Generative AI as a Tool for Speculative Urban Futures

ANONYMOUS AUTHOR(S)*

This paper looks at the use of four open-source generative algorithms for radically re-imagining urban structures to better facilitate human-nature cohabitation. It assess the possibility for these tools to be used in the course of speculative design to help people visual new uses for common structures that would allow these structures to be used by non-human residents of the urban landscape.

CCS Concepts: • **Human-centered computing** → *Visualization application domains*; **Human computer interaction (HCI)**;

Additional Key Words and Phrases: generative AI, speculative technology, urban structures, sustainable HCI

ACM Reference Format:

Anonymous Author(s). 2018. Generative AI as a Tool for Speculative Urban Futures. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

> This tool [a bus stop kiosk] may provide information for people riding the bus, or for tourists who need to consult a map. But it does not stand alone with a unified function apart from the nonhuman elements of the city. The structure may provide a home for birds (or may perhaps disrupt a previous home for birds). The light from the screen may affect moths and other nocturnal creatures. The kiosk could be designed to support water collection or to allow for a plant to grow on or in it to foster a more harmonious relationship among animals, humans, and technology [9].

When I first read the paper quote above, it prompted me to think more deeply about ways ordinary objects in the urban environment might be transformed to serve our non-human counterparts. Could the posts for street signs be redesigned to include bird houses? Could waste water from public fountains be collected to water nearby plants or animals? How might we transform other structures like benches or trashcans? As I began to imagine these transformations, I was also reminded of an article I had recently read about a new algorithm that could generate an armchair that looked like an avocado—and quite successfully if the pictures in the article were any indication [5]. I wondered, could these two things be combined?

Could a generative algorithm be used to create images of these re-imagined structures and be used as a speculative tool for sustainable urban design?

This paper explores this question.

Unfortunately, the algorithm generating beautiful avocados armchairs, Open AI's Dalle [7], is not publicly available. Therefore, I sought out some open source alternatives that could be used in a public manner, for instance in a participatory or speculative design session. This paper looks at four alternative algorithms. In the rest of this paper I very briefly introduce related work on the use of computer vision to study urban forms. I then introduce the parameters of the experiments performed with each of the four algorithms. For each algorithm, I present and discuss the results. Finally, I

Manuscript submitted to ACM

conclude with a general discussion on the results of the experiments and the implications for using generative AI as a speculative tool for urban futures.

2 RELATED WORK

There is a history of research using AI, in particular computer vision, to study urban forms. Examples include [11] which looked at environmental factors associated with urban safety, [8] which used satellite imagery to assess urban vitality, and [8] which used MLto identify good parkour spots in an urban environment. Of particular interest is [10], which used a general style transfer algorithm to study the difference between images of self-reported high and low well-being based on a national survey in Australia in order to see if there was a discernible environmental difference. As far as we can find, generative algorithms have not been used to create speculative urban structures in the way this paper explores.

3 METHODOLOGY

This paper looks at four open-source generative algorithms for radically re-imagining urban structures to better facilitate human-nature cohabitation. For the purposes of this paper, the example structure is a bus shelter with a garden growing on its roof, which actually exist in various forms in cities around the world [2**?** , 3]. This example was chosen because it requires fairly little reinterpretation of the shelter's main structure, so ostensibly it should be fairly easy to generate.

The four algorithms tested were: two versions of ruDalle (a Russian built Dalle recreation), Text2Image[1] and Optimized Image Prompts[2]; Big Sleep[3], a combination of OpenAI's CLIP and a BigGAN; and VQGAN+CLIP[4], the algorithm behind the popular Wambo Art app[5]. All of the algorithms were run on Google's Colab using a Pro subscription.Additional details about parameters for each algorithm are given in the respective sections.

Four prompt variations were used: *a bus shelter with a garden on top*; *a bus shelter with a garden roof*; *a bus shelter with a rooftop garden*; and *a city bus shelter with a garden on top*. All images for variations of the prompts and parameters are in the supplemental material[6]. To save space, this paper presents only the best (most real looking) and/or most instructive examples for each algorithm.

Fig. 1. Images created by ruDalle Text2Image for the prompt *a bus shelter with a garden on top.*

4 RUDALLE - TEXT2IMAGE

I start with RuDalle, because it purports to a 1.3 billion parameter model, which while quite large is still relatively small compared to the original Dalle's 12 billion parameters [7]. In fact, none of the results were promising. Figure 1 shows the results from the prompt *a bus shelter with a garden on top*. As this figure shows, the

[1]https://colab.research.google.com/github/tg-bomze/collection-of-notebooks/blob/master/Text2Image_v4.ipynb
[2]https://colab.research.google.com/drive/1S08bgB1h-la84_VFNfRXwpfrOQfQM9Wl
[3]https://colab.research.google.com/drive/1MEWKbm-driRNF8PrU7ogS5o3se-ePyPb
[4]https://colab.research.google.com/drive/1go6YwMFe5MX6XM9tv-cnQiSTU50N9EeT
[5]https://www.wombo.art/
[6]The supplemental material can be provided upon request.

algorithm has a tendency to produce cartoon-ish results, create an istock watermark (likely due to the training set), and to insert Russian text at various points throughout images. As can be seen in the supplemental material, the algorithm can produce a bus shelter and a garden on their own, and even does alright on a rooftop garden, but the combination alludes it.

5 RUDALLE - OPTIMIZED IMAGE PROMPTS

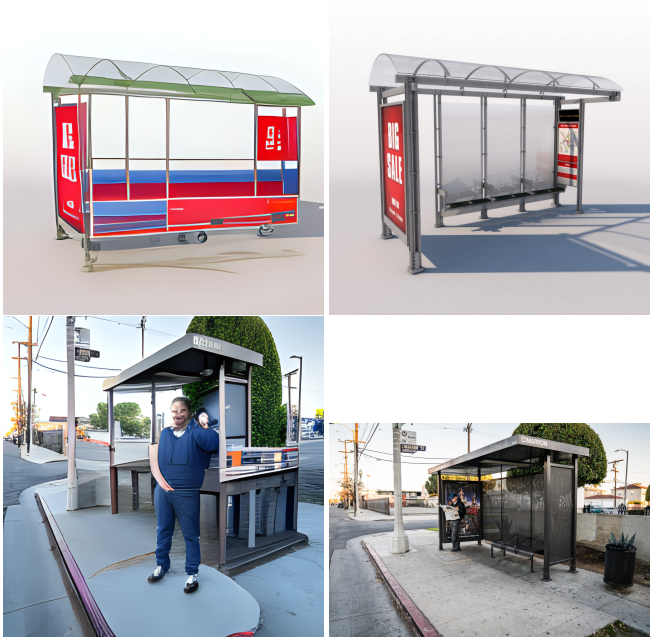

Fig. 2. Left: images produced by the optimized image prompt notebook for the prompt *a bus shelter with a garden roof*. Right: The associated image used as an initialization.

This setup allows the user to prompt RuDalle with an image to guide the generation. As Figure 2 shows, the image used to prime the algorithm makes a huge difference. In this case, the best results came from the prompt: *a bus shelter with a garden roof*. While both images came out somewhat successfully (strange distorted person aside), looking at the output next to the initial images it is clear that in the bottom case the algorithm basically morphed the tree behind the shelter to create the wall of green climbing over it. The first case also shows only minimal changes, which does not engender confidence in the algorithm's ability to create more complicated structures. Another drawback to this algorithm is that it requires a GPU to run.

6 BIGSLEEP

Figure 3 shows the image produced by big sleep for each of the four prompts. Aside from the one with the city tag, the prompts were also appended with the tag "realistic" (see the supplement for additional details) and "without blur and zoom" was also added.

While big sleep had moderate success overall, especially with the city prompt, a major drawback is that it takes nearly four hours to run 20 epochs of 1000 iterations each. It is not possible to run on the free tier of Colab without timing out of resources; a pro account is required.

7 VQGAN + CLIP

VQGAN+CLIP allows a choice of models trained on different datasets. Of the available datasets three are relevant for creating realistic urban scenes: Imagenet16384 [4], Coco [6], and ade20k [11].

As with the RuDalle Optimized Image Prompt, this algorithm also allows an initial image prompt. In order to avoid the morphing of background vegetation as in Figure 2, we used the top right image as our initial prompt. This algorithm

Fig. 3. BigSleep images of Bus Shelters with garden roofs generated by each of the four prompts.

also allows the use of a target image, but this did not improve results and so is only presented in the supplement. All images were produced using a run of 1000 iterations. As can be seen in the supplement, increasing the number of iterations substantially (also substantially increasing the run-time) does not significantly impact the image quality.

Fig. 4. VQGAN+CLIP images produced for the prompt *a bus shelter with a rooftop garden.* The left image is generated from text only. The right used an initial image prompt*.
*This image is from 850 iterations as the loss was spiking at 1000.

ade20k was the best dataset, which makes sense given it contains annotated outdoor scenes. The images produced from ImageNet and COCO can bee seen in the supplement. The prompt *a bus shelter with a rooftop garden* produced the most realistic results. Figure 4 shows the images for this prompt with no image prompt and with an initial image. This algorithm was perhaps the most successful, but still did not produce highly realistic results.

8 DISCUSSION

On the whole, it seems that at the open source level, generative algorithms are not good enough at producing realistic images that re-imagine urban structures to be used as a speculative tool. This is especially true in the case where there might be time or hardware constraints. However, I believe it is worth revisiting the question as new and more powerful models are released. In particular, if OpenAI's Dalle gets an open source or public API release generating more realistic images may be possible [7]. Other areas for future consideration are adjusting the various parameters as well as trying additional text prompts either in the bus shelter case or for other re-imagined structures. There may be a combination of algorithm, prompt, and parameters that produces useful results, but it seems clear there is no standard way to find such a combination.

REFERENCES

[1]]phillywatersheds [n. d.]. Green Roof Bus Shelter. http://archive.phillywatersheds.org/green-roof-bus-shelter

[2] 2018. Sheffield Bus Shelter. https://www.greenroofs.com/projects/sheffield-bus-shelter/

[3] 2019. Green bus stops ensure a future-proof city. https://www.sempergreen.com/en/about-us/news/green-bus-stops-ensure-a-future-proof-city

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[5] Will Douglas Heaven. 2021. This avocado armchair could be the future of ai. https://www.technologyreview.com/2021/01/05/1015754/avocado-armchair-future-ai-openai-deep-learning-nlp-gpt3-computer-vision-common-sense/

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[7] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[8] Sanja Scepanovic, Sagar Joglekar, Stephen Law, and Daniele Quercia. 2021. Jane Jacobs in the Sky: Predicting Urban Vitality with Open Satellite Data. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[9] Nancy Smith, Shaowen Bardzell, and Jeffrey Bardzell. 2017. Designing for cohabitation: Naturecultures, hybrids, and decentering the human in design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1714–1725.

[10] Jasper S Wijnands, Kerry A Nice, Jason Thompson, Haifeng Zhao, and Mark Stevenson. 2019. Streetscape augmentation using generative adversarial networks: Insights related to health and wellbeing. *Sustainable Cities and Society* 49 (2019), 101602.

[11] Haifeng Zhao, Jasper S Wijnands, Kerry A Nice, Jason Thompson, Gideon DPA Aschwanden, Mark Stevenson, and Jingqiu Guo. 2019. Unsupervised deep learning to explore streetscape factors associated with urban cyclist safety. In *Smart Transportation Systems 2019*. Springer, 155–164.