Transformations issue 36 (2022)
www.transformationsjournal.org

ISSN 1444-3775

AUTHOR BIO

Immanuel Koh is an assistant professor in both Architecture & Sustainable Design (ASD) and Design & Artificial Intelligence (DAI) at the Singapore University of Technology & Design (SUTD). He obtained his PhD at the École polytechnique fédérale de Lausanne (EPFL) in Switzerland, while doing transdisciplinary research between the School of Computer Sciences and the Institute of Architecture.
https://asd.sutd.edu.sg/people/faculty/immanuel-koh

AI-Urban-Sketching: Deep Learning and Automating Design Perception for Creativity

Immanuel Koh

ABSTRACT

The paper reconsiders style transfer with generative adversarial networks (GANs) as a powerful means towards a machinic extraction of perception, one that learns how to imitate how a human might spatially abstract, translate and eventually create designs. The aim is to investigate the potential of deep learning a mapping between two domains, one being the perceived reality of an urban scene, and the other, its representation on a sketch. The creative discipline under consideration in this paper is that of architecture.

KEYWORDS

deep learning, GANs, urban sketching, creativity, Google Street View

Introduction

Despite the ubiquity of digital tools today, sketching remains an important foundational course in most architecture schools worldwide. In fact, it is often an accreditation requirement to register as an architect. Its significance lies in educating the architect to learn to "see," in order to develop his/her own individual expression and thought (Bagnolo). More specifically, it is the form of observational drawing commonly known as urban sketching -- a more journalistic form of the artist's "en plein air" or travel sketch. This proposed mapping of reality to perception is unlike the naive image edge detection algorithms found in digital tools such as Adobe Photoshop, but one that arguably learns and automates the "ways of seeing" of an architect, thus a crucial step towards artificial creativity (Figure 1).

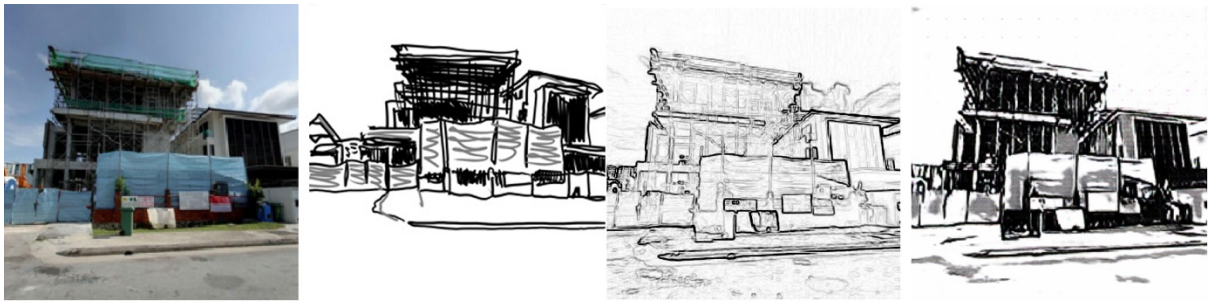

Fig. 1 A visual comparison (left-to-right): Google Street View image as a scene reference (training dataset A); human abstracting the scene as a sketch without tracing over (training dataset B); naive edge detection filter applied on the given Google Street View image in Adobe Photoshop; generated output from the trained generative adversarial networks given the Google Street View image as the input.

A similar mapping between two different visual domains could analogically be found in recent computer vision and artificial intelligence research, and in particular, research in style transfer using deep generative adversarial networks (GANs). The underlying mechanism of GANs is technically stochastic, and conceptually creative (at least in terms of its aim in generating novel outputs). In fact, its reception among AI artists attests to its potential capacity for artificial creativity. The paper, however, is not so much about generating novel images with GANs, but about the interplay between human abstraction and AI extraction as afforded by the creative appropriation of GANs. In this project, the artefacts used as a proxy to the architect's creative act are the hand-sketches that abstract Google Street View images. The GANs would, in turn, attempt to extract such an abstraction mapping in learning to see like the architect. Could machines experience such "seeing" too and perhaps begin to seed forms of artificial creativity (Figure 2)? The paper will first present the methods of the deep learning experiments. This is followed by a discussion and elaboration on the key ideas that underpin the design of the experiments and the ways in which they interact in pursuit of a possible artificial creativity, specifically through the lens of the urban sketching artefact. The results of the experiments will then be analysed at the end of the paper.

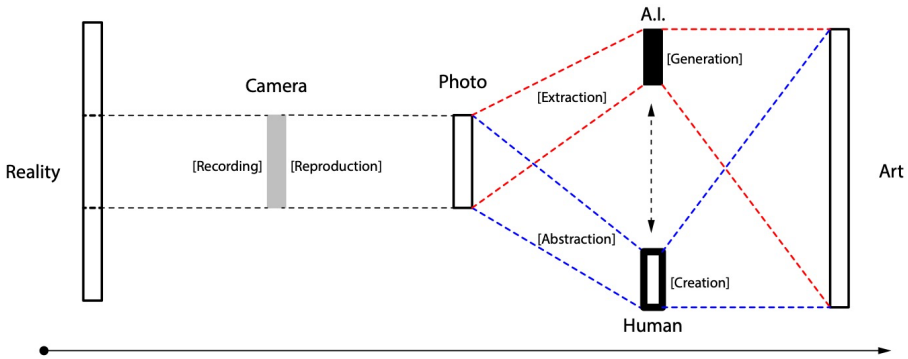

Fig. 2 In seeding artificial creativity, the paper proposes a parallel act of AI-Extraction and Human-Abstraction. The former with the disembodied Google Street View imagery (A) and the latter with the surrogate-embodied urban sketch (B) by the human. The combination of both domains is then realised with deep generative adversarial networks called CycleGANs, where an unsupervised mapping of both visual domains could be learned to generate new images (A') or sketches (B') as conditioned by B and A respectively.

Methods

To include the notion of embodied AI as suggested by the call for this issue of *Transformations*, the embodiment is here framed as a remote sensing photographic agent, rather than a physical actuating robot for creative production. Therefore, instead of creating a dataset of urban sketches done on location, Google Map's Street View is used as the source of scenes for the urban sketching. Without the availability of any datasets containing pairs of corresponding urban scenes and hand-sketches, a specific workflow is formulated in creating a new dataset for the experiment. Over a period of 7 weeks, a total of 2300 video clips have been logged, where each clip captures the sequential process of hand-sketching a different street view (Figure 3). Thumbnail sketching is a common practice for urban sketchers to quickly simplify a scene and pictorially frame objects in the surrounding space before working on the actual sketch. In doing so, it strips the scene down to its essences, where unnecessary details are omitted. For this paper, the AI model converts any given Google Street View into a feature-rich sketch, and vice versa. Using Singapore as its geographical site, pairs of latitude and longitude are sampled from contrasting planning areas, such as residential towns and central business districts (Figures 4 & 5). To ensure some degree of simplicity and uniformity of the training set, a limited palette is used. Stroke variables, such as colour and size, are deliberately standardised to introduce less variability during the data creation process (Figure 6).

Fig. 3 A logged sequence showing the sketching process from left to right.

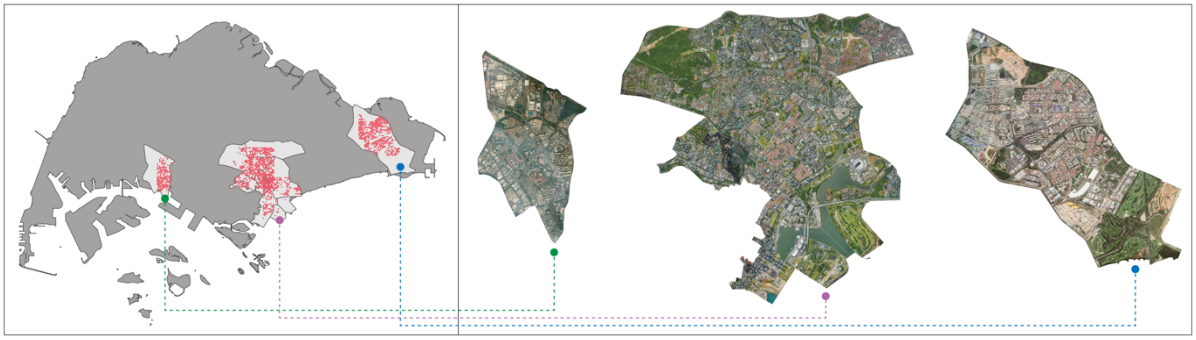

Fig. 4 (LEFT) Singapore as the geographical site to sample pairs of latitude and longitude from contrasting planning areas, such as residential towns and central business districts. Red dots in the grayscale diagram represent the geolocations of sampled urban scenes. (RIGHT) Cropped and zoomed-in satellite imagery views of the three respective urban regions where the urban scenes are sampled from.

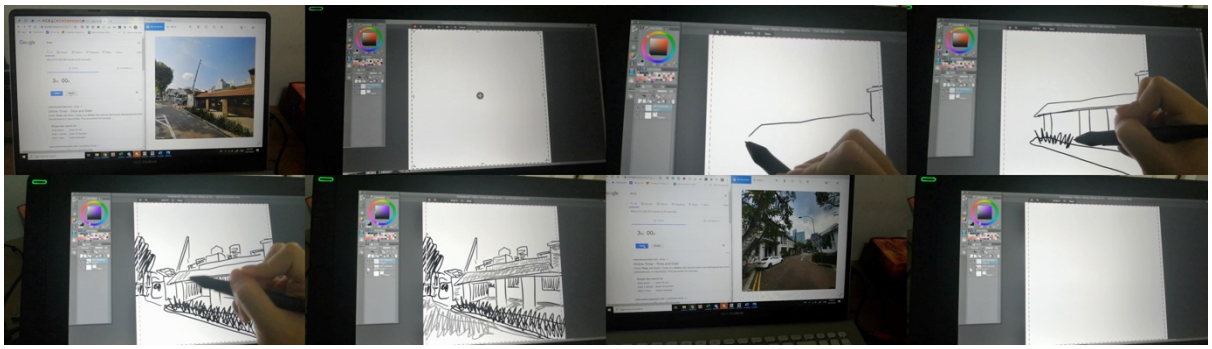

Fig. 5 Samples of training set used (left-to-right): Red dots represent locations of urban scene; scraped images from Google Street View as dataset A; corresponding hand-sketches as dataset B.

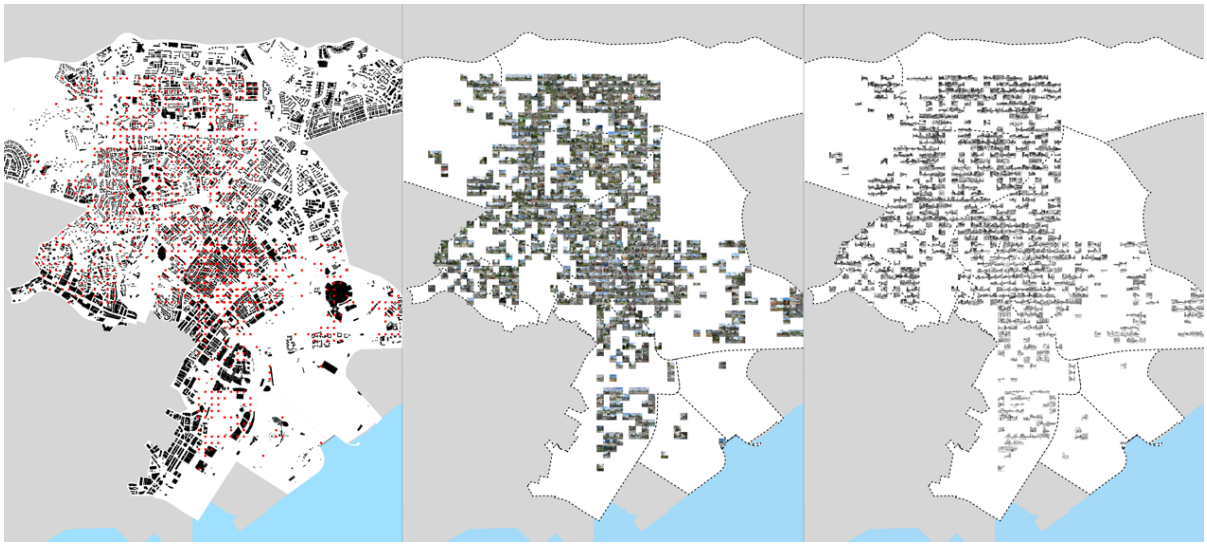

Fig. 6 A sequence of photographic images of the actual interface setup in action consisting of 2 screens – one for loading and viewing the Google Street View photos, and the other, for sketching on the touch screen tablet. Pen stroke colours are either black or white, or with an addition of a mid-tone. Brush stroke size is fixed. The omission of the eraser is to capture the cognitive traces of the sketch. Regardless of the orientation of the Google Street View image provided, the drawing area is the same, and the program records the strokes on the drawing canvas until the timer counts down to zero.

Two different GAN models, namely, the pix2pix model (Isola et al.) for paired image-to-image mapping and cycleGAN model (Zhu et al.) for unpaired image-to-image mapping, have been implemented and tested with the same datasets consisting of the sampled Google Map Street View images (dataset A) and their corresponding hand-sketches (dataset B). Both models use a convolutional neural network (CNN) architecture and maintain a similar idea of adversarial loss between the generator and discriminator found in the original GAN model design (Goodfellow et al.). Briefly, the first model uses a U-Net-based generator that downsamples batches of input images through a series of "convolution→batch normalization→leaky relu" blocks, and then upsamples them through yet another series of "convolution transpose→batch normalization→leaky relu" blocks (with dropouts applied for the first 3 layers), while maintaining a corresponding set of skip connections between both series. The discriminator, however, uses a PatchGAN architecture that downsamples the input image before mapping it to a one-dimensional output followed by a sigmoid function. Binary cross-entropy and the ADAM optimizer are the loss function and optimizer used respectively. Briefly, the second model uses a similar architecture as the first model, with the key exceptions that there are two generators and two discriminators that make use of an additional cycle-consistent loss. In this project, two pix2pix models have been trained, one mapping the urban sketches to Google Street View images (i.e., B→A'), and the other mapping it in reverse (i.e., A→B'). However, only one cyclegan model has been trained since it consists of two mappings by default (i.e., B→A' and A→B'). Analysis of the experimental outputs from these deep learning models will be made in the "Results" section.

Remote and Machinic "En plein air"

The *Urban Sketchers* movement was founded in 2007 with the mission to support a global community of predominantly artists and architects to sketch on-site in cities (Urban Sketchers). In this digital era, such a movement might seem counter-intuitive given the ubiquity of not only camera phones, but also that of consumer grade 360-degree cameras (e.g., GoPro), LIDAR (Light Detection and Ranging) sensors on autonomous vehicles and remote sensing satellite imagery (e.g., Google Earth). As indicative from its manifesto consisting of eight statements, the emphasis is on an embodied form of perception, abstraction, representation, and creation, where sketching (unlike photography) has the unique capacity to truthfully capture the intangible features of time, place and story. In a way, it recalls older embodied art practices of 17-18th century Grand Tour and 19th century Impressionism. Of course, a key difference for the urban sketchers movement is the communal sharing and non-profit online networking of like-minded urban sketchers distributed around the globe. Education is also at the heart of this endeavour, thus the emergence of several symposiums, workshops and publication worldwide. This is almost contrary to the projects by Jenny Odell who works alone at her desk with a digital stylus and tablet, cutting out and recomposing satellite imagery of urban infrastructures from Google Maps – a disembodied (or surrogate-embodied) artistic practice made possible by Google's

distributed and large-scale extraction apparatuses deployed simultaneously *on-site-on-ground* with Google's Street View cameras and *off-site-off-ground* with Google stitching of multiple aerial and satellite photographs. In the former, although the Google Street View vehicles are physically on site during the capturing of the urban spaces, their embodiment is radically different in the anthropometric sense. Google Street View was launched in 2007 (the same year as the release of the first iPhone), and by 2019, it has already captured 10 million miles of street imagery worldwide. Yet, despite having undergone multiple design iterations and technological upgrading in maximizing imagery coverage since its first fleet of Street View cars, its different spatial and temporal embodiment because of its machinic cone of vision and locomotion remains anything but human. It is three times the human's field of vision (much more if one is to exclude peripheral vision) with an elevated eye level (typically mounted on top of a car). If sketching is a means for capturing the invisible essences of a place in space and time, as claimed by the founder of the urban sketcher movement Gabriel Campanario (Urban Sketchers), could a remote and disembodied form of sketching be equally valid?

In this project, three regions of Singapore have been chosen for sampling a total of 2250 geolocated Google Street View images, with a split of 2000 as training set and 250 as test set. This set of imagery represents a distributed and non-anthropometric embodiment, which in turn, serves as the reference imagery for the manual hand sketching of another corresponding set of 2250 sketches. This other set of sketches represents a non-distributed and anthropometric embodiment. Together, the project's process of data creation engenders a strangely hybridized form of human-machine embodiment and disembodiment.

Sketchy as Being Generative

Are sketches inherently generative? Before answering this question, it is necessary to first clarify the role of sketches in the creative process. In her 1999 paper, when comparing images with sketches, Tversky writes that "drawings reveal people's conceptions of things, not their perceptions of things" (2). Unlike images (or for our purposes, photographs) which possess "a single, coherent point of view," (2) Tversky adds that "drawings, then, are representations of reality, not presentations of reality. Drawings can omit things that are actually there, they can distort things that are there, they can add things that are not there. They need not have a consistent point of view or a point of view at all" (3). In that sense, a sketch being imprecise, or just *sketchy*, provides the necessary material for creative reinterpretations as Tversky and Suwa put it in their paper "Thinking with Sketches": "a new idea, in turn, allowed him to reconfigure the sketch yet again, so that a positive cycle ensued: perceptual reorganization generating new conceptions and new conceptions generating perceptual reorganizations" (80). In fact, the possibility for meaningful reconceptions and reorganisation of the sketch has to do with how it reveals thoughts through the revelation of the segments of construction within the very sketch itself.

The sketch has the generative capacity for multiple reinterpretations; thus it should be understood as a means and not an end. Such a framing is obvious in architectural design pedagogy and practice where a sketch on paper is clearly not the building, but the becoming of a possible physical building, whether in the literal and metaphorical spatial sense. In fact, as early as 1973, architect Negroponte, who first directed MIT's *The Architecture Machine Group* before founding the MIT Media Lab and had worked on sketch recognition research with computer-aided design systems and artificial intelligence, defined sketching as follows: "Sketching can be considered both as a form of introspection, communicating with oneself, and as a form of presentation, communication with others" (663). In the former, sketching is then a means for iterative design ideations or even meditations. This is in contrast with much of the work today by AI researchers who often see the sketch as an end in and of itself. This mismatch of the "why" for the designers and "how" for the computer scientists could be seen in papers, such as "How Do Humans Sketch Objects?" (Eitz et al.). Not only are sketches framed as categorically recognizable (more specifically with only 250 classes to choose from), the latter community also often focuses on non-expert sketches, which are very different in their generative capacities than those by experts (or designers). More recent efforts in the deep learning of sketches from such a non-designer perspective can be seen at the first dedicated workshop called *Sketch-Oriented Deep Learning* at the top-tier AI conference CVPR 2021 (Computer Vision and Pattern Recognition). This line of research work aims to create deep generative model architectures for learning and synthesizing novel sketches with the most notable one being the paper "A Neural Representation of Sketch Drawings" (Ha and Eck) featuring the sketch-RNN model trained with a dataset of 50 million sketches with 345 categories (J. Jongejan et al.).

In this project, the AI model does not learn the technical neural representation of urban sketches as sequence of strokes but aims to learn the conceptual features of urban sketches as a proxy to artificial creativity.

Visual Accuracy: Shape or Conceptual?

In the 1997 paper titled "Why Can't Most People Draw What They See?", Cohen and Bennett propose a theoretical and empirical approach to understand the visual accuracy of drawings of photographs. Their operational definition states that:

> a visually accurate representation is one that can be recognized as a particular object at a particular time and in a particular space, rendered with little addition of visual detail that cannot be seen in the object represented or with little deletion of visual detail. (609)

For example, with such a definition, a photograph would be considered far more visually accurate than Picasso's *Guernica*. In a follow-up paper in 2016 titled "The Genesis of Errors in Drawing," it is found that the majority of studies on drawing accuracy are best defined in a qualitative manner, typically via independent observers' rating of accuracy, rather than in direct comparison with photographic stimuli (Chamberlain and Wagemans). Accordingly, the

weak mapping of shapes analysis (i.e., "pixel-for-pixel" comparison) to visual accuracy is inherently problematic, in view of other research that has already suggested how artists often distort geometric accuracy to better represent perceptual experience, such as in the work of Paul Cezanne (Robert Pepperell and Manuela Haertel; Joseph Baldwin et al.). The psychological study of sketching for the past two decades among researchers in cognitive science and neuroscience has focused on the sketch artefact as a key tangible manifestation of creativity related to visual perception and artistic processes. It is as if the sketch holds the invisible thinking processes of the artist and thus assumes a meta-representation or proxy of the artist's own creativity. If this is indeed the case, the pursuit of artificial creativity would then necessitate a pursuit in designing artificial intelligence systems that are capable of not just creating novel sketches, but more importantly, sketches that are re-interpretable with conceptual accuracy over shape accuracy. Urban sketches belong to the category of observational drawings that concerns spatial scenes, instead of simply objects without contexts. Recent neuroscience studies have shown that the reduction of photographic scenes into the simple line abstraction of sketches is sufficient to preserve the global scene structures and for functional MRI decoding (Walther et al.). In other words, seeing an urban sketch could be as good as seeing its real urban scene, since the former could trigger a similar activation in the human brain as the real stimuli.

In this project, rather than having the human subject directly tracing over a Google Street View imagery, the hand-eye coordination in transferring visual information is maintained. Although by doing the former, one would have guaranteed a "pixel-for-pixel" shape correspondence between the photographs and the derivative sketches, but the geometric and spatial distortion often emerged during a sketch might have been lost. In other words, it is more crucial to preserve visual accuracy in the concept than in the shapes.

Training Data: Artists or Non-Artists?

There is a major difference between a sketch done by an expert and that by a non-expert. More specifically for our purposes, the former refers to visual artists, designers, and architects. Out of the four drawing errors first articulated in 1997, namely, *misperception of the object*, *misperception of the drawing*, *motor skills* and *representational decisions*, it is the last error that is most relevant here, despite also being the error downplayed in their original study (Cohen and Bennett). Representational decisions involve having and using of a pictorial schema that could effectively represent the features of a given photograph in the form of a sketch. The Limited-Line Tracing Task experiment conducted by both Kozbelt et al. and Ostrofsky et al. made evident how the former group of subjects outperform the latter in the task of selecting the most significant information for depicting a given photograph. Each subject was given a clear plastic folder containing a grayscale photograph of an elephant. Using 30 pieces of dark brown tape as tracing line segments, the subject was to make careful decisions on which parts of the photograph to be rendered as a minimal line drawing within an allocated duration of 15 minutes. The result of the experiment confirms the superior feature extraction capability of the artist subjects. Accordingly, to train an artificially creative AI-urban-sketching

model, the dataset needed should consist of both photographic urban scenes and meaningfully abstracted sketches done by an expert.

In this project, a similar set of constraints is placed on the data creation process. The training dataset consists of a sketch done by a single architecturally-trained human subject, in either black or white tones, or with an addition of a mid-tone. The stylus brush stroke size is fixed. The omission of the eraser is to capture the cognitive traces of the sketch, which is often imperfect-looking. A duration of less than 2-5 minutes is allocated for each sketch. Regardless of the orientation of the Google Street View image provided, the drawing area is the same, and the program records the strokes on the drawing canvas until the timer counts down to zero.

The Image or Sketches of the City?

In the paper titled "What makes Paris look like Paris," the computer science researchers created a large dataset of geotagged Google Street View imagery and used a discriminative clustering algorithm to automatically discover geographically representative image elements of different urban scenes (Carl Doersch et al.). Their assumption is that the *image of a city* is a literal collection of image patches containing architectural elements of specific styles. For example, the presence of a particular style of windows or lamp posts would correspond to one of the urban centres of London, Paris or Prague. In the domain of architectural and urban design, there is a difference between these literal images scanned by the machine and the mental images that emerged in the human observer's mind. In the 1960 classic urban design book *The Image of the City,* Kevin Lynch listed five elements of such mental maps that more accurately represent the image of a city and formulated the term *imageability* as a guide to understanding and designing cities. The five elements of his *imageability* consist of *paths, edges, districts, nodes,* and *landmarks*. Accordingly, *paths* are those linear elements in the city which observers move through and are the predominant elements in their image. *Edges* are also linear elements in the city, but unlike *paths*, they are thresholds or boundaries (e.g., walls) between regions and serve to aid observers in organising features of the image. *Districts* are the two-dimensional extent within which observers experience being *inside* a zone that has a commonly identifiable character. It also serves as a reference to mentally contrast with other perceptible yet different districts. *Nodes* are the intensive foci in the city with which observers enter spatially, primarily in the form of junctions among paths (e.g., crossing and convergence) or simply points of concentration (e.g., squares and street corners). *Landmarks* are also point-references, but unlike *nodes*, they are external and not meant to be entered. They are physical objects that could be observed in the foreground, middle-ground, or background. They could be man-made (e.g., buildings, towers and signages) or natural (e.g., mountains and even the sun). Lynch remarks that, given the same physical reality of a city, different observers under different circumstances may see these elements interchangeably. For example, an expressway is a path for drivers, but could be an edge for the pedestrians. The mental interpretation of the same city element could thus be very different and one way to manifest these is through the sketch of the observer. For an urban sketcher, the *paths, edges* and *districts* are the linear elements, while *nodes*

and *landmarks* are the point elements, all of which can be representable as literal linear and point marks on their sketches, whether as in a perspective view (almost always the case) or as city maps. In other words, the specific type of urban sketch used in this paper – the simple thumbnail sketch, is sufficient to capture these cognitive demarcations observed in the image of the city. These perceptible elements could potentially depict the underlying morphological structures of a city. From a psychological perspective, Tversky and Suwa have similarly expressed that, at the abstract level (though with reference to route maps), "the primary elements … indicate concepts that are thought of as points, as lines, as areas, and as volumes. Design sketches also use these elements" (78).

In this project, in addition to the samples of Google Street View imagery, a set of corresponding thumbnail line urban sketches were also made. The motivation is that such sketches of interpreted photographs might also provide the underlying semantic demarcations which Lynch has expressed in his formulation of a city's *imageability*, thus capturing (as the urban sketcher movement manifesto puts it) the place, time and story of a city witnessed on-location.

Mapping Domains: Imagery and Sketching

Both *DeepDream* (Mordvintsev et al.) and *Neural Style Transfer* (Gatys et al.) have demonstrated that by manipulating the layers of a pre-trained convolutional neural network, one could alter and extract aspects of its perception to generate new visual imagery. This generative potential, especially with the Generative Adversarial Networks or GANs (Goodfellow et al.), has now even been taken up by world-renowned artist Pierre Huyghe in his 2018 *UUmwelt* exhibition at the Serpentine Gallery, as well as being auctioned off at Christie's in the same year for close to half a million dollars. GANs are deep generative models that learn via an adversarial process. Two models (i.e., generator G and discriminator D) are trained simultaneously, where G maximizes the probability of D making a mistake that it is generating samples from an approximated data distribution, in a similar manner to a minimax two-player game. As mentioned previously, the research in sketch-based deep learning has seen some progress in recent years, where the aim is often to train a model in either generating novel sketches from a dataset of sketches or performing sketch-to-image translation from photo-sketch pairs. In the case of the latter task, the *Sketchy Database* is among the most often used dataset, which despite its relatively impressive collection of 125 categories and 75,471 sketches of 12,500 objects, there exists no sketches of scenes *(Sangkloy, Burnell, et al.)*. The crowd-sourced sketches are always based on an object figure on a clean white background. Or, in other words, sketches that are non-spatial and without contexts. Although Sangkloy, Lu, et al. have created a dataset of 200,000 photo-sketch pairs of bedrooms (the only category that is relatively most spatial), the sketches were generated procedurally with boundary detection filter, instead of being hand-sketched by artists or at least crowdsourced from non-artists. The research domain's preference for contextless sketches is particularly evident, such as Chen and Hays who tried to eliminate the background noise that often arises from edge filters by training a deep model

that progressively learnt from another dataset of sketch-image pairs, instead of just edge-image pairs. In short, due to a similar objective of generating coloured photographic images from simple black and white line sketches, the AI research domain often sees the sketch artifact as simply a convenient conditional input for the synthesis of object imagery, and without any embedded conceptual value.

For this project, although similar deep neural network architectures were used, namely pix2pix (Isola et al.) and cycleGAN (Zhu et al.), the dataset was created from scratch and in a manner that defied those used (and preferred) by computer scientists in the research sub-field of sketch-based deep learning. The dataset of urban sketches is thus deliberately spatial (i.e., not object-based), contextual (i.e., with noise in the background or even foreground and middle-ground), manual (i.e., drawn by humans and not generated from edge filters) and expertly (i.e., drawn by a trained architect).

Results

Between the pix2pix and cycleGAN models, the former faces greater difficulty in arriving at a good structural mapping, regardless of whether it is A→B' or B→A' (Figure 7). This is due to its supervised training approach using paired input datasets, which places a greater constraint on an assumed "tracing" or one-to-one compositional similarity between the paired image A and sketch B during model training. Yet, there is no such exact pixel level superimposition, given that the original urban sketches are made, not by tracing over the Google Street View images on the tablet screen, but by "eyeballing" at them on one screen before transferring an abstracted composition of it onto a separate tablet screen.

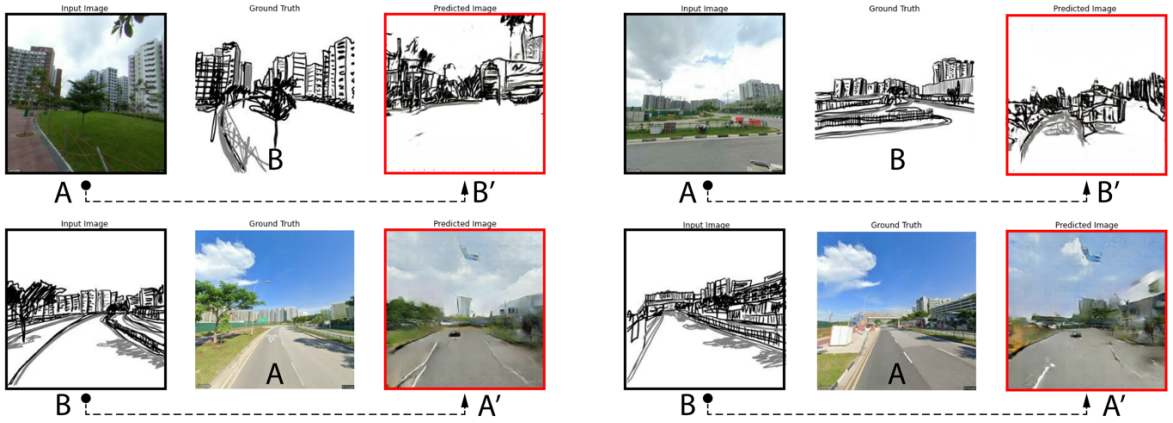

Fig. 7 Samples of generated images from the pix2pix model. (TOP) A→B': The generated sketch B' deviate substantially in its composition not only from the ground truth sketch B, but even the input image A. (BOTTOM) B→A': When compared with the ground truth image A, both generated image A' contain substantial amount of visual noise, such as the floating element in the sky, which might suggest some degree of mode collapse during the training.

Unlike the pix2pix model, the cycleGAN model yields significantly accurate structural mapping. whether it is A→B' or B→A'. This is due to its unsupervised training approach using unpaired input datasets, which places a looser constraint between the images from dataset A and sketches from dataset B during model training. By leveraging the additional cycle-consistency loss embedded in its architecture, the cycleGAN model is able to generalize well in learning a two-directional image-to-sketch mapping. As a result, it is also possible to perform a generative "reconstruction" with A→B'→A' and B→A'→B' (Figure 8). However, the visual quality of the cycleGAN's image-to-sketch mapping (A→B' or A'→B') is generally better since it is easier to map from a higher dimension in full photographic colours to a lower one with grayscale strokes. In a sense, the machine shares a similar mode of abstraction as the human architect, as it extracts a given input into its essences. The A→B' mapping is evidently more successful, as seen from a virtual drive-through along a road in Singapore where the street views (not part of the model's training set) are being "sketched" by the machine in real-time (Figure 9). In fact, abstraction of important architectural design expression can be directly observed from these generated sketches that include modern high-rise buildings and colonial low-rise shophouses.

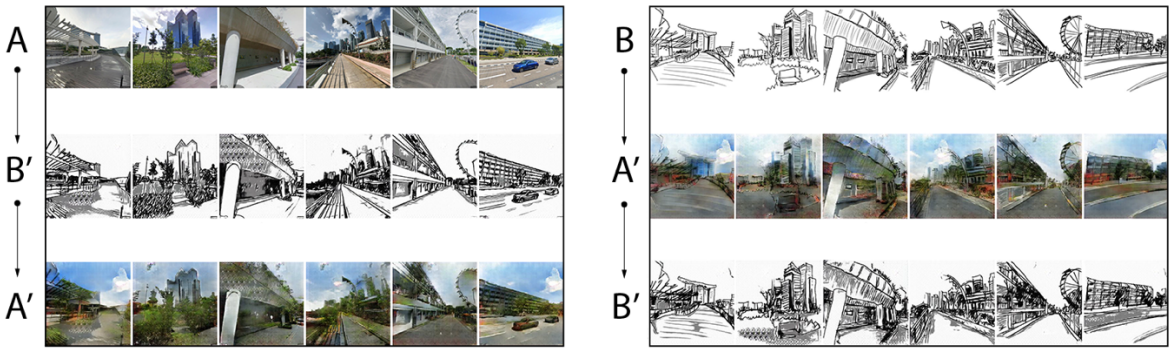

Fig. 8 Samples of generated outputs from the cycleGAN model: (LEFT) *A→B'→A'*: The original Google Street View images as image input *A* are used to condition the generation of sketch output *B'*, which in turn, are used to condition the generation of image output *A* as *A'* back. (RIGHT) *B→A'→B'*: The original hand-sketches of Google Street View images as sketch input *B* are used to condition the generation of image output *A'*, which in turn, are used to condition the generation of sketch output *B* as *B'* back.

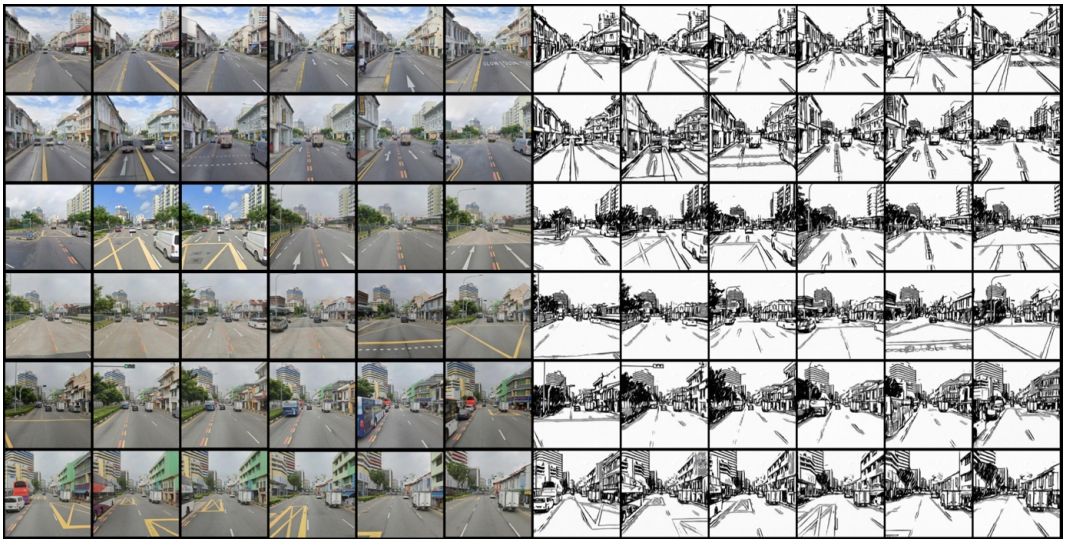

Fig. 9 Driving along a road in Singapore shown as a sequence of images from Google Street View imagery as inputs (A) on the left panel to the AI generated outputs as urban sketches (B') on the right panel.

Conclusion

Drawing from an array of disciplines, such as architecture, urban planning, psychology, cognitive science, computing, neuroscience, artificial intelligence and art, the paper has attempted to lay out the underlying ideas and empirical experiments needed to pursue a plausible artificial creativity with the construction of an automated AI-Urban-Sketcher. Deep learning models have been increasingly adopted in the fields of art, design and architecture, with the optimism that these new sets of pattern-recognition artificial eyes could help us in advancing our own understanding. As seen in their recent use for classifying styles in art (Elgammal et al.) and in architecture (Yoshimura et al.), questions on our traditional epistemological approaches will continued to be raised. Yet, without a critical discourse and artistic practice as attempted in the paper, the extent of artificial creativity afforded by these deep generative neural networks might not be easily discussed, and in the near future, steered in meaningful ways. The simple use of urban sketches as an artefact manifesting human creativity is one of many ways to address issues of human and machine agencies, and the ways in which both entities might collaboratively perceive and conceptualize new forms of design and art. The proposed automated AI-Urban-Sketcher in this paper is thus also a reflection on the possible forms of such interactions, in order to construct a viable understanding of creative automation that could harness the parallel mechanism of human abstraction and machine extraction.

Works Cited

Baldwin, Joseph et al. "Comparing Artistic and Geometrical Perspective Depictions of Space in the Visual Field." *I-Perception* 5.6 (2014): 536–47.

Carl Doersch, et al. "What Makes Paris Look like Paris?" *ACM Transactions on Graphics (SIGGRAPH)* 31.4 (2012).

Chamberlain, Rebecca, and Johan Wagemans. "The Genesis of Errors in Drawing." *Neuroscience & Biobehavioral Reviews* 65 (2016): 195–207. https://doi.org/10.1016/j.neubiorev.2016.04.002.

Chen, Wengling, and James Hays. "SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis." *ArXiv:1801.02753 [Cs]* (2018). http://arxiv.org/abs/1801.02753.

Cohen, Dale J., and Susan Bennett. "Why Can't Most People Draw What They See?" *Journal of Experimental Psychology: Human Perception and Performance* 23.3 (1997): 609–21. https://doi.org/10.1037/0096-1523.23.3.609.

Eitz, Mathias, et al. "How Do Humans Sketch Objects?" *ACM Transactions on Graphics* 31.4 (2012): 44:1-44:10. https://doi.org/10.1145/2185520.2185540.

Elgammal, Ahmed, et al. "The Shape of Art History in the Eyes of the Machine." *ArXiv:1801.07729 [Cs]* (2018). http://arxiv.org/abs/1801.07729.

Gatys, Leon A., et al. "A Neural Algorithm of Artistic Style." *ArXiv:1508.06576 [Cs, q-Bio]* (2015). http://arxiv.org/abs/1508.06576.

Goodfellow, Ian J., et al. "Generative Adversarial Networks." *ArXiv:1406.2661 [Cs, Stat]* (2014). http://arxiv.org/abs/1406.2661.

Ha, David, and Douglas Eck. *A Neural Representation of Sketch Drawings*. Feb. 2018. *openreview.net*. https://openreview.net/forum?id=Hy6GHpkCW.

Isola, Phillip, et al. "Image-to-Image Translation with Conditional Adversarial Networks." *ArXiv:1611.07004 [Cs]* (2018). http://arxiv.org/abs/1611.07004.

Jongejan, J. et al. *The Quick, Draw! - A.I. Experiment*. 2016, https://quickdraw.withgoogle.com/.

Kozbelt, Aaron, et al. "Visual Selection Contributes to Artists' Advantages in Realistic Drawing." *Psychology of Aesthetics, Creativity, and the Arts* 4.2 (2010): 93–102. https://doi.org/10.1037/a0017657.

Lynch, Kevin. *The Image of the City*. 1st ed. The MIT Press, 1960.

Mordvintsev, Alexander, et al. "Inceptionism: Going Deeper into Neural Networks." *Google AI Blog*, http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html. Accessed 13 Aug. 2019.

Negroponte, Nicholas. "Recent Advances in Sketch Recognition." *Proceedings of the June 4-8, 1973, National Computer Conference and Exposition on - AFIPS '73*, ACM Press, 1973. 663-675. https://doi.org/10.1145/1499586.1499747.

Odell, Jenny. "Projects." http://www.jennyodell.com/projects.html. Accessed 30 Jan. 2017.

Ostrofsky, Justin, et al. "Perceptual Constancies and Visual Selection as Predictors of Realistic Drawing Skill." *Psychology of Aesthetics, Creativity, and the Arts* 6.2 (2012): 124–36. https://doi.org/10.1037/a0026384.

Robert Pepperell and Manuela Haertel. "Do Artists Use Linear Perspective to Depict Visual Space?" *Perception* 43.5 (2014): 395–416.

Sangkloy, Patsorn, Jingwan Lu, et al. "Scribbler: Controlling Deep Image Synthesis with Sketch and Color." *ArXiv:1612.00835 [Cs]* (2016). http://arxiv.org/abs/1612.00835.

Sangkloy, Patsorn, Nathan Burnell, et al. "The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies." *ACM Transactions on Graphics*, vol. 35, no. 4, July 2016, pp. 1–12. *DOI.org (Crossref)*, https://doi.org/10.1145/2897824.2925954.

Tversky, Barbara. "What does drawing reveal about thinking?" *Visual and spatial reasoning in design. Eds.* J. S. Gero & B. Tversky. Sydney, Australia: Key Centre of Design Computing and Cognition, 1999.

Tversky, Barbara, and Masaki Suwa. "Thinking with Sketches." *Tools for Innovation: The Science behind the Practical Methods That Drive New Ideas*, Oxford University Press, 2009. 75–84.

Urban Sketchers. http://www.urbansketchers.org/. Accessed 31 May 2021.

Walther, D. B., et al. "Simple Line Drawings Suffice for Functional MRI Decoding of Natural Scene Categories." *Proceedings of the National Academy of Sciences* 108.23 (2011): 9661–66. https://doi.org/10.1073/pnas.1015666108.

Yoshimura, Yuji, et al. "Deep Learning Architect: Classification for Architectural Design through the Eye of Artificial Intelligence." *ArXiv:1812.01714 [Cs]* (2018). http://arxiv.org/abs/1812.01714.

Zhu, Jun-Yan, et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." *ArXiv:1703.10593 [Cs]* (2017). http://arxiv.org/abs/1703.10593.

Dataset

The dataset created for the project can be made available upon request for non-commercial use by sending an email to the author directly at Immanuel_koh@sutd.edu.sg.