Research Plan
AR3A010

Geert Coumans
Seyran Khademi
Casper van Engelenburg

William Kosta
Student no. 5941369

Addressing the Intangible:

Decomposing Architecture Atmospheres Using Foundation Models

Explore Lab 39
Nov 2024

Delft University of Technology
Faculty of Architecture and
the Built Environment

Table of Contents

1

Problem Statement & Research Question

Problem Statement

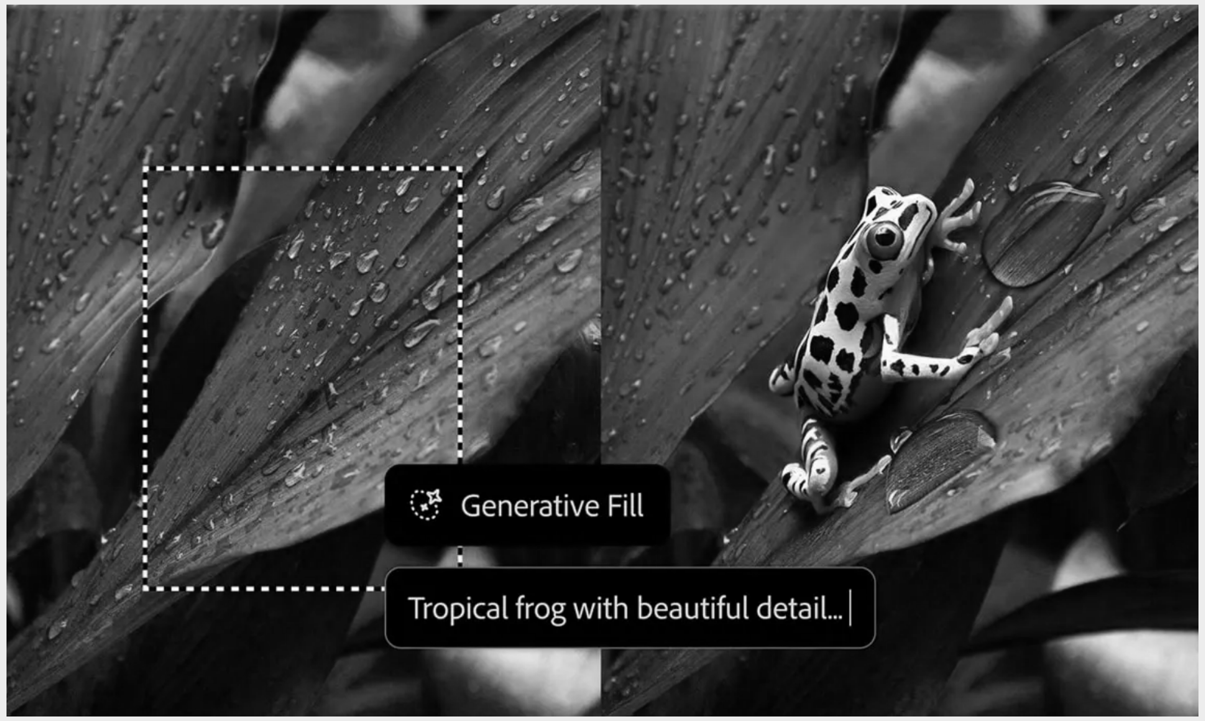


Figure 1. Adobe generative fill

Introduction

Recently, the prominence of artificial intelligence (AI) has been growing rapidly. The realisation that technology is able to perform tasks previously considered possible only for humans has sparked the trend of applying AI in various fields of work. Architecture is no exception to this. One of my first interactions with AI was through Adobe's generative fill (Figure 1), where the software fills or replaces a highlighted area within a canvas with a desired image specified by the user in a prompt. The accuracy and quality of the resulting fill sparked my interest in exploring the extent of AI and its capabilities, especially in architecture.

In architecture, on a superficial level, popular AI image-generation tools like DALL-E (Ramesh et al., 2021) can be used to seek inspiration. A more building-specific use of AI in architecture also include using models to predict building performance. So far, AI has been largely used to interact with the more pragmatic and quantifiable aspects of architecture.

However, architecture lies at the intersection of science and art. On the practical side, it must comply with structural and engineering requirements, but on the artistic side, good architecture should be able to "move" people. One quality that helps achieve this is atmosphere. This is a term that Peter Zumthor uses to describe this quality (Zumthor, 2006), which can also be loosely interpreted as the prevailing tone or mood of a space. Atmosphere is an emergent quality influenced by many factors. One of these factors is natural light. This is a factor I would like to focus on because it has dual nature: it is physically measurable, but in an atmospheric sense it is also evocative.

Figure 2. St. Pierre, Firminy, Le Corbusier

Problem Statement

There are far more architects whose designs aim at publicity and attention, compared to those who focus on atmospheres. At present, atmosphere-driven design, like that of Peter Zumthor, or the later work of Le Corbusier (Figure 2 & 3), is becoming a niche within architecture. When atmosphere is neglected, we have less control over it, leaving the outcome to chance. Without a systematic way to approach atmospheres, we risk losing the ability to create spaces with prominent atmospheres—such as those that approach the sublime, like churches or monuments—which are essential to human culture and expression. This decline in importance of atmosphere occurs for many reasons. One reason is that designing with atmosphere in mind is not easily transferable.

My research aims to solidify atmospheres as an important quality to be considered in design by making it more accessible. Because atmosphere is a fluid concept, I will take inspiration from *Anchoring the Design Process* (Van Dooren, 2020), where the author's goal is to analyse, break down, and work on the design process in architecture (which is also a fluid and implicit concept). In the text, the design process is abstracted, and a framework is created. With this framework, it becomes possible to address the fluid concept of the design process in a more accurate and meaningful way. It is under this parallel condition of the design process and atmospheres both being fluid and lacking vocabulary, framework and tools to be explored properly, that we can identify a shared problem of measurability and therefore a need to address them.

The nature of atmosphere being an abstract quality that humans feel, does not mean that it is born out of purely immeasurable elements. Atmosphere is conceived through a combination of measurable elements in the building that humans perceive. These perceived elements are then processed on an abstract level in the brain, which is then felt as atmosphere. This process of recognising implicit and abstract concepts out of measurable input is something that foundation models do well. Foundation models are AI models trained on broad data that can be adapted to a wide range of tasks (Bommasani et al., 2021). They effectively convert input data into representation that can be used for classification. This is done through an implicit training process rather than through hard coding or explicit instructions. In addition to that, foundation model gives access of investigating this topic in a large scale of data. This is important because there are patterns that can only become apparent when a large dataset is being used, such as similarities and trends. This makes foundation models an appropriate tool to explore for this research.

However, given the limited of examples of AI being applied to the intangible aspects of architecture, this leads to the following questions:

Figure 3. Colline Notre Dame du Haut
Ronchamp, Le Corbusier

Research Question

"To what extent are foundation models an effective tool to approach and address atmospheres and the role of natural light in architecture?"

Sub Questions

- How can we systematically approach the topic of atmospheres?

- Can atmospheres be clustered into different groups? What are the main groups?

- How can we effectively collect a large dataset of images that visually convey atmosphere?

- What are the different ways natural light can be used to contribute to the creation of atmospheres?

2

Theoretical Framework

Theoretical Framework (Conceptual)

In this conceptual section, the key concept of atmospheres, which the research tries to capture will be explained, through literary sources which speak about the topic.

Atmospheres

In his book *Atmospheres* (Zumthor, 2006), Peter Zumthor describes atmospheres as a quality in a building that manages to move people every single time. It is something people can sense within seconds of entering a building, much like a first impression. This sensation is closely linked to our primal survival instincts, through which we evolved to perceive and judge environments quickly—a contrast to our more logical and slower linear thinking. This suggests that atmosphere is something deeply embedded and natural to humans.

Zumthor discusses various factors he considers when attempting to create a particular atmosphere. This includes the body of architecture, material compatibility, the sound of space, the temperature of space, surrounding objects, between composure and seduction, tension between interior and exterior, levels of intimacy, the light on things, architecture as surroundings, coherence, and finally the beautiful form. These twelve points, according to Zumthor, can be seen as building blocks, which can be modified and mixed in order to achieve a certain atmosphere (figure 4).

The twelve points mentioned before are closely linked to the way humans experience a building, a theme also discussed in *The Eyes of the Skin* (Pallasmaa, 2005). In this text, the author argues for an emphasis on human senses beyond sight, such as touch, sound, and smell. Pallasmaa suggests that more attention to these senses will result in a richer architecture. He also proposes that sight can be understood as an extension of touch; when we see, we are not merely looking—our brain incorporates experiences associated with the visual input, connecting us more intimately with materials and textures, rather than perceiving them only as visual elements.

Another source that elaborates on the theme of atmospheres is *The Poetics of Space* (Bachelard, 2014). In this book, Bachelard discusses space in a phenomenological way. He introduces the concept of the "poetic image"—images linked to space that have the power to resonate deeply and emotionally with people. He argues that such images evoke universal feelings shared by many. For example, an

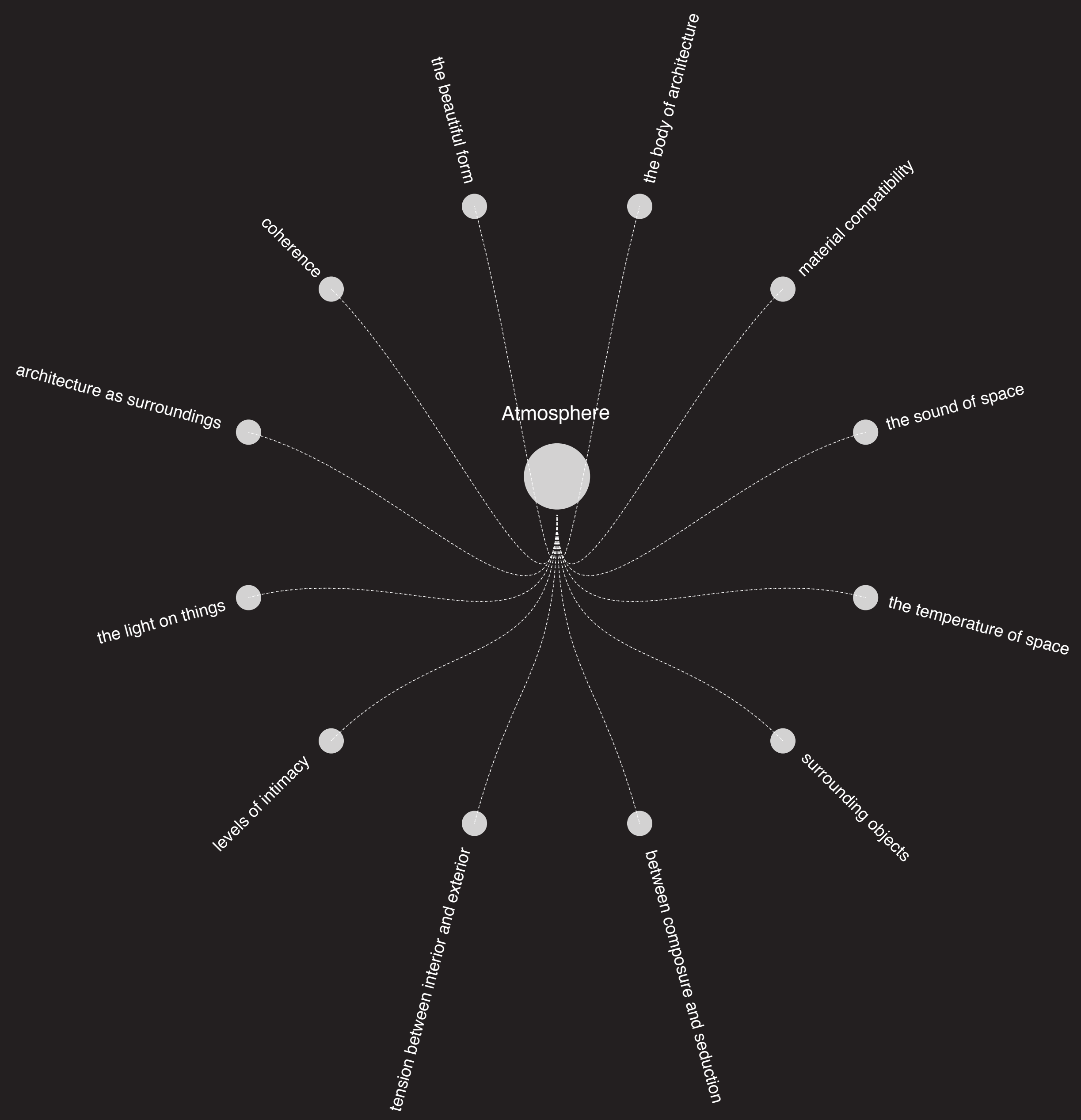


Figure 4. Factors affecting atmospheres

Theoretical Framework (Conceptual)

image conveying the idea of "home" is often associated with warmth, protection, and intimacy.

Drawing from the three sources mentioned before, it could be speculated that an image can provide a glimpse of a space's atmosphere. While it may not be as powerful as experiencing the building itself, the right images can convey significant (emotional) information about a space. In relation to this research, these sources would be helpful when creating an initial dataset of pictures that showcase a range of atmospheres in various spaces.

The Design Process

The design process in architecture is fluid and can vary from one designer to another. This fluidity makes the design process difficult to describe. In *Anchoring the Design Process* (Van Dooren, 2020), the author addresses the challenge of designing and teaching design. The author argues that designing is a complex skill, and performing a complex skill is primarily an implicit activity. To teach design, it is necessary to make the implicit become explicit. To approach the design process, a framework is introduced, consisting of five generic elements that typically exist in some form in any design process. These five elements include "experimenting," "guiding themes," "domains," "frame of reference," and "laboratory." This framework provides the tools and vocabulary needed to engage with the design process.

The nature of the design process can be seen as parallel to the nature of atmospheres, which is also fluid. Just as this framework helps anchor the design process, my research aims to provide vocabulary, classification, and tools for engaging with the intangible aspects of architecture, offering a starting point for interacting with atmospheres.

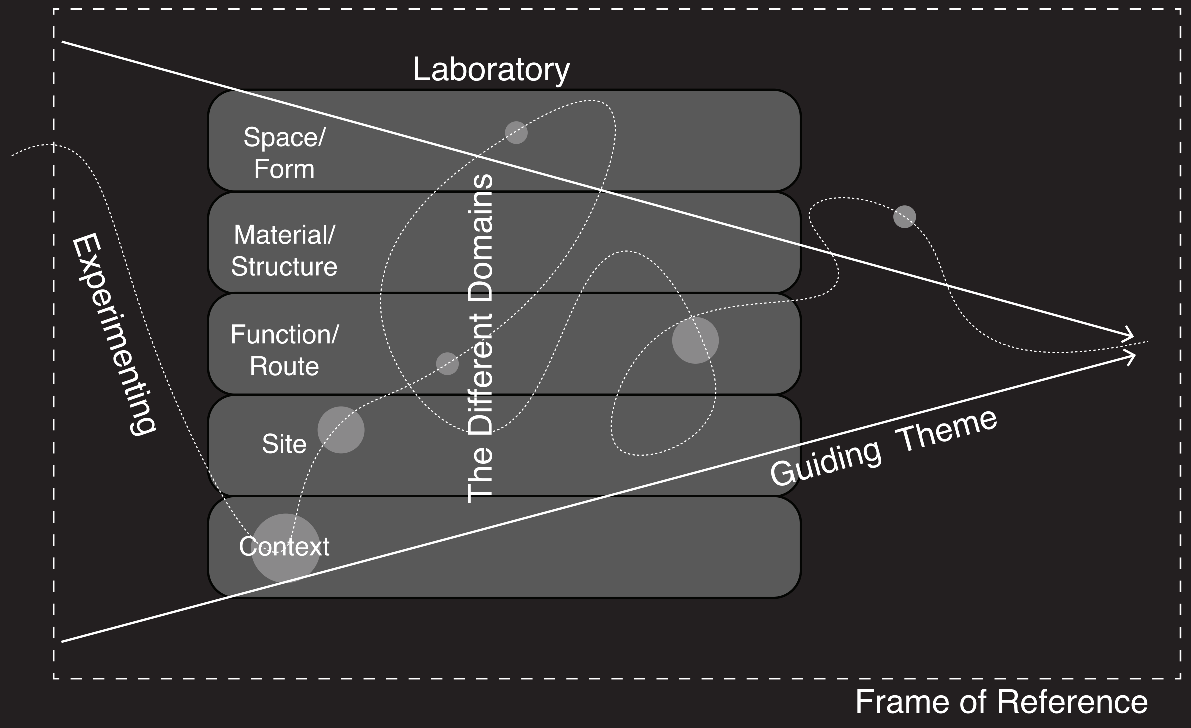


Figure 5. Design process framework diagram

Theoretical Framework (Technical)

Having set out the concept in the previous section, this section will explain the technical approach of the research and why it is being used to address atmospheres in the research.

Machine Learning (ML)

ML is a subset of AI, which involves training machines to perform complex tasks such as face verification (Sengupta et al., 2016), object detection (Redmon et al., 2015), and prompt-guided text generation (Radford et al., 2018). ML also includes Deep Learning (DL), which uses multi-layered neural networks, called deep neural networks (As & Basu, 2021).

The training needed to create a model (As & Basu, 2021) can be seen as parallel to how people train to become architects, relying primarily on examples and experience rather than explicit instructions (Van Dooren, 2020). For example, to train a model to recognise whether an image shows bricks or timber, it must be provided with a set of training images. With each image, the model examines ‘features’ similar to how a person might visually assess whether an object is brick or timber. In this example, features could include the color, texture, and shape of the object, among others.

Based on these features, the model assigns the image to a class. In this example, only two classes are needed: Class 1 for “bricks” and Class 2 for “timber.” If the input is an image of bricks, the correct output would be “1” for Class 1 and “0” for Class 2. However, if the output is incorrect, backpropagation allows the model to be adjusted for improved accuracy.

With DL (Figure 7), this learning process is automated (As & Basu, 2021): from raw data, to feature extraction, to classification, all steps are learned by the model. The model extracting features on its own is the reason why it is interesting to apply this to the intangible side of architecture, as there is currently no fixed and definitive classification method of atmospheres.

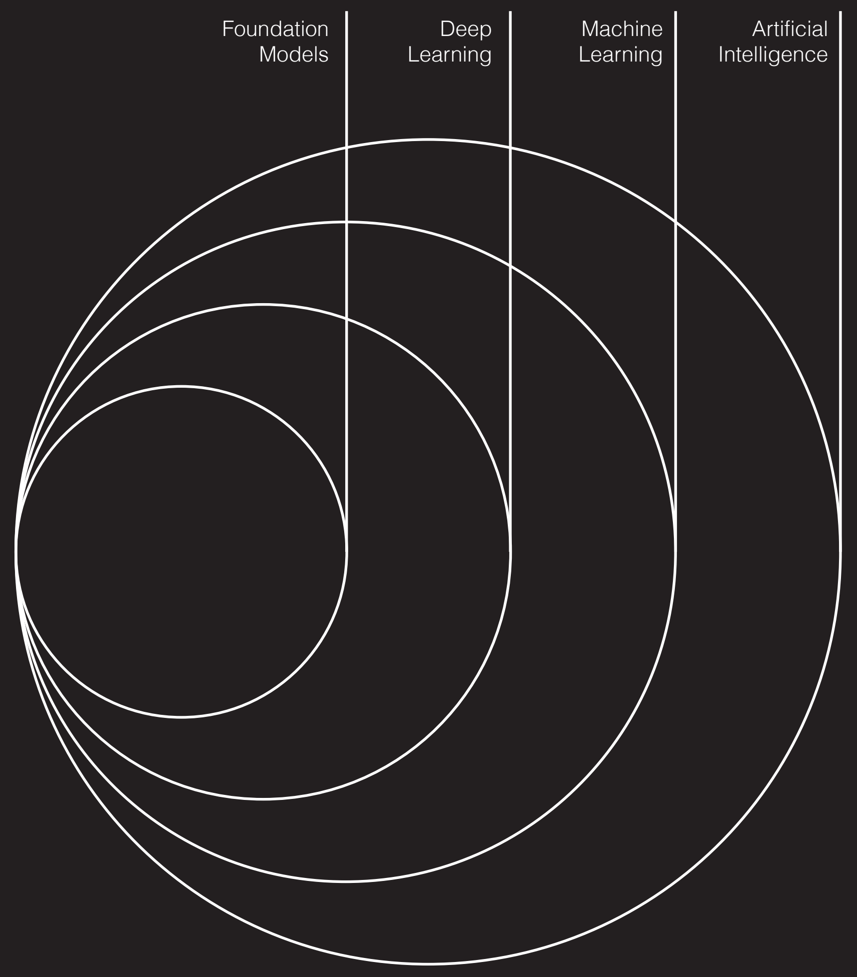


Figure 6. At present, all foundation models are deep learning based

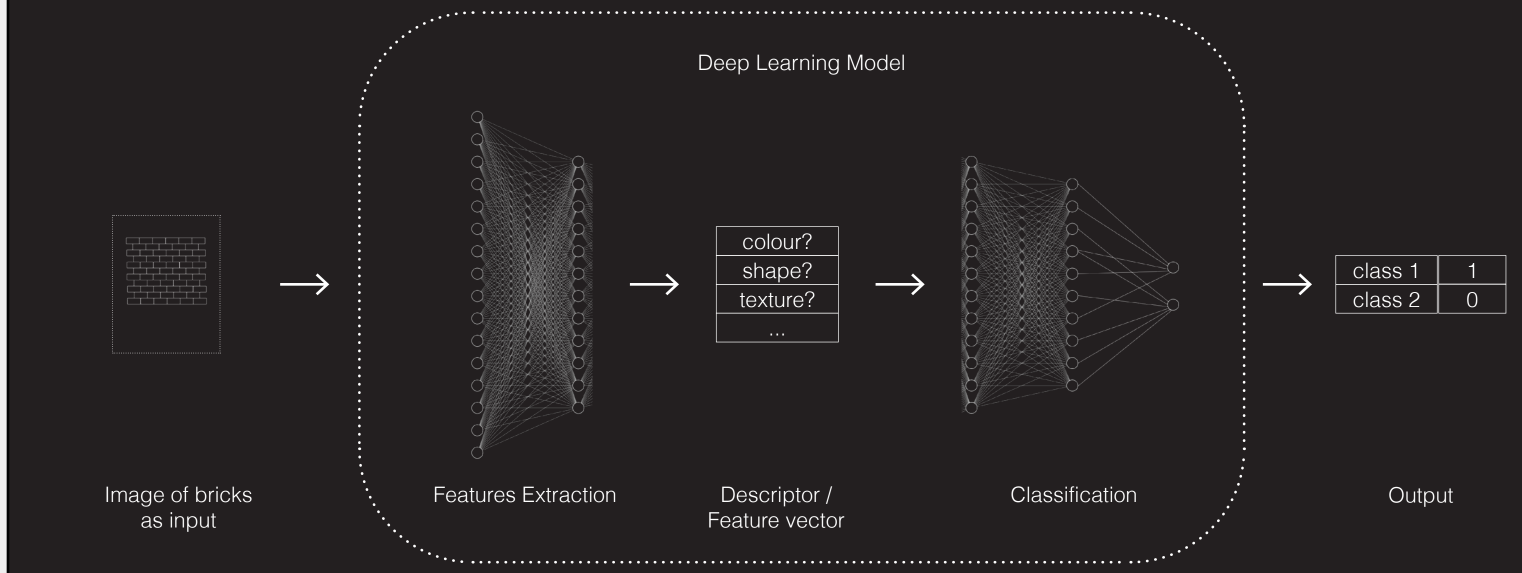


Figure 7. The fact that the models extract features and classifies on its own, based on a learning technique parallel to the way a humans train in architecture is why this method is relevant to the experiment

Theoretical Framework (Technical)

Foundation Models

Collecting and annotating data for specific tasks is labour intensive. Today, the vast amount of available data, the ability to perform large-scale computations, and the existence of proper optimisation algorithms (such as DL) make it possible to develop models that can generalise across specific tasks and domains. These models are referred to as foundation models. Popular examples include GPT (Radford et al., 2018) for language understanding and generation, and DALL E (Ramesh et al., 2021) for image generation.

DINOv2 (Caron et al., 2021), ViT (Dosovitskiy et al., 2020), YOLOv8 (Redmon et al., 2015), are open source foundation models that will be relevant to the research, due to its affinity with images. Because foundation models are trained on vast amounts of data, there are differences in the results when using different models.

Dimensionality Reduction

The output of most foundation models is high-dimensional, typically in the order of 1000 features. This situation in vector space suffers from 'the curse of dimensionality,' where, as the dimensionality grows, most of the 'space' becomes devoid of data points (Ananthaswamy, 2024). To visualise the dataset and analyse the relative distances of the dataset's instances in the outcome, we need a method to project the high-dimensional space into 2D or 3D.

To reduce the dimensions of the output, algorithms such as UMAP (McInnes et al., 2018) or t-SNE (Van Der Maaten & Hinton, 2008) can be used. t-SNE 'compresses' the high dimensional data without losing too much meaningful information. It preserves the local distances found in high dimensional data when expressing it in 2 or 3 dimensions. Taking the example of a model recognising hand written digits from 0 to 9 (MNIST dataset), this is useful because if the high dimensional data output clusters (local distance) a lot of pictures of the number "9" together (something that shows that the model is working), we can see it even when we compress the data in 2 dimensions. Whereas the information regarding where pictures of the digit "0" are clustered in relation to where the pictures of "9" are clustered, are not interesting to us and also not preserved by t-SNE (Figure 8).

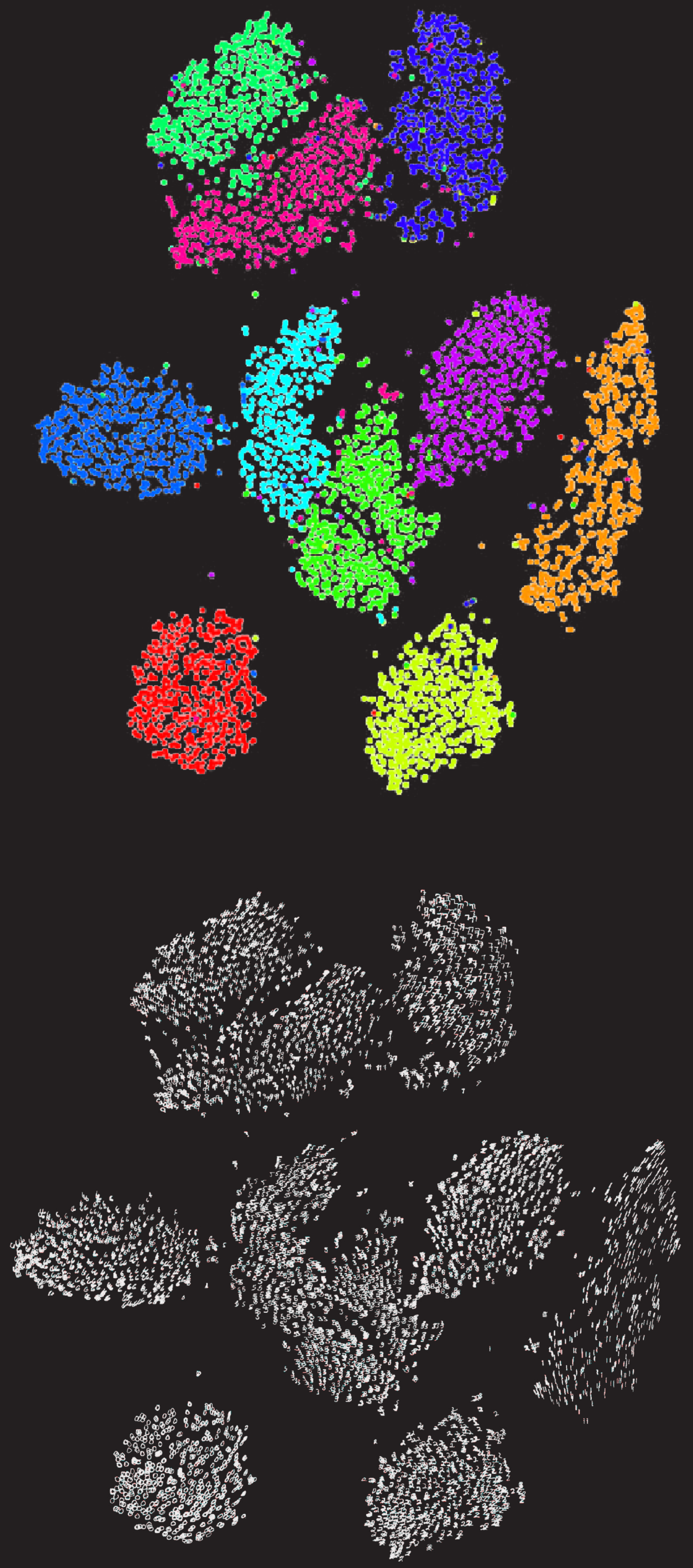


Figure 8. The graphs above are t-SNE visualisations of the MNIST dataset (Van Der Maaten & Hinton, 2008). With the large scale of the MNIST dataset, the visualisation by t-SNE reveals patterns, such as the frequency of how much people tend to write the digit "1" leaning to the right instead of left, or how similar the digit "3" and "5" is and how often it is indistinguishable. This is something that would be difficult to identify without scale and t-SNE's visualisation. The same insight is expected to be found with the topic of atmospheres in this research.

3

Method & Research Intent

Method

In order to use classify atmospheres and understand it better, using foundation model for its consistency and scale, I will be creating a 2D representation which maps different buildings based on atmosphere, creating a taxonomy of architecture atmosphere. To do that, I will be using the following method:

Data curation

In this first part of my research, I will prepare a relatively large dataset of architecture photos that display a variety of atmospheres. The method of collection will be both finding images by myself manually, and also setting up an automatic web scraper that collects photographs from websites such as Archdaily, Divisare, Pinterest, Flickr, and more.

Image processing and visualisation

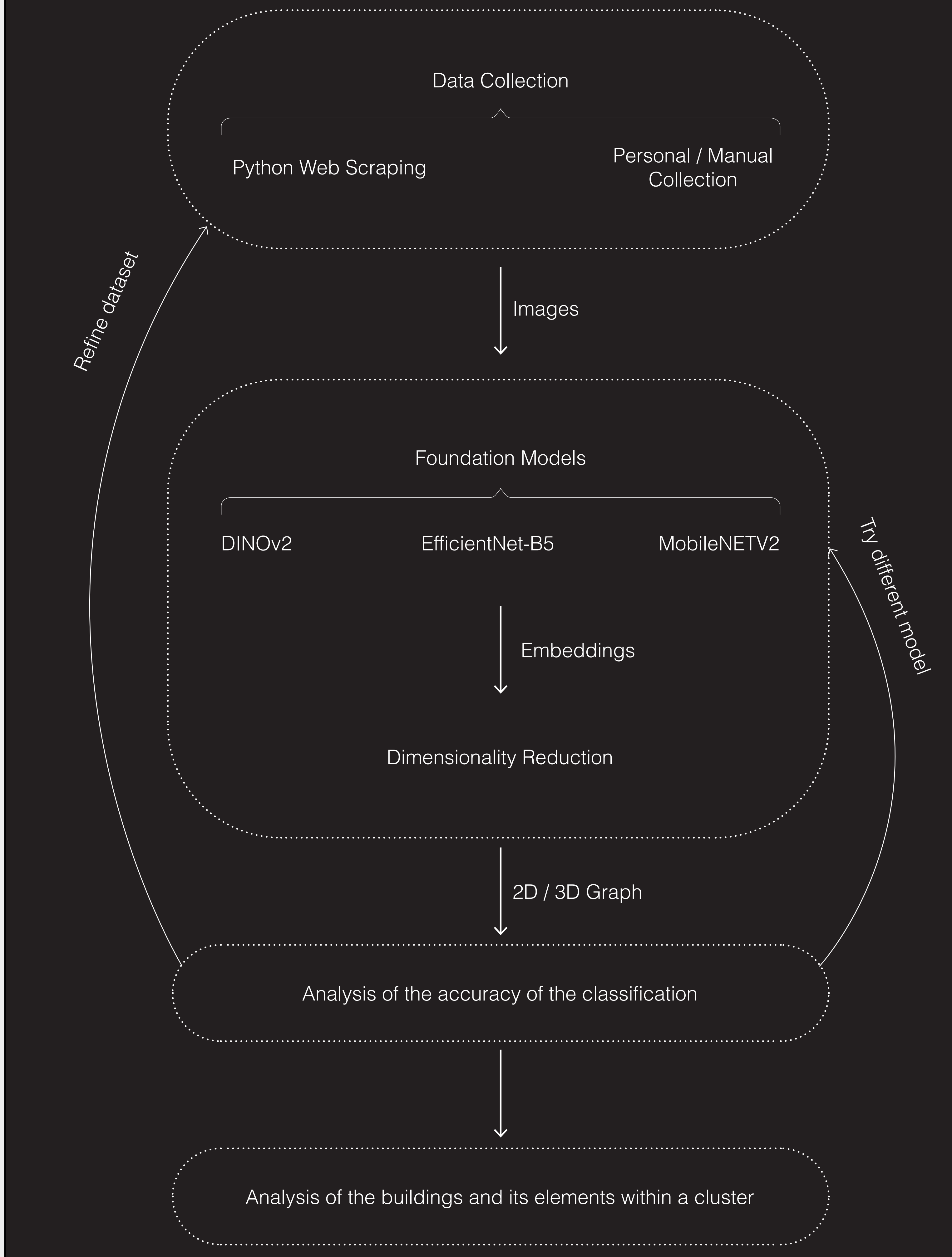
In the second part, I will investigate whether the foundation models available today are able to recognise atmospheres from an image and therefore cluster them into different groups. I will use several different foundation models such as DINOv2 (Caron et al., 2021), EfficientNet (Tan & Le, 2019) and MobileNet (Howard et al., 2017) to increase the chance of success. In doing so, we compute the feature vectors for each instance in the dataset. These are then passed to dimensionality reduction algorithms such as t-SNE (Van Der Maaten & Hinton, 2008) or UMAP (McInnes et al., 2018), so that the outcome can be presented in the form of a scatter plot (2D representation).

Analysis

In the third part of my research, I will analyse the resulting graph. In this graph, instances that are in a cluster are the images that the foundation model deems similar. The analysis will be done in two stages. The first stage is to judge the output to determine if the model is clustering the instances based on atmosphere or not. If not, the experiment will be repeated, changing variables such as the foundation model itself, isolating or culling parts of the dataset, etc.

When an acceptable result is achieved, the clusters will then be investigated, looking for patterns, overlap and trends. The corresponding buildings of the images that are clustered together will also be analysed to see if there are any architecture elements that are in common, which contributes to the creation of a specific atmosphere. Special attention will be given to natural light, its creation through architecture and its relation to atmosphere.

Figure 9.



Research Intent

Research Intent

The aim of this research is to address the intangible quality in architecture—atmosphere. This will be achieved by creating a classification and taxonomy of buildings based on their atmospheres. Foundation models will be used in the research because they excel at identifying implicit features and creating clusters (when represented in 2D) of images. Foundation models are also capable of processing large amounts of data, which will not only lead to a more accurate result, but also reveal trends and repetition or similarities that only become apparent with a large scale in the resulting 2D representation. Once the patterns and clusters are formed, architects can begin labelling the previously unnamed clusters of buildings, thereby creating a vocabulary for atmospheres and providing a starting point to study the topic in a consistent way (figure 9).

Precedents

Precedents of computer vision experiments aimed at addressing architectural features include "What Makes Paris Look Like Paris" (Doersch et al., 2015), which sought to identify distinctive visual elements representative of a city, and "Sight-Seeing in the Eyes of Deep Neural Networks" (Khademi et al., 2018), which used Convolutional Neural Networks to predict geolocations from images and visualise the learned attributes of the model, uncovering its process. My experiment draws inspiration from these studies, albeit in a more simplistic form.

Limitation

A limitation is the fact that this is done only in an academic setting, where data collection is made easy as it is now merely for research purposes. In a real world environment, the ethics of data collection will make this process not as straightforward.

Design Relation

The graduation design project will be a public building that allows for the display of dramatic atmospheres, especialy related to natural light. The knowledge gained from this research is expected to help find a clearer way of creating and deploying specific atmospheres in appropriate spaces.

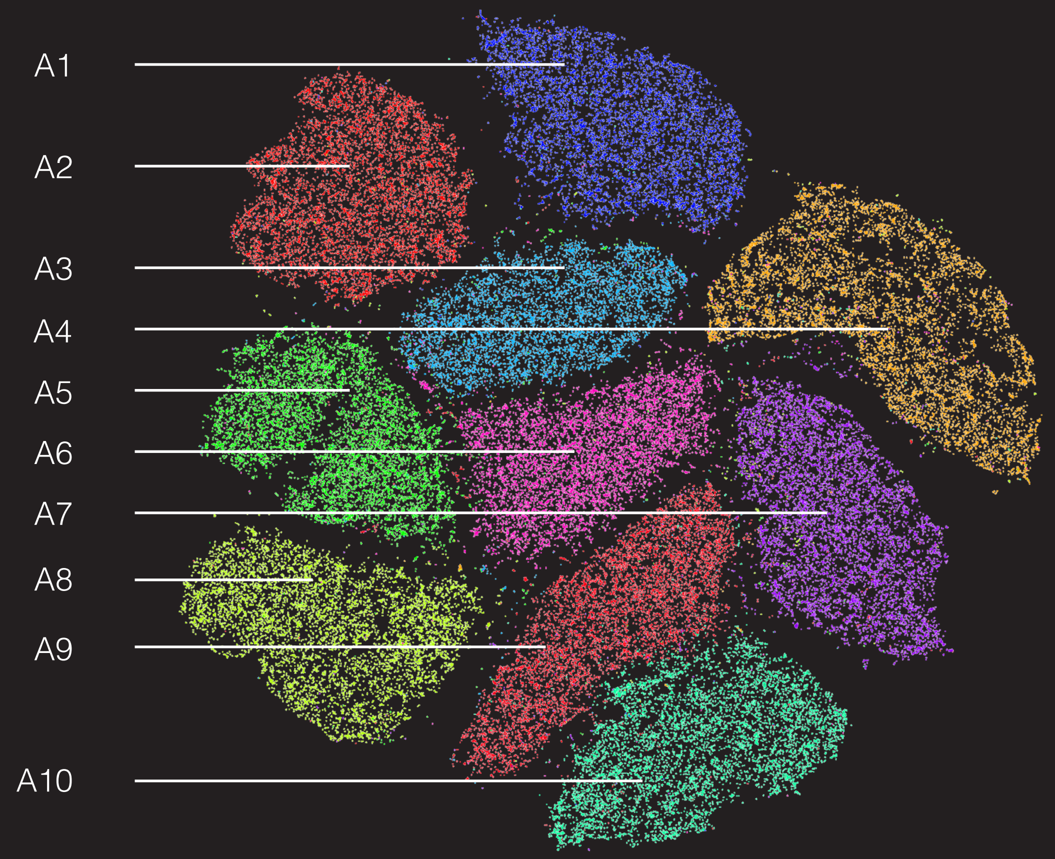
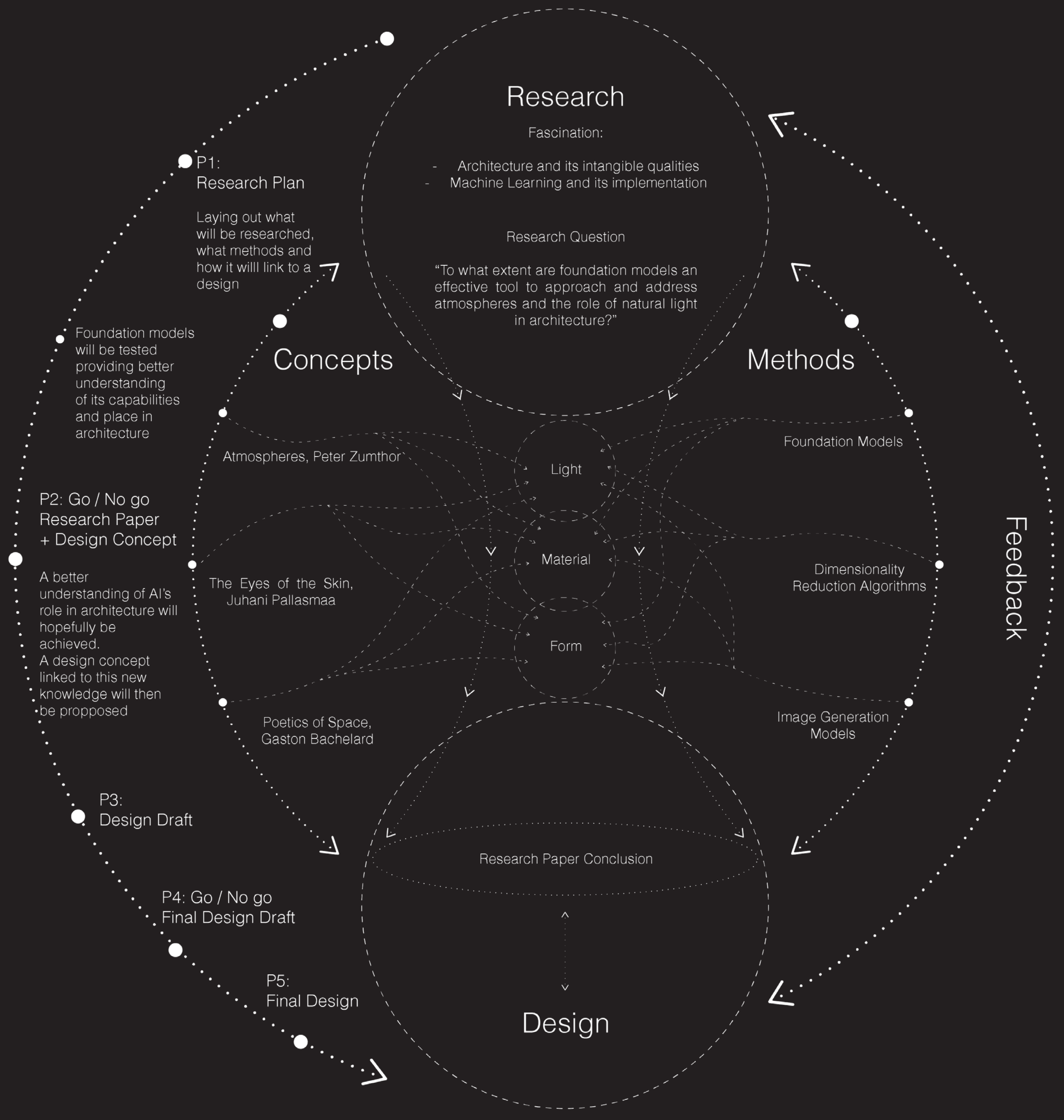


Figure 10. Above is a hypothetical outcome of a 2D representation. Unlabeled clusters from A1-A10 (or more) will form, which can be labelled after the process, to create a classification map or taxonomy of atmospheres. Using foundation models will allow consistency, and a larger scale of the data, and therefore clearer clusters and ease to summarise each cluster.

Plan Diagram

Figure 11.



Research

Fascination:

- Architecture and its intangible qualities
- Machine Learning and its implementation

Research Question

"To what extent are foundation models an effective tool to approach and address atmospheres and the role of natural light in architecture?"

P1:
Research Plan

Laying out what will be researched, what methods and how it willl link to a design

Foundation models will be tested providing better understanding of its capabilities and place in architecture

P2: Go / No go
Research Paper
+ Design Concept

A better understanding of AI's role in architecture will hopefully be achieved.
A design concept linked to this new knowledge will then be proppposed

P3:
Design Draft

P4: Go / No go
Final Design Draft

P5:
Final Design

Concepts

Atmospheres, Peter Zumthor

The Eyes of the Skin,
Juhani Pallasmaa

Poetics of Space,
Gaston Bachelard

Light

Material

Form

Methods

Foundation Models

Dimensionality
Reduction Algorithms

Image Generation
Models

Feedback

Research Paper Conclusion

Design

Bibliography

Ananthaswamy, A. (2024). Why machines learn: The Elegant Math Behind Modern AI. Penguin.

As, I., & Basu, P. (2021). The Routledge companion to artificial intelligence in architecture.

Bachelard, G. (2014). The Poetics of Space. Penguin.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Sydney, V. A., Bernstein, M. S., Bohg, J.,
Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., . . . Liang, P. (2021, August 16). On the Opportunities and Risks of Foundation Models. arXiv.org. https://arxiv.org/abs/2108.07258

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021, April 29). Emerging Properties in Self-Supervised Vision Transformers. arXiv.org. https://arxiv.org/abs/2104.14294

Carta, S. (2022). Machine learning and the city: Applications in Architecture and Urban Design. John Wiley & Sons.

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. A. (2015). What makes Paris look like Paris? Communications of the ACM, 58(12), 103–110. https://doi.org/10.1145/2830541

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020, October 22). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv.org. https://arxiv.org/abs/2010.11929?ref=labelbox.ghost.io

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2020, September 17). Multidimensional scaling, Sammon Mapping, and ISOMap: tutorial and survey. arXiv.org. https://arxiv.org/abs/2009.08136

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017, April 17). MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv.org. https://arxiv.org/abs/1704.04861

Khademi, S., Shi, X., Mager, T., Siebes, R., Hein, C., De Boer, V., & Van Gemert, J. (2018). Sight-Seeing in the eyes of deep neural networks: Vol. abs 1511 7247 (pp. 407–408). https://doi.org/10.1109/escience.2018.00125

McInnes, L., Healy, J., & Melville, J. (2018, February 9). UMAP: uniform manifold approximation and projection for dimension reduction. arXiv.org. https://arxiv.org/abs/1802.03426
Pallasmaa, J. (2005). The eyes of the skin: Architecture and the Senses. Academy Press.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by Generative Pre-Training. https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035#citing-papers

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021, February 24). Zero-Shot Text-to-Image Generation. arXiv.org. https://arxiv.org/abs/2102.12092

Reasons for the sensational in architecture. (2023, January 1). Domus. https://www.domusweb.it/en/architecture/2022/10/26/reasons-for-the-sensational.html

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015, June 8). You only look once: Unified, Real-Time Object Detection. arXiv.org. https://arxiv.org/abs/1506.02640

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021, December 20). High-Resolution Image Synthesis with Latent Diffusion Models. arXiv.org. https://arxiv.org/abs/2112.10752

Sengupta, S., Chen, J., Castillo, C., Patel, V. M., Chellappa, R., & Jacobs, D. W. (2016). Frontal to profile face verification in the wild. https://doi.org/10.1109/wacv.2016.7477558
Tan, M., & Le, Q., V. (2019, May 28). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv.org. https://arxiv.org/abs/1905.11946

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. https://jmlr.org/papers/v9/vandermaaten08a.html

Van Dooren, E. (2020). anchoring the design process: A framework to make the designerly way of thinking explicit in architectural design education. Architecture and the Built Environment, 17, 176. https://doi.org/10.7480/abe.2020.17.5351

Zumthor, P. (2006). Atmospheres: Architectural Environments, Surrounding Objects. Birkhaüser.

Zumthor, P. (2010). Thinking architecture. Birkhauser.

List of Figures