendation Engine

The Model Matrix & Ablation Studies

The "Semantic Snapper"

Offline Ingestion vs. Live Search

Pixels to Purchase

Beautiful renders are great.

But you can't buy a hallucinated sofa.

Introducing The Recommendation Engine

Introducing The Recommendation Engine

Bridging the gap between generative pixels and real-world e-commerce.

Three Specialized Models. One Unified Pipeline.

The Recommendation Engine: The Models Matrix

RT-DETR r50vd	SigLIP2 so400m	Qwen3-VL 4B Instruct
Function: Real-time Detection Transformer.	Function: Specialized Vision Encoder.	Function: Vision-Language Model (VLM).
Input: Unstructured RGB images (catalog or user upload).	Input: Isolated furniture crops.	Input: Isolated furniture crops + prompt instructions.
Output: Bounding box extraction, furniture crop generation, and confidence scores.	Output: 1152-dimensional normalized float vectors mapping visual similarity.	Output: Structured JSON taxonomy metadata (color, material, style, room fit) and natural-language descriptions.

But we didn't just guess which model to use.

The Recommendation Engine: The VLM Showdown (Ablation Study)

Model	Inference Latency	JSON Compliance	Verdict
2B Variant	9s	Failed	Prepends markdown fences; parsing fails at char 0.
Selected 4B Instruct	18s	Passed	Clean JSON output, highly accurate metadata. The optimal production balance.
8B Variant	23.5s	Passed	Marginally richer tags, but incurs a 31% latency penalty.

The Recommendation Engine: Beyond The YOLO Hype

YOLO

Complex Commercial Constraints



Many YOLO versions carry GPL/AGPL “copyleft” restrictions or require expensive commercial licenses for business applications.

Strategic Move for
Licensing Freedom

RT-DETR

Permissive Apache 2.0 Licensing



This license allows for unrestricted commercial use, modification, and redistribution without royalty obligations.

The Recommendation Engine: SigLip2: The Latency vs. Performance Sweet Spot

High-Precision 1152-Dimensional Visual Embeddings



1152-DIMENSIONAL
VECTOR

Converts furniture crops into ultra-detailed vectors for mathematically precise similarity matching.

Optimized for Commercial GPU Inference



12GB VRAM
CLOUD READY

The so400m architecture fits efficiently within 12GB VRAM footprints, allowing for cost-effective cloud deployment.

The Latency vs. Performance "Sweet Spot"



Larger models rejected to maintain the system's sub-60-second generation target without sacrificing search quality.

Powers Real-Time "Shop the Look"



Rapid embedding generation enables the system to fetch marketplace matches from the ChromaDB vector store instantly.

Why the so400m variant ??

Fixing The AI's Vocabulary

Fixing The AI's Vocabulary

AI models like to make up fancy color names

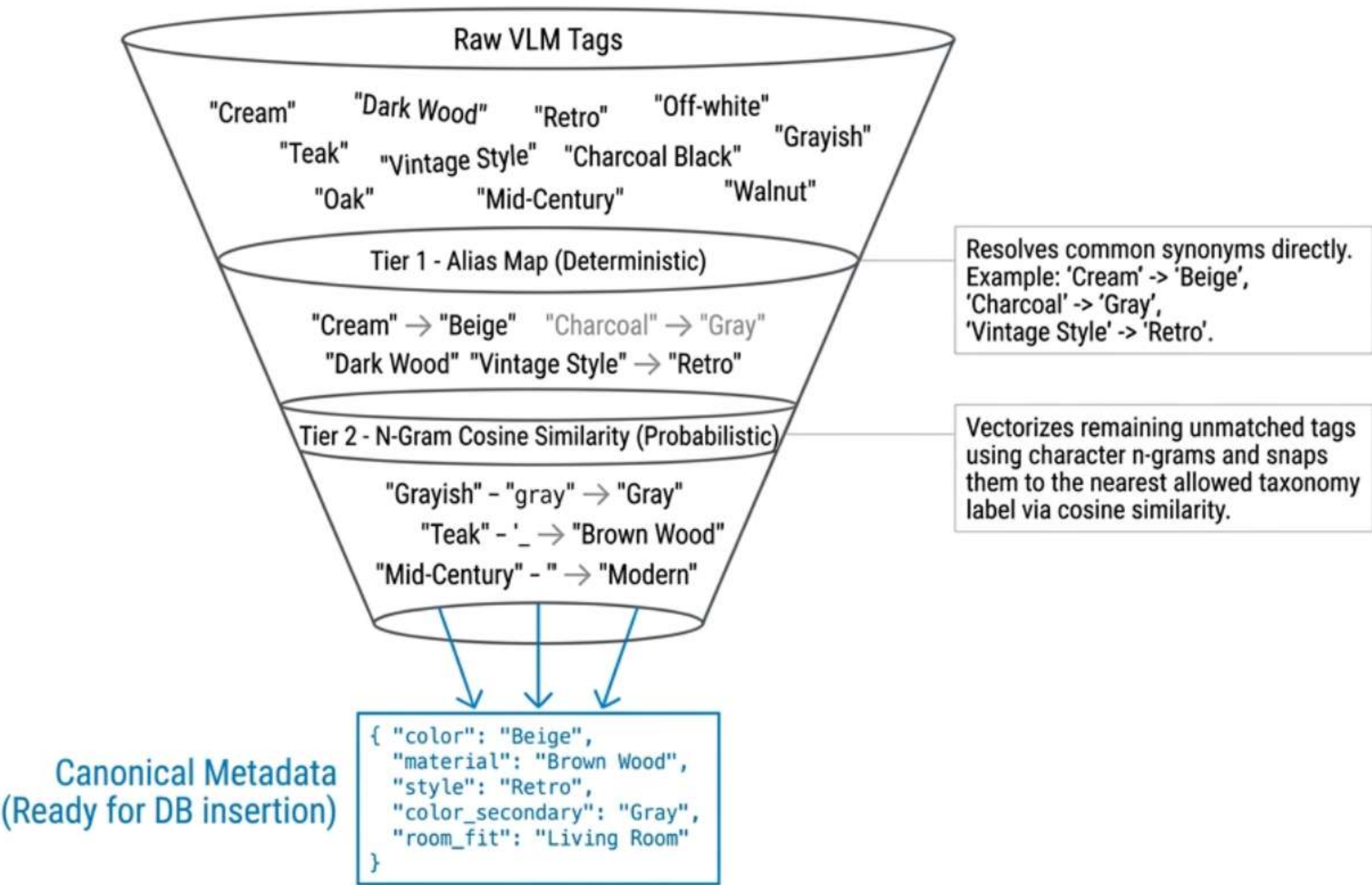
Fixing The AI's Vocabulary

AI models like to make up fancy color names

E-commerce databases **don't**

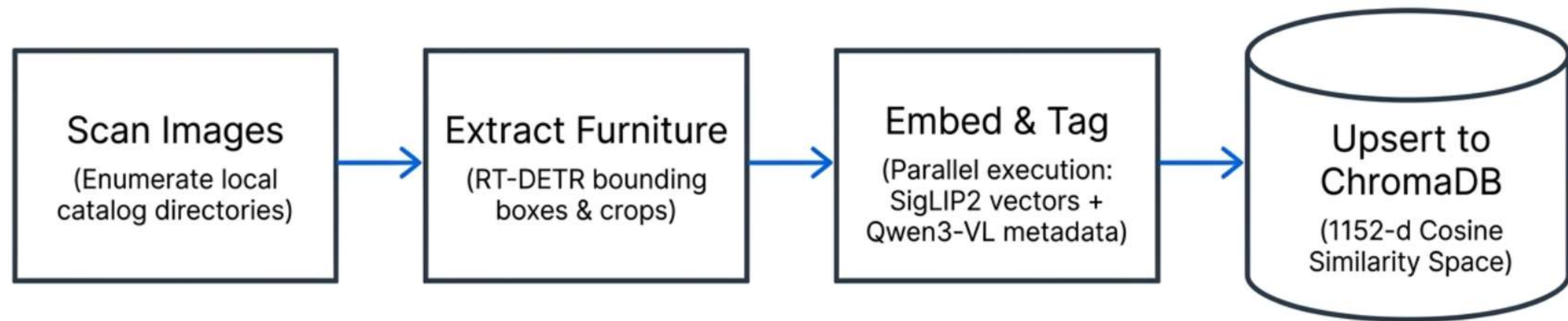
The Recommendation Engine: Metadata Normalization

The Semantic Snapper: Guaranteeing database hygiene by forcing VLM hallucinations into a canonical taxonomy.



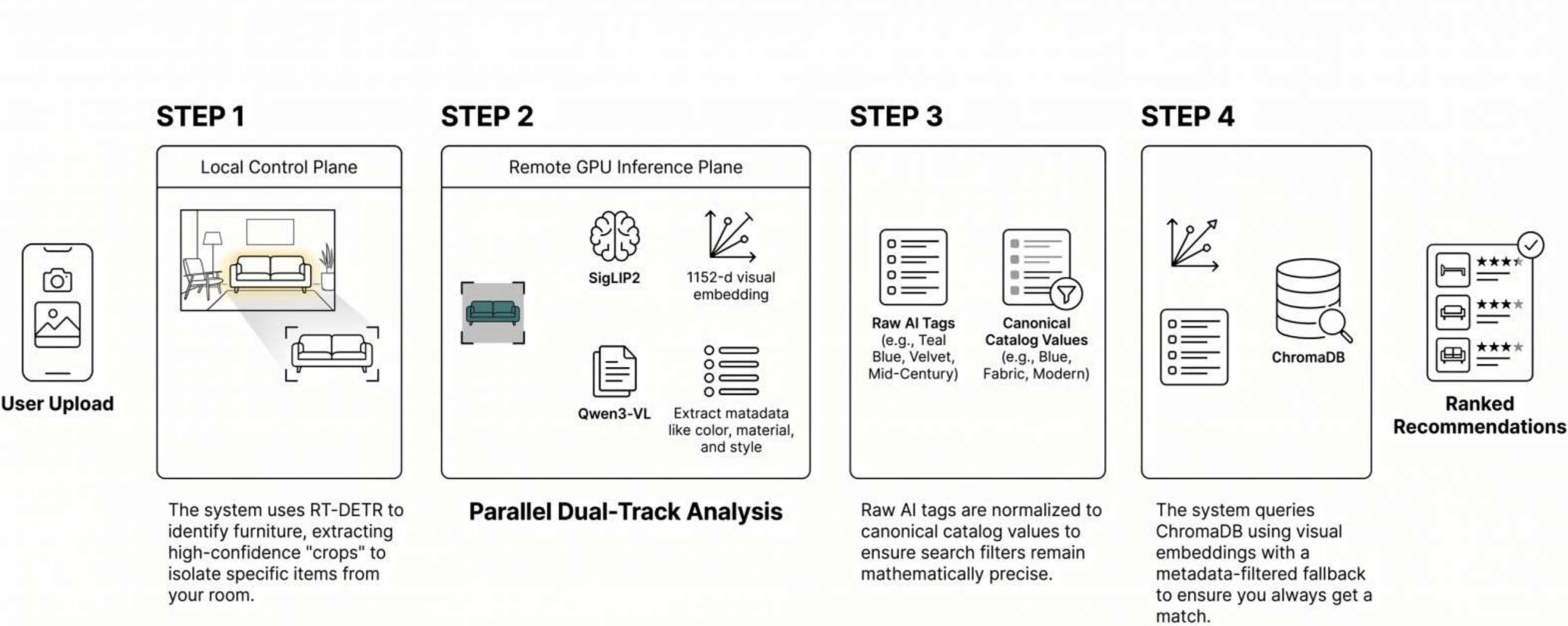
The Recommendation Engine: Offline-Ingestion (Digitizing The Catalog)

How we turn a real furniture catalog into an AI-searchable vector database.



The Recommendation Engine: Live Search (From Upload to Cart)

How the **Live Search** works



Bringing it all together !

The Recommendation Engine: Pixels to Purchase

The user **sees** the AI design, **clicks** a piece of furniture, and instantly **receives** the closest physical match from the store.





From latent space to your living space.