Surrealist Timekeeping with Generative AI: Diffusion Local Time

Lee Steven Butterman, Poeta ex Machina Labs
Benjamin Forest Fredericks, Yale Medical School

Abstract

Diffusion Local Time is the forefront of clock face technology, and solves the tyranny and the tedium of high-visibility low-computation low-power timekeeping like the household clock.

Through the use of Stable Diffusion 1.5, steered by a ControlNet, the clock face is a customizable generative AI image that renders the hours and minutes of the time with the clarity of spotting faces in clouds, relying on the viewer's pareidolia for them to read the time.

Diffusion Local Time lives in a place between 1) Christian Marclay's *The Clock*, a 24-hour film that is a montage of thousands of film and television images of clocks, edited together so that they show the actual time; 2) `xdaliclock`, a graphics-intensive timepiece based on a Mac 128K application, where digits morph into each other; 3) digital wristwatches, which are a pretty neat idea.

The code is at `https://github.com/lsb/diffusion-local-time`, and runs on everything from powerful GPUs to small Raspberry Pis.

Figure 1: 8:00 PM, as rendered by *Diffusion Local Time*.

Figure 2: 7:48 PM. Note that the legibility changes based on proximity.

1 Introduction

Stable Diffusion 1.5 [SD1.5] is a text-to-image model based on diffusion techniques. The output is conditioned by a prompt, and can also be steered by an additional ControlNet [ControlNet] model.

The control can be anything, from the depth of objects in the scene to be rendered, to edges of objects, to the brightness of the image. This last control, of the luminance of pixels, enables the output of photorealistic imagery that can be interpreted as digital imagery, like barcodes or QR codes [NHCIAO].

Stable Diffusion decomposes the problem of formulating an image into a sequence of denoising steps, often 50 or more. It is possible to distill the de-

Figure 3: A leafy park at sunrise in a large bustling city with tall buildings.

noising trajectory into a Latent Consistency Model [LCM], which can render a high-quality image into only 4 steps or fewer. In empirical tests, image quality suffered with fewer than 4 steps and a legible control image.

Because the readability of the time relies on the viewer's imagination through pareidolia, typographical choices are crucial for the project's usefulness. The typeface is Atkinson Hyperlegible [Atkinson], which is "designed to focus on letterform distinction to increase character recognition, ultimately improving readability."

The legibility of the image is a function of many variables: the prompt, the size of the image, the target distance from the viewer, the strength of the control, and the rest of the scene being rendered for a particular initial noise pattern. When changing the imagery, or the size of the image (based on hardware speeds), it can be useful to do a manual grid search of control strengths and find a strength that generates aesthetically pleasing images that are legible, like in Figure 4. Larger images allow for greater flexibility to find plausible objects in a scene for a prompt, so for a certain legibility, choosing a smaller images implies that the mean controlnet conditioning strength would be higher, and the variance would be higher as well, as in Figure 5.

It can be aesthetically pleasing to interpolate between frames of an animation, for a smooth transition. Sometimes, two frames of an animation can be projected into a conceptually smooth high-dimensional space for images, to interpolate between them in meaningful ways. In *Diffusion Local Time*, the frame rate on 2023-era clock-priced computers with 2023-era software can render one high quality 1440x810 frame on CPU in 45 seconds (or GPU in 10 seconds, risking transient GPU errors). Thus, latency requirements do not permit animations, though there is still an aesthetic need to keep the animation smooth from image to image. Thus, the seed only changes once an hour, which greatly increases the probability of the scene staying consistent from minute to minute as only the last digit of four changes.

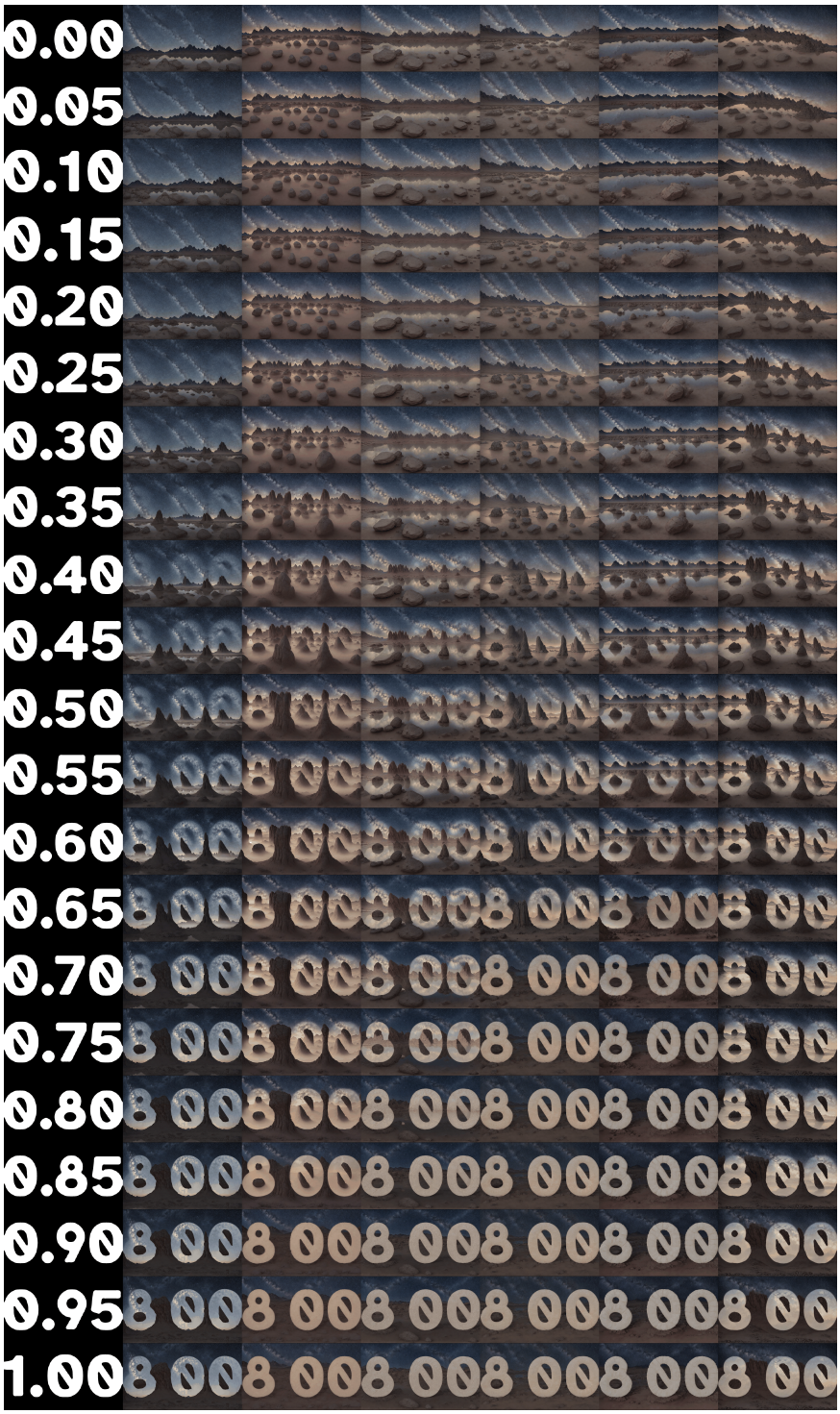

Figure 4: 8:00, for ControlNet conditioning strengths from 0.0 to 1.0, for six different random seeds, at size 1440×810. Note that strengths around 0.45 uniformly balance legibility and aesthetics for most seeds.

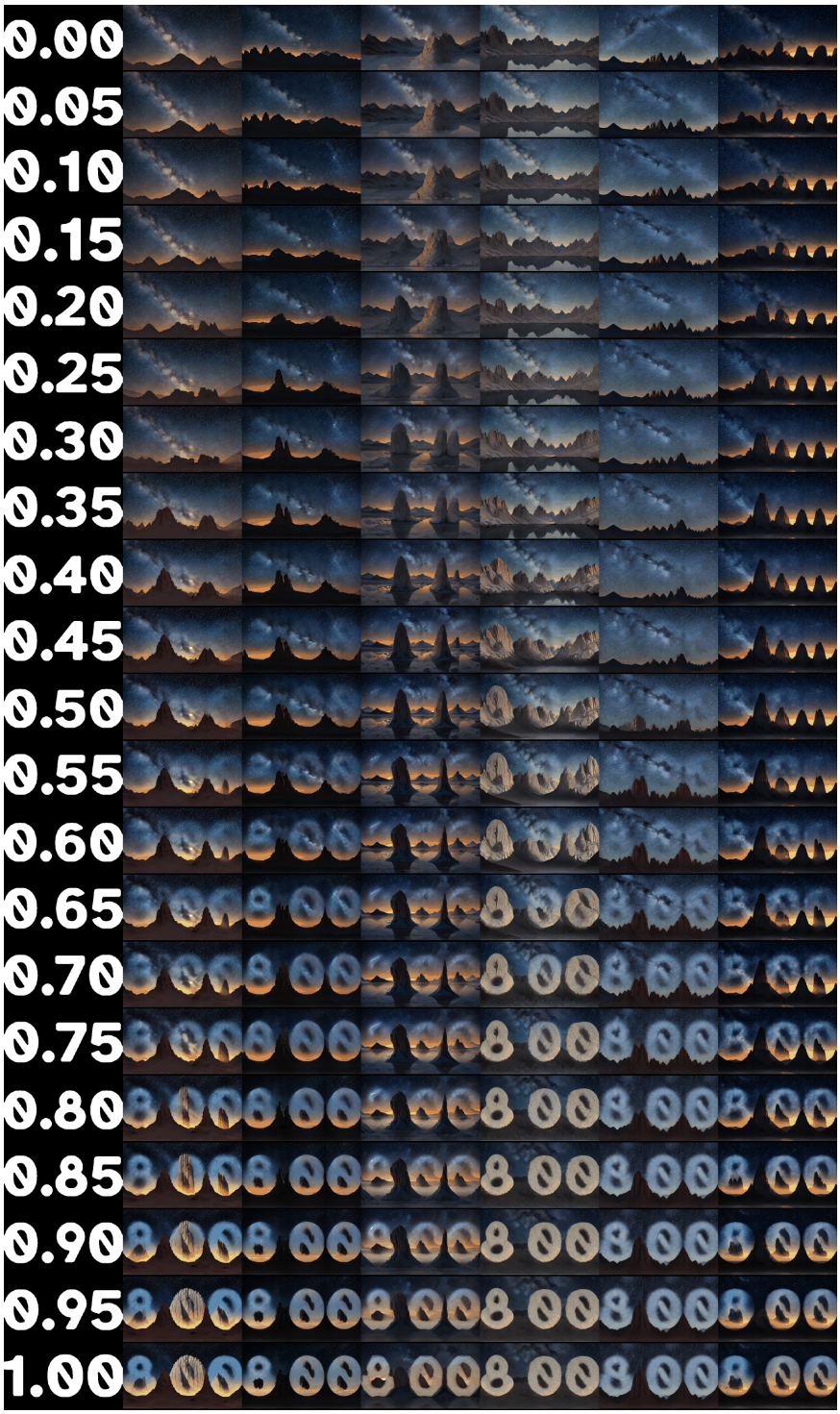

Figure 5: 8:00, for ControlNet conditioning strengths from 0.0 to 1.0, for six different random seeds, at size 480×270. Strengths around 0.7 for many of the seeds are optimal, but image plausibility has declined for a legible image, and the variance between seeds for legibility at a particular strength is higher.

2 Costs

2.1 Bill of Materials

- Raspberry Pi 5: -$80
- 32GB μSD Card: -$10
- 1920x1080 HDMI Monitor: -$110

2.2 Power

A typical screen will consume 8W of power, and a Raspberry Pi under load will consume 12W. A wall clock might consume 1 AA per year, which is 4Wh per 525600 minutes [Larson], or 7.6μW. Therefore, this artwork has 2-3 million times the power and impact of a household clock. Upgrading to a desktop with a cutting edge NVIDIA RTX 4090, which can render one *Diffusion Local Time* image every second, can consume up to 450W, with 40$\times$ the impact and power.

3 Related Work

3.1 Christopher Marclay's *The Clock*

The Clock is a 24-hour film that is a montage of over ten thousand clips from film and television of clocks, edited together so that they show the actual time. Marclay viewed the film as a memento mori, drawing attention to how much time the audience has spent watching it, compared to the escapism that cinema often provides. [Marclay]

In contrast, *Diffusion Local Time* quickly lets the audience lose themselves in the landscapes, and revels in the escapism that is a display that quite literally disappears on close inspection.

Further, *Diffusion Local Time* is field-servicable for the user's own tastes, and the time can optionally be synchronized with networked time servers on startup as needed. Advancements in text-to-video technology [Sora] make one hopeful that a future version of *The Clock* will be able to be generated on the fly, and be able to be customized to the user's tastes.

3.2 Steve Capps' and Jamie Zawinski's *Dali Clock*

The *Dali Clock* is a graphics-intensive timepiece, created originally by Steve Capps for the Xerox Alto and then rewritten for the Mac 128K, and then rewritten by Jamie Zawinski for XWindows as `xdaliclock` and also for web browsers in Javascript. [XDali]

The animation is smooth, even on a 7 MHz computer, morphing from HOUR:MINUTE:SECOND to HOUR:MINUTE:NEXT SECOND. The typography is fixed, and the background color changes slowly over time.

3.3 Digital wristwatches: a pretty neat idea

Digital watches came along at a time that, in other areas, we were trying to find ways of translating purely numeric data into graphic form so that the information leapt easily to the eye. For instance, we noticed that pie charts and bar graphs often told us more about the relationships between things than tables of numbers did. So we worked hard to make our computers capable of translating numbers into graphic displays. At the same time, we each had the world's most perfect pie chart machines strapped to our wrists, which we could read at a glance, and we suddenly got terribly excited at the idea of translating them back into numeric data, simply because we suddenly had the technology to do it. Compare  to **15:39**, especially in situations where seconds count at odd times, like internet flash sales, or flying commercial.

4 Diagnostic Usage

Diffusion Local Time relies on the viewer to intuit the visual illusion to recognize the time, and will usually display a wave of enthusiastic recognition when they perceive what is going on.

Pareidolia is commonly associated with seeing faces in clouds, imposing a meaningful interpretation on a nebulous stimulus. It is well established that part of the brain that recognizes faces can also activate when the brain recognizes novel objects like 'greebles' [Gauthier], and literacy is widespread in many nations, leading to the possibility of pareidolia presenting as seeing text in clouds or inkblots.

Pareidolias do not reflect visual hallucinations themselves but may reflect susceptibility to visual hallucinations, and increased hallucinations can be indicative of mild cognitive impairment or dementia. [UchiyamaDLB]

The state of the art is the Pareidolia test, which is a test for dementia with Lewy bodies, where the patient is shown a series of images and asked to identify if they see a face in the image. [MamiyaPT] The test is useful as one small part of a differential diagnosis, but one small problem with the current test is the large difference between positive and negative images, as shown in Figure 6.

Future work includes an interdisciplinary project between Neuroscience, Facilites Management, and potentially Catering Services, to redecorate hospital waiting areas with static or dynamic landscapes, to allow patients to self-administer a generative AI pareidolia test that provokes a comment to staff, either about the timepieces in the office, or potentially about the offer of food that they have read in the landscapes of corgis or seashore boulders. Small boxes of apple juice and graham crackers can be placed out of view in the waiting area, and the patient's request can be recorded for potential diagnostic value.

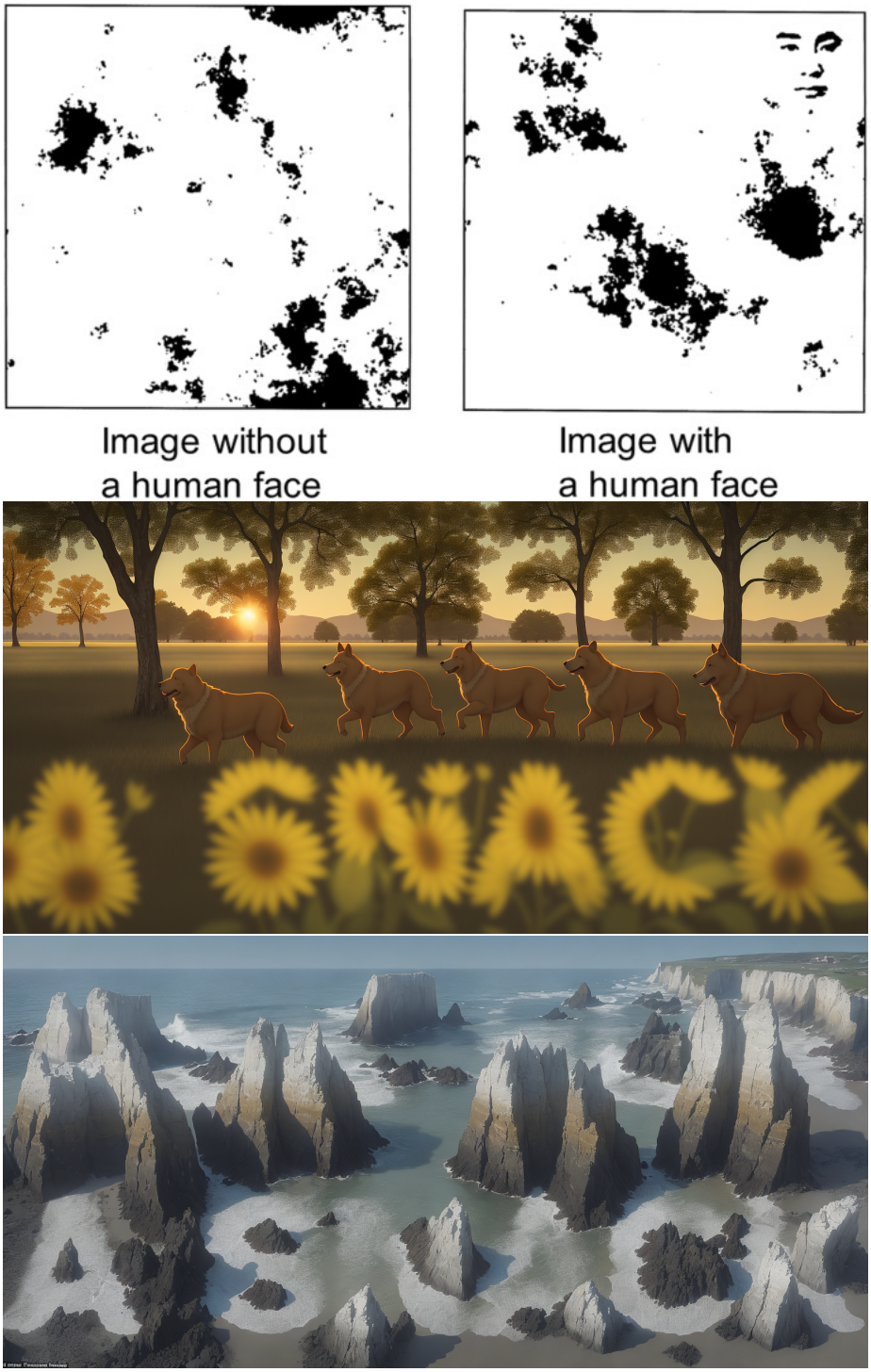

Figure 6: Sample image from the current pareidolia test, and two sample images with possible diagnostic value when patients are prompted to engage with staff and request a snack.

5 Bibliography

Atkinson: Braille Institute. Atkinson Hyperlegible Font. https://www.brailleinstitute.org/freefont

ControlNet: Lvmin Zhang, Anyi Rao, Maneesh Agrawala. Adding Conditional Control to Text-to-Image Diffusion Models. https://arxiv.org/abs/2302.05543

Gauthier: I Gauthier, M J Tarr, A W Anderson, P Skudlarski, J C Gore. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci. 1999 Jun;2(6):568-73. doi: 10.1038/9224. PMID: 10448223.

Larson: Jonathan Larson. Rent. 1996.

LCM: Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, Hang Zhao. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. https://arxiv.org/abs/2310.04378

MamiyaPT: Yasuyuki Mamiya, Yoshiyuki Nishio, Hiroyuki Watanabe, Kayoko Yokoi, Makoto Uchiyama, Toru Baba, Osamu Iizuka, Shigenori Kanno, Naoto Kamimura, Hiroaki Kazui, Mamoru Hashimoto, Manabu Ikeda, Chieko Takeshita, Tatsuo Shimomura, and Etsuro Mori. The Pareidolia Test: A Simple Neuropsychological Test Measuring Visual Hallucination-Like Illusions. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4865118/

Marclay: Christian Marclay. The Clock. https://www.tate.org.uk/whats-on/tate-modern/exhibition/christian-marclay-clock

NHCIAO: Benj Edwards. Redditor creates working anime QR codes using Stable Diffusion. https://arstechnica.com/information-technology/2023/06/redditor-creates-working-anime-qr-codes-using-stable-diffusion/

SD1.5: Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. https://arxiv.org/abs/2112.10752

Sora: Tim Brooks and Bill Peebles and Connor Holmes and Will DePue and Yufei Guo and Li Jing and David Schnurr and Joe Taylor and Troy Luhman and Eric Luhman and Clarence Ng and Ricky Wang and Aditya Ramesh. Video generation models as world simulators. https://openai.com/research/video-generation-models-as-world-simulators

UchiyamaDLB: Makoto Uchiyama, Yoshiyuki Nishio, Kayoko Yokoi, Kazumi Hirayama, Toru Imamura, Tatsuo Shimomura, Etsuro Mori. Pareidolias: complex visual illusions in dementia with Lewy bodies. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3407420/

XDali: Jamie Zawinski. xdaliclock. https://www.jwz.org/xdaliclock/