Computational Creativity in Architecture:

An Exploration of the Potential for Artificial Creativity using Generative Adversarial Networks to create Original Architectural Spaces from Text Descriptions.

Georgina Myers
WELSH SCHOOL OF ARCHITECTURE, CARDIFF UNIVERSITY
A dissertation submitted in partial fulfilment of the degree of M.Arch
2020

Contents

'Artificial Intelligence is the study of how to make computers do what real minds can' – Margaret Boden[1]

[1] Boden, Margaret A, The Creative Mind: Myths and Mechanisms, 2nd edn (Routledge, 2004)

Abstract

What is creativity in architecture, and can the process be carried out artificially? What could artificial architecture look like? This dissertation strives to understand and measure artificial creativity's potential to assist the architect's creative process, particularly in the generation of architectural concepts and ideas. This understanding was found by assessing human creativity, neural networks' underlying processes, and artificial intelligence's current creative capacity for architectural applications. To address this, a Generative Adversarial Network, based on AttnGAN, was trained to generate images of internal and external architectural space based on text descriptions of the space, exploring computational interpretations of architectural space.

The human and artificial creative processes were compared and put into their cultural and societal context to assess the output, process and repercussions of artificial architectural ideas on the domain. We interpret that creativity is a co-creative process, where we cannot decouple human input from the network's output. The artificial creative process is similar to our own and has been demonstrated by this paper to produce creative ideas; the model has created something humans could not. How we as humans choose to interpret this output and take it further into our design thinking is up to us.

Acknowledgements

I want to thank my supervisor, Ian Knight, for his excitement and interest in the experiments with this innovative technology, as well as for his guidance and help navigating this complex topic.

I would also like to thank Kovid for his knowledge and unfailing support. Lastly, I must thank my friends and family for their support, patience and open-mindedness.

List of Illustrations

Figure 12– *GAN Model*

Figure 13 - *Image-to-Image Translation with Conditional Adversarial Networks*
Isola, P, J Zhu, T Zhou, and A A Efros, 'Image-to-Image Translation with Conditional Adversarial Networks', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967–76 <https://doi.org/10.1109/CVPR.2017.632>

Figure 14 - *Cartoon GAN*
Chen, Yang, Yukun Lai, and Yong-Jin Liu, 'CartoonGAN: Generative Adversarial Networks for Photo Cartoonization', 2018 <https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf>

Figure 15 - *Adapted AttnGAN Model*
Diagram adapted by the author, original diagram from:
Xu, T, P Zhang, Q Huang, H Zhang, Z Gan, X Huang, and others, 'AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks', in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1316–24 <https://doi.org/10.1109/CVPR.2018.00143>

Figure 16 - *Initial Spaces Dataset Outputs*

Figure 17 - *Spaces Data-set initial Outputs in Grid*

Figure 18 - *Final Spaces Dataset Outputs - randomly generated outputs from different training epochs (by the author)*

Figure 19 - *Interpolation of Final Spaces Dataset*

Images by the author, photographs scraped from: 'ArchDaily', 2021 <www.archdaily.com/>

Figure 20 - *Large Images of Façade dataset Outputs*

Figure 21 - *Façade dataset output in grid form (see further images in Appendix 3.4)*

Figure 22 - *Interpolation of Façades 1*

Images by the author, photographs from Xu, Z, D Tao, Ya Zhang, J Wu, and A Tsoi, 'Architectural Style Classification Using Multinomial Latent Logistic Regression', in ECCV, 2014

Figure 23 - *Interpolation of Façades 2*

Images by the author, photographs from Xu, Z, D Tao, Ya Zhang, J Wu, and A Tsoi, 'Architectural Style Classification Using Multinomial Latent Logistic Regression', in ECCV, 2014

Figure 24 - *Bedroom Dataset Output*

All Images and Drawings belong to the author unless otherwise specified

1 Introduction

This dissertation explores whether computers can be considered creative. Creativity and computational tools are necessary - yet currently separate - parts of the architectural design process. Creativity is central to the architect's work, and computational tools are prevalent in design and society, allowing many diverse types of inputs: can we get a computational creative output that is useful to architects?

This research aims to understand and measure artificial creativity's potential to assist the architect in their creative process, especially in generating ideas, by assessing artificial intelligence's current creative capacity. To answer this aim, the original research undertaken in this study involves using and training a machine learning model to map abstract, text-based inputs to realistic visual outputs in the form of images. Artificial creativity will be approached from an architectural perspective through the understanding of exterior and interior spaces; this will reduce the problem to a manageable scale for the time available. While similar artificial neural networks have been applied in an architectural setting before, we will seek to help neural networks "understand" aesthetic and spatial values in a conceptual sense.

We will carry out an experiment on a neural network – based on the *AttnGAN*[2] model – which will model an artificial creative process in the architectural domain. Neural Networks, particularly Generative Adversarial Networks (GANs), are the most appropriate form of Artificial Intelligence (AI) for this study. Neural Networks (NNs) are conceptually the most similar to natural neural processes and have been used by psychologists to understand human creativity[3]. This research will serve as a study of artificial creativity, the outputs of which will then be analysed in the context of theoretical discourses about what defines creativity.

[2] T Xu and others, 'AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1316–24 <https://doi.org/10.1109/CVPR.2018.00143>.

[3] Herbert A Simon, 'Creativity and Motivation: A Response to Csikszentmihalyi', *New Ideas in Psychology*, 6.2 (1988), 177–81 <https://doi.org/10.1016/0732-118X(88)90002-5>.

1.1 Structure

The Literature Review, concerning creativity and NNs, contextualises the main body of research and establish a background for the later discussions of artificial creativity. Being able to define the concept and process of creativity is key to discussing and critiquing whether AI can be creative. Chapter 2.1 tries to understand the creative process psychologically and socially before moving on to the architectural creative process and its relationship with general creativity. This expanded definition will then be used to hypothesise and discuss what artificial creativity is, how we can define it and whether the NN in our experiment can be considered creative in an architectural sense.

In Chapter 2.2, the essential workings of machine learning (ML), and NNs will be explained along with their relevance. Several papers on ML and NNs will be studied to understand their evolution, the processes and functions behind them and why they matter regarding our own creativity and neurological functions. Then reflecting on the literature review, we will articulate the criteria by which it can be determined whether AI is capable of creativity in our experiment.

The experiment with the AttnGAN is explained in Chapter 3, Aims & Methodology before determining and discussing whether and how artificial creativity is possible, its potential limitations, and the case study's output in Chapter 4, Results and Discussion. The implication of artificial architectural creativity for the architectural domain will then be discussed, as we try to measure how and to what extent artificial creativity has been exhibited.

2 Literature Review

2.1 Creativity

This dissertation's focus is to discuss and attempt to show how AI, in the form of NNs, can be creative in an architectural setting. To discuss the possibility of artificial or computational creativity, it is necessary first to try to find a holistic definition and criteria for creativity in the human mind. Whilst looking at creativity through a human and societal lens, the current philosophical and psychological literature and research on creativity will be discussed. This holistic definition and process will be applied to architectural and spatial creativity, looking at what it means to be creative and innovative as an architect and designer. Lastly, in addition to this knowledge, different definitions and types of artificial creativity from Boden will be analysed to determine what artificial or computational creativity is and how it comes about.

2.1.1 Individual Creativity

To be human is to be creative; it is what differentiates us from every other species[4] , and it has been invaluable to our evolution[5]. Every human is creative, as creativity is grounded in our everyday abilities, including 'conceptual thinking, perception, memory and reflective self-criticism'[6]. Nevertheless, what is human creativity, and how is it possible?

Boden defines creativity as 'the ability to come up with ideas or artefacts that are new, surprising and valuable'[7]. Where these ideas come from is often romanticised and mysticised, meaning that most people are dubious and mistrusting of science's ability to

---

[4] René Víctor Valqui Vidal, 'To Be Human Is to Be Creative', *AI & Society*, 28.2 (2013), 237–48 <https://doi.org/10.1007/s00146-012-0415-1>.

[5] Mihaly Csikszentmihalyi, 'The Systems Model of Creativity and Its Applications', *The Wiley Handbook of Genius*, Wiley Online Books, 2014, pp. 533–45 <https://doi.org/https://doi.org/10.1002/9781118367377.ch25>.

[6] Margaret A Boden, *The Creative Mind: Myths and Mechanisms*, 2nd edn (Routledge, 2004).

[7] Boden, *The Creative Mind: Myths and Mechanisms*.p. 1

explain the creative process.[8] Before delving into the creative process itself, it is necessary to differentiate between small or personal creativity (P-Creativity) and big or historical creativity (H-Creativity). A P-Creative idea is surprising and new to the person who came up with it. In contrast, an H-Creative idea is a globally original thought that is new and valuable to broader society and history.[9]

There are several different models for the creative process. The four-stage process we will refer to was established by Wallas as a five-step process, then shortened by Hadamard: preparation, incubation, illumination and verification[10] (see Figure 1). This four-step process is usually referred to as the Wallas model for compatibility.

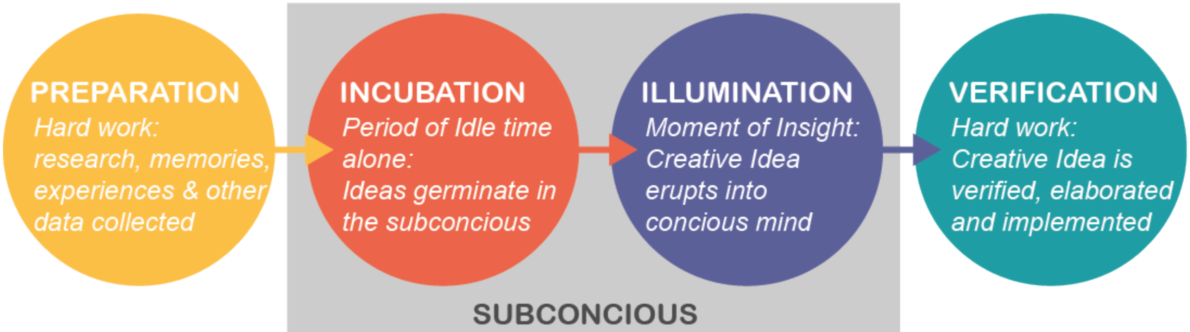

Figure 1 - The Wallas Model of the Creative Process

The Wallas Model starts at the *Preparation* phase where hard-work, research and data inform the idea that germinates in the subconscious processes. Although most of the work occurs in the conscious mind, the *Incubation* and *Illumination* stages occur in the subconscious. In the *Incubation* phase, new combinations of elements are repeatedly attempted until one is coherent, stable and realistic enough to emerge through to consciousness as insight in the *Illumination* phase[11]. After the idea has been created, it is verified and elaborated on in the conscious mind during the *Verification* phase to make

---

[8] Boden, *The Creative Mind: Myths and Mechanisms*. P.14

[9] Boden, *The Creative Mind: Myths and Mechanisms*.p.2

[10] Mihaly Csikszentmihalyi, 'The Systems Model of Creativity', 2014 <https://doi.org/10.1007/978-94-017-9085-7>.pp.67

[11] Mihaly Csikszentmihalyi and Keith Sawyer, 'Creative Insight: The Social Dimension of a Solitary Moment BT - The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi', ed. by Mihaly Csikszentmihalyi (Dordrecht: Springer Netherlands, 2014), pp. 73–98 <https://doi.org/10.1007/978-94-017-9085-7_7>. Pp. 73

sure it is a valuable idea. While the subconscious phases may be the most crucial to the creative process, they are impossible without the raw data and information from the *Preparation* phase of conscious work. Theorists disagree about what happens in the subconscious; it is somewhat of a 'black box'[12]. Although Hadamard argued that active guided processing occurs[13], most current researchers believe the process consists of chance combinations of thought processes below the conscious threshold[14]. This generation of subconscious random ideas, and the filter or threshold they must cross to be fit for the subsequent stages in the conscious mind, is at the heart of the creative process.

The Wallas model lays out the human creative process, of which the result is an idea or artefact that is novel, surprising and valuable. The conscious stages of the Wallas model – *Preparation* and *Verification* – are the result of hard work. The former relates to studying and analysing data for an extended period to get enough information to create a novel idea of value. The latter makes the conscious mind verify and question the new insight to ensure it is of value. However, if these processes were to be carried out independently of the idea's domain, the creative idea would lack context, social influence and data[15]. Creativity is not a reclusive process.

---

[12] Csikszentmihalyi, 'The Systems Model of Creativity'.

[13] Jacques Hadamard, *An Essay on the Psychology of Invention in the Mathematical Field.*, *An Essay on the Psychology of Invention in the Mathematical Field.* (Oxford, England: Princeton Univ. Press, 1949).

[14] Csikszentmihalyi and Sawyer. Pp.74

[15] Csikszentmihalyi, 'The Systems Model of Creativity'.

2.1.2 Creativity in Society

Individuals often think of creativity as a solitary act. After all, new insights and ideas occur when isolated, but creativity cannot exist in a vacuum[16]; it is surrounded and contextualised by societal and cultural context. Though external to the Wallas model's individual creative process, society and culture interact with the individual at the beginning and end of the creative process; the *Preparation* and *Verification* phases. Further evaluation is external to the individual process when the creative idea is contextualised and judged by those in the field. [17]

Csikszentmihalyi wrote 'Psychologists tend to see creativity exclusively as a mental process [but] creativity is as much a cultural and social as it is a psychological event.'[18] Creative ideas that have value in society impact society; a new H-creative idea or artefact adds knowledge to and builds on the domain. The systems model (see Figure 2), devised by Csikszentmihalyi and developed as a cultural model of evolution, occurs at the interface of three subsystems: the individual, who absorbs information from the domain/culture and uses it to come up with an idea, the field, who decides whether it should be included in the domain, and the domain, from where ideas are available to the next individual[19].

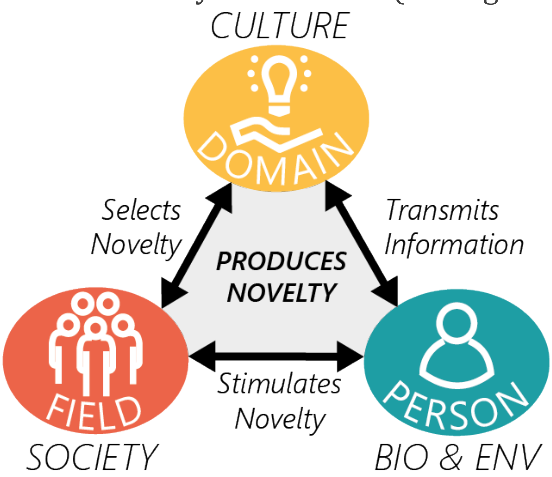

Figure 2 - Systems Model of Creativity

The field decides the value of an idea that seeks to join their domain; this is valid for all domains, including architecture. However, architectural design is a highly collaborative process: different individuals from different and overlapping domains, including

---

[16] Mihaly Csikszentmihalyi and Rustin Wolfe, 'New Conceptions and Research Approaches to Creativity: Implications of a Systems Perspective for Creativity in Education', 2014, 81–93 <https://doi.org/10.1007/978-94-017-9085-7_10>. Pp.159

[17] Csikszentmihalyi, 'The Systems Model of Creativity'. pp.72-76

[18] Csikszentmihalyi, 'The Systems Model of Creativity'.

[19] Csikszentmihalyi and Wolfe. Pp. 162

architects, engineers and clients, working together, with architects coordinating most of the work. Even when a design has been built the users of the design are continuously evaluating it.[20] Due to the architectural design process's collaborative nature, the architects' domain encompasses engineering, surveying, and many more fields. There is a wide variety and amount of information accumulated in the preparation stage and a large field, including the public, to validate the final artefact within this extended domain.

2.1.3 Architectural Creativity

The architect's process is steeped in creativity, yet day-to-day, architects do not pay much attention to their own processes, just the drawings or physical output. The architectural design process has two significant features according to Mozaffar and Khakzand[21]: it is firstly a creative effort, and secondly has a close association to drawing[22]. It is essential to break down the creative design process to understand how creativity functions in the architectural realm as architects create both artefacts and ideas.

Creativity in any form is centred on stochastic processes.[23] The mid-two stages of the Wallas Model, incubation and illumination, are the random solitary processes that usually happen in the subconscious[24], where they are less constrained by the real world making them a critical factor in abstract creativity. For the human architect, these stages

---

[20] Maii Emam, Dina Taha, and Zeyad ElSayad, 'Collaborative Pedagogy in Architectural Design Studio: A Case Study in Applying Collaborative Design', *Alexandria Engineering Journal*, 58.1 (2019), 163–70 <https://doi.org/10.1016/j.aej.2018.03.005>.

[21] Farhang Mozaffar and Mehdi Khakzand, 'ARCHITECTURAL DESIGN PROCESS IN TECHNOLOGY AGE', *IUST*, 19.6 (2009), 53–72 <http://ijiepm.iust.ac.ir/article-1-263-fa.html>.

[22] Myers, Georgina, Preliminary Literature Review, 2020

[23] Daniela Sirbu and I Dumitrache, 'A Conceptual Framework for Artificial Creativity in Visual Arts', *International Journal of Computers Communications & Control*, 12 (2017), 381 <https://doi.org/10.15837/ijccc.2017.3.2759>.

[24] Graham Wallas, *The Art of Thought,* (New York: Harcourt, Brace and Company, 1926); Georgina Myers, *Preliminary Literature Review*, 2020.

often happen during the drawing and conceptual process, where we are influenced by our experiences, memories and precedents from the preparation stage.

While separate from the creative models mentioned above, Boden defines three main types of creativity that interact with the domain in different ways: exploratory creativity, transformational creativity, and combinational creativity. Exploratory creativity explores an existing conceptual space by creating unfamiliar ideas in a familiar space, like an artist making a new painting in a familiar style. Combinational creativity is the 'unfamiliar combination of familiar ideas'[25], which requires in-depth knowledge of different conceptual spaces (imagine the artist combining textiles and painting into textile art). Transformational creativity transforms a conceptual space by 'thinking something which, with respect to the conceptual spaces in their minds, they couldn't have thought of before'[26]; the artist has created a novel style of painting. Combinational and exploratory creativity are the most prevalent types in the architectural field, as they operate within the confines of the known conceptual space.

The architect Daniel Liebeskind is credited with saying 'To provide meaningful architecture is not to parody history, but to articulate it'[27]. This quote emphasises the preparation stage of creativity and the importance of the domain and field in coming up with new architectural ideas. Precedents and existing designs are hugely important in architecture; influencing exploratory ideas at the conceptual stage of the design process and combinational ideas during the manifestation and enactment of the design. Architectural creativity is rarely coming up with an 'original' or transformational creative idea; it recalls and combines different existing design solutions, ideas and precedents into a new P-creative idea or artefact.

The beauty of architecture is its constraints, or as architect Walter Gropius famously stated, 'Limitation makes the creative mind inventive'[28]. Originality in architecture is

---

[25] Boden, *The Creative Mind: Myths and Mechanisms*. Pp.3

[26] Boden, *The Creative Mind: Myths and Mechanisms*.

[27] Jean-Claude Dubost and Jean-François Gonthier, *Architecture for the Future* (Paris: Paris : Terrail, 1996).

[28] Walter Gropius, *Re:Education of an Architect*, Gropius Papers, 1939.

when a genuinely transformational idea occurs, usually an H-creative idea. However, the building is still a combinational idea to an extent due to the physical and structural constraints of architecture itself. One must question whether spaces and buildings are even able to be exhibit genuine transformational creativity. Architectures' constraints differentiate it from art. A building must have walls and be structurally sound, but it also has the same limitation to be creative; useful, surprising and valuable. H-Creative architecture is rare, but can arguably be seen in the works of Zaha Hadid, Bjarke Ingels and Antonio Gaudi (see Figure 3), their creative artefacts and ideas challenge the limitations of and redefines what architecture can be.

Figure 3 - Amager Bakke (BIG), Casa Batllo & Sagrada Familia (Antoni Gaudí), and The Heydar Aliyev Center (Zaha Hadid)

2.1.4 Artificial (computational) Creativity

Artificial creativity is a relatively new concept in the creative realm; still, it has the potential to change and greatly aid our creative process. The idea of artificial creativity seems to go against our innate understanding of what creativity is. As Lady Lovelace said in the mid-19th century, “The analytical engine has no pretensions whatever to originate anything. It can do [only] whatever we know how to order it to perform.”[29]. However, we now know that the first sentence is false; although the machine can only do what we order it to perform, we can “order” the machine to be original.

There are four stages to the development of artificial creativity as defined by Boden in the Lovelace questions[30]:

1. ‘Can computational ideas help us understand how human creativity is possible?’
2. ‘Can computers appear to be creative?’
3. ‘Can computers appear to recognise creativity?’
4. ‘Can computers ever really be creative?’

Questions one and two have been answered in the affirmative, with examples of question two cropping up in the sciences and the arts[31]. The first three questions are empirical scientific questions. Question four is a philosophical question that is much harder to answer and invokes questions of its own; how does one judge a computer to be creative?

To be creative, the computer would need first to appear creative and recognise creativity. However, does the computer need to be independently creative and carry out its creativity without prompting from the programmer? Could a joint process still be deemed creative? One major problem is how AI can be pushed towards being both novel

[29] Christopher Hollings, Ursula Martin, and Adrian Rice, ‘Ada Lovelace and the Analytical Engine’, *Oxford Blog*, 2018, p. https://blogs.bodleian.ox.ac.uk/adalovelace/2018/0.

[30] Boden, *The Creative Mind: Myths and Mechanisms*. Pp. 16

[31] Boden, *The Creative Mind: Myths and Mechanisms*.

and useful, rather than just surprising and random. For now, humans are always involved at some point or another in the process to guide the AI.

For both the second and fourth question, one needs to ask what is being creative here: creator, machine, or society? As the programmer still has to define value in creativity to the computer. Even training a computer to recognise creativity requires question three to be true. As John Smith, a researcher at IBM, wrote 'It is easy for AI to come up with something novel just randomly. But it's very hard to come up with something that is novel and unexpected and useful.'[32]

We will focus on the second and fourth Lovelace question. If a computer appears creative, it is worth questioning why this is and how it is actually creative. AI can be creative in the same three ways as humans: combinational, explorational and transformational creativity.[33] However, before recent revelations in NNs, AI has been most successful in exploratory creativity[34]. The hurdle for artificial combinational creativity must be that it struggles to identify the values and lacks the rich associative memory of humans; on the other hand, we struggle to express our values computationally.

AI is designed to mimic the way organic brains function; the structure of Artificial Neural Networks (ANNs) is based on our natural neural networks[35]. Our own life experiences inspire everything that humans create; places we have been, precedents studied, and conversations had. By this effect, everything that a machine creates is inspired by its database of experiences (the data that we train it on), which is equitable to our natural database of experiences. The two-mid stages of creativity in the subconscious are equitable with how the randomised weights in neural networks work;

---

[32] Ellen Cornillon, 'The Quest for AI Creativity', *IBM*, 2016, p. 1 <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-creativity.html> [accessed 6 November 2020].

[33] Margaret A Boden, 'Creativity and Artificial Intelligence', *Artificial Intelligence*, 103.1 (1998), 347–56 <https://doi.org/https://doi.org/10.1016/S0004-3702(98)00055-1>.

[34] Boden, 'Creativity and Artificial Intelligence'.

[35] Nikolaus Kriegeskorte, 'Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing', *Annual Review of Vision Science*, 1.1 (2015), 417–46 <https://doi.org/10.1146/annurev-vision-082114-035447>. Pp.418

our ability to be creative gets better the more we train it. Even if artificial networks are different from natural ones, it is not a limitation on whether they could be creative.

2.1.5 Concluding Creativity

The main ideas for creativity have been collated in Table 1 below. These criteria for creativity give us a basis to judge what kind of AI best suits our exploration of creativity and how we may judge it. A key point in defining artificial creativity will be looking at what decides value, in both process and output and categorising the type and stage of creativity. The Lovelace questions then set up a further philosophical debate about what it means to be creative and challenge the notion that creativity is a purely human endeavour. NNs will be explored in the following section to determine which are best applied to creating artificial creativity, which will later be measured by the below criteria.

Idea	Criteria	PROS	CONS
Boden's definition	'Ability to come up with ideas or artefacts that are new, surprising and valuable.'	Analyse the idea or artefact itself, with minimal criteria.	How to determine value is unclear.
Boden's Types of Creativity	Three main types of creativity: exploratory, transformational, and combinational creativity.	Allows us to differentiate between different types of creativity.	It risks forced categorisation and needs defining in both process and output.
Systems Model	The field decides whether the individual's creative idea belongs in the domain.	The field decides what is of value and connection to society is established.	It does not analyse the process and risks being a shallow analysis due to simple criteria.
Wallas Model	Is each stage occurring? Preparation, incubation, illumination, verification.	Analyse the process itself Clear stages that need to all occur.	Is it safe to assume that humans and machine have the same process?
The Lovelace Questions	Can computers: help us understand human creativity, appear to be creative, appear to recognise creativity, ever really be creative?	Specific to artificial intelligence- it helps us measure the extent to which it is being creative rather than treating creativity as binary.	The criteria are not clear cut as the questions are somewhat philosophical.

Table 1 – Definitions for Creativity

2.2 Neural Networks

2.2.1 How Does the Architecture of Neural Networks Enable them to Be Creative?

Artificial neural networks (ANNs), named after and inspired by biological NNs, is a branch of machine learning with strong parallels to natural neurological architecture. Using massive data sets, ANNs are trained to learn the rules and data parameters in datasets to discover patterns and relationships that may not have occurred directly to humans[36][37]. The complexity of these networks has exploded in recent years: shallow networks (networks with one hidden layer, see Figure 4) are barely used; instead, deep neural networks (DNNs) with multiple layers are favoured as they allow the representation of many complex functions more concisely. DNNs have even surpassed human-level performance in several areas including visual object classification, speech recognition and machine translation. While there are some examples of artificial creativity, these are contested and limited to specific network architectures.

ML is separate from traditional AI; in classic AI, it is the programmer's responsibility to explicitly program the system's algorithm, in contrast to the ML's systems that 'learn' (extract structural/statistical regularities) from the input data.[38] While other forms of AI follow a psychological approach, ANNs follow a neuroscientific approach using natural neuron-like elements.[39]

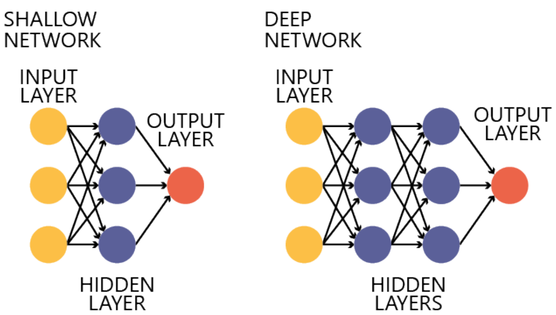

Figure 4 – Shallow vs Deep Neural Networks

---

[36] Imdat As, Siddharth Pal, and Prithwish Basu, 'Artificial Intelligence in Architecture: Generating Conceptual Design via Deep Learning', *International Journal of Architectural Computing*, 16.4 (2018), 306–27 <https://doi.org/10.1177/1478077118800982>.

[37] Myers.

[38] Anthony M Zador, 'A Critique of Pure Learning and What Artificial Neural Networks Can Learn from Animal Brains', *Nature Communications*, 10.1 (2019), 3770 <https://doi.org/10.1038/s41467-019-11786-6>. Pp.2

[39] Zador.

On a higher level, ANNs automatically learn associations between inputs and outputs[40]. ANNs have three stages of layers: an *input layer* consisting of the input data, this layer feeds into multiple *hidden layers* that use mathematical manipulation to turn the input into the *output layer*, which produces the outcome[41][42] (see Figure 5).

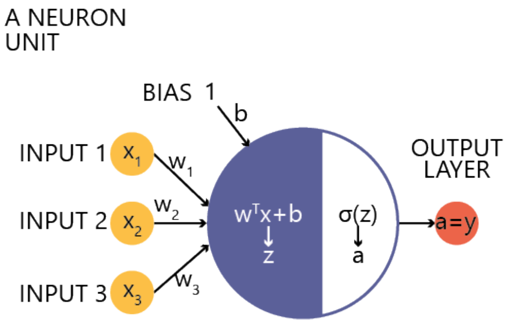

Figure 5 – Neuron in an ANN

When trying to understand a NN, it is best to start small with a single neuron (see Figure 5). The learning mechanisms of which are: the feature inputs (inputs x_1, x_2 and x_3), the randomly initialised weights (w_1, w_2 and w_3), the bias term (b), the summation (z), and the activation (a)[43]. A neuron takes several inputs, computes a weighted sum z, and imposes a threshold using an activation function a (sigmoid activation function, σ, is used in this example); learning to match inputs to the correct output.[44] Activation functions model non-linear relationships of data, which allows an understanding of more complex patterns like that of real-world data[45]. The neuron's activation has parallels with natural neurons: the neuron's response to inputs and the single binary output parallel the biological hardware of categorisation[46] and spiking neurons (when a neuron is triggered by an incoming electrical impulse and reaches its action potential)[47].

[40] Niloy Purkait, *Hands-On Neural Networks with Keras: Design and Create Neural Networks Using Deep Learning and Artificial Intelligence Principles* (Birmingham: Birmingham: Packt Publishing, Limited, 2019).

[41] As, Pal, and Basu; Roger Lustig, 'Margaret Boden, the Creative Mind: Myths and Mechanisms', *Artificial Intelligence*, 79.1 (1995), 83–96 <https://doi.org/10.1016/0004-3702(95)90025-X>.

[42] Myers.

[43] Purkait.

[44] Kriegeskorte.

[45] Purkait.

[46] Kriegeskorte.

[47] Purkait.

When single neurons are stacked as hidden layers in a DNN, they can model complex functions with immense expressive power[48]. However, they are still not as powerful as natural mechanisms in applications such as abstract thought and creativity[49].

Moving forward through the hidden layers as described above, we have completed one forward pass of *forward propagation*; additionally, *backpropagation* is the key to an effective and reactive NN. Backpropagation uses the calculated errors from the forward propagation to repeatedly adjust the network's weights, minimising the distance between the desired output value and the actual output value: backpropagation minimises the network's error. Forward propagation and backpropagation result in layers that develop for a particular task domain[50]: they are trained on the dataset.

Where and how information is processed roughly mimics the human brain's processing strategies[51][52]; NNs are the type of AI that functions the closest to how our brains function.[53] Therefore, the non-linear design approach of ANNs is undoubtedly the best way of artificially implementing the creative process.

---

[48] Kriegeskorte.

[49] Zador.

[50] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams, 'Learning Representations by Back-Propagating Errors', *Nature (London)*, 323.6088 (1986), 533–36 <https://doi.org/10.1038/323533a0>.

[51] As, Pal, and Basu.

[52] Myers.

[53] Kriegeskorte.

2.2.2 To What Extent Can Neural Networks Understand Meaning?

Data science is the scientific domain that deals with generating meaning from raw data[54]. Through iterative observations of real-world problems, the overall phenomena are quantified into distinctive *features* and dimensions to predict future outcomes. The quantification of the real world is known as encoding: the process of converting real-world data from variables to a numerical format that neural networks can interpret or understand[55]. This concept of *encoding* meaning into numerical formats is best demonstrated through examples of ANN configurations.

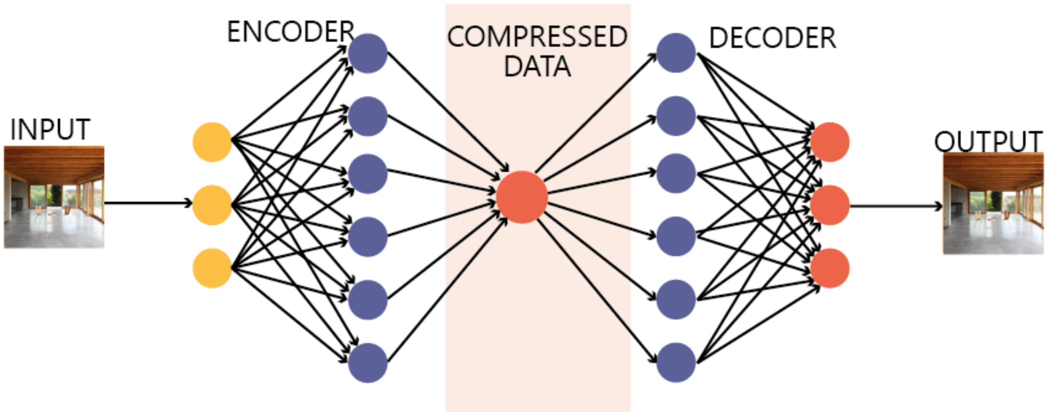

Figure 6 - Standard Autoencoder Diagram

Encoding is implemented practically though autoencoders: autoencoders learn the encoding for sets of data, such as images, natural language data and other data types that would benefit from a dimensionality reduction. They find low-dimensional representations of high-dimensional data while preserving their core attributes non-linearly; they encode and compress data[56]. The second part of the autoencoder is the decoder, which reconstructs the output using the generated code by transforming the latent space (see Figure 6). In the process of learning to reconstruct the original image,

[54] Purkait.

[55] Paul Lavrakas, 'Encyclopedia of Survey Research Methods' (Thousand Oaks, California, 2008) <https://doi.org/10.4135/9781412963947 NV - 0>.

[56] Purkait.

a loss function is devised to measure the distance between the original image and the reconstructed image, which the model learns to minimize. In minimizing this reconstruction loss, a trained model *must* have learned weights that are meaningful for reconstruction, or in other words, encoded the features of the original input in a way that is meaningful to the network for its task.

Similarly, we can also map non-numerical units of meaning, such as words, to compressed encodings of their meanings. Let us represent each word as a vector of numerical attributes. We then train a network that uses these vectors to compute a prediction for a masked word in a sentence, given the word-vectors around it (its *context*). When this network is trained, each word-vector's attributes are encoded such that similar words map to similar vectors.[57] Figure 7 illustrates 3-dimensional word-vectors produced by a trained network and their relationship to tangential words.

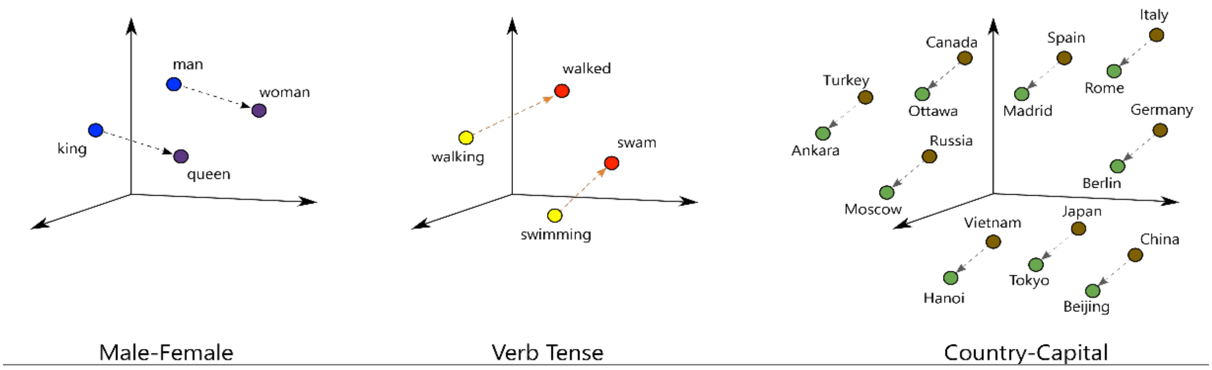

Figure 7 - Word-Vectors: score of meanings and plotting in latent space

As we have discussed, there is real flexibility in the way meaning can be baked into encodings. All networks encode their inputs in the intermediate layers, an encoding can take on different meanings depending on how the input data, network architecture and loss function are defined. The way we set up the network can radically change the meaning of an encoding; networks condense and compress inputs to minimise the loss function and self-improve.

This compression of data forces the network to assign meaning to information; this meaning is comparable to understanding, without which the artificial creative process

[57] Tomas Mikolov and others, 'Efficient Estimation of Word Representations in Vector Space', 2013.

would not be possible. Understanding is key to the preparatory and evaluation stages of creativity: firstly understanding the raw data and secondly, using meaning, judging an idea or artefact's value. The way encoding stores meaning in lower-dimensions is coincidentally how the human mind also holds information[58]

2.2.3 What Types of Neural Networks have the Most Creative Potential and How Do They Function?

As covered above, NNs' structural similarities to natural neural networks give them the most potential for creativity of the different types of AI. However, particular NN architectures and combinations are considered to have more creative potential than others.

In any NN, neurons are organised into at least three layers: input, hidden and output layers (see Figure 4). There are multiple possible connection patterns between the input and output layers, which fall into two directional categories: feedforward networks and recurrent networks[59]. Feedforward networks connect unidirectionally and can include fully connected layers (each neuron connects to a neuron in the next layer) or pooling (a group of neurons connect to a single neuron in the next layer)[60]. Recurrent network (RNNs) can retain 'memory' from previous neurons by allowing connections between neurons in the same layer or a prior layer in a cyclic manner. RNNs can base their predictions on learned context, much like how we can create ideas to fit into a domain based on our existing knowledge of the context.

[58] Purkait.

[59] Purkait.

[60] Purkait.

A convolutional neural network (CNN) is a feedforward deep learning algorithm[61] that is the most widely used structure to analyse and process visual data[62]. CNNs take an input, assign importance to various features in the input and differentiates between varying features[63]. A network is convolutional when it contains convolutional layers, which extract features from an input image, see Figure 8. The network reuses neurons to recognise specific data patterns regardless of their location[64]; an example feature extraction network is shown in Figure 9.

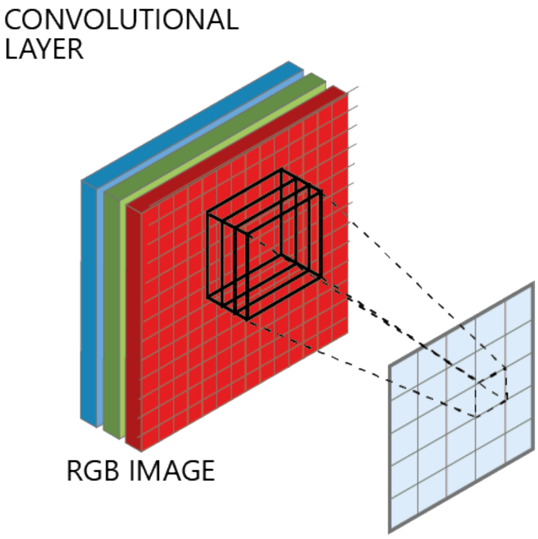

Figure 8 – Convolutional Layer

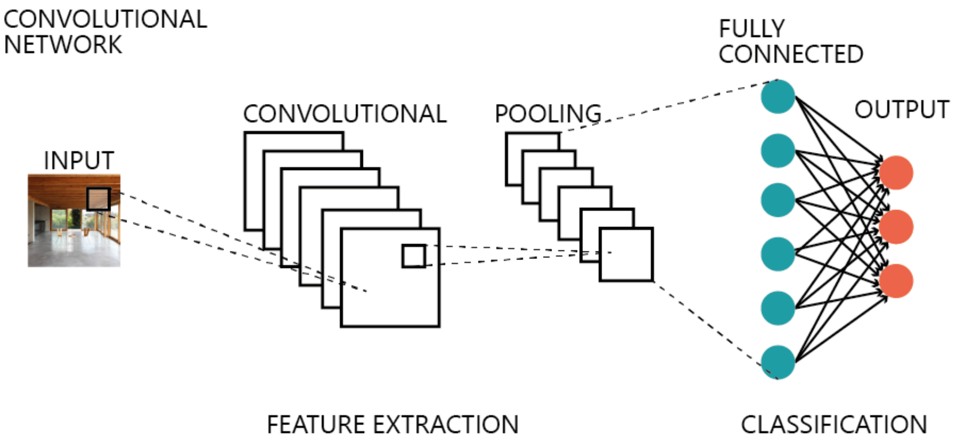

Figure 9 – A Feature Extraction CNN

[61] Purkait.

[62] Matthew D Zeiler and Rob Fergus, 'Visualizing and Understanding Convolutional Networks', 2013 <https://arxiv.org/abs/1311.2901>.

[63] Kriegeskorte.

[64] Purkait.

CNNs encode an image with a structure analogous to natural neurons' connectivity patterns; the inspiration for CNNs was partly the visual receptive fields and visual cortex structures [65]. These networks work by discovering patterns in images by way of edge detection, in order of increasing complexity[66], as shown in Figure 10.

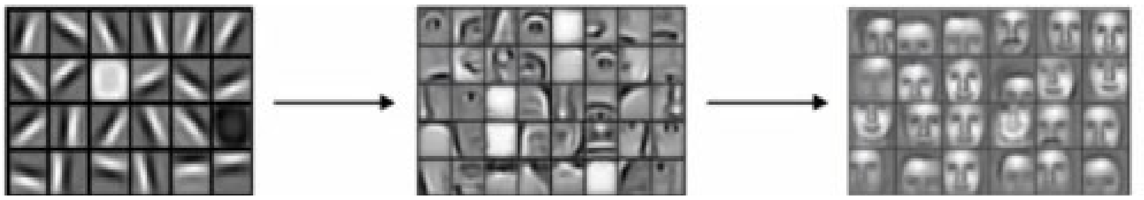

Figure 8 - Edge Detection

CNNs have been adopted for use by creative designs because of their success at visual outputs that can track, identify and describe objects. The outcomes of Google's Deepdream[67] (see Figure 12) are notable examples of artificial intelligence's creative potential. Deepdream uses CNNs to find and enhance images' patterns, incidentally, creating very imaginative artwork and drawing acclaim from scientists and artists alike. However, there is a notable lack of directionality or meaning in DeepDream's outputs.

[65] Kunihiko Fukushima, 'Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position', *Biological Cybernetics*, 36.4 (1980), 193–202 <https://doi.org/10.1007/bf00344251>.

[66] Andrew Ng, 'Machine Learning | Coursera', *Coursera*, 2013; Zeiler and Fergus.

[67] Alexander Mordvintsev, Christopher Olah, and Mike Tyka, 'Deepdream-a Code Example for Visualizing Neural Networks', *Google Research*, 2.5 (2015); Alexander Mordvintsev, Christopher Olah, and Mike Tyka, 'Iterative_Places205-GoogLeNet_buildings', *Google AI Blog*, 2015 <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>.

Figure 9–Neural Net Dreams – generated from random noise using Google Deepdream trained on buildings

For the artificial implementation of the creative process, we will be specifically looking at Generative Adversarial Networks; what this network could produce, the different components, and the network's inner workings.

2.2.4 What are Generative Adversarial Networks, and which Have the Most Architectural Creative Potential?

Generative Adversarial Networks (GANs) are a group of NNs that together form a cyclic process. There have been many recently proposed new GAN models, but to understand these, and how they are creative, we first need to understand their concepts and workings. First conceptualized in 2014 by Ian Goodfellow, GANs are a class of DNNs that simultaneously train two NNs: the generator G generates fake data based on the data distribution, and the discriminator D tries to discern whether G's fake data belongs in the data distribution[68]. Through this competitive joint training process, both the generator and the discriminator learn to perform their roles better- hence the term

[68] Ian J Goodfellow and others, 'Generative Adversarial Networks', 2014.

adversarial[69]. G aims to maximise D's probability of making a mistake; the generator learns to produce more accurate and convincing outputs while the discriminator becomes more astute at discriminating between real and generated outputs (see Figure 13 for the GAN's structure).

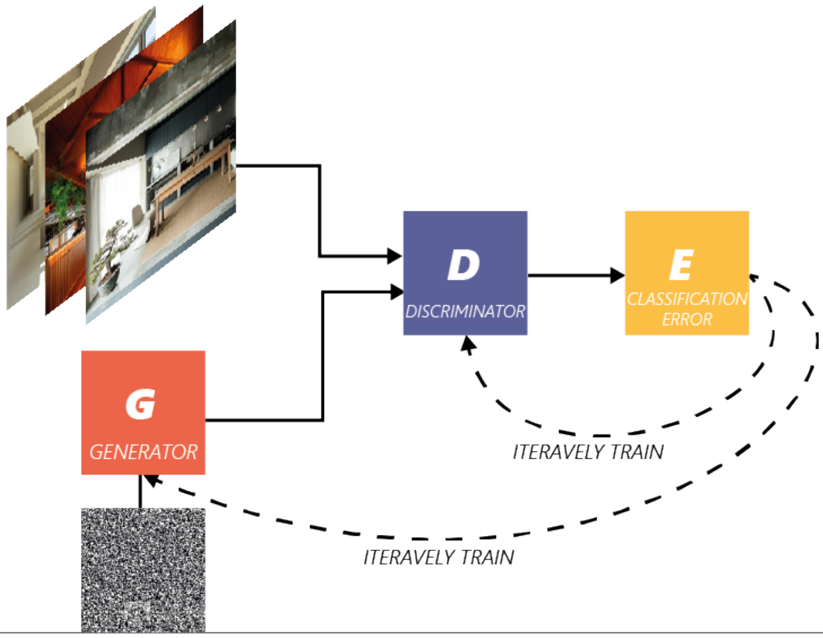

Figure 10– GAN Model

Once trained, only the generator is retained, which is then used to generate the final image: the overall purpose of the model is to create an image that fits into the training data[70]. The training process of GANs is conceptually like that of the architectural design process; we design and then critique, improving as we develop our work.[71]

GANs are used in several generative design frameworks to generate a creative or imaginative output[72]. The GAN generated images' novelty and creativity is down to the randomness, meaning it will always produce a novel image. GANs are superior to other algorithms as they can 'understand' what is in the image through vectors and semantic

[69] Jakub Langr, *GANs in Action Deep Learning with Generative Adversarial Networks* (Manning Publications, 2019).

[70] Langr.

[71] Langr.

[72] Purkait; As, Pal, and Basu; *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Cham: Cham: Springer International Publishing, 2019), <https://doi.org/10.1007/978-3-030-28954-6>.

meaning; after all, novel, valuable ideas only come through understanding and knowledge. The creativity of GANs can be classified as combinational and exploratory creativity, as it combines ideas from the data in new ways to create a novel image in an attempt to fit into the dataset and explores the conceptual space of the dataset. This is much like how the architect combines ideas from different buildings and precedents to create a novel building.

Numerous GAN models should be taken into consideration when attempting to recreate artificial creativity with an architectural application. Image-to-Image Translation with Conditional Adversarial Networks[73] is one example, where labelled and colour-coded images result in photo-realistic images (see Figure 13). However, this is a network that takes an existing visual input and changes the graphic style; functioning as a visual translational tool rather than an example of artificial creativity. CartoonGAN[74] is another example of a GAN with creative potential (see Figure 14), but it applies a learned style to an existing image. Rather than inputting a ready-made visual to transform, we want to give the network an abstract description or brief from which the model will formulate its idea or artefact.

A few models fall into this category, MirrorGAN[75], StackGAN++[76] and AttnGAN[77], which all use image-to-text generation to create semantic consistency and smooth the generative processes. AttnGAN is the best of these models; StackGAN++ is more focused on realism, and MirrorGAN exceeds domestic computing constraints as it uses redescription. AttnGAN focuses on semantic consistency, ensuring that the input brief is reflected in the final

---

[73] P Isola and others, 'Image-to-Image Translation with Conditional Adversarial Networks', in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–76 <https://doi.org/10.1109/CVPR.2017.632>.

[74] Yang Chen, Yukun Lai, and Yong-Jin Liu, 'CartoonGAN: Generative Adversarial Networks for Photo Cartoonization', 2018 <https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf>.

[75] Tingting Qiao and others, *MirrorGAN: Learning Text-to-Image Generation by Redescription*, 2019.

[76] Han Zhang and others, 'StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP (2017) <https://doi.org/10.1109/TPAMI.2018.2856256>.

[77] T Xu and others.

image/idea. This visual representation of meaning is a far more challenging criterion for creativity than realism. It ensures that the output is not only new and surprising but valuable and meaningful in its context.

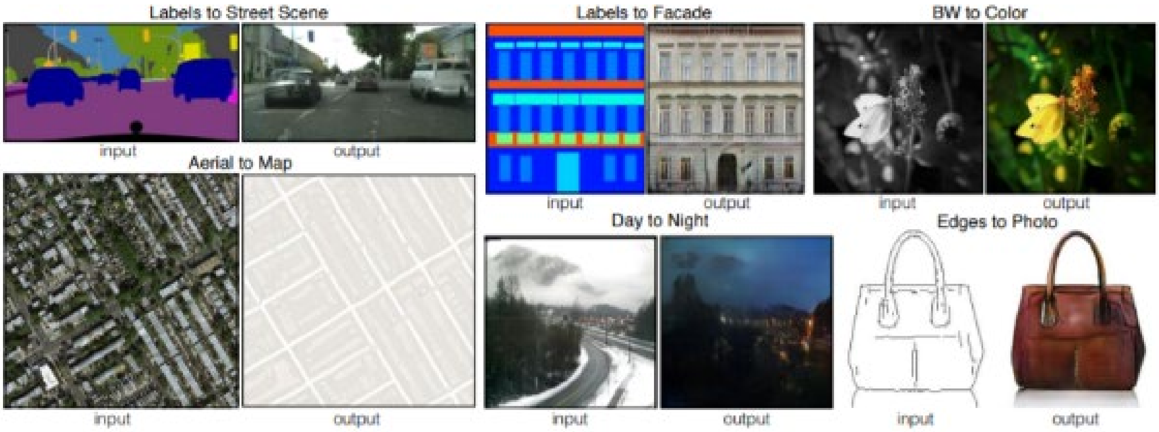

Figure 11 - Image-to-Image Translation with Conditional Adversarial Networks

Figure 12 - Cartoon GAN

3 Aims & Methodology

3.1 A Criteria for Artificial Creativity

As established above, GANs are the type of NN with the most creative potential. Suppose a generator generates an image that can convince the discriminator it is real. Could it also convince a human that it is real? Could this idea pass through the subconscious filter into the conscious realm as the output? Usually, when assessing AI's performance, it is measured against humans' results as a 'gold standard[78][79]. Consequently, a good measure for artificial creativity would be comparing it to human creativity; how does artificial creativity perform in both process and output compared to a human?[80]

However, we need to return to creativity's basic definition: to create something novel and valuable. As previously established, is it easy for a GAN to create something novel, but to make sure it is something of value is more complicated. There must then be a part of the model that tests and ensures that the idea has value at the verification phase.

As architects, we have looked at the philosophical, psychological and social domains regarding creativity, understood what is defined as creativity and brought those over without comment or question to see where those definitions fit into or are capable of being applied to the object of study (see Table 1). From this table and our analysis of neural networks, we form the criteria to assess the artificial creative process:

1. To what extent does the machine appear to be creative in its output?
2. Do the stages of creativity mimic those of human creativity, and what type of creativity is it?
3. To what extent did humans aid this process?
4. Is the machine recognising and determining its own creativity and value?

---

[78] Daniel Jurafsky and James Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2008, II.

[79] Lan Huang and others, 'Learning a Concept-based Document Similarity Measure', *Journal of the American Society for Information Science and Technology*, 63.8 (2012), 1593–1608 <https://doi.org/10.1002/asi.22689>.

[80] Anton Oleinik, 'What Are Neural Networks Not Good at? On Artificial Creativity', *Big Data & Society*, 6.1 (2019), 2053951719839433 <https://doi.org/10.1177/2053951719839433>.

5. How does the output fit into the domain?

This dissertation will attempt to explore these criteria by running an experiment: using a GAN to create original spatial image interpretations of ideas and semantic meaning. By creating original images that have meaning to the field, the GAN will show novelty and value, hopefully evoking surprise from the field. The GAN's workings will be explained to interpret and discuss the results thoroughly alongside the criteria established above.

3.2 Method Overview

This research aims to understand and measure the current potential for computational techniques and methods to produce creative outputs in an architectural sense by assessing artificial intelligence's creative ability in the domain of architecture. GANs have been shown to produce creative outputs in the sciences and the arts: they are demonstrably the most creative neural network[81]. GANs are vastly superior to other techniques of synthesising data[82] and have had substantial recent advances, as exemplified in Chapter 2.2. Out of the potentially applicable GAN variants to explore the criteria for artificial creativity, the AttnGAN model is most suitable, due to its ability to generate realistic images from text-based inputs, while enforcing semantic consistency between the input and output.

We undertake an experiment where the object of study is a NN; we only look at, apply, and analyse one type of GAN to understand artificial creativity's potential. The experiment does this by creating and training a machine learning model to quantitively transform abstract, text-based descriptions of spatial qualities into visual outputs in the form of images, which will be qualitatively interpreted. We will later discuss whether the image outputs or the process itself fall into any aspect of creativity. On its own, the study will not prove that artificial creativity is statistically or generally applicable in architecture, as there are not yet enough studies to show general applicability.

---

[81] Ahmed Elgammal and others, 'CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms', 2017.

[82] Langr.

The purpose of GANs is to mimic a target distribution, so we must use a conditional GAN like AttnGAN where the inputs fed into the model will predictably affect the type of output in a valuable way to the field. The input and output's semantic consistency gives the output value and direction, guiding the generative process in a particular direction that is *conditional* on the input.

To ensure conditionality, AttnGAN utilises text-to-image generation(T2I). As text and images are in entirely separate realms, the way the two can be mapped is interesting and challenging; enabling outputs to take on more surprising creative forms, in contrast to stylising or mimicking existing images. To give the GAN more independence, we move away from the supervised learning and hand-labelling of data used in AttnGAN, towards unsupervised learning. We want the GAN to make its own relatively objective interpretation of space, rather than humans making subjective assumptions about what AI would understand.

Specific to our aim and dataset, it is essential to predict and understand how a GAN independently interprets space. The abstractness of space influenced our choice of using unsupervised learning. Space is challenging to put into words and define, as it is interlinked with feeling and emotion; therefore, the way we approach it mathematically should not be limited to our vocabulary. However, there is a noticeable hierarchy of space in architecture going from global features, structure and shape, to local features, details, openings, and furnishings. We can utilise this concept to define our space, the workings of this will be explained in the following section.

The trained network produces an image that is both realistic and semantically consistent with an initial description given. By creating a system with relatively unconstrained parameters, the system's output should be independently novel and authentic[83]: producing a more abstract form of creativity that we rarely see from computers. We will discuss how manipulating the inputs between two different outputs allows us to interpolate between them, to visualise spatial and stylistic understanding.

[83] Langr.

The image outputs produced will then be analysed against the creative criteria defined above, to conclude if the image produced meets any of these criteria.

3.3 Training the GAN

Creating a working GAN to output images of architectural space was an iterative process using state-of-the-art technology that took many attempts, techniques and datasets. The first natural step was to construct and reorganise the AttnGAN model to make sure it worked for the original bird dataset, as demonstrated by the original paper[84]. The first architectural dataset was a hand-picked collection of about 1000 spatial images from ArchDaily that were manually labelled with abstract words describing the space; this took a significant amount of time to put together. However, space is subjective and challenging to describe qualitatively; the dataset's small size and actual abstract words are the hypothesised reasons why the model did not train appropriately using this initial dataset. Pressured by the time constraints and the need to objectively label space, two unlabelled datasets were found online, for building façades[85] and for bedrooms[86]. A clustering model was created to automatically and objectively label the datasets. The AttnGAN model produced interpretable image outputs for these two datasets; however, the façades dataset was much more successful than the bedrooms dataset (see Appendix 3.1). This difference in output is most likely due to the clear object focus of the building outline and structure compared to the spaces, where it is hard to know what the GAN is focused on generating.

3.3.1 Training Input

The final images for the datasets used for the training process were web-scraped from ArchDaily and then manually sorted by a set of criteria, which has been developed by trial-and-error testing. For the spaces dataset, the images were sorted by a clustering

[84] T Xu and others.

[85] Z Xu and others, 'Architectural Style Classification Using Multinomial Latent Logistic Regression', in *ECCV*, 2014.

[86] Yu, F, Y Zhang, Shuran Song, Ari Seff, and J Xiao, 'LSUN: Construction of a Large-Scale Image Dataset Using Deep Learning with Humans in the Loop', ArXiv, abs/1506.03365 (2015)

algorithm, which groups similar photos, allowing us to remove all the floorplans, sections and other non-realistic drawings form the dataset automatically. This dataset is then sorted manually to remove any whole building, material and detail images and further purge any photographs not abiding by the below criteria.

- The image must be taken with a straight angle to the wall, rather than looking down, to avoid perspective distortion.
- The image must depict a space, two of the following must be visible: walls, roof or floor.

However, these criteria were not strict enough, resulting in very abstract and unrealistic images of space from the AttnGAN model, so further criteria were needed:

- The image must depict a space with all the following visible: walls, roof and floor.
- The space depicted must be minimally furnished to clarify the focus of the image.

The landscape images will be cropped to two square images, and the portrait images will be centre cropped. Then each image will be flipped: thus, quadrupling our dataset size through data augmentation.

If manually labelled, the experiment results would be susceptible to prior assumptions about space; the type of descriptions corresponding to space is subjective to the labeller's interpretation. The words mean and symbolise something, they embody a particular idea, and it is this meaning that the GAN ultimately needs to understand for artificial creativity to occur.

As such, we use a clustering algorithm to overcome human bias in labelling the images. Each image in the final image dataset is assigned seven word-vector descriptions to create the text and image dataset required; these descriptions are not real words, but they carry semantic meaning within our dataset. The clustering process can be described as followed: each image in the dataset is clustered seven times, and each time one word-vector is assigned to each cluster. After each clustering iteration, the total number of clusters halves from the previous number, and therefore the number of images within each cluster increases. The smaller groups have both global features, like size and shape, and local features in common and the bigger groups have less in common but have similar global features. These images and seven word-vector pairs are our training inputs for the AttnGAN. We also input a single sentence vector that summarises the seven words (see Appendix 2.5 for the full clustering algorithm).

3.3.2 Process

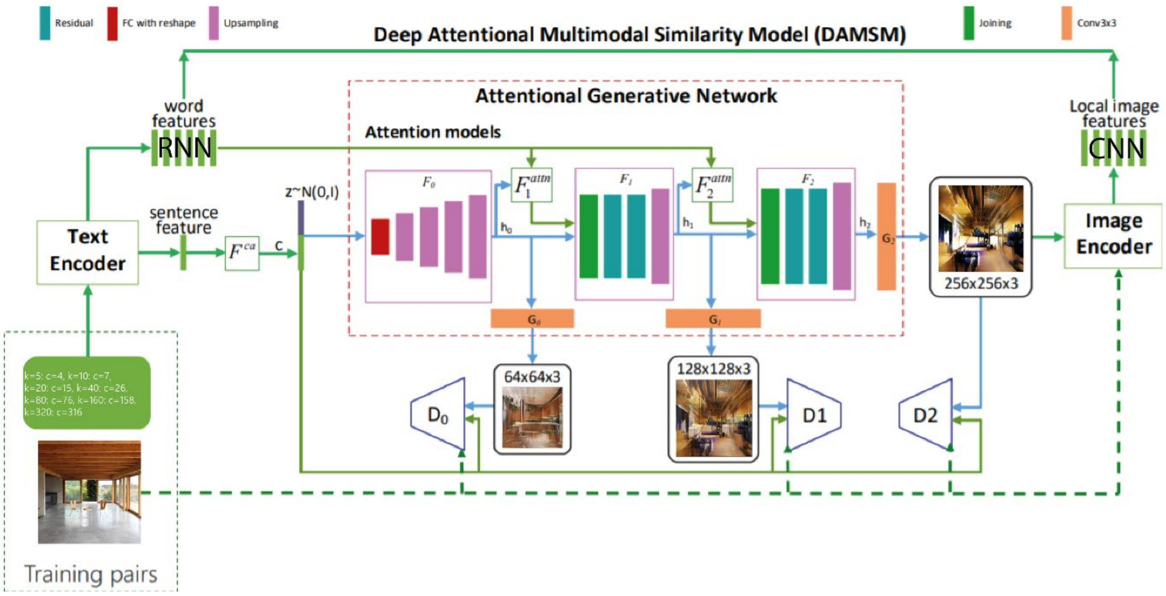

Figure 13 - Adapted AttnGAN Model

We first encode the text description of each image into word-vectors and a sentence-vector, then concatenate the sentence-vector with a randomly generated noise vector (z), which ensures variability and to an extent novelty in the image outputs. These are then fed into the Attentional Generative Network (AGN), which progressively produces higher resolution images. We feed these images to the three separate discriminators at three particular resolutions (64x64, 128x128, and 256x256 pixels). These predict whether the image is real or fake (*D1, D2 and D3),* and the prediction values are backpropagated to update the parameters of both Discriminator and Generator. Between each jump in resolution, the AGN uses *attention models* to influence the next generated output (Figure 13). Much like how humans can pay attention to particular areas of an image during interpretation, the attention mechanism allows the network to learn the alignment between sub-regions of images and those images' text descriptions.

The final component is the Deep Attentional Multimodal Similarity Model (DAMSM) which is key to computing semantic consistency (see Appendix 2.4). The DAMSM does this by re-encoding the highest resolution generated image into a collection of vectors and computing alignment between these vectors and the word-vectors from the original description (see Appendix 2.3).

In essence, by allowing attention to play a role in forming images at each resolution – and using a loss-function that critiques the alignment between generated images and descriptions – the network provides the pathways for backpropagation to produce weights. These weights will ensure consistency with the input word-vectors and their images (see Appendix 2.2 for further details). The "realism" losses from the three Discriminators in the AGN and the "semantic consistency" loss from the DAMSM combine together to form the network's overall loss function. Therefore, in minimizing this loss through the training process, the network is forced to learn weights that encourage realism and semantic consistency in its images throughout the generation process.

3.3.3 Output

When we have our final output-image based on the word-vector inputs, we can use interpolation to manipulate the image output towards different building and spatial styles. Interpolation works by swapping the word-vectors in the inputted captions. By swapping the word-vectors out with those from another image, it is possible to change and morph the building or spatial style, to appear more like the other image. For example, changing the global variables will alter the building or space's size and shape, and changing the local variable will change the local features and details of the building.

4 Results and Discussion

Though the final GAN model creates novel images, defining whether a particular output or the process of the GAN model is creative is another matter entirely. After seeing the output of our GAN model, the extent to which it is creative in its' output and process will be measured by the creativity criteria established in Chapter 3.1 and the interpretation of output data. We will look at the GAN results from different datasets and analyse them before looking back at the theory and understanding the artificial creative process.

4.1 Interior Spaces

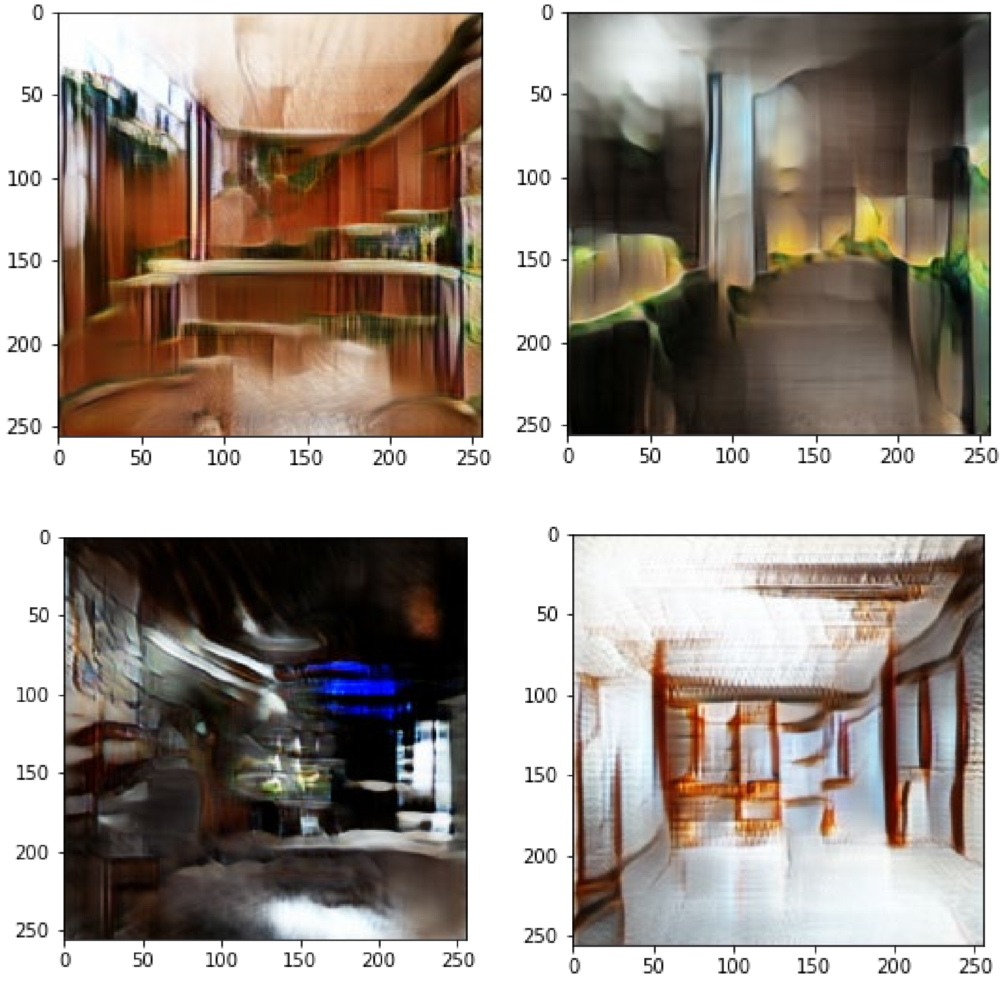

Figure 14 - Initial Spaces Dataset Outputs

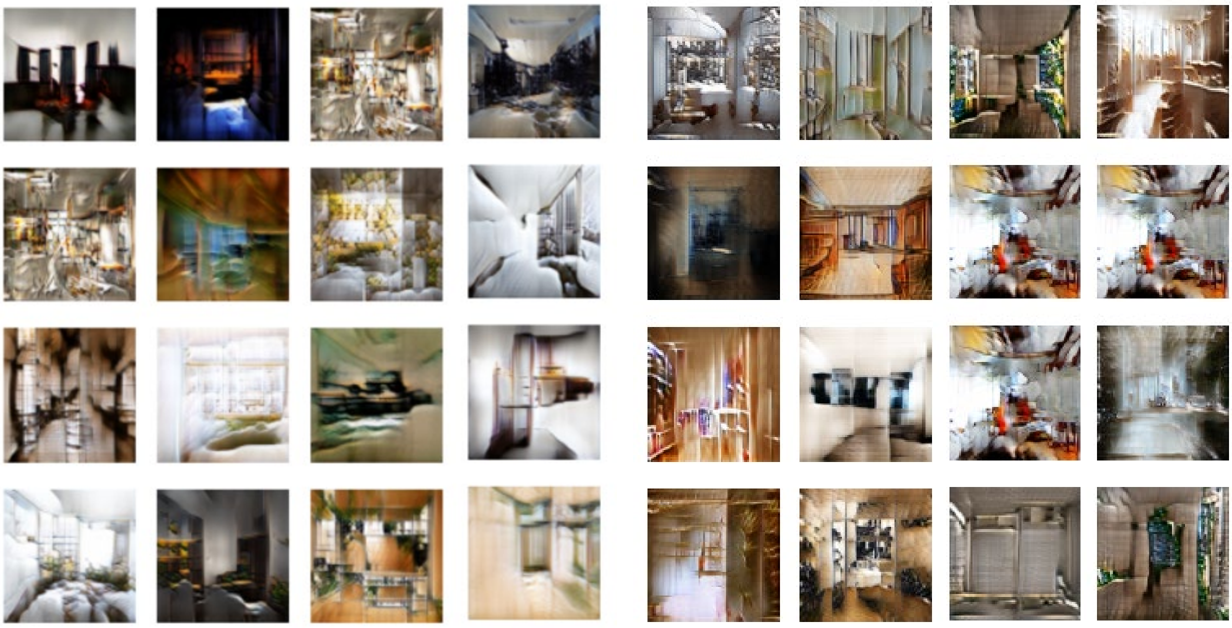

Figure 15 - Spaces Data-set initial Outputs in Grid

The initial spaces dataset outputs (see Figure 16 and Figure 17) were successful but abstract, leaving much room for interpretation. The last epoch (when the loss function indicated the model was done) indicates that the GAN is focused on particular forms and perpendicular geometric shapes, as shown in Figure 16. The GAN likely picked up on the lines between walls, floors, ceilings and openings, as well as some furnishing and simplified them.

Because creativity is to an extent subjective, as defined by the philosophical definition of creativity, the author has acted as the arbiter as to whether something is creative or not. The author felt that the outputs from this initial space's dataset were flat and uninteresting from an architectural perspective. The initial dataset outputs fulfil part of the definition of creativity, but they were not felt to be valuable from an architectural or aesthetic viewpoint. Additionally, the purpose of a GAN is to fit the distribution; however, the dataset's output does not appear to do so, indicating that the dataset size was too small.

Figure 16 - Final Spaces Dataset Outputs - randomly generated outputs from different training epochs

The effect of the larger dataset was quite drastic. The resulting outputs show the emergence of 3D forms, depth and complex forms such as stairs, doors and bookcases (see Figure 18). The images are much more interpretable as space and fit the distribution better. As can be seen, the GAN's ability to better fit the distribution exhibits the preparation phase's effects on the creative process's success; the GAN can draw from more information and knowledge. Much like how an architect with years of experience can design a much better building that a first-year architecture student. As this output is more valuable for architectural applications, the author regards it as more creative than the initial spaces dataset's outputs.

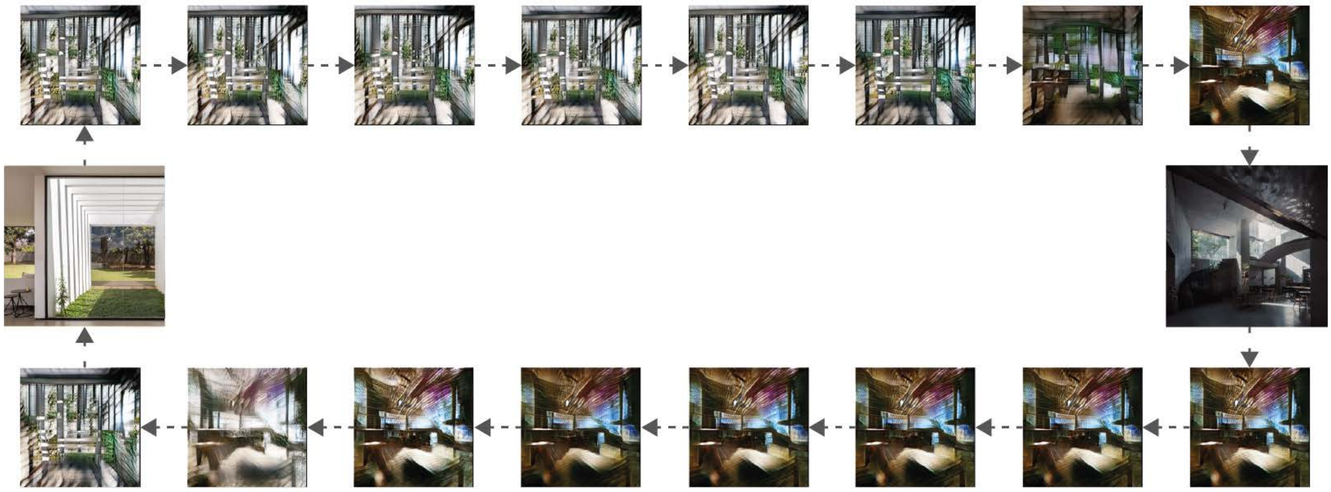

Figure 17 - Interpolation of Final Spaces Dataset

The interpolation images above show the progressive shift in image output produced by swapping word-vectors in the two captions. The interpolation of the space's dataset exposes the weaknesses of identifying space and aesthetics rather than objects. Like most DNN models, the clustering algorithm is trained on objects, so it struggled to cluster and identify global features. The weakness of clustering spaces is evident in the lack of changes when swapping captions in Figure 19.

The interpolation allows us to control the GAN and manipulate the output. From the same 10,000 images, by swapping the word-vectors, the machine could produce completely different solutions from the input images and something that a human designer could not produce. This ingenuity and manipulability gives the output value and novelty, surprising the author and providing a solid claim to creativity

4.2 Façades

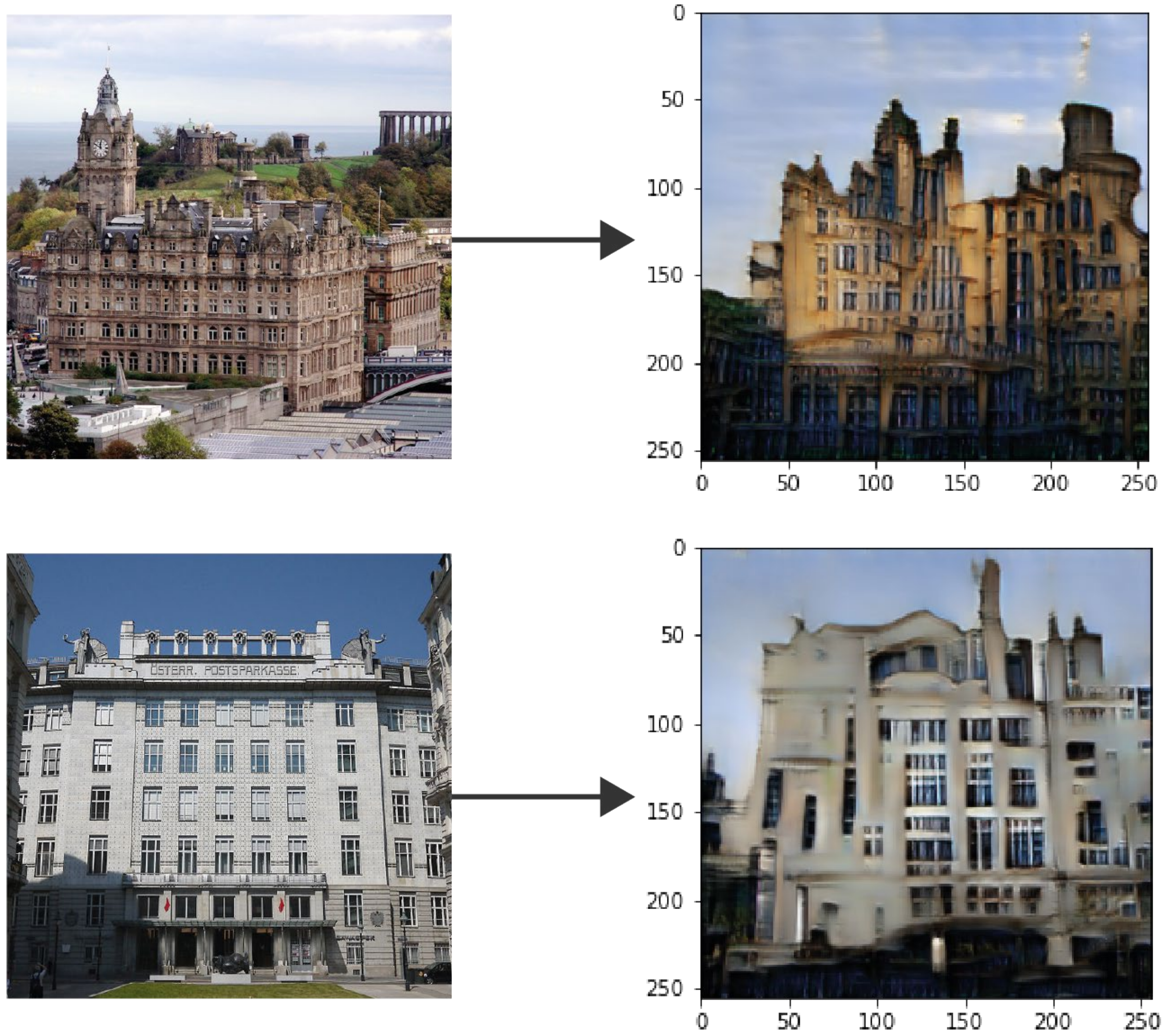

Figure 18 - Large Images of Façade dataset Outputs

The easily identifiable façade outline made the building façades dataset's output images even more realistic that the spatial outputs, showing both conventional and unconventional architectures. Some novel and exciting buildings are shown that are surprising and of value, like the later spatial images; these are genuinely creative outputs.

Figure 19 - Façade dataset output in grid form (see further images in Appendix 3.4)

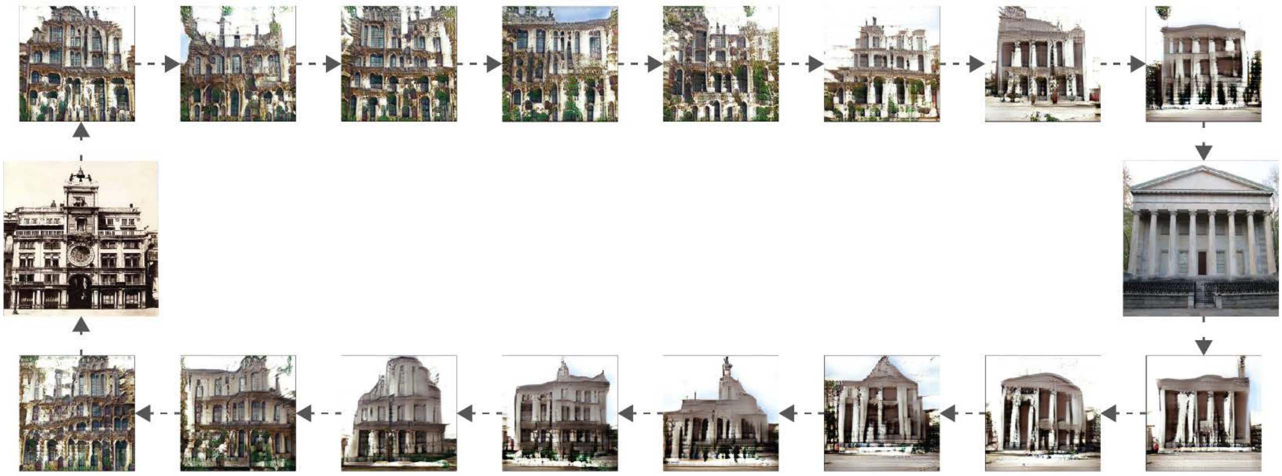

Figure 20 - Interpolation of Façades 1

Figure 21 - Interpolation of Façades 2

The hypothesised reason why the façades dataset performed much better than the spaces dataset in interpolation is since the clustering algorithm is trained on objects. Hence, façades are much easier to identify. Again, with interpolation, the creative value comes from being able to specify the type of façade output. Interpolation allows combinational creativity that combines and merges different architectural styles in unexpected ways.

Although the images are valuable against the authors' spatial reference and give some insights into how the input images can be combined to produce different spaces, the value is in being able to specify the type of spatial output; the GAN's use of text-to-image generation gives the output more creative value.

4.3 Creative Output

As the example images above show, the GAN will usually output novel images due to its stochastic nature, but whether the outcome is surprising or of value is another question that is key to determining the extent of the GANs creativity. We need to understand how the model interpreted the data and what outputs to expect. The data pairs fed into the model explain architectural space to a neural network; this way, the meaning of architecture and space is left to the model to interpret. Although we initially started with word-vector inputs defined by humans, allowing the model to independently interpret the image was a more natural and successful approach. To understand how the model interprets space, humans must try to understand space through the lens of the machine; using object identification and semantic meaning.

Classical or symbolic AI performs well in reasoning and can help us interpret real-world objects; symbols represent meaningful objects. Perlovsky and Ilin use the term situation-symbol in this regard. For example, the 'situation "office" is characterised by the presence of a chair, a computer, a desk, a book and a bookshelf'[87]. The same would theoretically be true of an architectural space, but what it is characterised by is less clear. 'The principal difficulty is that irrelevant objects are present in every situation'[88]; one issue of spaces as situations is that we may perceive multiple objects, only some of which are relevant to a situation, leading us to conclude that the objects in the situation should not define the space. We must focus on how to visually interpret objects and ideas; spaces need to be characterised by something, in the case of GANs, this is the distinct lines and edges that form the image.

Rather than focusing on the quantitative material qualities of buildings, which are much more straightforward for computation software to understand, we have instead directed our focus to space's immaterial qualities. As Louis Kahn said, "Architecture is the thoughtful making of space"[89]; space is integral to our understanding of architecture and our experience in the world as humans. The outputs appear to exhibit a preliminary understanding of space and form, generating outputs that were not just manipulating pre-existing data but instead exciting and creative spaces and façades.

---

[87] Leonid Perlovsky and Roman Ilin, 'Brain. Conscious and Unconscious Mechanisms of Cognition, Emotions, and Language', *Brain Sciences*, 2.4 (2012), 790–834 <https://doi.org/10.3390/brainsci2040790>.

[88] Perlovsky and Ilin.

[89] Louis Kahn, 'Architecture Is the Thoughtful Making of Spaces', *Perspecta*, 4 (1957), 2–3.

4.4 The Artificial Creative Process

It is crucial to understand the specific artificial creative process of the AttnGAN model in our experiment. For this analysis, we must compare the artificial creative GAN process to the Wallas model and human creativity to categorise how and what type of creativity occurs and determine its interaction with and its place in the Systems model.

The artificial creative process of the AttnGAN model can be put into the context of the Wallas Model relatively easily. The preparation stage happens in the clustering model and the text encoder, where the data is being processed and understood in meaningful terms as word-vectors. Thaler discusses artificial creativity driven by neural networks and sees the source of its creativity in perturbations and random noise[90]; like the random subconscious stages of the Wallas Model. This randomness was visible in the model's outputs; the facades were hugely varied despite an exact input vector.

There are two scales at which one can see parallels with the creative process: at a micro-scale and a macro-scale. On a micro-scale, the model is going through constant creative cycles using forward and backwards propagation, continually learning and verifying the 'ideas' using backpropagation. At a macro-scale, the incubation stage can be seen as occurring in the AGN; the attention models each generate an image from a randomisation of pixels, and quality check the meaning of the image to make sure it has value. The verification phase happens at several parts of the model: the discriminator, which checks the model is progressing well, and the DAMSM, making sure it has consistent semantic meaning or value. When the idea surpassed all these thresholds, it is much like an idea passing the consciousness threshold into conscious thought. We can also see a final verification stage occurring externally to the model; the human ensures the output is novel, valuable and surprising, and worthy of entering the human domain. Here, the author has acted as the arbiter and decided the final outputs were worthy of being shared further. Interpolation can be seen as a further part of the verification stage, as an elaboration into understanding an idea, manipulating it in the conscious

[90] Stephen L Thaler, 'The Emerging Intelligence and Its Critical Look at Us', *Journal of Near-Death Studies*, 17.1 (1998), 21–29 <https://doi.org/10.1023/A:1022990118714>.

mind. Despite the internal thresholds in the model, it is clear the model cannot function independently of its human creator.

4.5 Whose Creativity is it?

The human aids the preparation phase by finding and preparing example data or precedents and programming the machine to understand and interpret this input. By designing and developing the model, humans have aided this process and inevitably the outcome; however, evolution also 'designed' our neurological processes. The critical difference is that humans have free will to choose what to research and a vast dataset of memories and experience to use in the preparation phase. However, the Wallas model's middle two stages are a black box for both human and machine. Humans' most significant influence on the machine's idea is finding the training data and deciding parameters to feed into the network. However, the interpolation of the images shows the humans can dictate the type of output even further. For both natural and artificial creativity, humans are in charge of the preparation stage, and their design decisions decide the type of outcome.

The incubation and illumination stages of artificial creativity are mostly independent processes. In these stages, the model's random seed has a considerable influence on the output. The human's only role here is to ensure the parameters are set correctly and ensure the model is adequately trained on good data to produce a valuable output. In deciding what model and what training data is used, humans set the standard for a creative idea and value. While it is guided, the machine, much like our subconscious, creates the final creative idea somewhat independently. The final verification stage is a collaborative process; by deciding the number of epochs (training time) the human chooses how well trained the model is. By choosing the input word-vectors, the human is deciding the type of spatial output. These different factors mean the GAN's potential outputs are infinitely variable, and, due to the stochastic nature of NNs, only partially under the human's control.

As all that is digital is variable, architecture's digital reincarnation means that questioning the identicality and authorship of design is more critical than ever. This line of questioning is crucial when discussing semi-independent algorithms like neural networks, where the goal is to avoid identicality all together. 'Insofar the objectile is, technically, an open-ended algorithm, and a generative, incomplete notation, the objectile's designer will "authorize" some general norms to determine aspects common to a range of variable and individual events'[91]. This understanding of algorithms reflects the outputs of our experiment. The human designs and directs the system as to what kind of output is required; therefore, the outcome is still under human authorship. The artificial creative process is blended with the natural creative process and becomes a co-creative process between human and machine; however, the human is still holds much responsibility and is ultimately in control.

The artificial architectural creative process should roughly follow the natural creative architectural process; however, as has been argued, there is even more collaboration. The key to artificial creativity functioning is this collaboration; the sharing and exchange of skills, data and ideas between human and machine.

4.6 Purposeful Creativity

We must ask whether the GAN was able to recognise and determine its own creativity and value. The discriminator ensures realism in the image, while the DAMSM computes the outcome's accuracy to the input words, making the output valuable to and controllable by humans and the field. There is currently no quantitative method to determine whether the output is novel, surprising or valuable by such subjective criteria and for such a varied field.

---

[91] Mario Carpo, *The Alphabet and the Algorithm*, 1st edn (The MIT Press, 2011).

Neural networks have inherent limits; the reliance on linear and logistic regression limits neural networks' possibilities and abilities[92]. They are only good at identifying patterns if the training data has a structured character. They cannot challenge and rebel against the domain in the same way humans can as their limitations derive from their foundational principles[93]; they are programmed to follow the rules we set.

Creativity is the opposite of merely training a network; it undermines the 'normal' operations of neural networks by creating something novel rather than finding and abiding by existing data patterns.[94] Due to the image generation nature of the GANs, the type of ideas they can come up with are limited to the data that has been fed into them in training; like humans, any novel idea comes from existing memory, data and research. Although interpolation highlights the similarities when specific word-vectors are asked for, the outputs are substantially different and creative. The architectural limitations on creativity alongside the outputs lead to the conclusion that GANs exhibit combinational creativity. While the GAN must detect the data's rules, it is unlikely to understand the structural and physical rules by which real-life architecture must abide. Its' ideas and outputs are arguably less inhibited by real-world constraints and more explorative than those of the human architect. This is especially visible in the façade outputs, where some outputs seem to defy gravity, with parts of the buildings appearing to float in the air (see Figure 21).

The creativity exhibited by the GAN is both combinational and exploratory; it combines ideas and styles learnt from and explores the existing conceptual space provided by its dataset. There is further combinational co-creativity occurring during interpolation, as the human swaps the word-vectors combining different ideas into new images. Some examples of combinational creativity are arguably also transformational; these terms are not mutually exclusive. There is no way for the human to predetermine that the idea the GAN comes up with will be transformational; just as there is no way for a human to decide to have a transformative idea. The machine or its creator cannot know how

[92] Oleinik.

[93] Oleinik.

[94] Thaler.

creative its idea will be in advance. However, a human knows in hindsight and can analyse their idea independently, whereas the machine can only measure how far this new idea is from the precedents; the machine cannot know whether it is original beyond its dataset.

Another discussion is to what extent is combinational creativity innovative. Creativity is not black and white; it is a spectrum[95] which we can measure using Boden's three criteria for creativity: new, surprising and valuable. However, a problem arises with artificial creativity versus natural creativity, the term "surprising". To be surprised is an emotion and as we well know, machines are not yet and may never be capable of emotion. Csikszentmihalyi argues that emotions are necessary for creativity, and "the [artificial] creative process does not include affect, motivation, and curiosity, and hence could not be said to replicate what goes on in the mind of a person confronting a problem creatively"[96]. The machine does not need to have affect, motivation or curiosity to analyse data and develop novel ideas; it generates novel ideas because that is what it is programmed to do. However, just because the machine may not be surprised and cannot measure its success does not mean that humans would not be able to do so or be surprised by the outputs. Surprise is arguably not a deciding factor of the creative process. However, surprise can still leave a creative impression on the observer, making them more inclined to see the output as creative. The GAN does not need to recognise its novelty or creativity, or be purposeful to be creative; creativity is judged by the field of the domain in which the output belongs.

---

[95] Dean Keith Simonton, 'Quantifying Creativity: Can Measures Span the Spectrum?', *Dialogues in Clinical Neuroscience*, 14.1 (2012), 100–104.

[96] Mihaly Csikszentmihalyi, 'Solving a Problem Is Not Finding a New One: A Reply to Herbert Simon BT - The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi', ed. by Mihaly Csikszentmihalyi (Dordrecht: Springer Netherlands, 2014), pp. 63–66 <https://doi.org/10.1007/978-94-017-9085-7_5>.

4.7 Entering the Domain

A key issue in assessing creativity is determining what counts as a 'valuable' novelty. NNs are easily trained to detect and appreciate novel patterns[97], but determining whether they are valuable is challenging. 'Value is not found by science [or arts] but negotiated by social groups'[98], so accepting novel ideas as valuable requires social action.

To find the model's place in the Systems Model of Creativity,[99] we must determine and assign who the individual is, what the domain is and of whom the field consists. Firstly, we have determined in Chapter 4.4 that creativity is a collaborative process; therefore, the co-creator and machine will be referred to as the individual in this systems model. Secondly, the idea exists in the architectural domain, as we are generating fictional architectural spaces. Therefore, according to Csikszentmihalyi, the field and the gatekeepers that must approve the creative idea for use in the domain are architects.

As we already know that the idea is novel, we are only trying to determine whether the outputs have value to and surprise architects. The individual interacts with the domain: the human is part of and has knowledge of the architectural domain, this knowledge is passed to the machine, interacting indirectly with the domain. However, the field's perception of the individual's ability and potential to succeed will influence the determination as to whether the individual's contribution will be accepted into the domain. The acceptance of the idea into the domain will depend on the field's perception of the individual: the author and AI. However, like most of society, architects are sceptical of artificial creativity as they romanticise the idea of natural creativity[100] and may be biased about the image outputs. AI is not yet developed enough to match human creative ability. AI's potential is undermined by humans and a society that believes that creativity is a fundamentally human endeavour. However, humans should

---

[97] Thaler.

[98] Margaret A Boden, 'Computer Models of Creativity', *The AI Magazine*, 30.3 (2009), 23 <https://doi.org/10.1609/aimag.v30i3.2254>.

[99] Csikszentmihalyi, 'The Systems Model of Creativity'.

[100] Boden, *The Creative Mind: Myths and Mechanisms*.

not have a monopoly on creativity, especially when AI evolves many degrees faster than natural neural evolution; the field may hold back its potential.

If the artificial spaces generated are valuable and surprising to the field, then arguably it will always fit into the domain. Further, as the output image is as subjective and as hard to describe as the input images were, one should argue that the perception of creativity is individual to each viewer; everyone has differing opinions about what they see as valuable and are surprised by different things. Creativity is subjective and, therefore, if the GAN's ideas or images are seen as valuable, novel and surprise a knowledgeable audience or even just a few actors within that audience; the outputs are creative as they have managed to re-enter and impact the domain.;.

5 Conclusion

From specific definitions for creativity, this work has shown that computers can be creative. The research aimed to understand artificial creativity's potential in an architectural setting, using DNNs to interpret and generate spatial architectural ideas. The modified AttnGAN model exhibited a creative process comparable to the natural creative process; however, this turned out to be a co-creative process with human and machine working together. Although there was no way for the machine to assess its creativity, the GAN exhibited exploratory and combinational creativity in its image outputs, which were novel, valuable, and surprising to the author. Based on these conclusions, the model exhibits artificial creativity and fits into the architectural domain as a creative tool, process and output. The research carried out illustrates the immense potential for implementing artificial intelligence in the architectural domain.

This dissertation's findings challenge the notion that creativity is a purely human endeavour. It shows that AI can follow the Wallas Model's creative process, fit into the Systems Model of Creativity and be a useful generative idea and design tool for architects and designers. Architects can achieve more working with AI, but the technology is not yet where AI can independently create useful outputs. Further, the results indicate that AI struggles to identify and portray abstract objects, such as space and atmosphere, visually and realistically. Nevertheless, the GAN's abstract interpretations and ideas are sources of inspiration and indicate a basic understanding of architectural space and the creative architectural process.

However, the current limits of this systems mean that the co-creator determines the start of the process and must feed in instructions, parameters and enormous amounts of data for the GAN. Beyond the input images, the co-creator was able to guide the process and control the aesthetics or feelings of space that she wanted: making the output more valuable to the design process. The model can take the same dataset and come up with very different final solutions depending on how the weights and parameters are set up.

At this present moment in time, we must conclude that there appear to be the inklings of the possibility of creativity coming out of AI. However, it is very much a co-creative process as computers are still not capable of analysing their outputs reflectively against the task they were initially set to do. On the other hand, humans cannot function with the speed, precision and scale of AI, drawing ideas from and making connections between thousands of images. The model has created something humans could not, and this can only be described as creative. How we as humans choose to interpret this output and take it further into our design thinking is up to us.

The research raised philosophical questions about the creative process itself, what it means to be an architect and designer in the 21st century, and what role AI will play in the future of architecture. Based on these conclusions, architects should consider the impact that AI will have on the future of the architectural domain and profession and think realistically about a future where we are designing using machines not just as tools, but as co-designers in our creative processes.

To better understand the implications of these results, future studies could address other DNNs and their potential creative applications in architecture and the potential for more qualitative generative design. Having concluded that the output can be considered creative, in limited aspect as what creativity is defined as, we must look at the practical uses for this model. In practical terms, an architect can build up a massive dataset to feed into the model, add specific images clients may want, and use these word-vectors to produce an image from which the client can indicate what kind of space they are looking for.

Word Count (excluding preliminary pages, footnotes and bibliography) = 10,954

6 Appendices

6.1 Appendix 1– The Dataset

Like the AttnGAN model, our dataset consists of two data types: colour images and word vector embeddings. This machine learning model's dataset consists of thousands of photographs of different architectural spaces. Each has a strong sense of identity and character that have been web-scraped from ArchDaily and then manually sorted (see Figure 14 Data Examples). These images will be tagged using our caption generation system to tag them with seven abstract word vectors describing the image from global to local. They will be cropped as 2 square images, and then each image will be flipped: thus, quadrupling our data set through data augmentation.

Some potential difficulties will be making sure the data set is varied and accurately labelled. The AttnGAN dataset consisted of 10,000[101] images of birds; this should be the maximum size our dataset will have to be. The façade dataset[102] was of about 7000 images, and the bedrooms dataset[103] was of about 10,000. A 2000 image dataset will first be used, the second image dataset iteration was of 6000 images, and the third was of over 10,000 images of space.

[101] T Xu and others.

[102] Xu, Z, D Tao, Ya Zhang, J Wu, and A Tsoi, 'Architectural Style Classification Using Multinomial Latent Logistic Regression', in ECCV, 2014

[103] Yu, F, Y Zhang, Shuran Song, Ari Seff, and J Xiao, 'LSUN: Construction of a Large-Scale Image Dataset Using Deep Learning with Humans in the Loop', ArXiv, abs/1506.03365 (2015)

6.2 Appendix 2 – Training the GAN

The deep learning model will use T2I generation by deep multi-modal similarity, primarily taking direction from the AttnGAN paper by Xu and others[104]. Text-to-encoding generation is used with image-to-encoding generation to create semantic consistency and smooth the generative processes. We will broadly follow the same approach, but we will be using different datasets that use abstract word-vectors instead of English text and interpolating the outputs.

The machine learning framework we will be using can be summarised in 4 stages, as shown in Figure 15.

6.2.1 Appendix 2.1 Text Encoder

i. The semantic text embedding determines our input words' meaning by turning them into word vectors that are abstract, specific to the dataset and do not map to real vocabulary; they are placeholder words that the GAN will use to map to meaningful embeddings. The text encoder takes a list of n words and embeds them into n+1 vectors (word vectors plus global sentence vector).

6.2.2 Appendix 2.2 Attentional Generative Network (AGN)

i. The AGN is a GAN that uses an additional attention mechanism to learn what sub-regions of the image correspond to what word from the text encoder in three stages, progressively yielding a higher resolution image.
    i. This attention mechanism gives meaning on a word-by-word, attribute-by-attribute basis, through word vectors and a summary or sentence vector. The first stage uses the sentence vector to generate a low-resolution image. The sentence vector is concatenated with a noise vector and then through a series of upsampling and convolutional layers, progressively reshaped from a long, dense vector into an image sized shorter vector. This short vector is initially a $2,2,k$ vector, where k is significantly larger and is progressively reshaped into a $64,64,v$ where v is significantly shorter than k (larger in the spatial dimension, shorter in the depth dimension).

[104] T Xu and others.

ii. Then the following two stages use word-vectors and an attention layer to form a word-context vector. These stages are interspaced by the attentional part of the model. Each subregion of the image produces a weighted context vector from all words in the corresponding caption representing a weighted average of how well each word aligns to this image region. The point of residual blocks is that they do not change the image's spatial dimensions; they merge words and image.
    i. The discriminator is the image encoder: three separate, isolated discriminators function independently at each resolution stage, from the input image they output a sigmoid number between 0 and 1. That gets fed back into the generator through a binary cross-entropy loss function.
iii. The Loss Function used was the Min-Max GAN Loss from Ian Goodfellow's Generative Adversarial Networks Paper[105]

$$\nabla\theta_g \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)$$

6.2.3 Appendix 2.3 Image Encoder

iv. The image encoder takes in final highest resolution output of the GAN and produces two embeddings: a regional feature embedding and a global feature embedding, from the ResNet pre-trained model) It is first reshaped to the same size, by matrix multiplication, to make sure GAN embeddings semantically aligned with the embeddings from ResNet.

6.2.4 Appendix 2.4 Deep Attentional Multimodal Similarity Model (DAMSM)

v. The DAMSM is pre-trained before the GAN to compute the similarity between the generated image and the sentence[106]. It works by aligning the word embeddings from the text encoder and the feature vectors from the image encoder: each image feature vector (17,17,289) is passed through a linear layer to make it $(17,17,k)$, where k is the same dimensionality as the word vectors in the caption of the image. For each 17x17 pixel vector, local word vectors get compared against local image vectors to compute a weighted context vector of all words in the caption resulting in a vector that is $1,1,k$. This concatenated image and word vector is computed for all image regions, and this process is then repeated across the batch of images.

[105] Goodfellow and others.

[106] T Xu and others.

vi. The loss function of DAMSM is constructed so that the loss is very low when the alignment between image 1 and caption 1 is very high, and the alignment between image 1 and any other captions is very low. This way, the DAMSM learns that image 1 needs its corresponding caption and not the others; this alignment between two different vectors and spaces results in semi-supervised learning.

   i. Additionally, a mask is applied to ensure that image 1 has a low alignment to every other caption; we must mask every image of the same class to make sure it does not operate on images similar to itself.

vii. The DAMSM Loss

   i. The probability of sentence D_i matching image Q_i is computed, where w stands for word, by dividing the similarity between sentence D_i and image Q_i by the similarity between a different sentence D_j and image Q_i.

$$P(Q_i|D_i) = \frac{\exp(\gamma_3 R(Q_i, D_i))}{\sum_{j=1}^{M} \exp(\gamma_3 R(Q_i, D_j))}$$

   ii. 'The negative log posterior probability that the images are matched with their corresponding text descriptions (ground truth)'[107]

$$L_1^w = -\sum_{i=1}^{M} \log P(Q_i|D_i)$$

   iii. This same equation is applied to L_2^w, L_1^s and L_2^s and DAMSM loss is defined as the sum of L_1^w, L_2^w, L_1^s and L_2^s, where s stands for sentence.

6.2.5 Appendix 2.5 Caption Generation

viii. To ensure that the image descriptions are not limited to our vocabulary, unique captions are generated by taking a pre-trained ResNet model and taking a late-stage average pooling layer that produces a 512-dimensional embedding for each image. Now all image features have vector representations; however, we need to compare how different the image's embedding is. To make sure the embedding of an image is comparable to others, we needed to overcome the curse of dimensionality - everything is too far apart and therefore, incomparable. This should be avoided by ensuring the embeddings were not too far apart. A dimensionality reduction

[107] T Xu and others.

technique, Uniform Manifold Approximation and Projection (UMAP)[108], reduces dimensionality and shrinks the embeddings into dense 64-dimensional image embeddings. UMAP is faster than other techniques and preserves the local and global structure and features well, encoding both the image's structure and features.

ix. To capture local and global spectra, agglomerative structuring was used to create bottom-up hierarchical clustering by taking the cosine distance between each embedding as a measure of similarity. This process, known as clustering, starts every image off in its own cluster, the closest clusters get connected first, with the clusters progressively get larger until we are left with one cluster. We stop this process at various defined levels and assign one word to each cluster at each stage (in our case, a maximum of 800 vocab and a minimum of 7 clusters found through trial and error). Finally, each image ends up with seven captions total; each word ranges from local to global each representing a cluster. Through training these vectors will take on meaning, this works because words are just strings of letters and as long as similar images have similar labels, the model will learn to map these to the right images.

[108] Leland McInnes and others, 'UMAP: Uniform Manifold Approximation and Projection', *Journal of Open Source Software*, 3 (2018), 861 <https://doi.org/10.21105/joss.00861>.

6.3 Appendix 3 – Extra GAN Output Images

6.3.1 Appendix 3.1 - Bedrooms Dataset Output

Figure 22 - Bedroom Dataset Output

7 Bibliography

As, Imdat, Siddharth Pal, and Prithwish Basu, 'Artificial Intelligence in Architecture: Generating Conceptual Design via Deep Learning', *International Journal of Architectural Computing*, 16.4 (2018), 306–27 <https://doi.org/10.1177/1478077118800982>

Boden, Margaret A, 'Computer Models of Creativity', *The AI Magazine*, 30.3 (2009), 23 <https://doi.org/10.1609/aimag.v30i3.2254>

———, 'Creativity and Artificial Intelligence', *Artificial Intelligence*, 103.1 (1998), 347–56 <https://doi.org/https://doi.org/10.1016/S0004-3702(98)00055-1>

———, *The Creative Mind: Myths and Mechanisms*, 2nd edn (Routledge, 2004)

Carpo, Mario, *The Alphabet and the Algorithm*, 1st edn (The MIT Press, 2011)

Chen, Yang, Yukun Lai, and Yong-Jin Liu, 'CartoonGAN: Generative Adversarial Networks for Photo Cartoonization', 2018 <https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf>

Cornillon, Ellen, 'The Quest for AI Creativity', *IBM*, 2016, p. 1 <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-creativity.html> [accessed 6 November 2020]

Csikszentmihalyi, Mihaly, 'Solving a Problem Is Not Finding a New One: A Reply to Herbert Simon BT - The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi', ed. by Mihaly Csikszentmihalyi (Dordrecht: Springer Netherlands, 2014), pp. 63–66 <https://doi.org/10.1007/978-94-017-9085-7_5>

———, 'The Systems Model of Creativity', 2014 <https://doi.org/10.1007/978-94-017-9085-7>

———, 'The Systems Model of Creativity and Its Applications', *The Wiley Handbook of Genius*, Wiley Online Books, 2014, pp. 533–45 <https://doi.org/https://doi.org/10.1002/9781118367377.ch25>

Csikszentmihalyi, Mihaly, and Keith Sawyer, 'Creative Insight: The Social Dimension of a Solitary Moment BT - The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi', ed. by Mihaly Csikszentmihalyi (Dordrecht: Springer Netherlands, 2014), pp. 73–98 <https://doi.org/10.1007/978-94-017-9085-7_7>

Csikszentmihalyi, Mihaly, and Rustin Wolfe, 'New Conceptions and Research Approaches to Creativity: Implications of a Systems Perspective for Creativity in Education', 2014, 81–93 <https://doi.org/10.1007/978-94-017-9085-7_10>

Dubost, Jean-Claude, and Jean-François Gonthier, *Architecture for the Future* (Paris: Paris : Terrail, 1996)

Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone, ‘CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms’, 2017

Emam, Maii, Dina Taha, and Zeyad ElSayad, ‘Collaborative Pedagogy in Architectural Design Studio: A Case Study in Applying Collaborative Design’, *Alexandria Engineering Journal*, 58.1 (2019), 163–70 <https://doi.org/10.1016/j.aej.2018.03.005>

Explainable AI: Interpreting, Explaining and Visualizing Deep Learning (Cham: Cham: Springer International Publishing, 2019), <https://doi.org/10.1007/978-3-030-28954-6>

Fukushima, Kunihiko, ‘Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position’, *Biological Cybernetics*, 36.4 (1980), 193–202 <https://doi.org/10.1007/bf00344251>

Goodfellow, Ian J, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, and others, ‘Generative Adversarial Networks’, 2014

Gropius, Walter, *Re:Education of an Architect*, Gropius Papers, 1939

Hadamard, Jacques, *An Essay on the Psychology of Invention in the Mathematical Field.*, *An Essay on the Psychology of Invention in the Mathematical Field.* (Oxford, England: Princeton Univ. Press, 1949)

Hollings, Christopher, Ursula Martin, and Adrian Rice, ‘Ada Lovelace and the Analytical Engine’, *Oxford Blog*, 2018, p. https://blogs.bodleian.ox.ac.uk/adalovelace/2018/0

Huang, Lan, David Milne, Eibe Frank, and Ian H Witten, ‘Learning a Concept-based Document Similarity Measure’, *Journal of the American Society for Information Science and Technology*, 63.8 (2012), 1593–1608 <https://doi.org/10.1002/asi.22689>

Isola, P, J Zhu, T Zhou, and A A Efros, ‘Image-to-Image Translation with Conditional Adversarial Networks’, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–76 <https://doi.org/10.1109/CVPR.2017.632>

Jurafsky, Daniel, and James Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2008, II

Kahn, Louis, ‘Architecture Is the Thoughtful Making of Spaces’, *Perspecta*, 4 (1957), 2–3

Kriegeskorte, Nikolaus, ‘Deep Neural Networks: A New Framework for Modeling

Biological Vision and Brain Information Processing', *Annual Review of Vision Science*, 1.1 (2015), 417–46 <https://doi.org/10.1146/annurev-vision-082114-035447>

Langr, Jakub, *GANs in Action Deep Learning with Generative Adversarial Networks* (Manning Publications, 2019)

Lavrakas, Paul, 'Encyclopedia of Survey Research Methods' (Thousand Oaks, California, 2008) <https://doi.org/10.4135/9781412963947 NV - 0>

Lustig, Roger, 'Margaret Boden, the Creative Mind: Myths and Mechanisms', *Artificial Intelligence*, 79.1 (1995), 83–96 <https://doi.org/10.1016/0004-3702(95)90025-X>

McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Grossberger, 'UMAP: Uniform Manifold Approximation and Projection', *Journal of Open Source Software*, 3 (2018), 861 <https://doi.org/10.21105/joss.00861>

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, 'Efficient Estimation of Word Representations in Vector Space', 2013

Mordvintsev, Alexander, Christopher Olah, and Mike Tyka, 'Deepdream-a Code Example for Visualizing Neural Networks', *Google Research*, 2.5 (2015)

———, 'Iterative_Places205-GoogLeNet_buildings', *Google AI Blog*, 2015 <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>

Mozaffar, Farhang, and Mehdi Khakzand, 'ARCHITECTURAL DESIGN PROCESS IN TECHNOLOGY AGE', *IUST*, 19.6 (2009), 53–72 <http://ijiepm.iust.ac.ir/article-1-263-fa.html>

Myers, Georgina, *Preliminary Literature Review*, 2020

Ng, Andrew, 'Machine Learning | Coursera', *Coursera*, 2013

Oleinik, Anton, 'What Are Neural Networks Not Good at? On Artificial Creativity', *Big Data & Society*, 6.1 (2019), 2053951719839433 <https://doi.org/10.1177/2053951719839433>

Perlovsky, Leonid, and Roman Ilin, 'Brain. Conscious and Unconscious Mechanisms of Cognition, Emotions, and Language', *Brain Sciences*, 2.4 (2012), 790–834 <https://doi.org/10.3390/brainsci2040790>

Purkait, Niloy, *Hands-On Neural Networks with Keras: Design and Create Neural Networks Using Deep Learning and Artificial Intelligence Principles* (Birmingham: Birmingham: Packt Publishing, Limited, 2019)

Qiao, Tingting, Jing Zhang, Duanqing Xu, and Dacheng Tao, *MirrorGAN: Learning Text-to-Image Generation by Redescription*, 2019

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams, 'Learning Representations by Back-Propagating Errors', *Nature (London)*, 323.6088 (1986), 533–36 <https://doi.org/10.1038/323533a0>

Simon, Herbert A, 'Creativity and Motivation: A Response to Csikszentmihalyi', *New Ideas in Psychology*, 6.2 (1988), 177–81 <https://doi.org/10.1016/0732-118X(88)90002-5>

Simonton, Dean Keith, 'Quantifying Creativity: Can Measures Span the Spectrum?', *Dialogues in Clinical Neuroscience*, 14.1 (2012), 100–104

Sirbu, Daniela, and I Dumitrache, 'A Conceptual Framework for Artificial Creativity in Visual Arts', *International Journal of Computers Communications & Control*, 12 (2017), 381 <https://doi.org/10.15837/ijccc.2017.3.2759>

Thaler, Stephen L, 'The Emerging Intelligence and Its Critical Look at Us', *Journal of Near-Death Studies*, 17.1 (1998), 21–29 <https://doi.org/10.1023/A:1022990118714>

Vidal, René Víctor Valqui, 'To Be Human Is to Be Creative', *AI & Society*, 28.2 (2013), 237–48 <https://doi.org/10.1007/s00146-012-0415-1>

Wallas, Graham, *The Art of Thought,* (New York: Harcourt, Brace and Company, 1926)

Xu, T, P Zhang, Q Huang, H Zhang, Z Gan, X Huang, and others, 'AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1316–24 <https://doi.org/10.1109/CVPR.2018.00143>

Xu, Z, D Tao, Ya Zhang, J Wu, and A Tsoi, 'Architectural Style Classification Using Multinomial Latent Logistic Regression', in *ECCV*, 2014

Zador, Anthony M, 'A Critique of Pure Learning and What Artificial Neural Networks Can Learn from Animal Brains', *Nature Communications*, 10.1 (2019), 3770 <https://doi.org/10.1038/s41467-019-11786-6>

Zeiler, Matthew D, and Rob Fergus, 'Visualizing and Understanding Convolutional Networks', 2013 <https://arxiv.org/abs/1311.2901>

Zhang, Han, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and others, 'StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP (2017) <https://doi.org/10.1109/TPAMI.2018.2856256>