REGIONS
by DARYL ANSELMO
— JUNE 2024

ABSTRACT

From tape to stencils, and friskets to dodging - masking techniques have been widely used across all forms of traditional image production for centuries, allowing artists to protect certain parts of an image, while isolating others for specific changes.

For AI to be useful as a production tool, precise artistic control is required. While inpainting and outpainting are now common techniques, specifying distinct regions to be diffused with different visual information, up-front, is powerful and not as widely practiced.

For his 10th consecutive 13-week deep-dive, *'Regions'*, Anselmo focuses his latest exploration on the application of these methods to control image diffusion.

Fig 1.1: *unsettling greek revival, with pink clouds and corpse flowers - June 23, 2024*

Fig 1.2: *misty seascapes are vignetted around cycladic ruins - May 5, 2024*
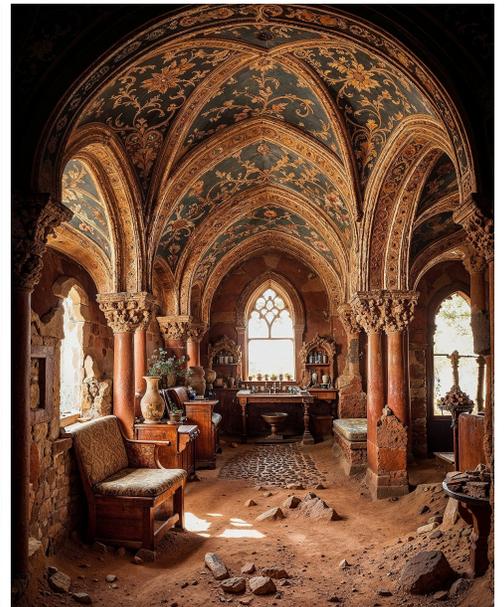
Fig 1.3: *a vertical gradient of medieval rosemaling over coober pedy - May 21, 2024*

PROCESS

At the most basic level, **regional prompting** involves creating a digital mask, constructing prompts for each region of the mask, then conditioning and combining the inputs used to guide diffusion.
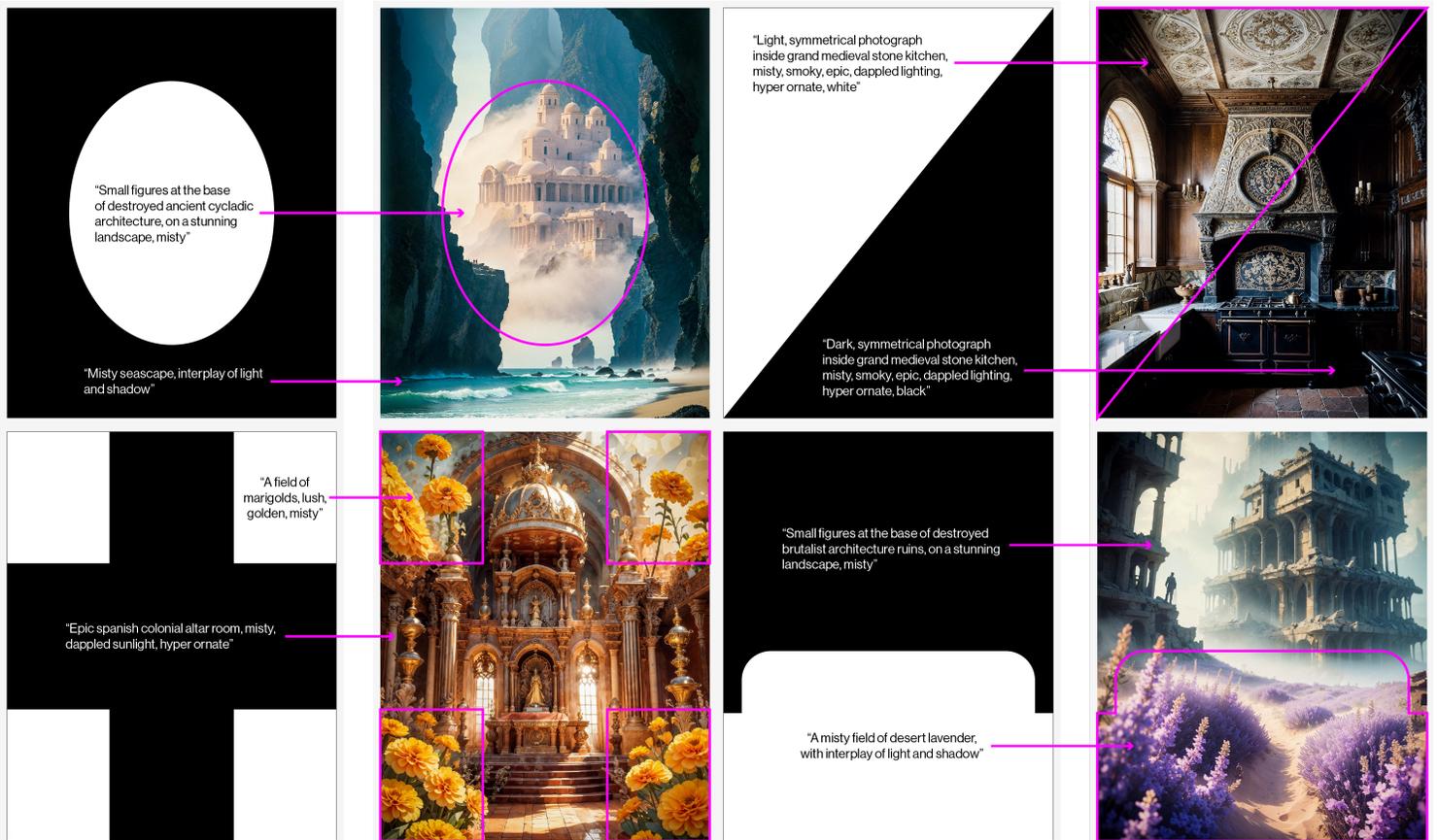
Fig 2: Examples of simple regional prompting

The amount of influence for a region can be controlled by a strength parameter on the Conditioning (Set Mask) node.
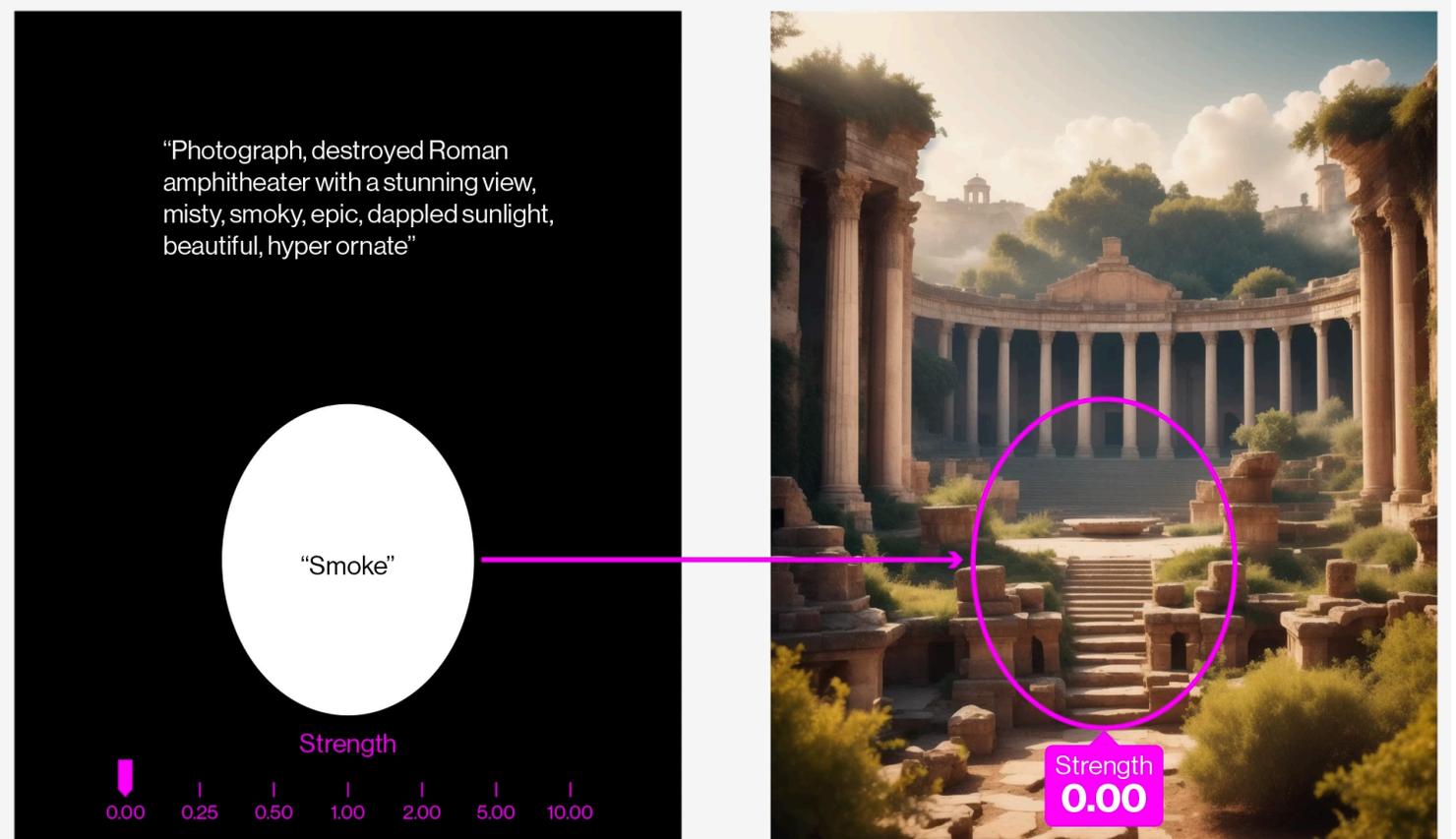
Images are injected into the process using ipAdapters. Attention masks (in addition to mask conditioning) are used to further diffuse information from a source image into the target region. The strength parameter of the ipAdapter, when combined with the strength parameter in the conditioning mask, produces very concentrated results.

Data for ipAdapters came from other generations (often from Midjourney), or, was generated on the fly and directly injected into the workflow. Data produced for ipAdapters was catalogued for future use.

Fig 4: Source images are converted into *attention masked ipAdapters*
From: unsettling greek revival, with pink clouds and corpse flowers - June 23, 2024

Gaussian blur and gradients are used in masked regions to soften edge transitions and blend visual information, creating more surreal worlds.

"An epic distant shot, Icelandic turf cottage in classic farmhouse style, made of white stone on a lazy river, dappled lighting"

"Green and yellow slime landscape"

Fig 5.1: Example using a Gaussian Blur node in the mask, used to feather the transition between two different prompted regions

"Symmetrical kodachrome photograph inside a grand, enormous medieval stone rosemaling bedroom, misty, smoky, epic, beautiful, dappled sunlight, hyper ornate"

"Coober Pedy"

"Symmetrical photograph inside a charming cabin, with bauermealeni, misty, smoky, beautiful dappled sunlight, hyper ornate, epic"
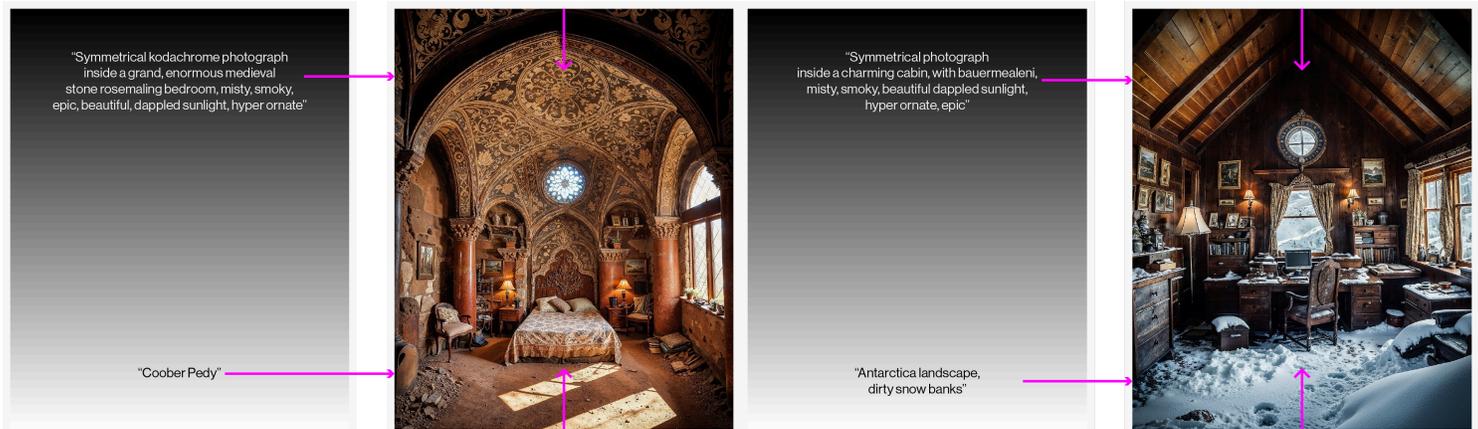
"Antarctica landscape, dirty snow banks"

Fig 5.2, 5.3: Examples using a Gradient Mask in regional prompting, blending two different visual concepts from top to bottom

Images generated throughout the project typically used between 2 and 5 masks, although theoretically the sky is the limit. Masks are often RGB channel-packed - a technique similar to one used in modern gamedev optimization. Channels are extracted from a channel-packed image using the *Convert To Mask* node.
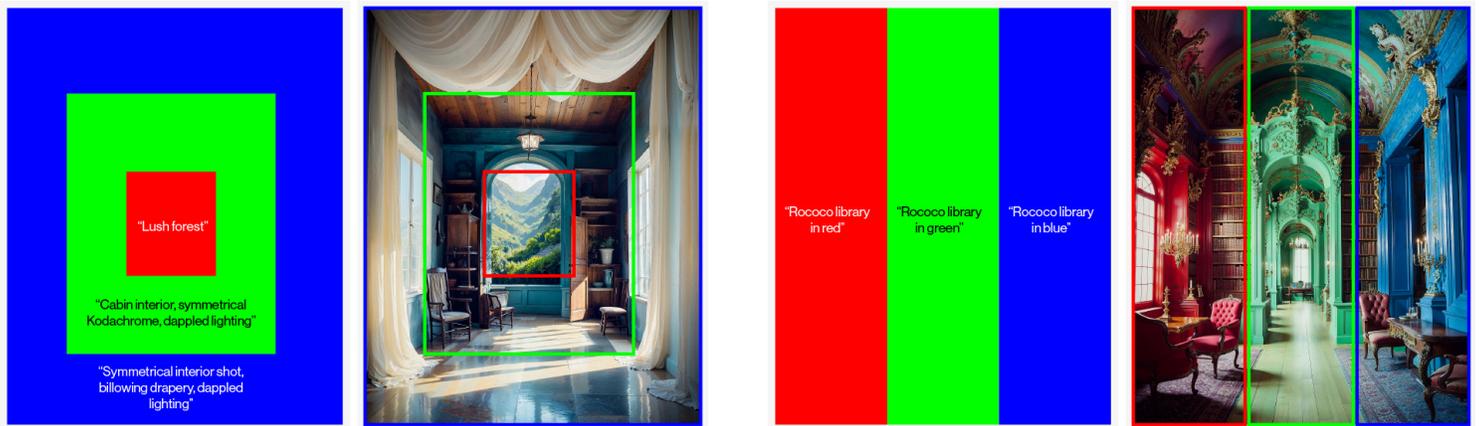
Fig 6: Examples of channel packing masks

Masks were either created offline (in Photoshop, Illustrator, Midjourney, After Effects) then loaded into the workflow, or, generated on the fly using various nodes. Latter stages of the project experimented more with generating masks on the fly using a variety of nodes, including `ComfyRoll` (by Suzie1 and RockOfFire), and `KJNodes` (by kijai).
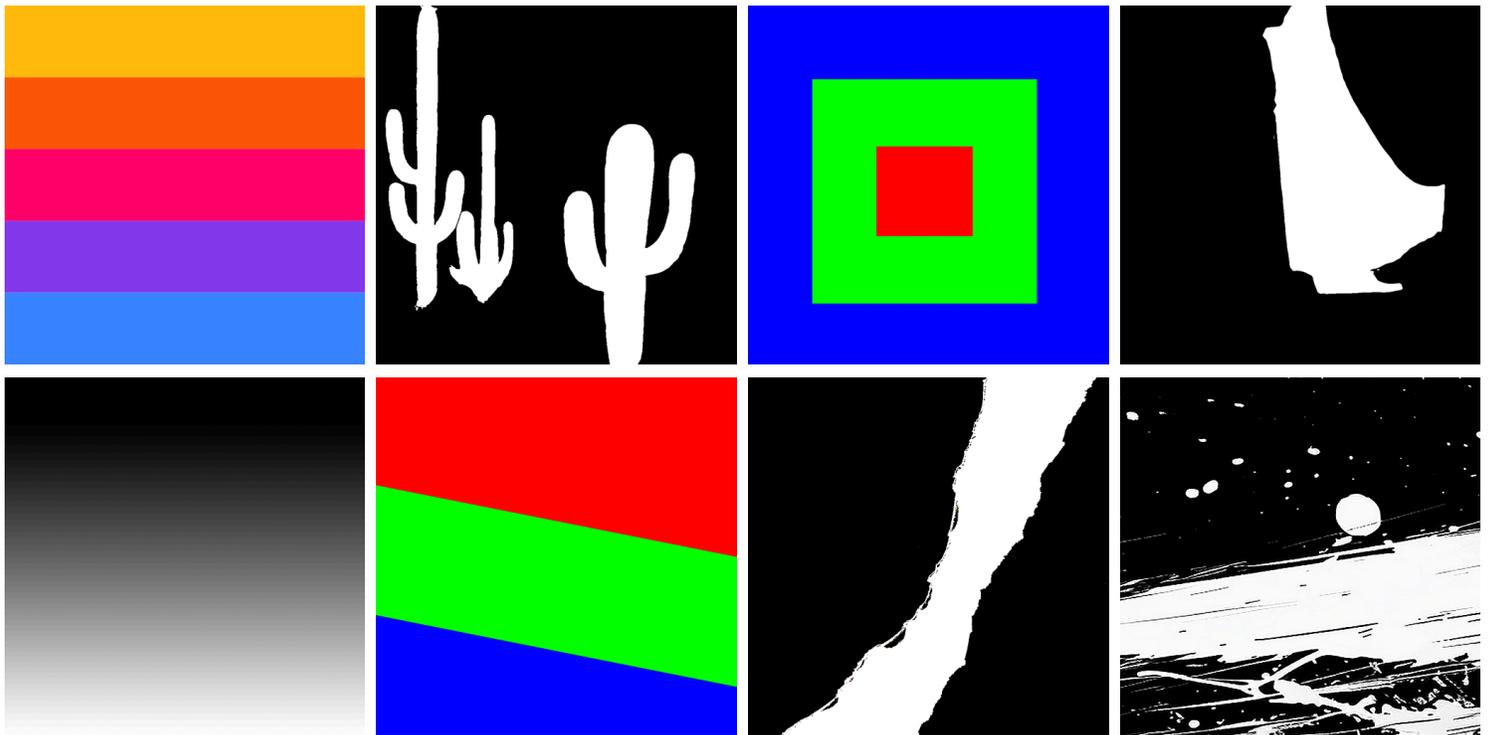

Fig 7: Examples of masks used throughout the project

DISCOVERIES   Z-depth based image stylization is a technique used in gamedev across a number of applications including object isolation, background removal, and visual stylization. One phase of 'Regions' used Zoe-depth estimation to generate masks, which are then used to simulate Z-depth based image stylization (colloquially called the *plunge* workflow.)

For demonstration, *plunged* images appear in a simplified abstract expressionist style in the foreground, and become progressively more photorealistic the deeper the image goes. Admittedly, at time of production, (and especially at time of writing) there are better depth estimation techniques for this task.

| BASE GENERATION | AB-EX STYLE TRANSFER | DEPTH ESTIMATION | BREAKDOWN |

Fig 8: Example of the *Plunge* workflow, using Zoe depth estimation to simulate a z-depth based image stylization
From: tropical ruins, abstract in foreground, more realistic with depth - April 16, 2024

Depth masks are refined throughout the project using a variety of techniques, including a demonstration of automated in-workflow background replacement.

| BASE GENERATION | REFINED DEPTH MASK | REPLACEMENT BACKGROUND | BREAKDOWN |

Fig 9: A depth mask is estimated and refined, then used directly in the workflow, in a background replacement operation.
From: dark red cavern drapery, the background is removed and replaced with a tuscan landscape - May 30, 2024

SAM and GroundingDino are used to isolate specific objects for further processing. In this case, the base image is generated, and style transferred into a second image. SAM is used to segment, and GroundingDino used to isolate the 'cactus' objects from a simple prompt. The identified cactus from the stylized image are pasted back over the photorealistic image. This technique has

many applications and could be expanded for use in transparent object generations, automated object replacement, or layered workflows.

| BASE GEN | STYLE TRANSFER | SEGS MASK | SEGS | SEGS ALPHA | BREAKDOWN |

Fig 10: A demonstration of SAM and GroundingDino used to create transparent layers.
From: painterly cactus inserted into hyperrealistic desert landscapes, with ruins - April 15, 2024

Stable Cascade was used for a handful of experimental images. When used in concert with the LCM sampler, clean geometric shapes and lines can be achieved. In 'Regions', all Stable Cascade generations were eventually style-transferred to other architectures using a Multi-ControlNet setup. Snippets were adopted from `The Raven's Workflow` (by u/-ellary-).

Fig 11.1: *white, geometric renaissance interior depth-plunged over a lush dreamscape - June 4, 2024*

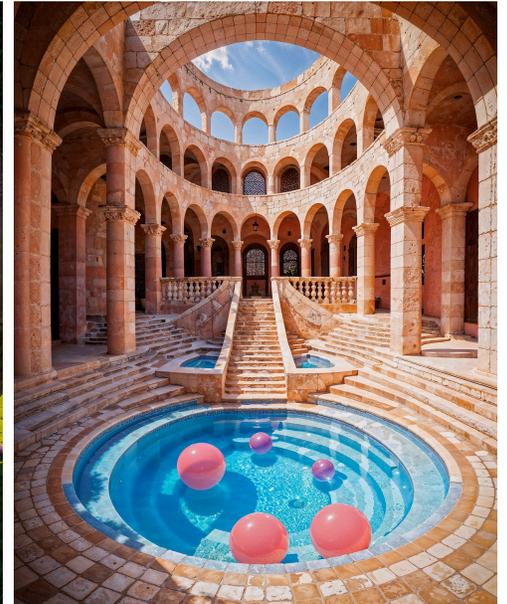Fig 11.2: *"manicurved" geometric landscapes - June 5, 2024*

Fig 11.3: *geometric pools with cycladic influences and spheres, in pink and blue - June 17, 2024*

It wasn't until very late in the project, Anselmo began breaking some of the unspoken rules of diffusion. Masked regions became more clearly defined, and styles more crisply distinct from each other. The main discovery was that the core workflow was suited to collage surrealism, and deliberate violations of hi-res fix were employed. In these cases, sampling to shrinks visual information into smaller relative regions and fills the gaps with details where they don't exist, resulting in distinctive, high resolution, chaotic visuals that embrace 'ai-ness'.

Fig 12.1: *processing an overwhelming amount of information - June 27, 2024*

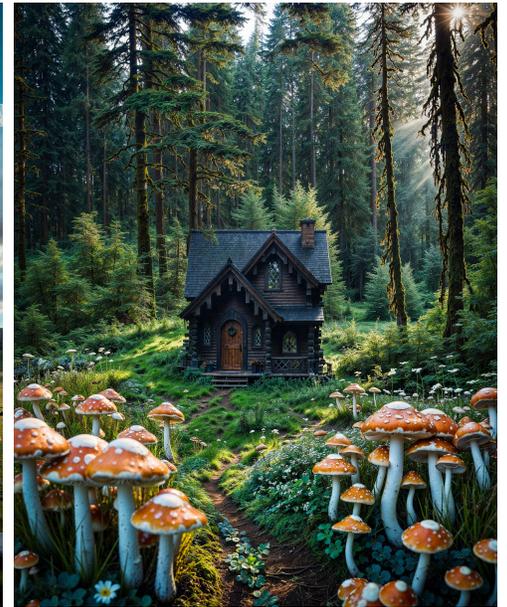Fig 12.2: *modernism is blended into a chaotic pile of stone boulders - June 28, 2024*

Fig 12.3: *dark gothic cabin centered in a 3x3 grid, with two boxes of mushrooms in the corners, and the other six set in the pacific northwest forest - June 29, 2024*

COLLECTION   The entire collection can be viewed here.

BIO

Daryl Anselmo is a Canadian-American artist, director, advisor, and founder. He is the co-creator of the original NBA Street and Def Jam franchises for Electronic Arts, was the Art/Creative Director for FarmVille 2 at Zynga, and served for many years as a Director of Art for The Walt Disney Company.

Now an artist and proponent for the creative use of AI-based workflows, Daryl has lectured at numerous institutions including Stanford University, Siggraph, UC Berkeley, and Google. His work was showcased on the Main Stage at Ted 2023.

Currently splitting his time between San Francisco and Vancouver, Daryl is obsessed with technology and writes his own code. He is currently deepening his art practice and providing consulting and creative services for various clients.

INFO

- 91 sets of images, released daily between March 31 and June 29, 2024
- 935 total images. 623 in portrait aspect ratio (3584 x 4608), 171 in landscape (5376 x 3072), and 141 in 'ultrawide' landscape (6144 x 2560)
- Checkpoints used: realvisXL40 (35), realisticVision_v5.1 (35), lahMysteriousSDXL_v40 (34), stable_cascade_stage_c/b (14), realisticVisionV60B1 (12), leosamsHelloWorldXL50 (8), juggernaut-x-v10 (5), epicRealism_naturalSinRC1 (1)
- 265 total hours spent
- Sample workflows and data are available for download here.
- Download a PDF of this whitepaper here.
- View the entire collection here.

A limited run of signed prints are available. Contact the artist here.