

From Where We Came



Group 7
Janet Li

Personal Background

I'm currently in my fourth year as a Bachelor's/Master's (for the Integrated Master's Bachelor's program) student studying Electrical and Computer Engineering in the College of Engineering (CIT). I have not taken any machine learning-related courses prior to this class.

DESCRIPTION

Inspired by Anna Ridler's work, "The Shell Record," [1] which we discussed during class, I thought of my own collection of 23 various rocks and shells that I've collected from U.S. national parks and beaches from 2016 to 2022. However, unlike Ridler's work, which commented about collecting, currency, and value, I wanted to draw attention to the geologic time scale and the place which the modern world around me has in it, particularly how everything on this planet was once nothing but water or stone. In doing so, I hoped to put our tiny and short-lived existence into perspective.

Using Canva, I created two collages of my 23 rocks and shells and used them as style images in the Style Transfer algorithm we studied in class, applying the collages to various photos I had taken. However, lack of control over the basic style transfer sometimes led to undesirable or artistically lacking results. Thus, I improved upon the original results by implementing a spatially-guided version of style transfer based on [a paper by Kolkin et al](#) [3], where it's possible to map regions of the content and style images onto each other, allowing for greater artistic control.

Concept

I was initially inspired by "[The Shell Record](#)" by Anna Ridler [1], which we examined in class as an example of how the dataset images themselves can also be impactful. The 23 rocks and shells used in this project were hand-collected by me from various U.S. national parks, beaches, and streams over a span of six years, from 2016 to 2022. I used this collection to form two collages for this project as style image inputs for the style transfer algorithm we learned in class. I enjoy collecting rocks/shells from these locations because these rocks or shells have all been in or near bodies of water and have likely been shaped by the water over an unimaginably long period of time. After taking a geology course and learning about the geologic time scale, I was reminded of the following analogy: that if all of Earth's history were condensed into a 24-hour clock, with Earth's formation starting exactly at 12:00am, modern humans would not arrive until 12:00am, exactly 24 hours after Earth's formation. To put this into perspective, I found a great image online:



Images of my original collection of rocks and shells



Collage #1, used as the style image for style transfer. Made in Canva.



Collage #2, used as the style image for style transfer. Made in Canva.

The second collage was created after seeing the results of the style transfer algorithm with the first collage, as an experiment to see whether the reduced scale of individual rocks and the addition of an overall “pebble-like” style would produce better results than the first collage. All images generated used a content weight of 0.5 and a maximum image width of 512 pixels.

However, the results of the standard style transfer were not all artistically satisfying – in many of the content images, I wanted specific portions to map to specific sections of the style image collage (and to specific rocks or shells). Thus, by recommendation of our TA, Peter Schaldenbrand, I implemented the spatially-guided version of the style transfer algorithm, where the user can specify region-to-region guidance mapping from the style image to the content image, from the same paper by Kolkin et. al [3].



Unconstrained (top middle) versus guided style transfer (bottom middle). Bottom left and bottom right images represent the user-chosen color mappings between the content and style images. From Kolkien et. al [3].

Although the code used for spatially-guided style transfer was made publicly available by the authors of the paper, many parts of it were deprecated, improperly formatted, buggy, and poorly maintained (academics don't usually get paid to maintain their old research code!). Additionally, I could not find any publicly available and easy-to-use notebooks or code repositories of the spatially-guided version of style transfer. Thus, I cloned the repo and made my own edits so that the repository would work smoothly with Google CoLab (code links can be found at the end of this report). For the spatially-guided version of style transfer, I chose two content images to experiment with, and created their style-to-content color/region mappings by hand using Goodnotes.



Pair #1 for spatially-guided style transfer: Guidance images (top row) versus input content (bottom left) and style (bottom right) images demonstrating region-to-region mapping based on colors. Made in Goodnotes.



Pair #2 for spatially-guided style transfer: Guidance images (top row) versus input content (bottom left) and style (bottom right) images demonstrating region-to-region mapping based on colors. Made in Goodnotes.

For spatially-guided style transfer, I chose to only use the first collage with the enlarged individual rocks because I felt that, with spatial guidance, the second collage with many repeated patterns of the same arrangement would simply result in an overall monotonous style over the original content image. These two content images were chosen for spatial guidance because I felt that the content image in Pair #1 had good composition and was thematically close to the natural world segment of this project's concept,

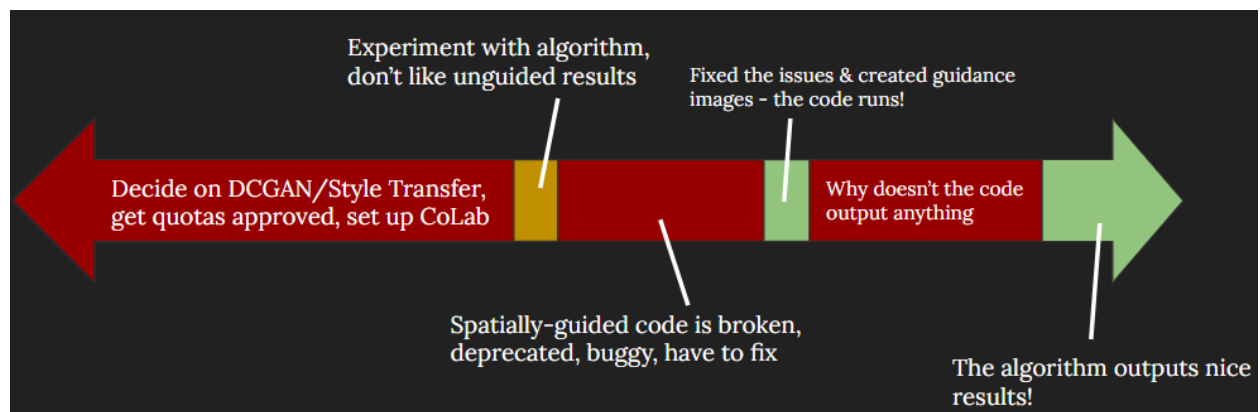
while the content image in Pair #2, focused more on the bond between a man and his dog, held more emotional interest and was thematically closer to the modern world segment of this project's concept. Additionally, the content image in Pair #2 is more dynamic with the objects in the scene, and I was curious as to how this might influence the resulting image from the spatially-guided style transfer.

As for the technique of choosing which rocks in the style image to map to which regions, I loosely tried to keep to similar colors of the original image (for example, choosing more tan shells or rocks for the man and his dog). However, I also chose mappings based on interesting textures that I thought would result (for example, the waviness of the tan shell onto the man and his dog).

I initially tried using Kolkin's original repository for spatial guidance by importing it into CoLab. However, the initial repository was deprecated, had numerous bugs, and didn't format results properly. Thus, I cloned Kolkin's repository and made my own fixes. The details of these bugs will be discussed more in the "Process" section, but overall 75 commits were made on top of Kolkin's original code. Each run of the spatially-guided model to generate an image took anywhere from 400 to 550 seconds.

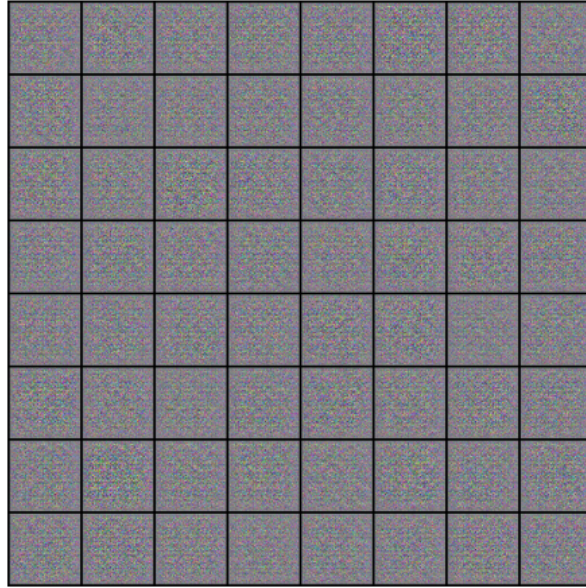
Process

An overall timeline of the process of this project can be seen below:



Overall timeline of this project

Multiple days were spent at the beginning of this project simply to get the proper permissions for Google Cloud Platform from CMU as well as enough GPU quota from Google Compute Engine. Initially, I also thought about a project much more similar to Anna Ridler's: using DCGAN to generate new, unique rocks and shells based on the dataset of my collection. I spent about a day and a half trying to work the DCGAN code to fit my dataset and to set up my CoLab environment, but ultimately ended up with just noise from the DCGAN code.



Initial result from DCGAN algorithm on this dataset

At the same time, I ultimately decided that, rather than generating new images of rocks or shells, I wanted to compose/influence images of the real world using this collection. This is when I fully fleshed out the concept I wanted to attempt for myproject.

I attempted this composition of images from rocks with style transfer, where the style image was a collage of the collected dataset of rocks and the content image was something we might see in everyday life. I first tried this out with four images: the ocean, a desk in a library, a picture of CMU's campus with the sky, and a building in downtown Pittsburgh. However, with each of these images, I found that the resulting image was largely similar to just a collage of the style image overlaid on top of the pictures. This was mostly due to large spaces in the content images (typically the sky or the ocean). And while this made some results artistically interesting (like in Try #1 below), most of the other results looked tacky.

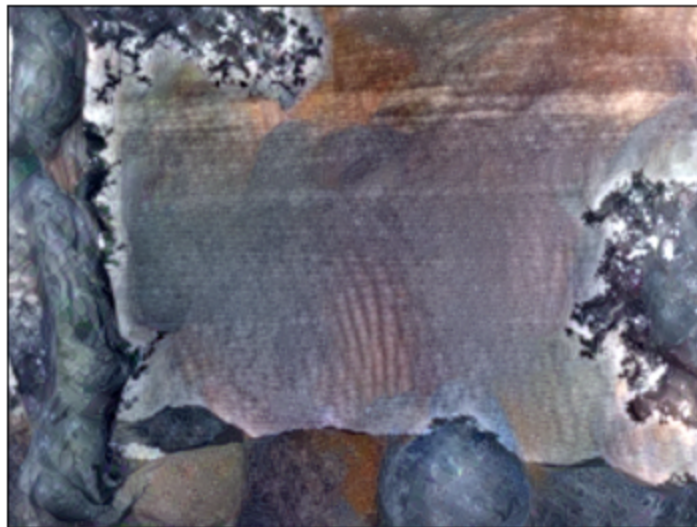
Content



Style



Try #1: Content image was an ocean image from Northern California



Resulting image from style transfer at resolution [384, 512]

Content



Style



Try #2: Content image was a desk from Sorrells Library at CMU



Resulting image from style transfer at resolution [512, 465]

Content



Style



Try #3: Content image was the sky on CMU's campus



Resulting image from style transfer at resolution [512, 384]



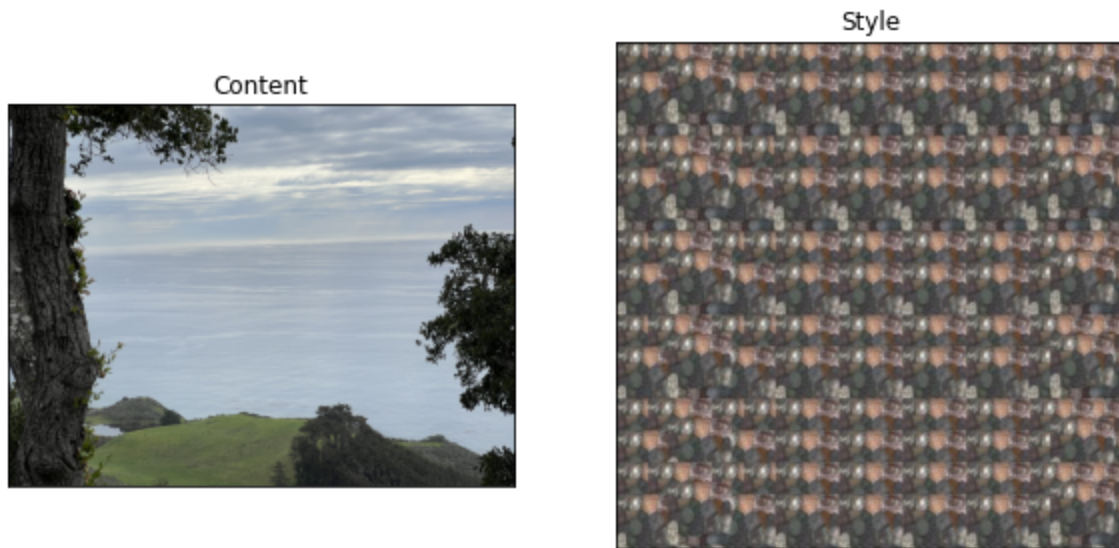
Try #4: Content image was a building in downtown Pittsburgh



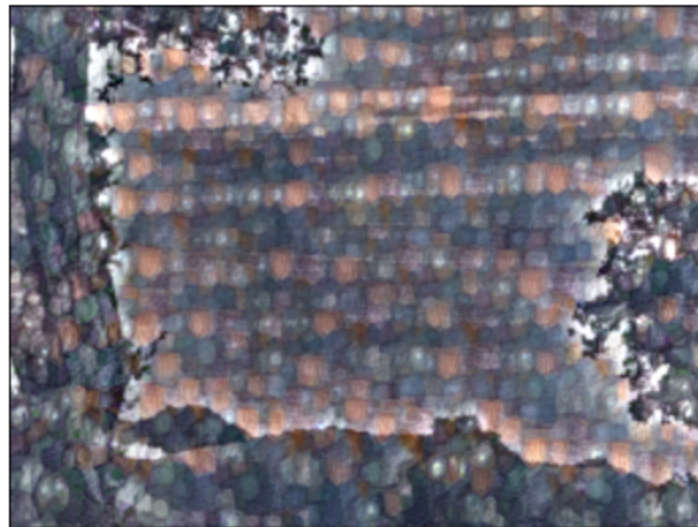
Resulting image from style transfer at resolution [384, 512]

In each of these images, I found that there were still parts I quite liked how the style transfer applied. For the ocean image, I enjoyed the transformation that the bark of the tree went through, and in the last try, I also enjoyed how the building transformed.

In an experiment, I tried to see how a larger collage composed of sub-collages of the initial one might affect the resulting images, so I tried using two images again with this new, larger collage with smaller individual rock components.



Try #5: Ocean content image with a much larger collage and smaller individual rocks



Resulting image from style transfer at resolution [384, 512]

For the less flat foreground elements, such as the trees and the cliffs, I actually enjoyed this result with the larger collage quite a bit more. However, it definitely exacerbated the previous issue with flatter images, where the ocean and sky now look like the collage was pasted on top.

Then, I tried an image that had many more dynamic elements – I had taken a picture of a man holding his dog in Big Sur during winter break, and I felt that this image's composition had lots of interesting, dynamic objects. I wanted to see how this image might look with the new style collage image.

Content



Style



Try #6: Content image with less flat space using the larger collage



Resulting image from style transfer at resolution [512, 384]

This was by far my favorite resulting image up until this point! I think that the content image having smaller areas of flatter spaces definitely helped to avoid the artificial, overlaying effect seen in the previous results. However, save for Try #1, I didn't particularly find any of these resulting images artistically or conceptually satisfactory. By recommendation of our TA, Peter Schaldenbrand, I decided to implement the spatially-guided version of style transfer by Kolkin et. al [3], where user-input region

mapping allows for greater stylistic and artistic control of the style transfer algorithm. An example from Kolkin et. al of this can be seen below:



Unconstrained (top middle) versus guided style transfer (bottom middle). Bottom left and bottom right images represent the user-chosen color mappings between the content and style images. From Kolkin et. al [3].

where the purple, red, blue, and gold regions map respectively to each other from the content and style images. You can see in the result of the spatially-guided style transfer image, in the bottom middle, allowed for greater artistic control by the user, keeping each owl mostly with monochromatic colors and choosing the same color of eyes for each owl. Thus, I felt that style transfer with spatial guidance would be a great artistic tool to implement for this project.

I chose two content images to use for the spatially-guided version of style transfer: the ocean scene and the man with his dog. As stated in the “Technique” section of this report, I felt that both represented opposing parts of my theme, with the ocean scene reflecting the natural state of the world and the man with his dog showing our modern world today. I imported both images into GoodNotes and manually created region/color mappings with the same collage style image using the paint tool, which can be seen below:



Pair #1 for spatially-guided style transfer: Guidance images (top row) versus input content (bottom left) and style (bottom right) images demonstrating region-to-region mapping based on colors. Made in Goodnotes.



Pair #2 for spatially-guided style transfer: Guidance images (top row) versus input content (bottom left) and style (bottom right) images demonstrating region-to-region mapping based on colors. Made in Goodnotes.

However, the original repository from Kolkin was deprecated, buggy, and had improper formatting issues. Much of the time spent on this project was solely dedicated to trying to resolve these bugs and errors, which I did in my own cloned GitHub repository (you can find the link at the end of this report). The largest changes I made to the original code include removing relative imports between modules, properly formatting the resulting image array to ensure that the image values were between 0 to 255

and of type `numpy.uint8`, and fixing a major bug in the `__main__` function of `styleTransfer.py`, which wasn't built to properly parse command-line arguments from two ways of calling `styleTransfer.py` from the shell with differing number of arguments. That last bug was particularly tricky, and alone took about two to three hours to discover and resolve. GitHub user 'futscdav' made a much faster and more memory-efficient version of Kolkin's STROTSS algorithm [4], however it did not support spatial guidance. If I had more time and technical knowledge, I would have liked to have supported futscdav's repository with spatial-guidance functionality, and this may become a future personal project. In total, fixing up Kolkin's spatial guidance code to work with CoLab took about nine to ten hours of continuous work. However, once the code was properly fixed, I got two (very different, but interesting) results that I will discuss below.

Reflection

As a result of the two spatial guidance style transfer experiments, the following two images were generated:

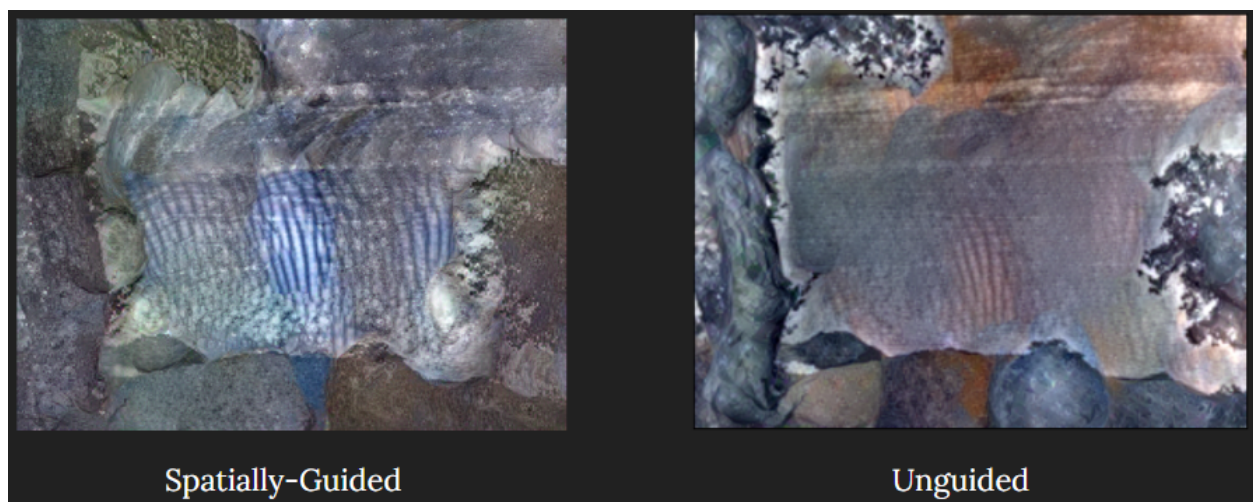


Generated images from spatial guidance

I ultimately found the top image of the man with his dog to be very artistically interesting. Every time I look at the image, I seem to find something else. For example, in the top left corner, it seems like there is almost a galaxy or nebula from the applied style. I thought it was very interesting that the man's two legs are different colors entirely, or that the dog's body is a different texture from the dog's head, even though the same mapping was overlaid on top of the entire man and dog. Slightly above the man's head, I can almost make out the image of another dog's head (similar to a Shiba's), and there seems to be

almost a figure with a face at the very right side of the image. Overall, I also enjoyed the mix of textures, from the grainy/pebble-like texture near the bottom of the image to the wave and cloud-like texture towards the top. It felt like the image was composed of stones, which achieved my overall goal for this project's concept. Additionally, it definitely no longer felt tacky, as if the original collage had simply been pasted over the original content image.

As for the second generated image from the ocean scene, I was surprised to find that I didn't enjoy this result as much as the unguided image. Although it's possible that I simply didn't pick my region mappings well enough, I found the spatially-guided ocean scene to be quite messy, as opposed to the original unguided result, and came to the surprising conclusion that spatially-guided style transfer may not always guarantee better results than unguided style transfer.



Side-by-side comparison of the spatially-guided ocean scene versus the unguided ocean scene from style transfer

Overall, for both of the spatially-guided results, I was also quite surprised by the resulting colors. Both images have a dream-like, blue-dominated and cool-toned color scheme, which was not as present in the unguided versions of style transfer, even though all the rocks from the style collage were mapped to a region in the content images. Given more time, I would have liked to investigate why this was the case.

Thus, because of the reasons above, for the two spatially-guided results, I only chose the generated image of the man with his dog as one of the results for this project. The other result I chose for this project was the unguided scene of the ocean. In terms of artwork I would have enjoyed hanging in my own household, I enjoyed the aesthetics and artistic style of this piece the most, with its dreamlike textures and compositions of the trees and the foreground cliffs. The resulting color palette for the ocean and sky also turned out surprisingly beautiful, with both warm and cool color tones. While the generated image of the man with the dog was dynamic, with interesting figures popping up everywhere in the background, I found that the “style” of the unguided ocean scene was the most pleasing to me.

This led to another reflection of mine: what is “style” in the first place?



What's the difference in "style" between the two style collages I used?

Comparing the two style collages of rocks I used, you could argue that the style of the second image is grainy or pebble-like. But as for artistic style, it personally falls a little flat, due to its repeated patterns. It's not something I would generally think of when describing artistic styles as prominent as Van Gogh's or Monet's. However, the style of the first collage is intriguing, due to the lack of shadows on the rocks (since the backgrounds were simply removed and the rocks pasted on top of one another), creating a sort of eerie and surreal effect. Could this be classified as a type of style, or do some images inherently have more style than others? I felt that this project left me thinking about this concept for quite some time.

Overall, I felt quite satisfied with my ideas and techniques for this project. If I had more than zero technical background in machine learning (and more time), I would have liked to experiment more drastically with some of the algorithms used. I felt that the spatial guidance technique for style transfer could be a huge boon for digital artists who primarily rely on ML models to generate their artwork, since it provides greater artistic and stylistic control. Again, I would have liked to implement this algorithm to be faster and more memory-efficient, since it did take an unnecessarily long amount of time to generate these images (about 400-550 seconds each time). An overwhelming amount of time was spent on simply debugging the deprecated and poorly-maintained repository from Kolkin et. al, and it made me realize how unfortunate it is that academics aren't paid to maintain their old research code. Additionally, a large chunk of time was spent at the beginning of this project simply getting used to the CoLab environment and receiving enough GPU quota, so I'm glad that I won't have to waste that time again for future assignments in this class. I learned to start early, decide early-on the concept I want to approach this assignment with, and that there are indeed scenarios where zero control over the model may produce results (subjectively) better than when the user has some control over the result (as in the case of the ocean scene).

RESULT

Overall, I chose two results for this project: one was made using the spatially-guided version of style transfer, and the other was unguided.

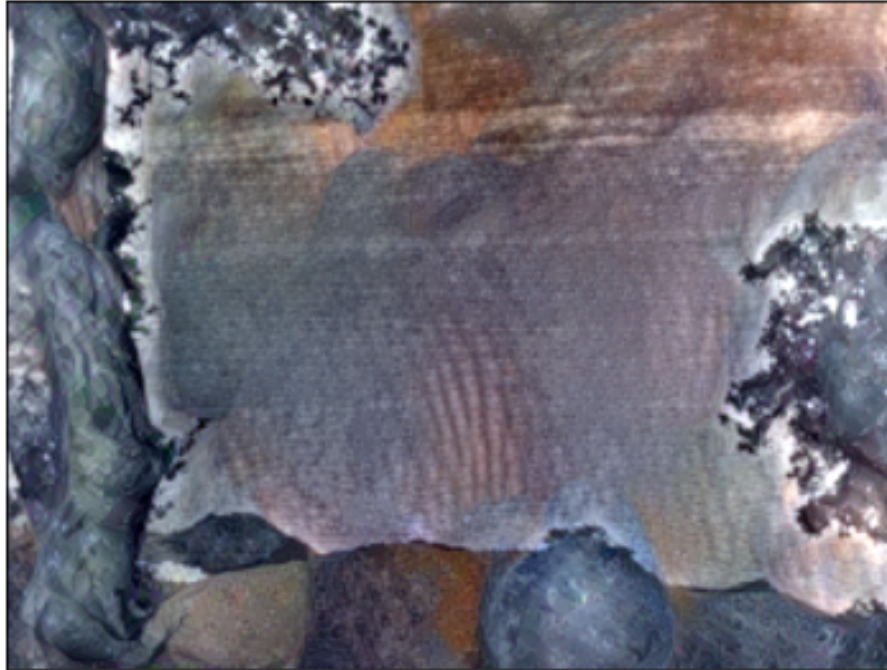


Result #1: Created using spatial guidance

The above result was created using spatially-guided style transfer with the following guidance images:



Pair #2 for spatially-guided style transfer: Guidance images (top row) versus input content (bottom left) and style (bottom right) images demonstrating region-to-region mapping based on colors. Made in Goodnotes.



Result #2: Created using unguided style transfer



Original content and style image used to generate Result #2

In the “Reflection” section of this report, I already discuss why these images were chosen as my results. I found both to be artistically satisfactory and conceptually intriguing.

CODE

Github Link: <https://github.com/janet-li/STROTSS>

Google CoLab Link:

https://colab.research.google.com/drive/1Vqg_XTcKv-xYAuSCB_JAs2xW8sxHAcBf?usp=sharing

REFERENCE

[1] Ridler, Anna, 2021. *The Shell Record*

<http://annaridler.com/the-shell-record-2021>

[2] MacPhee, Larry, 2020. *The History of Life on Earth*

https://jan.ucc.nau.edu/lrm22/lessons/timeline/24_hours.html

[3] Kolkin et. al, 2019. *Style Transfer by Relaxed Optimal Transport and Self-Similarity*

<https://arxiv.org/pdf/1904.12785.pdf>

[4] futscdav, 2020. *Optimized STROTSS*

<https://github.com/futscdav/strotss>

[5] Nicholas Kolkin, 2019. *STROTSS*

<https://github.com/nkolkin13/STROTSS>

[6] Peter Schaldenbrand, 2022. *STROTSS Style Transfer Notebook*

https://colab.research.google.com/drive/1pBfUg0HAiLkhqW_vqzR353TA5RGIBW98?usp=sharing#scrollTo=JihukNAZk7bJ