Supplementary Material
Generative Powers of Ten
powers-of-ten.github.io

A. Additional comparisons

We show additional qualitative comparisons with super resolution and outpainting models in Fig. 2. In Fig. 1, we compare with the super resolution model for photograph-based zoom.

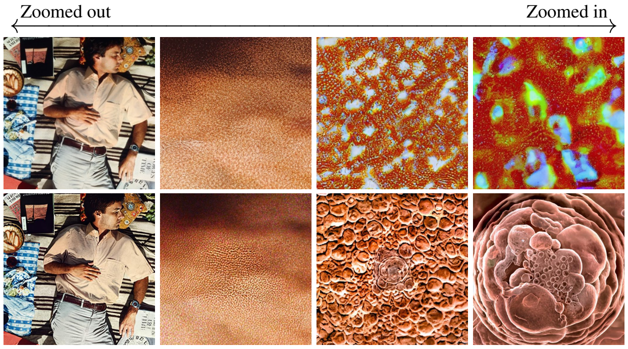


Figure 1. Comparison between the Stable Diffusion super-resolution model (top) and our method (bottom), zooming into a scene defined by a provided real input image (left).

B. Quantitative evaluations

We conduct a user study involving 38 participants who were presented with a set of 18 pairwise comparisons of our method and one of the these two baselines. Participants were asked to select one of the two options in response to the question, "Which [...] looks like a camera zooming into a consistent scene?"—our method was chosen in 92% of 684 responses.

In addition, we report (1) CLIP scores which measure text-image alignment, and (2) CLIP aesthetic scores (from MultiDiffusion [1]), which measure image aesthetic quality on the generated images, using our method and baseline methods. The scores are shown in Tab. 1.

	CLIP-score ↑	CLIP-aesthetic ↑
SR	29.18	4.89
Outpainting	30.08	5.51
Ours	**31.39**	**5.65**

Table 1. Quantitative evaluation compared with baselines. Metrics computed at all prompt scales and averaged across all examples.

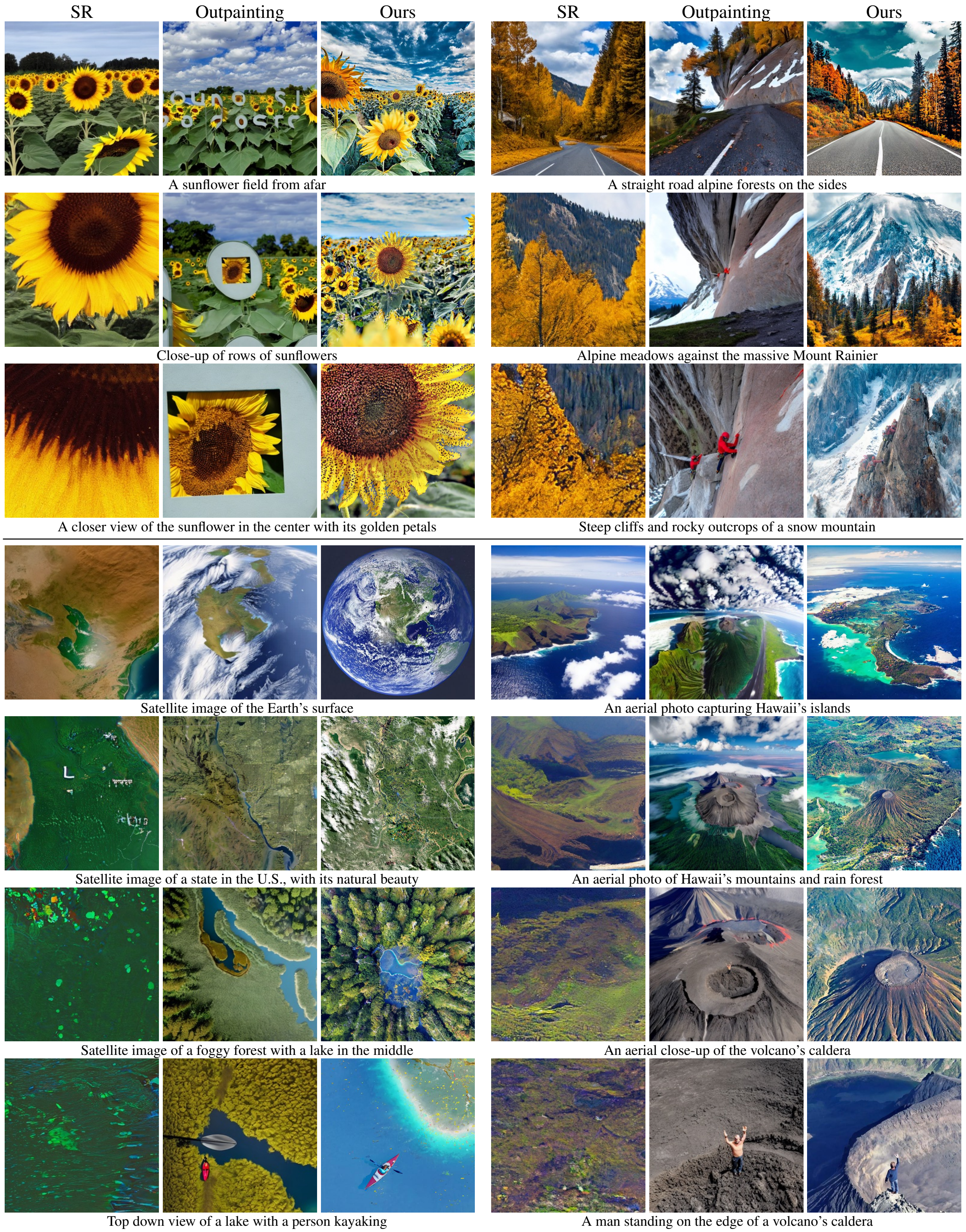
C. Text prompts generation

As mentioned in the main paper, large language models are a viable option for generating text prompts that describe a scene at various zoom levels, but their outputs are often imperfect—either describing scales that do not perfectly correspond to the scale factors used in sampling, or describing content with text phrases that do not match the learned distribution of the text-to-image model. In these cases, we often find it necessary to make manual adjustments to the set of text prompts. We show a comparison of the prompts generated by ChatGPT and the corresponding manually refined prompts (which were used to generate our zooming videos) in Tab. 6. Some sequences were not generated automatically—these are shown in Tabs. 2, 3, 5, and 4.

D. Effect of prompts

In Fig. 3, we compare sequences generated using the ChatGPT-generated prompts and our refined prompts (Tab. 6). The differences are usually subtle, *e.g.*, the ChatGPT prompts for *Sunflower* do not align with the relative object scales, so while the zoom stack images all look plausible, the object scales in the video are jarring (though adding an extra intermediate scale solves this); but sometimes they are catastrophic, *e.g.*, in *Forest*, the zoomed-out prompts describe images from viewpoints that are incompatible with other levels.

To visualize the effects of user control, we additionally provide results with edited prompts in Fig. 3. In *Forest*, we change the innermost level from "bark with cracks, lichen and insects" to "a woodpecker on top of the bark", resulting in a camouflaged woodpecker (see (c), bottom). In *Sunflowers*, we change the outermost prompt from "sunny day" to "sunset time"—we see this affects all other zoom levels as well (see (f)). We find that certain edits require changing the prompt at multiple adjacent zoom levels—otherwise coarser



A sunflower field from afar

A straight road alpine forests on the sides

Close-up of rows of sunflowers

Alpine meadows against the massive Mount Rainier

A closer view of the sunflower in the center with its golden petals

Steep cliffs and rocky outcrops of a snow mountain

Satellite image of the Earth's surface

An aerial photo capturing Hawaii's islands

Satellite image of a state in the U.S., with its natural beauty

An aerial photo of Hawaii's mountains and rain forest

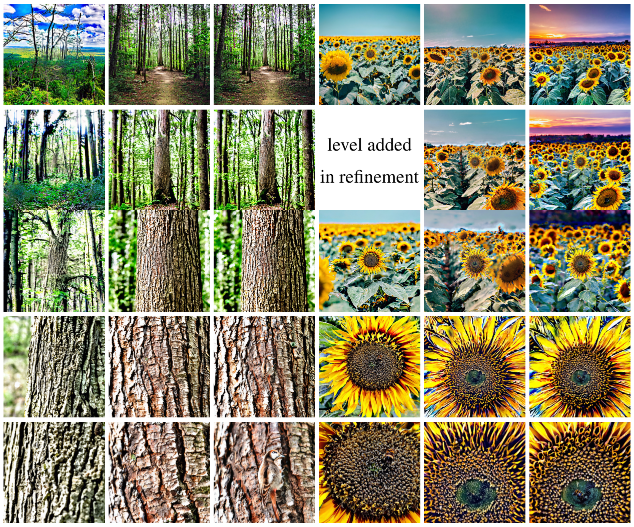
Satellite image of a foggy forest with a lake in the middle

An aerial close-up of the volcano's caldera

Top down view of a lake with a person kayaking

A man standing on the edge of a volcano's caldera

Figure 2. Comparisons with Stable Diffusion Outpainting and super-resolution (SR) models.



(a) ChatGPT (b) Ours (c) Ours (d) ChatGPT (e) Ours (f) Ours

Figure 3. Images generated with our method using: **(a,d)** prompts initially generated from ChatGPT, **(b,e)** prompts improved with manual refinement, **(c,f)** same as (b,e), with one edited prompt.

priors may overwhelm the creation of finer-level content (*e.g.*, in the woodpecker example).

E. Failure cases

Our method relies on the text-to-image diffusion model producing images of a scene at a particular set of scales from a particular viewpoint, and finding the exact set of text prompts that produce this can often be difficult. In Fig. 4, we show examples of cases where (1) the relative scale between a set of layers does not match the distribution of images that the model intends to create, and (2) the model intends to create images from different viewpoints across different zoom levels. As mentioned in the main paper, one possible improvement could be to optimize for suitable geometric transformations between successive zoom levels. These transformations could include translation, rotation, and even scale, to find better alignment between the zoom levels and the prompts.

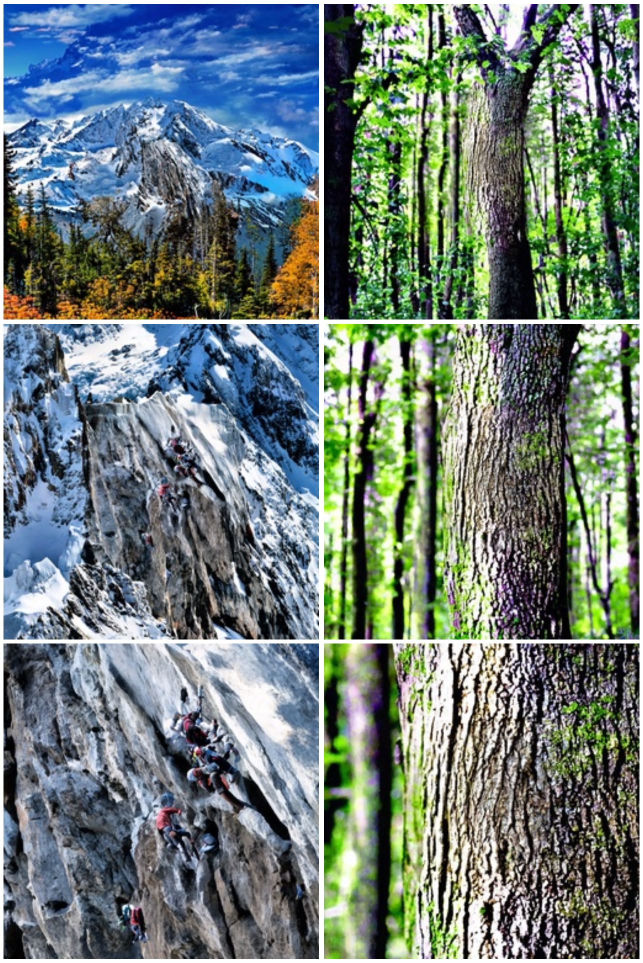


Figure 4. **Failure cases.** *Left*: an example where the predicted images from different levels observe the scene from different viewpoints (initially from a nearly horizontal view, but finally from an oblique upward-facing view). Right: an example where image priors do not correspond to the relative scale between zoom levels, as seen in the fact that multiple scales of the bark texture exist at a single zoom level.

- *A straight road in the middle with alpine forests on the sides under the blue sky with clouds; autumn season*
- *A photo capturing the tranquil serenity of a secluded alpine forest road with Mount Rainier in the far end; blue sky; autumn season*
- *A photo of serene alpine meadows against the massive Mount Rainier*
- *Extreme close-up of the steep cliffs and rocky outcrops of a snow mountain occupying the entire image; tight framing*
- *Extreme close-up of the steep cliffs and rocky outcrops of a snow mountain occupying the entire image; tight framing*
- *A team of climbers with red clothes climbing on the rugged cliffs; low camera angle*

Table 2. Complete prompts for the **Mount Rainier** example (column 4 in Fig. 7) with relative scale $p = 2$.

- *Small galaxy far away surrounded by large starry dark sky, millions of sparkling stars against dark background and vast emptiness*
- *Beautiful, high quality photo of Andromeda Galaxy*
- *Galactic core, tight framing*
- *Galactic core, tight framing*
- *Thousands of stars against dark space in the background*
- *Dark starry sky*
- *Dark starry sky with a foreign solar system in the middle*
- *Far view of alien solar system with a star and multiple exoplanets. Smaller stars in the background*
- *Alien solar system with one of the exoplanets in the center*
- *An exoplanet of a foreign solar system*
- *A close-up of an exoplanet in a foreign solar system, revealing a dry and arid climate*
- *Very high up top-down aerial image of deserted continents with reddish-hued soil in an alien planet revealing a dry and arid climate*
- *High up top-down aerial image of deserted continents with reddish-hued soil in an alien planet revealing a dry and arid climate*
- *Top-down photorealistic aerial image of a continent with a lot of deserts in an alien planet*
- *Top-down photorealistic aerial image of a desert with an alien outpost in the middle*
- *Top-down view of an alien outpost as seen directly above*

Table 3. Complete prompts for the **Galaxy** example (column 1 in Fig. 7) with relative scale $p = 2$.

- *A girl is holding a maple leaf in front of her face, partially obscuring it*
- *A brightly colored autumn maple leaf. The leaf is a rich blend of red and yellow hue and partially covering the face behind it; tight framing*
- *A brightly colored autumn maple leaf*
- *Orange maple leaf texture with lots of veins; macrophotography*
- *Macrophotograph showing the magnified veins pattern on the orange maple leaf texture; macrophotography*
- *High resolution macrophotograph showing the magnified veins pattern on the orange maple leaf texture; macrophotography*

Table 4. Complete prompts for the **Maple Leaf** example (column 2 in Fig. 6) with relative scale $p = 2$.

- *An aerial view of a man lying on the picnic blanket with his hand in the center of the image*
- *A close-up realistic photo showing the back side of a men's hand; uniform lighting; this lying person's hand should be put on top of light faded white shirt*
- *A close-up photo capturing the surface of skin of the back hand; uniform lighting*
- *Photo taken through a light microscope of skin's epidermal layer. The outermost layer, the stratum corneum, becomes apparent; Multiple rows of dense tiny skin cells becomes visible in the middle.*
- *Photo taken through a light microscope of a close up of skin's epidermal layer consisting multiple rows of dense tiny skin cells*
- *Photo taken through a light microscope showcasing several skin cells with similar sizes; with one cell in the center*
- *Photo taken through a light microscope of a single round skin cell with its nucleus in the center*
- *Photo taken through a light microscope of a nucleus within a single cell*

Table 5. Complete prompts for the **Hand** example (column 1 in Fig. 6) with relative scale $p = 4$.

ChatGPT generated	Manually refined
Forest, $p = 2$	
• *View of a vast forest from a hilltop*	<level removed in refinement>
• *Path leading to the dense forest from open land*	• *Path leading to the dense forest from open land*
• *Entrance of a forest with sunlight filtering through the trees*	• *Entrance of a forest leading into an oak tree in the middle with sunlight filtering through the trees*
• *Heart of a forest filled with tree trunks, leaves, vines, and undergrowth*	• *Heart of a forest with a tall oak tree in the middle, filled with tree trunks, leaves, vines, and undergrowth*
• *Single oak tree towering above the rest of the forest*	• *Textured tree trunk of a tall oak tree in the middle of a forest*
• *Close-up of a textured oak tree trunk and branches*	• *Close-up of a textured oak tree trunk in a forest*
<level added in refinement>	• *Close-up of a textured oak tree trunk in a forest*
• *Detailed view of an oak tree bark showing ridges and groove*	• *Detailed view of an oak tree bark showing ridges and groove*
• *Close-up of tree bark showing small cracks, lichen, and insects*	• *Close-up of tree bark showing small cracks, lichen, and insects*
Hawaii, $p = 2$	
• *An aerial photo capturing Hawaii's islands surrounded by the vast Pacific Ocean from above*	• *A aerial photo capturing Hawaii's islands surrounded by the vast Pacific Ocean from above*
• *An aerial photo showcasing Hawaii's rugged coastlines and pristine beaches*	• *An aerial photo showcasing Hawaii's rugged coastlines and pristine beaches*
• *An aerial photo revealing Hawaii's majestic mountains and lush rainforests*	• *An aerial photo revealing Hawaii's majestic mountains and lush rainforests*
• *An aerial shot of Hawaii's dramatic crater ridges and expansive lava fields*	• *An aerial shot of Hawaii's dramatic crater ridges and expansive lava fields*
• *Aerial view of surreal steam vents and sulphuric fumaroles within Hawaii's volcanic landscape*	• *An aerial close-up photo of the volcano's caldera*
• *Aerial perspective capturing the raw power and natural beauty of the volcano's caldera*	• *An aerial close-up photo of the rim of a volcano's caldera, with a man standing on the edge.*
<level added in refinement>	• *A top down shot of a man standing on the edge of a volcano's caldera, waving at the camera.*
Sunflowers, $p = 2$	
• *A sunflower field from afar*	• *A sunflower field from afar*
<level added in refinement>	• *A sunflower field*
• *Move closer to the sunflower field; individual sunflowers becoming more defined, swaying gently in the breeze*	• *Close-up of rows of sunflowers of the same size facing front and swaying gently in the breeze; with one in the center*
• *Zooms in on a specific sunflower at the field's edge*	• *Zooms in on a single front-facing sunflower in the center at the field's edge*
• *Closer view of the sunflower. Emphasize the sunflower's golden petals and the intricate details*	• *Closer view of the sunflower in the center. Emphasize the sunflower's golden petals and the intricate details*
• *An image focusing solely on the center of the sunflower Showcase the dark, velvety disc florets, and capture the honey bee sipping nectar and transferring pollen*	• *An extreme close-up of the center of the sunflower Showcase the dark, velvety disc florets, and capture the honey bee sipping nectar and transferring pollen*
Earth, $p = 4$	
• *A distant view of Earth, showing continents and oceans*	• *Satellite image of the Earth's surface showing a landmass in the middle as seen from space*
• *Zooming in on a continent, with major geographical features visible*	• *Satellite image of landmass of the Earth's foggy surface*
• *A focused view on a specific region, highlighting rivers and landscapes*	• *Satellite image of a state in the U.S., showing the state's natural beauty with rivers, forests, and towns scattered across*
• *Narrowing down to a dense forest area, showcasing the canopy and terrain*	• *Satellite image of a quaint American countryside surrounded by forests and rivers in a foggy morning*
• *Zooming in on a specific lake, surrounded by the forest.*	• *Satellite image of a foggy forest with a lake in the middle shoot directly from above*
• *Close-up of the lake's surface, with surrounding vegetation*	• *Satellite image of a lake surrounded by a forest shoot directly from above*
• *Top-down view of a person kayaking in the lake, amidst the forest.*	• *Top down view of a lake with a person kayaking shoot directly from above*

Table 6. Generated prompts from ChatGPT vs. our manually refined prompts. We (1) removed prompts which are view inconsistent with others, (2) add more levels to make the relative scale correct, (3) add description to give more context about the entire scene.

References

[1] Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. Multidiffusion: Fusing diffusion paths for controlled image generation. *arXiv preprint arXiv:2302.08113*, 2023. 1