

Theme Worlds

Wide panoramas whose **left and right edges are the same edge** — roll one sideways and no seam exists.

Stable Diffusion v1.5 only.

Zero source images. Zero fine-tuning.

One notebook, one free Colab T4.

Marlen Melis

KAIST · Fall 2026



volcanic_lava_world + flooded_metropolis · seed 201



cyberpunk + underwater · seed 1001



fantasy + medieval · seed 1002

SD v1.5 paints 512x512 — so never ask it for more

1 · MULTIDIFFUSION, RE-IMPLEMENTED

Ask SD v1.5 directly for 2048x512 and it paints **four 512-ish scenes glued together** — duplicated skylines, warped geometry. Instead: keep **one shared latent canvas**, cut overlapping 64x64-latent (512 px) windows, run one ordinary DDIM step on each, and **average every window's opinion back into the canvas at every step**.

$$x_{t-1} = (\sum_i S_i^T \Phi(S_i x_t)) \oslash (\sum_i S_i^T 1)$$



Stride 16 latent cells → **16 windows per step**, every latent cell covered ≥ 4x. No weight is touched and no tensor exceeds 64x64 latents — which is why 2048x512 fits on a free T4. **DDIM is required: scheduler.step()** runs per window, out of order, and DDIM is stateless (SD v1.5's default PNDM is not). fp16 UNet, **fp32 fusion accumulator**, ~400 batched UNet calls per candidate.

2 · CIRCULAR WINDOWS = THE LOOP



The **amber window runs off the right edge and continues on the left** — `idx = arange(w0,w1) % W`. Both ends share windows from step 1 and are denoised as one surface; **nothing is stitched afterwards**. The VAE also gets ±8 latent columns of circular padding, cropped after decoding, so the decoder never sees a boundary either.

Measured over **12 paired runs**: seam ratio **7.94** with wrapping off versus **1.05** with it on — a mean **7.6x** improvement on identical seeds, never below 5.4x.

3 · TWO THEMES, NO TRUNCATION



A fused two-theme prompt measures **97 tokens**; everything past 77 is silently discarded — and what is cut is the tail, i.e. the quality suffix, every time.

So I never concatenate. The sampler takes **K conditioning branches**, each with its own embedding *per window*; one UNet call carries the unconditional pass plus all K branches:

$$\hat{\epsilon} = \epsilon_u + g \cdot \sum_k w_k (\epsilon_k - \epsilon_u)$$

compose — both themes on every pixel (costs 1.53x)
regions — each theme owns a side of the loop, via a circular raised-cosine field
blend — one averaged embedding everywhere
single_prompt — the naive concatenation, kept as a control

CONTROLS — CYBERPUNK + UNDERWATER, ONE SEED, 1024 × 512, 30 STEPS, IDENTICAL IN EVERY OTHER RESPECT



circular OFF · rolled 50% · seam **6.39**



circular ON · same seed · seam **0.97**



single_prompt — dry skyline, reef gone



blend — neither city nor reef



regions — reef centre, towers at the ends



compose — coral-encrusted neon towers

Read it left to right: wrapping the windows removes the join entirely (6.39 → 0.97 on the same seed); then the four ways of combining two themes — from the naive concatenation that **deletes one theme outright**, through an averaged embedding and a region split, to noise-space composition that fuses them into one hybrid world.

Does it actually loop? Roll it, then measure it.



As generated — cyberpunk, seed 1000, 2048 × 512



Rolled 50% — the former left/right join now sits dead centre. Nothing to see. Seam ratio **1.16**

The metric. $R = \text{mean}|\text{first col} - \text{last col}| \div \text{mean}|\text{neighbouring col difference}|$ — $R \approx 1$ means the wrap-around join is statistically indistinguishable from any other pair of adjacent columns, while a strip merely cropped to size blows up.

REPRODUCIBLE TO THE BYTE

Two Colab sessions, different days, same seeds → **byte-identical PNGs** (MD5 f737a887..., 16f0c451..., 8a6e4445...). Initial noise is drawn on the **CPU** generator, whose RNG is identical across GPUs.

A BUG FOUND IN THE REFERENCE

diffusers' deprecated panorama pipeline scatters windows by *canvas* row index, so circular padding breaks on any canvas taller than one window. Mine never addresses a window by canvas coordinates — which is why the 768-tall worlds work.

WHAT I LEARNED

Tokenisation, not attention, is what kills fused prompts — measuring the prompt beat staring at the image. Composition is seed-sensitive, so the notebook renders several seeds and ranks them on a **worst-of-both-themes** CLIP score.

1.12

Mean seam ratio

28 panoramas, 12 configurations; 26 of 28 below 1.5

7.6×

Circular on vs off

12 paired runs: 7.94 → 1.05 on identical seeds; never below 5.4×

1.53×

Cost of compose

K+1 passes per window, 9 paired runs; theory said 1.5×

0.99

vs diffusers' own sampler

Reference 0.95 ± 0.22 vs mine 0.99 ± 0.22 over 12 pairs

CIRCULAR ABLATION — SAME SEED, SAME STEPS, SAME PROMPT

ink_wash_mtns + dragon_spire	14.63	1.36
fantasy + medieval	11.91	1.11
volcanic_lava + flooded_city	8.51	0.78
cyberpunk + underwater	6.39	0.97
...12 runs, mean	7.94	1.05

Everything loaded, everything set, nothing borrowed

PRETRAINED WEIGHTS — THE COMPLETE LIST

SD v1.5 (fp16) stable-diffusion-v1-5/ stable-diffusion-v1-5	Generates every submitted pixel. Ships its own CLIP ViT-L/14 text encoder + VAE; scheduler swapped to DDIM.
CLIP ViT-L/14 openai/clip-vit-large-patch14	Scores finished images only — generates nothing. The allowed "pretrained reward model" case; switchable off.
Source images	Zero of the 10 allowed — no image input, no upload step, no data.zip , no training of any kind.

CANDIDATE SELECTION

CLIP scores **six square crops taken around the loop**; a theme's score is its best crop and the candidate's score is the **worst of the two themes** — which punishes a render that quietly dropped one. Seam ratio > 4 is disqualified. Every run writes a JSON log with checkpoint id, prompts, seeds and scores.

SETTINGS

Canvas	2048 × 512 · also 1536 × 768
Window / stride	64 × 64 latents / 16 cells
Sampler	DDIM · 50 steps · CFG 7.5
Windows per step	16 · cells covered ≥ 4×
Precision	fp16 UNet · fp32 accumulator
Fusion	compose λ 0.5 · regions σ 2.0
Decode	± 8 latent cols wrapped
Control render (1024×512, 30 st.)	32 – 36 s
...same render, compose	49 – 57 s

Prompt = "{theme}, seamless wide panorama, sweeping vista, single horizon, ... digital matte painting" (≤ 55 tokens); the negative prompt targets seams, tiling, duplicate objects and text (66 tokens).

RULE COMPLIANCE

- ✓ Every pixel from SD v1.5
- ✓ No SDXL / FLUX, no third-party checkpoint, LoRA or DreamBooth
- ✓ No training or fine-tuning
- ✓ Zero source images → no **data.zip**
- ✓ CLIP used only as a judge
- ✓ Single **.ipynb**, end to end on a Colab T4, all cell outputs saved
- ✓ Every seed fixed and logged; CPU-drawn noise reproduces across GPUs
- ✓ Every label drawn with PIL on finished pixels — no text asked of the model
- ✓ Static PNG output



giant_tree_civilization + biopunk_organic_city



ink_wash_mountains + dragon_spire



arctic_ice_cave + cozy_village_snowfall

The takeaway: a sampling schedule, not a bigger model — circular MultiDiffusion turns SD v1.5's 512 px window into a seamless 2048 px world, and multi-branch conditioning lets two themes share it without ever truncating a prompt. MultiDiffusion: Bar-Tal et al., ICML 2023 · Composable Diffusion: Liu et al., ECCV 2022 · CLIP: Radford et al., ICML 2021