Technology|Architecture + Design

ISSN: 2475-1448 (Print) 2475-143X (Online) Journal homepage: www.tandfonline.com/journals/utad20

On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs

Jeffrey Huang, Mikhael Johanes, Frederick Chando Kim, Christina Doumpioti & Georg-Christoph Holz

To cite this article: Jeffrey Huang, Mikhael Johanes, Frederick Chando Kim, Christina Doumpioti & Georg-Christoph Holz (2021) On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs, Technology|Architecture + Design, 5:2, 207-224, DOI: 10.1080/24751448.2021.1967060

To link to this article: https://doi.org/10.1080/24751448.2021.1967060

Jeffrey Huang
Swiss Federal Institute of Technology, Lausanne (EPFL)

Mikhael Johanes
Swiss Federal Institute of Technology, Lausanne (EPFL)

Frederick Chando Kim
Swiss Federal Institute of Technology, Lausanne (EPFL)

Christina Doumpioti
Swiss Federal Institute of Technology, Lausanne (EPFL)

Georg-Christoph Holz
Swiss Federal Institute of Technology, Lausanne (EPFL)

On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs

Recent advances in Generative Adversarial Networks (GANs) hold considerable promise in architecture, especially in the early, creative stages of design. However, while GANs are capable of producing infinite numbers of new designs based on a given dataset, the architectural relevance and meaningfulness of the results have been questionable. This paper presents an experimental research method to examine *how human and artificial intelligences can inform each other to generate new designs that are culturally and architecturally meaningful*. The paper contributes to our understanding of GANs in architecture by describing the nuances of different GAN models (SAGAN vs DCGAN) for the generation of new designs, and the use of Natural Language Processing (NLP) for the conceptual analysis of results.

◁ Opening Figure. Excerpt of dataset preparation and GAN training. (Credit: Media x Design Laboratory, EPFL)

Keywords: Machine Learning, Architectural Design, Generative Adversarial Networks, Natural Language Processing, Precedents, Ambiguity

Introduction

The introduction of Generative Adversarial Networks (GANs) has been embraced with considerable enthusiasm by the design community, because of their capacity in learning and recognizing patterns in images and extrapolating for the generation of new images (Goodfellow et al. 2014). These advances in artificial intelligence (AI) research are particularly promising for architecture, since *images* have always played a key cognitive role in the creative process of architectural idea generation (Lawson 2009). Images have been used by architects as inspirations, metaphors, analogies, and justifications (Goldschmidt and Smolkov 2006; Purcell and Gero 1996). Images are consciously and subconsciously in the thought process of architects, representing their perspectives, memories, and origins and revealing the inherent quality of their architecture (Olgiati 2013).

Entire design schools and architectural movements have used images as points of departure in the design process. For example, the movement "Analog Architecture" in Switzerland, which has been a significant influence in the development of contemporary Swiss architecture (Caruso 2009), used images and analogies as references and drivers of architectural designs (Bressani and Sprecher 2019). The use of images in the "Analog Architecture" movement was due to the influence of Aldo Rossi in the 1970s and was carried forward by well-known contemporary architects, such as Valerio Olgiati and Christian Kerez (Caruso 2009).

Images not only play a role as external references and cultural precedents but also act as internal heuristic catalysts. Developing an internal mental image is the first step of the three design activities: imaging, presenting, and testing (Zeisel and Eberhard 2006). In this sense, the image refers to the mental image of the designer. By comparing the mental image and the evolving design, designers can modify the image while improving the design, which creates a subjective knowledge for developing and organizing ideas in the image itself. Visual imagery thus supports creativity. Finke's experiments—which are based on his principles of visual imagery: implicit encoding, perceptual equivalence, spatial equivalence, transformational equivalence, and structural equivalence—proved statistically how creative thinking and discovery can be achieved with visual imagery (Roskos-Ewoldsen et al.1993).

In architecture, the use of images as precedents not only helps designers jumpstart their designs and overcome inertia, but also grounds the resulting design in a geographical context and gives it a cultural relevance and meaning, as typically the images chosen and used as precedents for a design implicitly relate to the geographical site and program of the design problem and reflect the Zeitgeist and cultural preferences prevalent at the time of the design.

This research draws on the use of images/precedents by human designers as powerful drivers of design in architecture, and explores the following research question:

How can human and AI (GAN and Natural Language Processing) be combined into a collective intelligence to generate new designs from pictorial training data that are culturally and architecturally meaningful?

This research contributes to the understanding of GANs in architecture by comparing alternative GAN models (DCGAN vs SAGAN) for the generation of new designs, and Natural Language Processing (NLP) for human verbal reaction analysis, and by proposing a design workflow that combines collective machine and human intelligence.

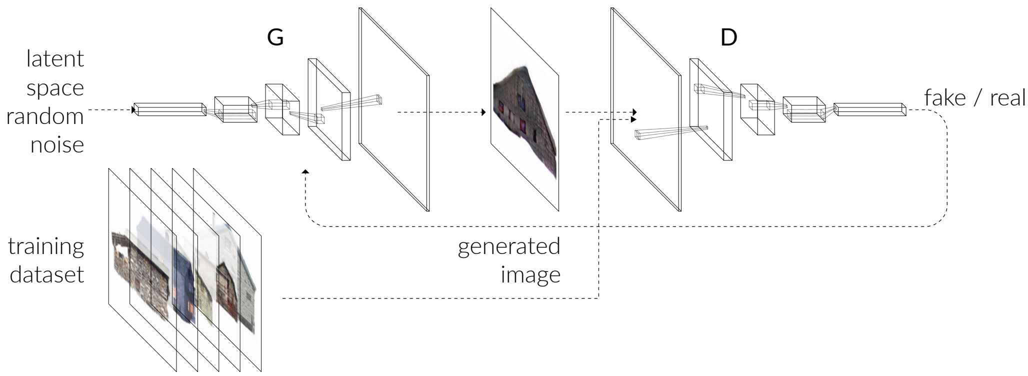
State of the Art

Overview of Generative Adversarial Network

Generative Adversarial Network is a type of generative neural network model for unsupervised machine learning (Goodfellow et al. 2014) which has drawn much attention in the arts and design due to its capacity to learn and generate creative products that are usually attributed to human endeavor. The basic architecture of GANs comprises two competing networks (Figure 1): a generator ***G*** and a discriminator ***D***. During the training, the generator ***G*** learns to map a high dimensional latent space ***z*** to generate new sample data. Inversely, the discriminator ***D*** learns to distinguish the fake samples from the real samples in the dataset, then gives feedback to the generator. With this adversarial mechanism, their capacity to distinguish and generate fake samples improves along with the training. In this way, by exposing GANs to a particular set of arts or design examples in the training process, the networks can automatically learn the underlying pattern, or probabilistic distribution, of the given data and generate novel samples that share similar properties to the original images.

Numerous GANs have been proposed with different directions and contexts since its invention (Hindupur 2018). The development of deep convolutional GAN (DCGAN) was one of the first improvements made by adopting deep convolutional neural networks that already have a successful application in image classification and object detection tasks (Radford et al. 2016). This development enabled DCGAN to work with images and provided a basic model for the later advancement of GANs. Some developments focused on the *loss function* of the discriminator part, a small piece of calculation that determines how the discriminator would distinguish the real data from the fake samples. The aim was to improve the training stability and diversity of the generated samples. A notable development was the Wasserstein GAN (WGAN) and its improved version (WGAN-GP) (Arjovsky et al. 2017; Gulrajani et al. 2017). WGAN gave the discriminator a gradual score from fake to real, usually a binary in regular GANs (either fake or real). It changed the *discriminator* from virtually a forgery expert to a *critic* giving a score to the generated samples. However, while WGAN and WGAN-GP provided more training stability than DCGAN, the slower training time hindered the implementation of these GANs more widely.

Self-attention GAN (SAGAN) has drawn notable recognition due to its ability to capture long-range spatial relationships from the training data to generate structurally coherent images (Zhang et al. 2019). It was achieved by inserting attention-maps between



◁ Figure 1. The basic architecture of GANs, comprising two competing networks: generator G and discriminator D. (Credit: Media x Design Laboratory, EPFL)

the convolutional neural network layers that gave the networks different receptive field-sizes akin to human attention during the training. As a result, SAGAN was able to generate images that had correct structures from the learned samples, which could be used to generalize object anatomy from different perspectives. A variant with potential useful applications is *pix2pix*, a conditional GAN (cGAN) that provides general-purpose solutions to image-to-image translation problems (Isola et al. 2018). In cGANs, the generated images/data are conditional to an input. In the case of *pix2pix*, the input is an image. The training of *pix2pix* requires labeled pairs of both input image and output image for the GAN to learn the translation patterns. It has potential practical applications for translating satellite images to maps, outlining drawings to colored drawings, and coloring b/w photos.

Initially, the resolution of generated images from GANs was limited to 256 x 256 pixels due to training instability in higher resolutions. Progressive-growing GAN (ProGAN) increased the resolution of generated images by training the GAN in multiple phases, from a shallow resolution then quadrupling it into target resolution (Karras et al. 2018). This multiphase training process in ProGAN allowed GAN to generate HD images and provided a basis for the development of StyleGAN, the current state of the art GAN. StyleGAN allows the disentanglement and control of the generated image's style (Karras et al. 2019). It provides the ability to control several learned features separately, such as perspective, color, and pattern. However, in order to gain such detailed control using StyleGAN, a large amount of data is needed to provide sufficient information for the GAN to be able to generate desirable results.

Generative Adversarial Network in Architecture

Newton provided the first comprehensive overview for potential applications of GANs in architecture by using several exploratory experiments (2019). The exploration of the potential uses in architectural design showed that GANs were more suitable for learning and generating 2D images such as building façades and floor plans due to the exponential increase of computational cost in 3D model applications. The experiments using DCGAN and WGAN for generating architectural design from a relatively small-sized dataset of building façades and floor plan images (100-1000 images) showed that the selection of the GAN variant and curation in the training dataset affected the fidelity and diversity of generated design. DCGAN was found to be more suitable for a larger dataset of photographic images, while WGAN was found to work better with a smaller and narrower specific dataset such as floor plan drawings of a particular architect or building typology. The research demonstrated the importance of data curation, especially the curation of size (the number of images) and content of the dataset, when working with GANs. The larger the dataset, the better the generated images in terms of photographic quality. The diversity of content within the dataset also was found to impact the generative capacity of GANs directly. Smaller variation in the dataset produced a more focused and better image, while more variation in the dataset created more diverse results, but usually with less accuracy.

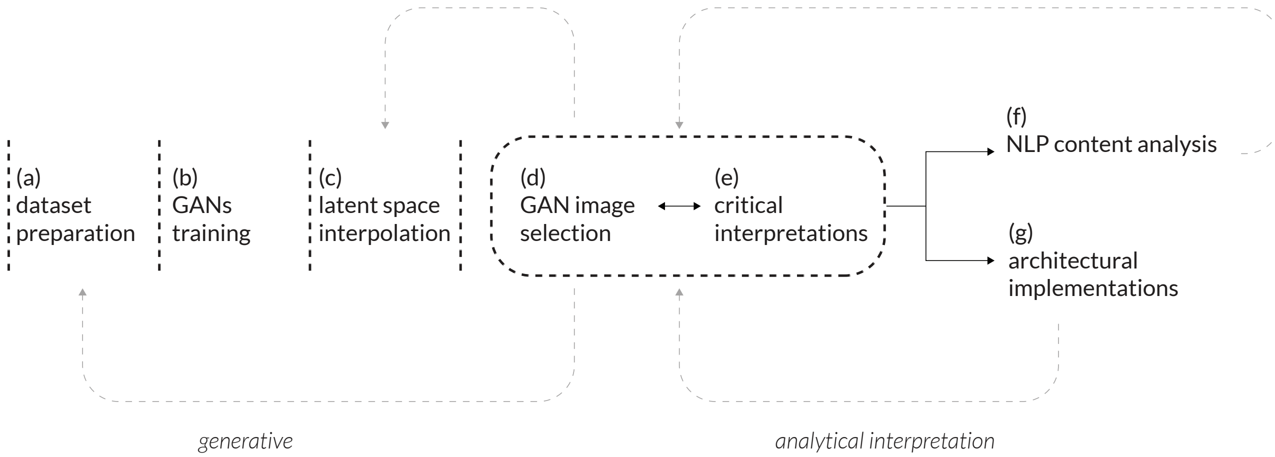
Another architecturally relevant experiment showed that StyleGAN could work with satellite images and large datasets (30,000 images) (Zhang 2019). The use of StyleGAN demonstrated a way to learn and control obvious land features of satellite images such as forest, buildings, rivers, streets, coastlines, and sea. In contrast, the training results of StyleGAN using 2,000 or fewer images of floor plans showed hardly readable results.

An interesting application of GANs for floor plan images was demonstrated by the use of *pix2pix*, which enabled supervised, controlled training of floor plan generation (Chaillou 2019). The focus was on translations among different information categories in the floor plans, such as functional programs, window and door positions, building footprints, site, and furniture. The proposed framework offered a way to generate floor plans in a staged manner, from a land plot to a furnished plan, step by step. *Pix2pix* has also been explored as a surrogate model to replace computational-cost spatial analysis (Tarabishy et al. 2019).

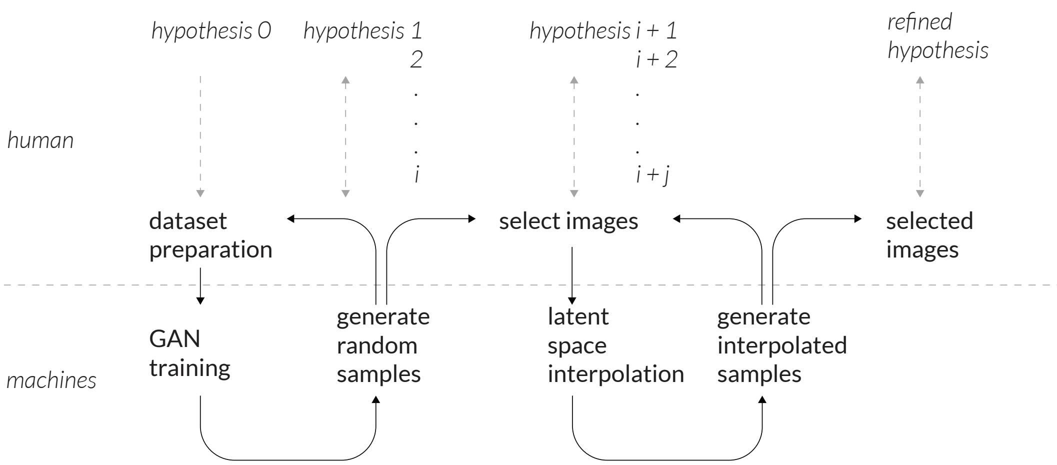
GANs work natively with data that can be represented in numerical values. 2D image data fits perfectly with GANs because each pixel of an image is represented in RGB channels, essentially a three-dimensional matrix of numbers. However, GANs are shown to be able to work with graph representations of space by projecting the graph into vector space (Wang et al. 2017). Based on this principle, a framework for generating graph-based conceptual design was proposed to explore the use of GANs in a conceptual level representation (As et al. 2018).

The above mentioned experiments showed a vast potential of GANs in architectural design involving several modes of representation. In general, GANs are potentially useful in a design processes involving images as design references. GANs provide the unsupervised means for generating/synthesizing images that are

PEER REVIEW / INTELLIGENCE



△ Figure 2. The research design with two distinctive process loops: the generative (left) and the analytical interpretation process (right). (Credit: Media x Design Laboratory, EPFL)



△ Figure 3. Generalization of the human-machine partnership in the GAN-based design process. (Credit: Media x Design Laboratory, EPFL)

statistically relevant to the learned images. The previous exploration of GANs in architecture shows that GANs provide novel creative processes using images by means of dataset curation and selection of GANs (Newton 2019). This research attempts to extend the explored possibility into a complete design process and a specific design context. By seeing images or architectural appearance as the manifestation of a particular architectural culture, GANs can be used as a data-driven, artificial design inspiration for balancing a particular architectural tradition and novelty. The ability of GANs to learn from a given set of pictorial data and generate new images can be seen as analogous to the human designer's learning from precedents to capture the Zeitgeist and prevalent cultural preferences in a given context. The above review of the different variants of GANs shows that the generative capacity of GANs is still in the stage of development and experimentation.

Natural Language Processing

Natural language processing (NLP) refers to the use of the computer to process natural human language to glean useful information from text (Srinivasa-Desikan 2018). Machine learning has advanced language modeling used in NLP techniques by embedding the language into high-dimensional vector spaces. Hence the semantic relationships between words can be effectively inferred from by the use of neural networks (Mikolov et al. 2013). NLP models have become ubiquitous in our everyday lives, appearing in language translation, speech recognition, and personal assistants. As computational techniques in architecture have shifted from quantitative methods to more nuanced, qualitative domains of human lives, NLP techniques are beginning to show their relevance in the analysis of the semantic dimension of human-machine design processes.

The language of design can relate to the enactment of design in the conceptualization, development, and evaluation of ideas by acting as operator in negotiating the outcomes and registering its process (Dong 2007). NLP techniques have been used to analyze and extract the pattern of linguistic behavior in collaborative design processes, given that verbal expression is still an essential vehicle in design and communication processes (Ungureanu and

Hartmann 2021). The coupling of NLP and computational generative design enables a new interface between machines and humans in the different design stages, combining GANs with the capacity of NLP techniques to analyze the semantic design expression, thus giving the process a qualitative breadth. This research explores such a possibility by investigating a future design workflow that combines machine-human intelligence.

Methodology

To address the research question stated above, an experimental research method is used, consisting of a generative process and an analytical interpretation process.

The generative process is composed of the following subprocesses: *(a) dataset preparation that includes data collection, data cleaning, and data filtering; (b) GAN training that includes comparative experiments with SAGAN and DCGAN; (c) latent space interpolation; and (d) selection of GAN images*. The objective of the generative process is to yield a collection of visual imageries to be analyzed and developed as architectural projects.

The analytical interpretation process involves (e) *critical interpretations* of the GAN images that include the explicit articulation of the subject's architectural selection criteria using textual descriptions, raster drawings and vector diagrams, followed by two separate stages: (f) *NLP content analysis*: an objective validation process to extract the conceptual evidence from the subjects' textual interpretations using NLP content analysis; and (g) *architectural implementations*: a qualitative validation of the architectural potential by translating the 2D interpretations of the GAN images into 3D geometrical forms and architectural designs with a program, site, and tectonics.

The experimental research method consists of two distinctive process loops (Figure 2): the generative process (left) and the analytical interpretation process (right). The validation of the research happens in the NLP content analysis (f) and in the architectural implementations (g).

To examine the role of human intelligence and the interplay between human and machines in the GAN-based design process, the methodology introduces punctual human interventions in the process, namely in the selection and preparation of the datasets, tuning the hyperparameters, guiding and highlighting GAN images in the exploration of latent spaces, making final GAN image selections, and in the analytical interpretation and exploitation of the results (Figure 3).

As the GAN is trained and learns from the images, so do the participating humans learn and evolve. In the experiment, humans, based on their mental image of "Swissness" (Zeisel and Eberhard 2006), make an initial hypothesis about "Swissness" that guides their initial selection and preparation of the image dataset. As they interact with the GAN and perceptively absorb the generated images (in a continuous feedback loop), they gradually adjust their mental image to the evolving GAN images and refine their hypothesis about "Swissness."

In the next phase, NLP content analysis was used to extract conceptual evidence and externalize the implicitly hypothesized selection criteria used by the subjects in the previous steps. The method of "text-mining" was employed by using the off-the-shelf Python libraries NLTK and SpaCy as a set of techniques to identify non-trivial trends in the text data (Razno and Khairova 2020). The employed techniques were composed of *preprocessing*, POS *(Part-of-Speech)* tagging, and co-occurrence network analysis. The *preprocessing* stage involved the tabulation, normalization, *tokenization*, and filtering out of stop words from the text. It was followed by a *POS tagging* to identify and count the adjectives, verbs, and nouns used to express architectural concepts or criteria. Finally, the co-occurrence network of the identified concepts was used to discover the relation of one concept to another and also highlight its importance (Matsuo and Ishizuka 2004).

Experiment and Results

The seven stages of the experiment include: (a) dataset preparation, (b) GAN training, (c) latent space interpolation and (d) selection of GAN images, (e) critical interpretations of the GAN images, (f) NLP content analysis, and (g) architectural implementations (Figure 4).

Dataset Preparation

In this first stage of the research, the required data was collected and preprocessed through data cleaning and data curating (Figure 5). Human subjects collected a broad dataset as a starting point for the GAN: in total, over 6,000 images of 900+ Swiss mountain alpine refuges were collected from three online databases (SAC, Capanneti, and Hikr) by using custom scripts to automate the scraping of the images from the website metadata. To automatically detect the building and crop the images, a custom script was written using a pretrained model Faster RCNN+Inception ResNet V2 on Open Images Dataset V4 (Kuznetsova et al. 2020; Huang et al. 2017). An extra manual cleaning process was completed by the subjects to isolate the building part of the image from the environment. As a result, the main dataset was reduced to 4,000+ images of Swiss alpine refuges.

From that, several curated/filtered datasets consisting of 600 to 4,000 images were derived for each subject group based on their initial hypotheses about Swissness. The specificity of the datasets was inversely proportional to the dataset size, usually constrained by the required perspective or specific formal and material composition.

The selection criteria for collecting, filtering, and curating derived from the human subject's initial hypothesis (mental image) about Swissness. This hypothesis about Swissness is not static and evolves in parallel to the machines who learn about Swissness through their exposure to thousands of images used as training data. Figure 6 demonstrates a sample set of selection criteria and decision tree used by one of the participants.

Generative Adversarial Network Training

Both DCGAN and SAGAN were applied to the curated datasets, allowing the comparison of how the choice of GAN could affect the generated outcome (Figure 7). Each was implemented in a Google Colab notebook, which enabled a small-scale machine learning experiment using Google's cloud server.

While both GANs have shown the capacity to learn and generate images of Swiss refuges, different reactions to the dataset and learning results were noted. For small and highly curated datasets (600–1,000 images as in the dataset 1–3) DCGAN was preferred,

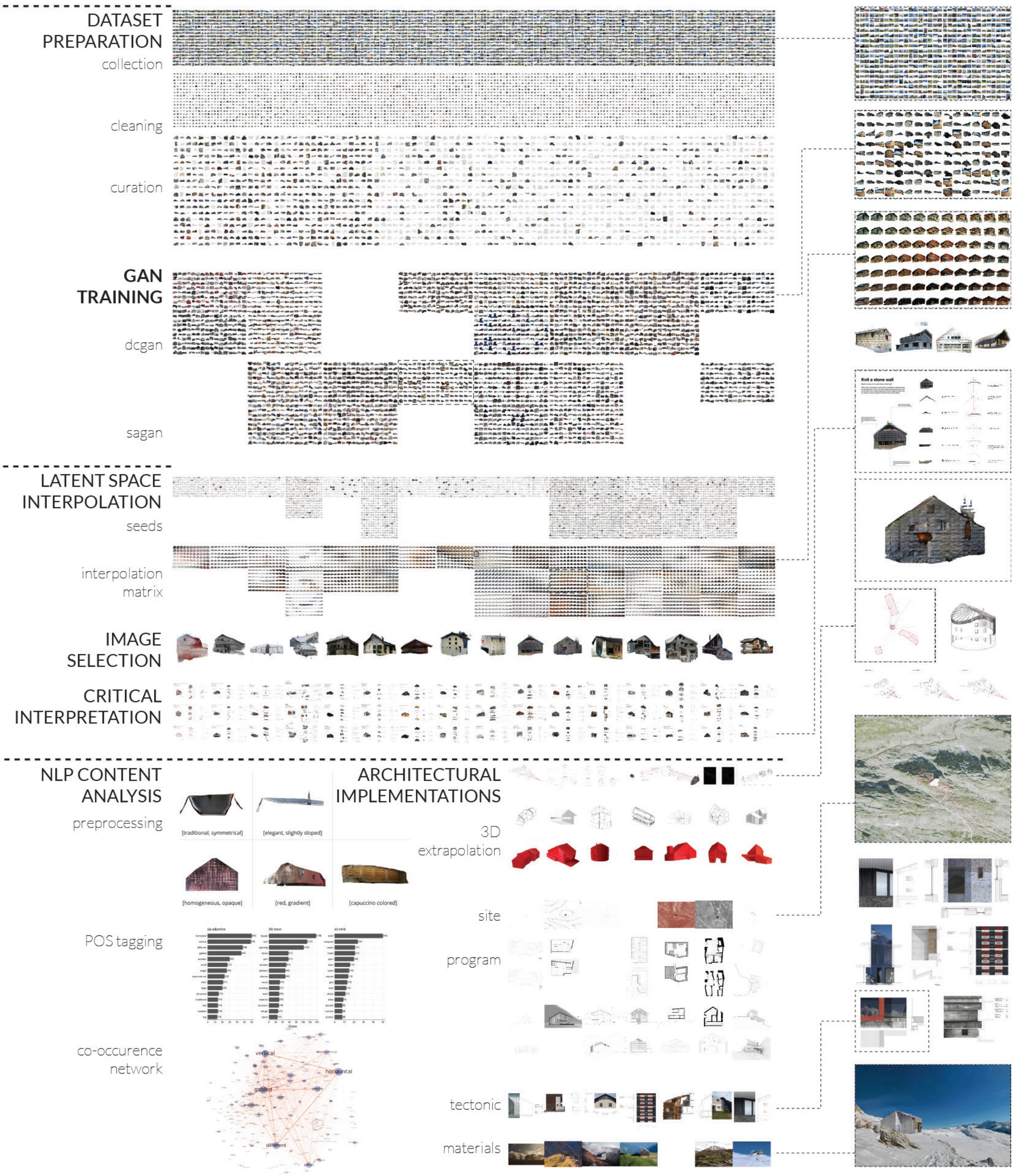
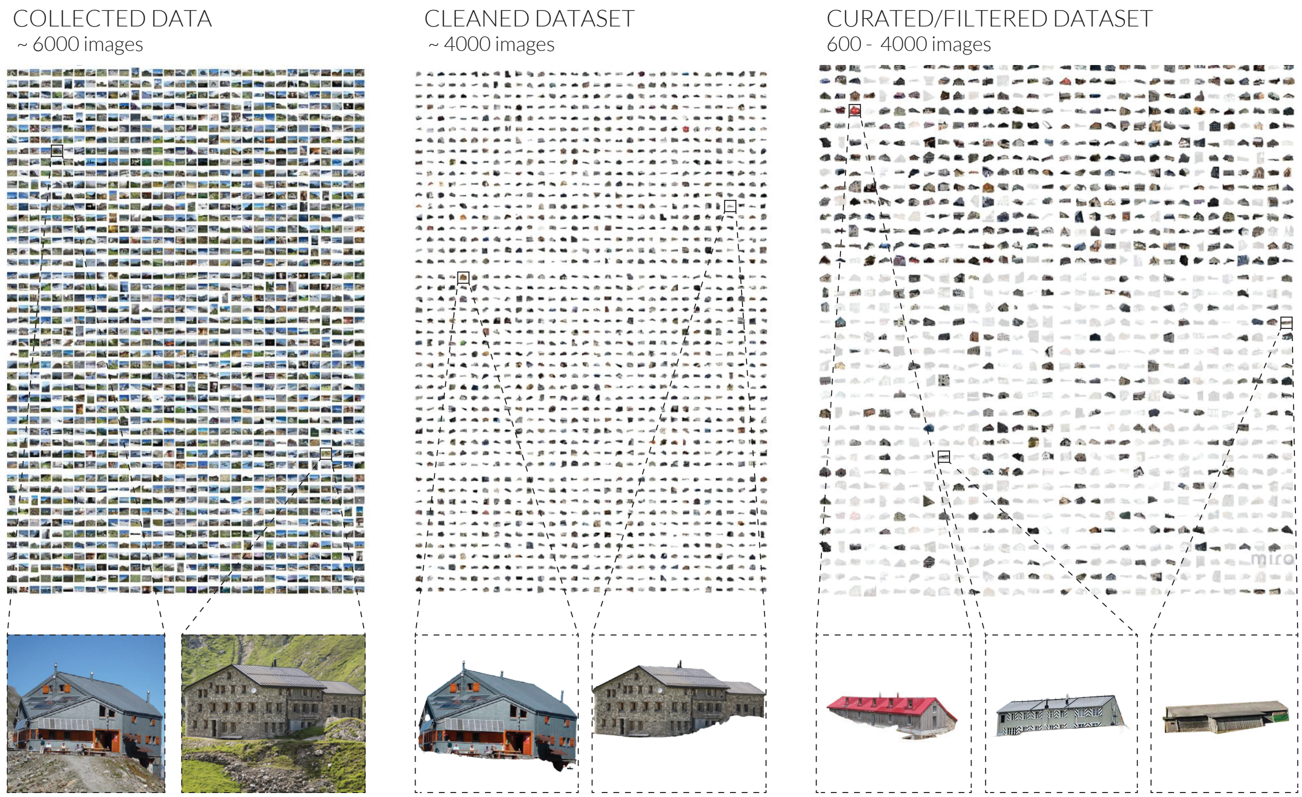
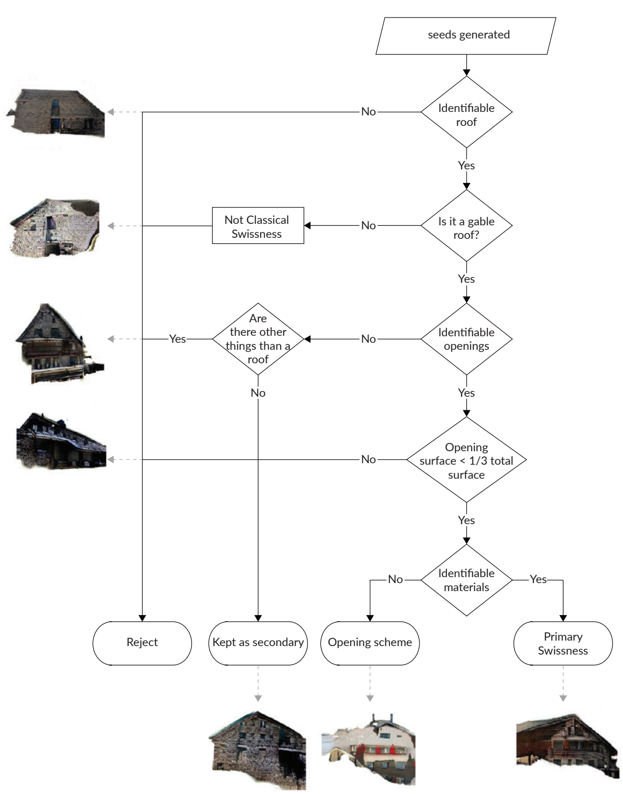


Figure 4. Overview of the experiment and results. (Credit: Media x Design Laboratory, EPFL)



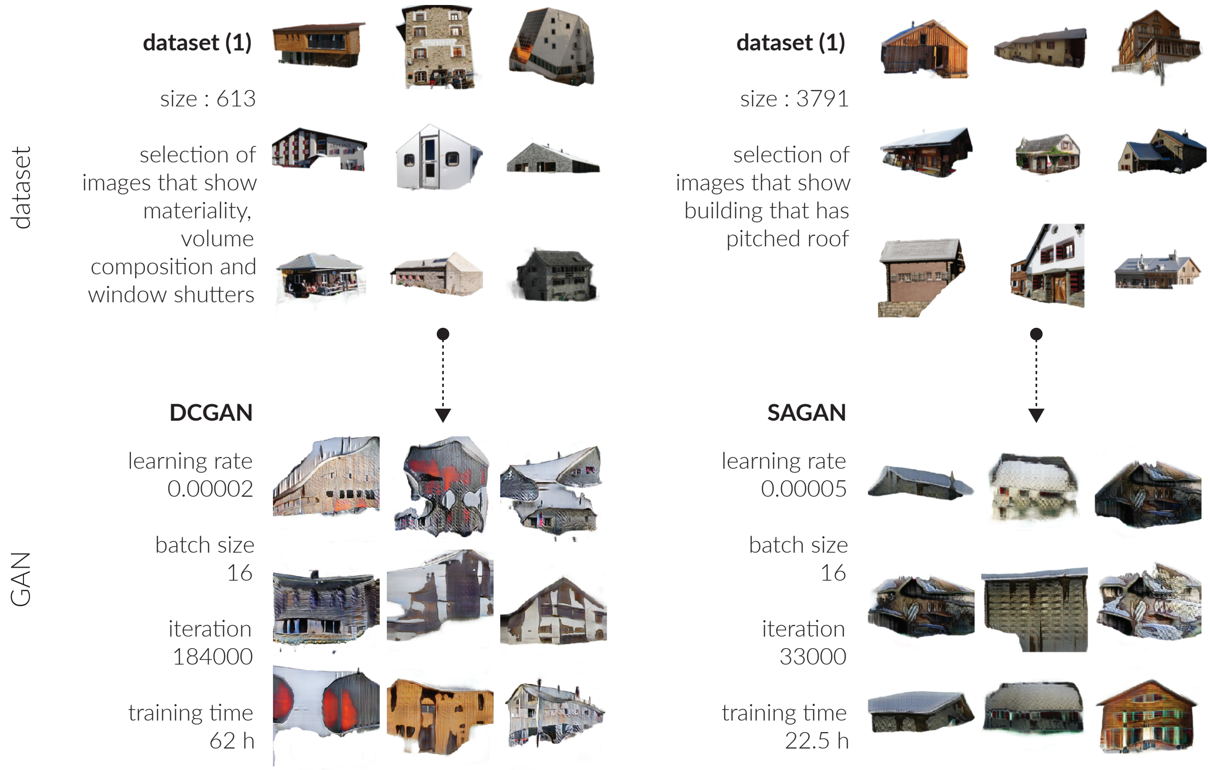
△ Figure 5. The three phases of dataset preparation. (Credit: Media x Design Laboratory, EPFL)



△ Figure 6. Decision tree showing the selection criteria (derived from the subject's mental image/initial hypothesis) and sample visual results. (Credit: Media x Design Laboratory, EPFL)

while for a more generic and larger dataset (from 3,700 to 4,500 images as in dataset 4 and 5) SAGAN was preferred. SAGAN was better able to learn from the greater variations that occur in larger datasets. The generated images from DCGAN were relatively ambiguous in their general form, but in some cases able to show smaller architectural elements in detail, such as windows with diagonal patterns and frames (DCGAN 1–3). On the other hand, SAGANs were able to generalize the bigger patterns of the building, such as the orthogonality and symmetry of a form, yet tended to homogenize textures and overlook architectural details (SAGAN 1–3). It appears that DCGANs mostly learned architectural information that is consistent in the dataset and tend to be blurry in the parts where the given information is more heterogeneous, while SAGANs were capable of capturing and focusing on particular architectural information that was prominent in a large dataset. This intrinsically different behavior of the two GANs was reflected in and influenced noticeably the formal expression of the later derived architectural projects (Figure 8).

The difference in reaction to the dataset between DCGAN and SAGAN could be explained by the addition of the attentional layer of SAGAN. Ordinary DCGANs use growing convolutional layers that can learn the hierarchical representation of the general images as a composition of detailed features learned by the convolutional feature maps (Radford et al. 2016): the generator learns specific representation for major scene components that are repetitively present in the data. Consistency of parts to whole spatial relationships in a small dataset allows the GANs to overfit the dataset, making the GANs able to generate images yet lack generality.



△ Figure 7. Comparison of several training results from DCGAN and SAGAN with different dataset. (Credit: Media x Design Laboratory, EPFL)

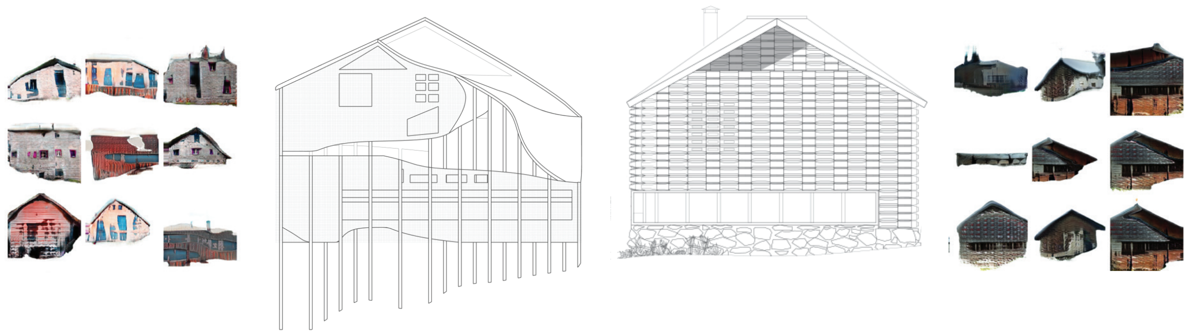
Compared with SAGAN (Figure 9a, 9b), DCGAN can generate a meaningful set of images from a small dataset. However, in a larger but noisier dataset, the inconsistent spatial relationships prevent the convolutional layers from converging and learning features from the dataset (Figure 9c).

Attentional layers in the SAGAN allow the networks to focus on different parts of the image, better capture details and map the global relationship from the images in order to help the network to get the global structure of the images correctly (Zhang et al. 2019). That ability allows SAGANs to work with noisier data and extract the structure of the images. The original experiment by Zhang et al. (2019) also reported that SAGANs are lower-performing with images that are distinguished more by texture than by geometry, which corroborates this experiment. Training with a noisy dataset requires additional hyperparameter tuning. Smaller batch size could have an adversarial effect on the training (Figure 9d), while having higher batch size could potentially improve the variation and quality of the generated images (Figure 9e). In general, having a lower learning rate seems to assure less failure in the training but also increases the time and resources required for computation.

Latent Space Interpolations

Latent space interpolation unfolded the results of the GAN in space, enabling an overview and comparative perceptual understanding of the results. Based on their perception of the generated random images, the human subjects determined which images corresponded to their evolving understanding and mental image of Swissness and guided the generative process accordingly by highlighting specific points/images in the latent space. The corresponding vectors of the highlighted images could then be interpolated to generate another set of GAN images. This process was used iteratively to further refine the hypotheses.

The exploration of GANs' *latent space* was shown in two ways, by using 2D interpolation matrix and linear morphing animations (Figure 10). In principle, a trained GAN model is a map of a point in *latent space* to an image. Generating an image from GAN occurred by feeding the GAN a point from the *latent space* to get the corresponding image. The exploration of GAN's latent space began by yielding numerous random points (144 points) from the latent space to generate images that were visualized in a seed matrix. From that matrix, four to five images were selected alongside with their corresponding points. The selection was made by the human subjects based on their evolving hypotheses about Swissness. The points were then interpolated and the corresponding images generated in a matrix. The initial images could be similar or diverse. The latent space offered a way to study the subtle variations that happen within similar instances and enable a more nuanced selection of a particular image within the gradual spectrum. Alternatively, with strongly diverse images, the latent space offered a visual instrument to explore possible new variations and interpolations. By using the same principle, the interpolation trajectory from the selected images was visualized in the form of an



Project (1) : DCGAN

Project (2) : SAGAN

△ Figure 8. Comparison of project's elevation developed from DCGAN and SAGAN. (Credit: Media x Design Laboratory, EPFL)

(a) SAGAN: failed attempt generating images with a small dataset (242 images).

(b) DCGAN model trained with the same dataset (242 images) and small learning rate (2 e-5).

(c) DCGAN: failed attempt training with 4480 uncurated images.

(d) SAGAN model trained with 4480 uncurated images in small batch size (8) resulting in disfigured images.

(e) Increasing the batch size (32) improves the variation and quality of the generated images in SAGAN model trained with 4480 uncurated images.

◁ Figure 9. Failed attempts in training GANs and different strategies to improve it. (Credit: Media x Design Laboratory, EPFL) animation video. The animation made explicit the transformation from one instance to another as a function of time, recalling the animation based *morphing* design strategy as described by Greg Lynn (Lynn 1999). Accordingly, extra information could be derived from the *interpolatory* trajectory of the elements to enrich the interpretation of the image.

GAN Image Selection

The human subjects selected specific GAN images in the latent space using selection criteria derived from their evolving Swissness hypotheses which continued to be refined in the iterative generate perceive-highlight-interpolate feedback loops (Figure 11).

The results of the GAN image selection process revealed a wide spectrum of options that range from photorealism (close resemblance to real Swiss architecture), neutral abstraction, to almost absurd estrangement: some of the selected GAN images exhibited clearly delineated architectural structures while other GAN images exhibited high degree of blurriness, ambiguity, and glitches for creative imagination.

Critical Interpretation

In this important stage in the research, the selected GAN images were decomposed and interpreted by the subject (Figure 12). After selecting the GAN images, the human subjects performed a critical analysis of the images to reflect upon and externalize their selection criteria and detect further information and intentions that were embedded in the images (Olgiati 2013). GAN images were interpreted in three steps in relation to their hypotheses of *Swissness*: 1) a textual analysis to describe the quality of Swissness in the GAN image; 2) a decomposition of architectural elements from a rastered GAN image; and 3) a reading of the GAN image through diagrammatic analysis using vectorial drawings. The textual description and the drawings made explicit the hypotheses used for the selection criteria. The results of this critical interpretation fed, in parallel, the NLP content analysis and the architectural evidence.

Through the cognitive deconstruction of images, architectural elements such as roof, scale, windows, shingles, colors, materiality, and textures were identified. And a geometrical analysis with grids, measurements, and proportions led to the discovery of architectural qualities such as symmetry, repetition, proportions, and rhythm. In this stage, the research produced a catalog of Swissness, drawn from the textual, raster, and geometric analysis derived from the critical interpretation of a total of 70 GAN images.

NLP Content Analysis

The corpus of critical interpretation was then analyzed further by using NLP techniques to reveal the most frequently used keywords (adjectives, nouns, and verbs) and map their relations to the individual GAN image (Figure 13). From the NLP content analysis, it was shown how the quality of Swissness could be made explicit and the criteria for selection post-rationalized and externalized. The adjectives provided a description of the qualities for *Swissness*, the nouns provided the most frequently mentioned architectural elements, and the verbs provided a reading of the more performative aspects of the architecture.

The three most frequent adjectives (*horizontal*, *vertical*, *different*) in the analysis were considered to be common descriptors of architectural images and not specific to the notion of *Swissness*. They were omitted in the t-SNE visualization in order to emphasize the more specific adjectives (*gabled*, *wooden*, *small*, *single*, *asymmetrical*, etc.). Analogously, the most common nouns (*façade*, *roof*, *opening*) represented the more generic elements of architecture, while the following elements (*stone*, *part*, *window*, *volume*, *wood*, *level*, *material*, etc.) revealed the more particular elements in Swiss refuge architecture.

By projecting the keywords from different projects to the network (Figure 14), the descriptive criteria of images were revealed. While the main nodes expressed the generic elements of architecture such as *roof*, *façade*, and *opening*, a more specific set of descriptors consisting of specific adjectives such as *playful*, *mineral*, *soft*, and *raster* provided additional breadth to the reading of the images. The use of NLP analytical techniques enabled a systematic reading of the qualitative aspect of the GAN generated images.

Architectural Implementations

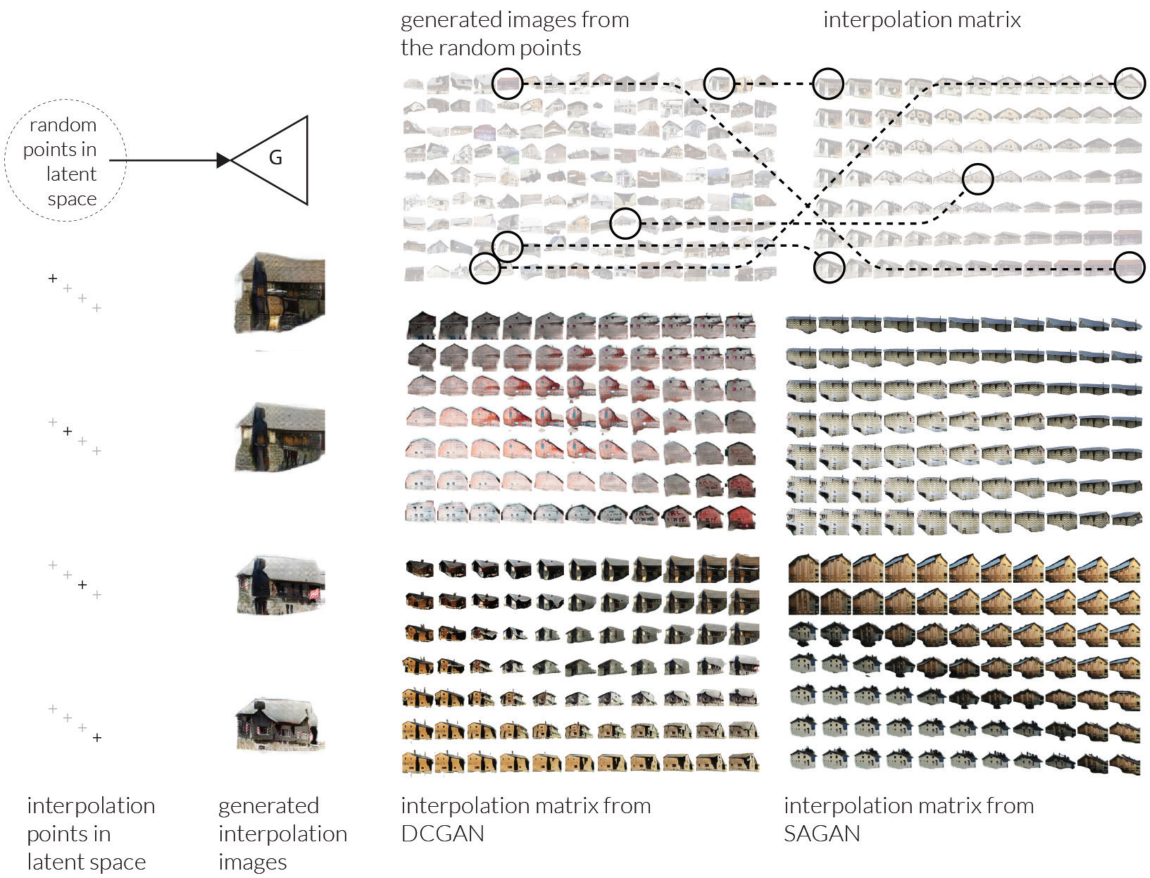
This stage constitutes a transition from research into design. The aim of this stage was to investigate the "architecturalness" of the GAN images, by testing how the GAN images could be translated into the architectural reality of program, site, and tectonics.

The first step was the transformation of the selected GAN image into a 3D architectural volume using methods of projective geometry and input parameters: *Parallel projection* (Figure 15, left) allowed extrusion of the image into a three-dimensional volume. *Inverse projection* (Figure 15, center) used Taylor's theorem to generate orthographic drawings from the perspectival images (Andersen 1992). The method used three vanishing points of the image, to produce two vertical and one horizontal projections for construction of a 3D volume. This approach produced the most accurate 3D information with less creative interpretation by the designer. *Anamorphic projection* (Figure 15, right) was used to reverse engineer the 3D form from a single point of view. It allowed an interpretation by placing a series of construction planes in a 3D space to project the architectural elements from an image. This method produced results that were true to original GAN images with a high degree of freedom in overall form.

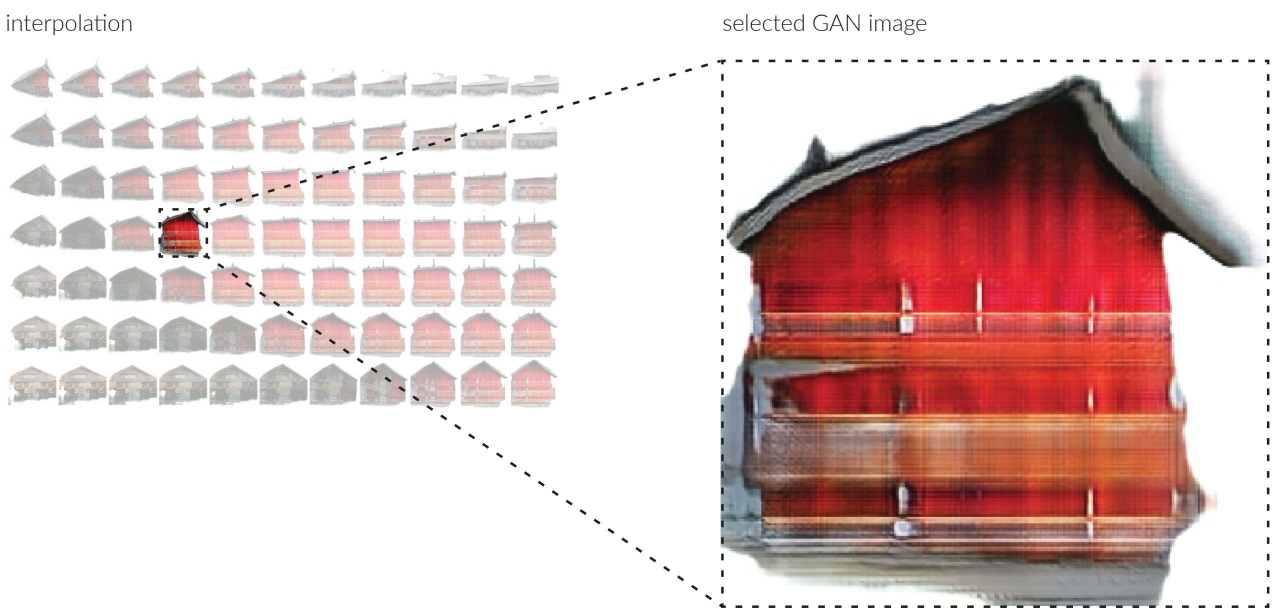
These projective methods allowed for different degrees of generation and interpretation of architectural volumes (Figure 16). The 3D forms were then situated on sites to contextualize architectural instantiations, and incorporated a realistic architectural program of a refuge, a typical Swiss alpine shelter. Finally, the development of material and construction systems was derived as a response to the GAN's expression of materiality (Figure 17). While some GAN images suggested a more conventional construction system, other GAN images, featuring glitches or blurriness, were exploited for proposing new construction systems.

Discussion

This research investigates the combination of human perceptual intelligence with the generative and analytical intelligence of the machines (GANs in general, and DCGAN and SAGAN in particular, and NLP) in the creation of meaningful architectural designs.

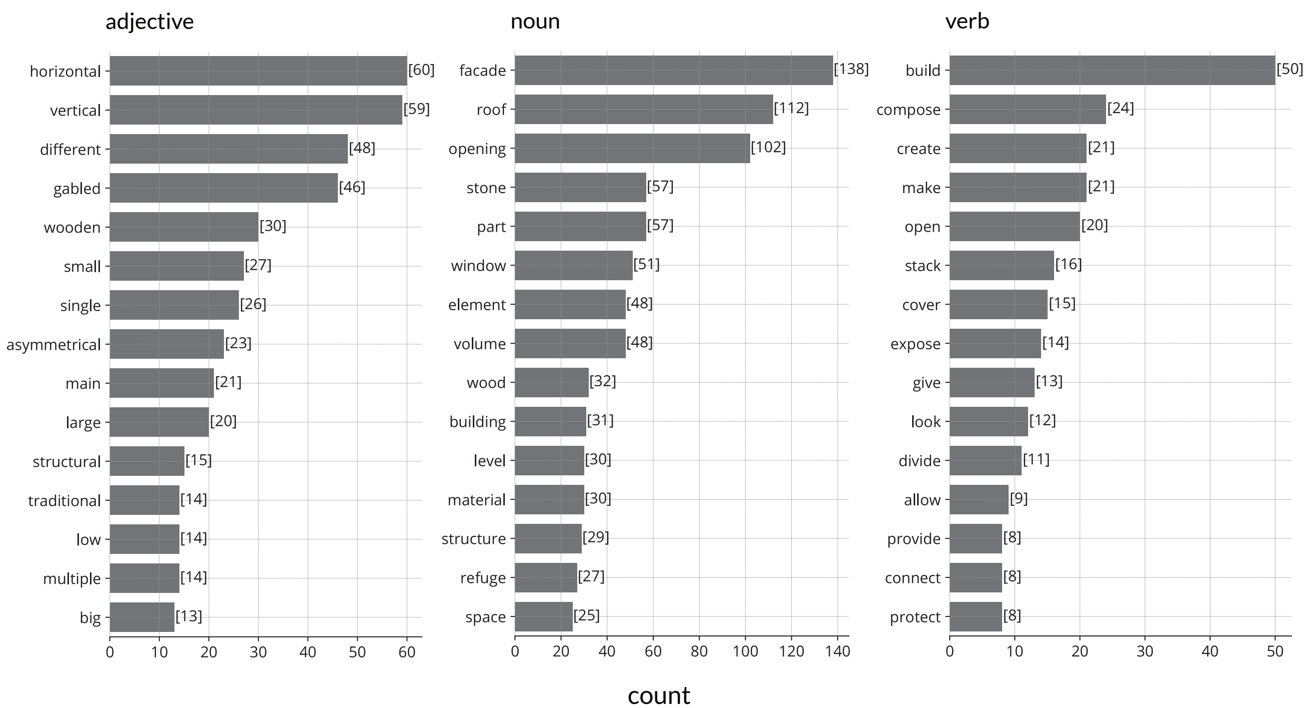


△ Figure 10. Several examples of the interpolation matrix from both SAGAN and DCGAN. (Credit: Media x Design Laboratory, EPFL)

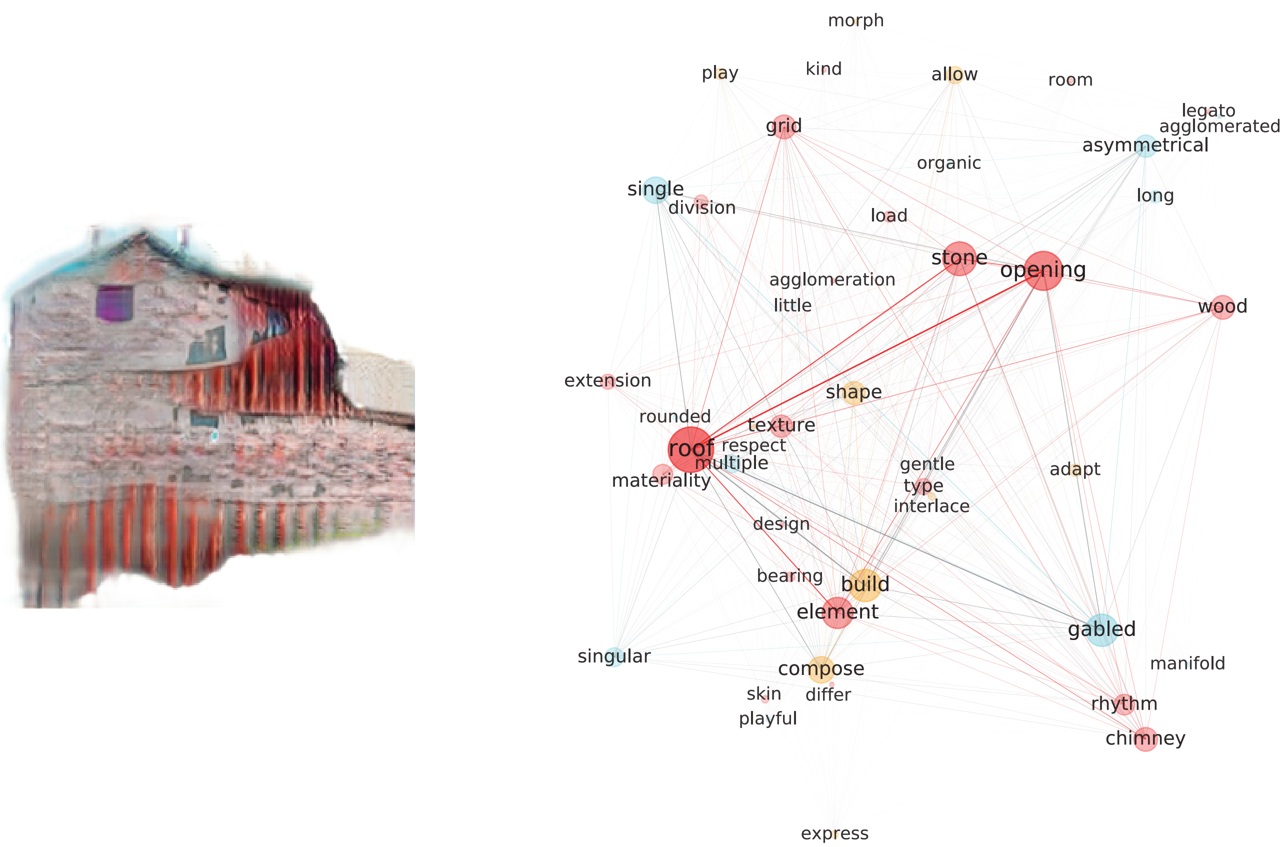


△ Figure 11. Selected GAN generated image. (Credit: Media x Design Laboratory, EPFL)

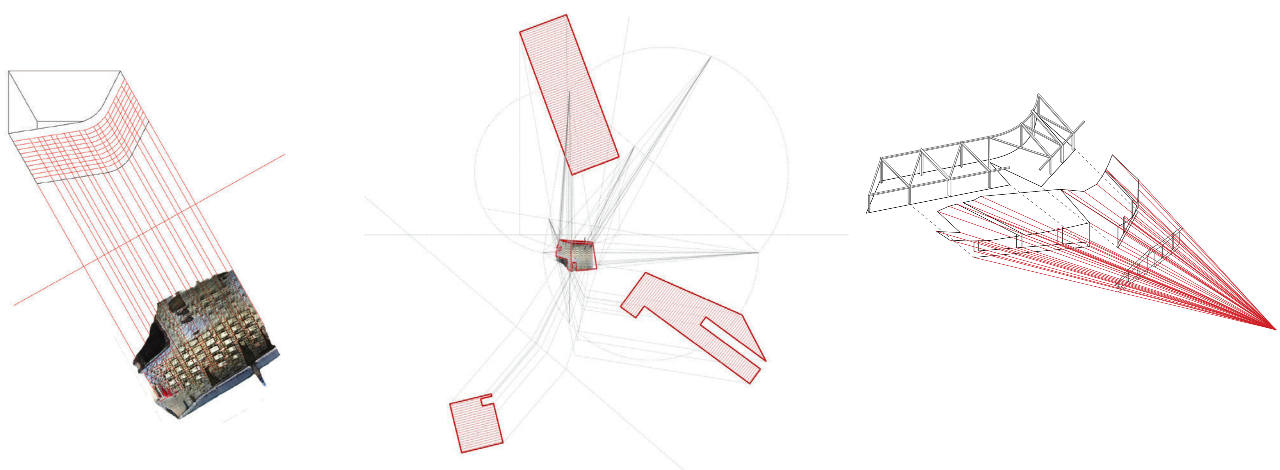
△ Figure 12. The critical interpretation of a GAN generated image. (Credit: Media x Design Laboratory, EPFL)



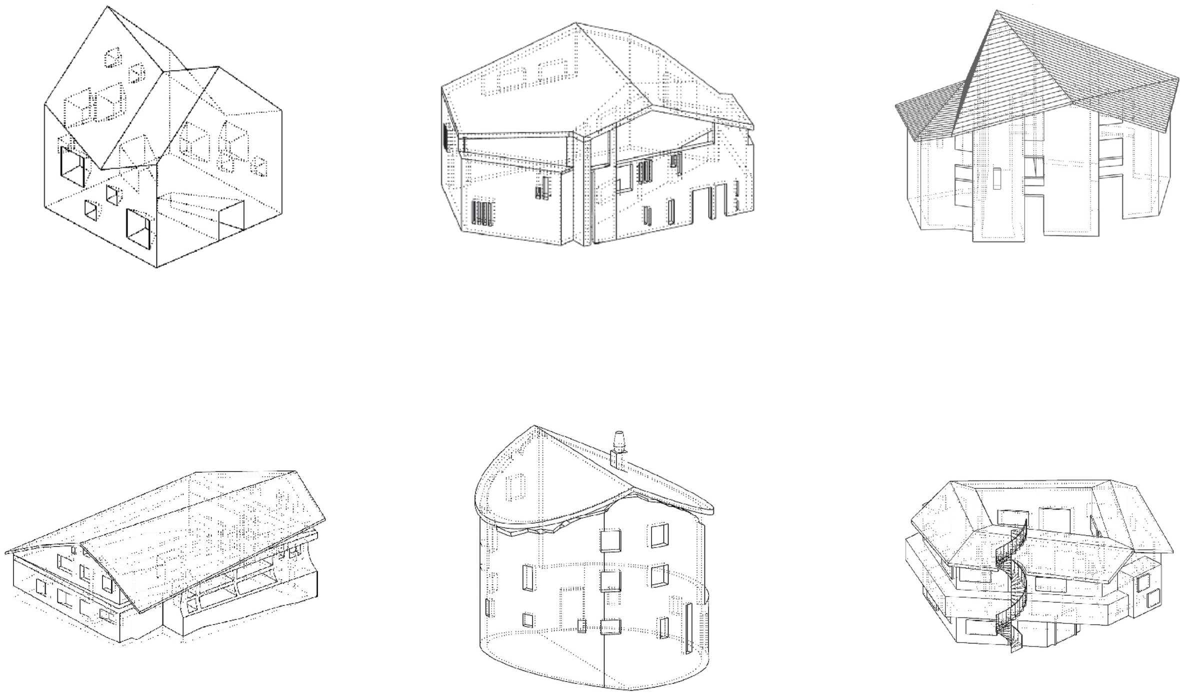
△ Figure 13. The histogram shows the frequency of identified adjectives, nouns, and verbs gathered from NLP content analysis of critical interpretations of the GAN image. (Credit: Media x Design Laboratory, EPFL)



△ Figure 14. Constellation of keywords extracted from NLP analysis of a GAN image. (Credit: Media x Design Laboratory, EPFL)



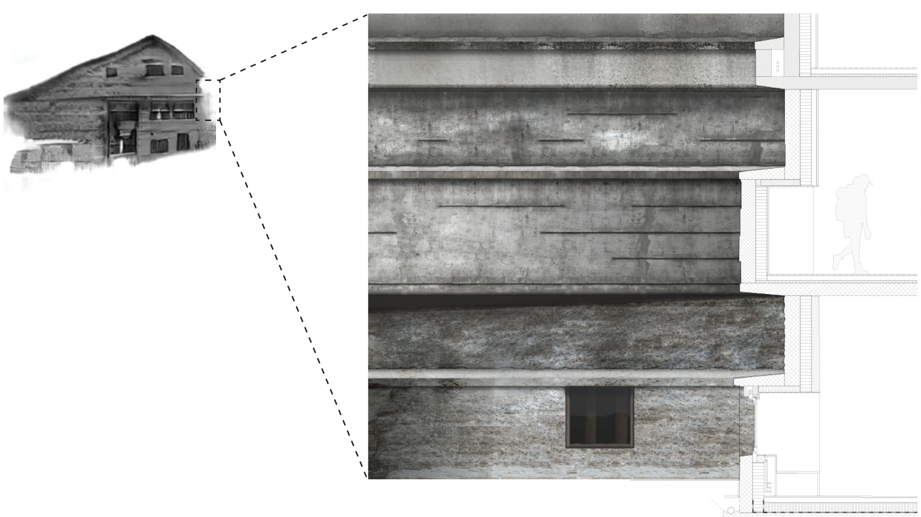
△ Figure 15. The projective approaches in translating 2D images to 3D volumes. Parallel projection (left), inverse projection (center), and anamorphic projection (right). (Credit: Media x Design Laboratory, EPFL)



△ Figure 16. The resulting 3D models in turntable animations. (Credit: Media x Design Laboratory, EPFL)

GAN image

façade section



Wall construction for upper floors

Reinforced concrete:	250 mm
Thermal insulation, vaporproof	200 mm
Vertical battens, wood cement	60 mm

Floor construction

Ready-to-lay parquet flooring	20 mm
Screed	80 mm
Separating layer, sound insulation	40 mm
Concrete slab	200 mm

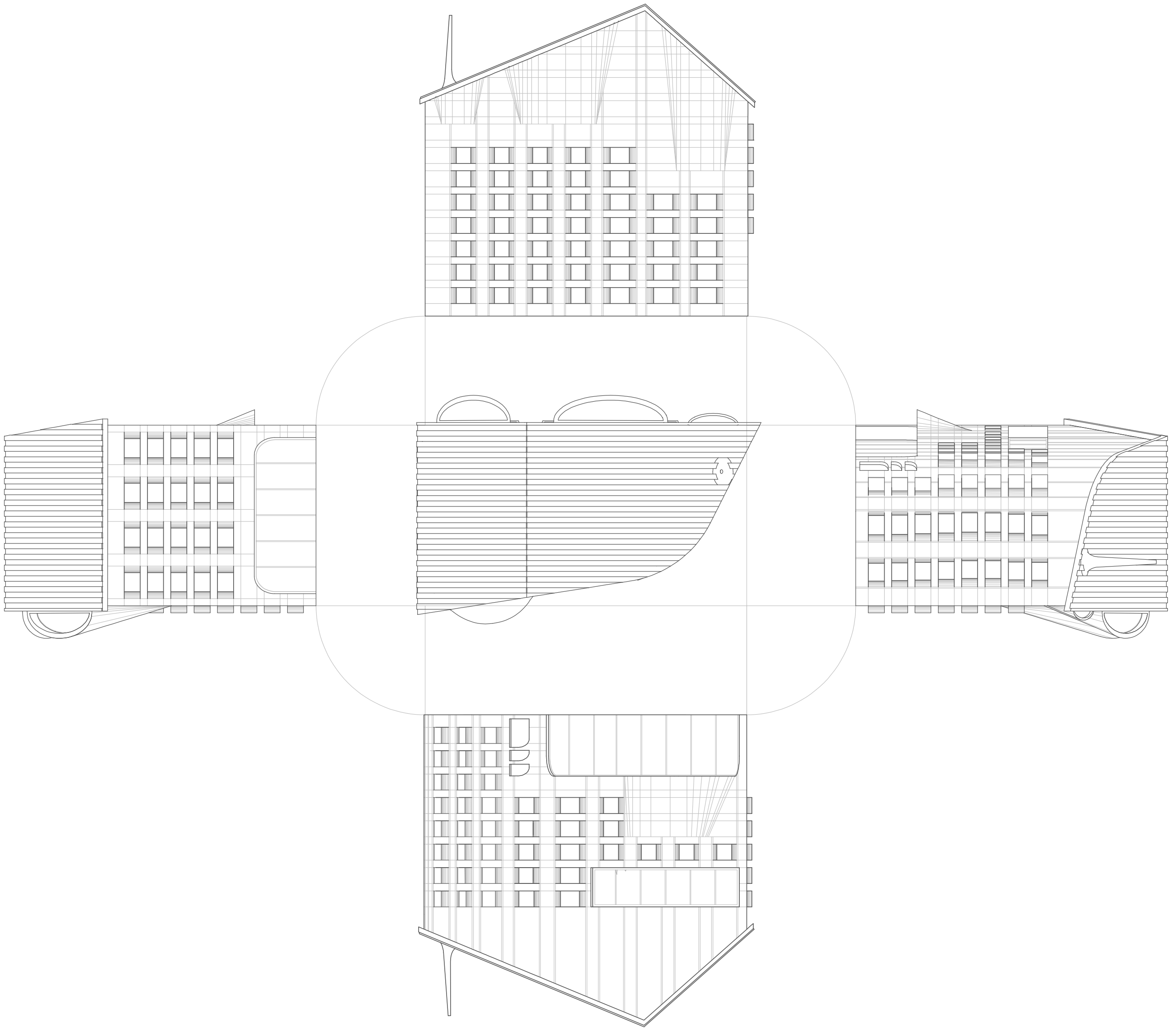
Wall construction ground floor

Reinforced concrete	250 mm
Thermal insulation, vaporproof	200 mm
Concrete elements polished finish	120 mm

Foundation construction

Floor covering stone tiles	15 mm
Screed	80 mm
Thermal insulation	150 mm
Damp-proof membrane	2 mm
Concrete slab	200 mm

△ Figure 17. Façade section in detail at 1:20 scale. (Credit: Media x Design Laboratory, EPFL)



△ Figure 18. Example of an architectural project cogenerated by humans and machines, exposing the creative tension between contextuality (Swissness) and inventiveness (architectural novelty) perpetuated by the SAGAN. (Credit: Media x Design Laboratory, EPFL)

Limitations

While the use of a narrow topic (Swissness, Alpine architecture) allowed for a relatively controlled experiment and systematic observation of the cause and effect of different interventions, there are certain limitations of this approach. The main limitation relates to the size of the data set and access to quality data. For this study, data was scrapped from open data sources on the Web. Future research could expand the size and quality of the data set, which would increase selectivity and curability.

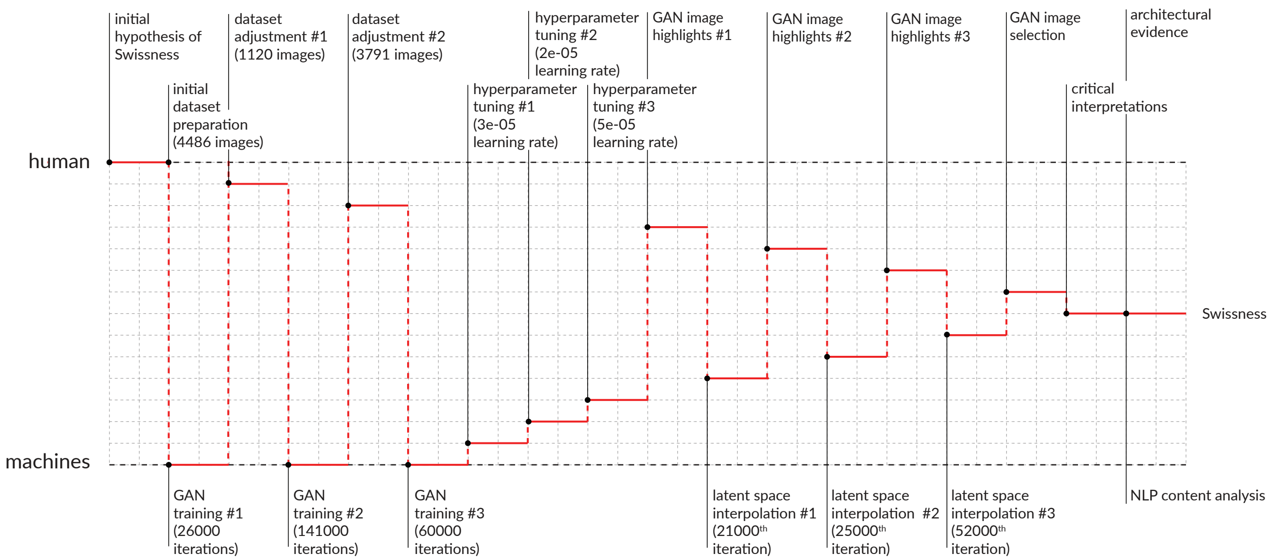
Additionally, the small number of human subjects involved in the experiment (20) limits the generalizability of the research. Future research could expand on the size and the diversity in the composition of the subjects. Finally, there are limits related to the GAN technology.

Reflections

Despite these limitations, several observations can be made that contribute to our understanding of the interplay between GANs, NLP, and architectural design.

(1) The choice of GAN techniques matters. Various GAN models possess different learning abilities, thus the choice of the technique directly informs design results. The experiments using DCGAN and SAGAN demonstrated that the former worked better with fine textures and small details, while the latter performed better in generalizing the building form.

(2) GANs generate designs with creative ambiguity. GAN



△ Figure 19. Overview diagram illustrating the interplay between machines and humans in a typical GAN image generation and selection workflow across all stages of the research design. (Credit: Media x Design Laboratory, EPFL

images balance between being contextual and innovative (Figure 18). GANs can learn from precedents and inherit cultural and regional visual qualities (Frampton 1983), yet open up room for invention by introducing imprecisions and glitches (Austin and Matthews 2018). The imprecisions and ambiguity which disrupt the normative disciplinary representation entail the discovery of new spatial and formal arrangements and new material combinations, facilitating creative interpretations by human perception.

(3) GANs suggest new relationships between human and machine (Figure 19). The use of AI technologies fundamentally questions the notion of authorship in design. The shift of human agency from a designer to a curator of datasets, a training partner, and a design jury can be both uneasy as well as liberating and exciting.

(4) The training of GANs involves a close causality between the characteristic of the dataset, the algorithm, and hyperparameters which need to be optimized to generate a desirable result. The hyperparameter optimization of the GAN training was achieved through an iterative process and qualitative evaluation of generated images, thus linking the performance of GAN to the experience and assumptions of the subjects. The current quantitative method to measure generative models is still a suboptimal solution for measuring the quality of generated images that are more nuanced than other machine learning problems (Barratt and Sharma 2018). It is clear that quantitative measurement alone will not be sufficient in measuring the more creative nuances of GAN models.

Conclusions

This research combines the generative capacity of GANs with the analytical and reflective capacity of NLP methods. A future research in streamlining a more nuanced qualitative-quantitative approach in measuring machine learning performance could potentially be used to address the aesthetic and ethical dimensions of the machine learning approach and to anticipate its undesirable consequences.

Additionally, the research experiment articulated potential future design workflows that combine machine-human intelligences. The NLP methods used in the experiment allow a systematic analysis of human interpretation of GAN generated images in order to provide a more holistic measurement of the use of GANs in architecture. The research could be extended by involving a crowdsourced human evaluation of GAN generated images in which NLP methods could be leveraged to analyze a large corpus of evaluations. The coupling of generative (GAN) and analytical (NLP) capacities of machines, and their combination with the perceptual intelligence of humans into a collective intelligence, could usher in entirely new approaches to the design of future architectures and cities, where the boundaries between human and artificial agency have become increasingly blurred.

Acknowledgements
Experimental results described in this research include contributions by Samuel Aeschimann, Enzo Arioli, Aina Rodríguez Bofill, Romain Claus, Philippe Audemard d'Alançon, Núria Fàbrega, Leslie Faisan, Basil Ferrand, Christophe Gautier, Thomas Gobet, Mathieu Iffland, Simone Izzo, Yasmine El Karamoudi, Clara de Lapuerta, Tamara Lobo, Nicolas Mahren, Benoît Meugniot, Eric Nardini, Kimberlyne Nguyen, Lucía Quintela, Salla Sivunen, and Ruojun Zhao. The authors also wish to thank Michael Hensel, Tamotsu Ito, Immanuel Koh, Eric Lapierre, Enriqueta Llabres-Valls, Nicole McIntosh, Peter Ortner, Philipp Schaerer, Adrien Verschuere, and Andrew Witt for helpful discussions and review of the work described in architectural implementations.

Data Statement
Data available on request from the authors.

References

Andersen, K. 1992. *Brook Taylor's Work on Linear Perspective: A Study of Taylor's Role in the History of Perspective Geometry.* Springer Science & Business Media.

Arjovsky, M., S. Chintala, and L. Bottou. 2017. "Wasserstein GAN." ArXiv Preprint. ArXiv:1701.07875. December. https://arxiv.org/abs/1701.07875.

As, I., S. Pal, and P. Basu. 2018. "Artificial Intelligence in Architecture: Generating Conceptual Design via Deep Learning." *International Journal of Architectural Computing* 16:4: 306–27. https://doi.org/10.1177/1478077118800982.

Austin, M. and L. Matthews. 2018. "Drawing Imprecision: The Digital Drawing as Bits and Pixels." In *Recalibration on Imprecision and Infidelity—Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2018*, 36–45.

Barratt, S. and R. Sharma. 2018. "A Note on the Inception Score." ArXiv:1801.01973 [Cs, Stat]. June. http://arxiv.org/abs/1801.01973.

Bressani, M. and A. Sprecher. 2019. "Atmospheres." *Journal of Architectural Education* 73:1: 2–4. https://doi.org/10.1080/10464883.2019.1560792.

Caruso, A. 2009. "Whatever Happened to Analogue Architecture." *AA Files* 59: 74–75.

Chaillou, S. 2019. "AI + Architecture: Towards a New Approach." Thesis, Harvard University Graduate School of Design.

Dong, A. 2007. "The Enactment of Design through Language." *Design Studies* 28:1: 5–21. https://doi.org/10.1016/j.destud.2006.07.001.

Frampton, K. 1983. "Towards a Critical Regionalism: Six Points for an Architecture of Resistance." In *The Anti-Aesthetic: Essays on Postmodern Culture*, edited by Hal Foster, 1st ed., 16–30. Port Townsend, WA: Bay Press.

Goldschmidt, G. and M. Smolkov. 2006. "Variances in the Impact of Visual Stimuli on Design Problem Solving Performance." *Design Studies* 27:5: 549–69. https://doi.org/10.1016/j.destud.2006.01.002.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems*, 2672–80.

Gulrajani, I., F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. 2017. "Improved Training of Wasserstein GANs." In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5769–79. Red Hook, NY: Curran Associates Inc.

Hindupur, A. 2018. "The GAN Zoo." *Medium* (blog). Accessed June 26, 2021. https://deephunt.in/the-gan-zoo-79597dc8c347.

Huang, J., V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, et al. 2017. "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors." ArXiv:1611.10012 [Cs]. April. http://arxiv.org/abs/1611.10012.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros. 2018. "Image-to-Image Translation with Conditional Adversarial Networks." ArXiv:1611.07004 [Cs]. November. http://arxiv.org/abs/1611.07004.

Karras, T., T. Aila, S. Laine, and J. Lehtinen. 2018. "Progressive Growing of GANs for Improved Quality, Stability, and Variation." ArXiv:1710.10196 [Cs, Stat]. February. http://arxiv.org/abs/1710.10196.

Karras, T., S. Laine, and T. Aila. 2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." ArXiv:1812.04948 [Cs, Stat]. March. http://arxiv.org/abs/1812.04948.

Kuznetsova, A., H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, et al. 2020. "The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale." *International Journal of Computer Vision* 128:7: 1956–1981. https://doi.org/10.1007/s11263-020-01316-z.

Lawson, B. 2009. *What Designers Know.* Repr. Amsterdam: Elsevier/Architectural Press.

Lynn, G. 1999. *Animate Form.* 1st ed.. New York: Princeton Architectural Press.

Matsuo, Y. and M. Ishizuka. 2004. "Keyword Extraction from a Single Document Using Word Co-occurrence Statistical Information." *International Journal on Artificial Intelligence Tools* 13:01: 157–69. https://doi.org/10.1142/S0218213004001466.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. 2013. "Efficient Estimation of Word Representations in Vector Space." In *ICLR*, ArXiv Preprint. arXiv:1301.3781. September. https://arxiv.org/abs/1301.3781.

Newton, D. 2019. "Generative Deep Learning in Architectural Design." *Technology | Architecture + Design* 3:2: 176–89. https://doi.org/10.1080/24751448.2019.1640536.

Olgiati, V. 2013. *The Images of Architects: 44 Collections by Unique Architects.* Luzern, Switzerland: Quart Verlag.

Purcell, A.T. and J. S. Gero. 1996. "Design and Other Types of Fixation." *Design Studies* 17:4: 363–83. https://doi.org/10.1016/S0142-694X(96)00023-3.

Radford, A., L. Metz, and S. Chintala. 2016. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." ArXiv:1511.06434 [Cs]. January. http://arxiv.org/abs/1511.06434.

Razno, M. and N. Khairova. 2020. "Using NLP Python Tools in Methods of Content Analysis." In *Computational Linguistics and Intelligent Systems: Proceedings of the 4nd International Conference (2), 2020*, 240–42. Lviv Polytechnic National University.

Roskos-Ewoldsen, B, M. J. Intons-Peterson, and R. E Anderson. 1993. *Imagery, Creativity, and Discovery: A Cognitive Perspective.* Burlington: Elsevier.

Srinivasa-Desikan, B. 2018. *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, SpaCy, and Keras*. Expert Insight. Birmingham Mumbai: Packt.

Tarabishy, S., S. Psarras, M. Kosicki, and M. Tsigkari. 2019. "Deep Learning Surrogate Models for Spatial and Visual Connectivity." *International Journal of Architectural Computing* 18:1: 53–66. https://doi.org/10.1177/1478077119894483.

Ungureanu, L.-C. and T. Hartmann. 2021. "Analysing Frequent Natural Language Expressions from Design Conversations." *Design Studies* 72 (January): 100987. https://doi.org/10.1016/j.destud.2020.100987.

Wang, H., J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo. 2017. "GraphGAN: Graph Representation Learning with Generative Adversarial Nets." ArXiv:1711.08267 [Cs, Stat]. November. http://arxiv.org/abs/1711.08267.

Zeisel, J. and J.P. Eberhard. 2006. *Inquiry by Design: Environment/Behavior/Neuroscience in Architecture, Interiors, Landscape, and Planning*. Revised edition. New York: W. W. Norton.

Zhang, H. 2019. "3D Model Generation on Architectural Plan and Section Training through Machine Learning." *Technologies* 7:4: 82. https://doi.org/10.3390/technologies7040082.

Zhang, H., I. Goodfellow, D. Metaxas, and A. Odena. 2019. "Self-Attention Generative Adversarial Networks." ArXiv:1805.08318 [Cs, Stat]. June. http://arxiv.org/abs/1805.08318.

Jeffrey Huang is the Director of the Institute of Architecture at EPFL, Head of the Media x Design Lab, and a Full Professor in Architecture and Computer Science at EPFL. His research examines the convergence of physical and digital architecture. His recent work on Artificial Design (Design Brain) will be featured at the Seoul Biennale of Architecture and Urbanism 2021.

Mikhael Johanes is a Doctoral Assistant at Media x Design Laboratory EPFL. His works examine the computational medium with its generative and analytical possibilities in architectural design. He won Outstanding Papers and Design Works in the UIA 2017 Seoul World Architect Congress and has also published in several international journals.

Frederick Chando Kim is a doctoral candidate at Media x Design Laboratory at EPFL. His research focuses on form, geometry, and machine learning. His work in the Slovenian Alps has been awarded an AIA New England Design Honor Award and a nomination for the EU Mies Award.

Christina Doumpioti is a Ph.D. researcher and educator at EPFL's Media x Design Lab. Her research and teaching focus on new connections between design and mechanic intelligence related to intensive spatial attributes and microclimate. Since 2012 she has been a member of the collaborative design network OCEAN net.

Georg-Christoph Holz is a German architect based in Lausanne Switzerland. He studied at London Metropolitan University, Brighton University, the Royal Danish Academy, and graduated in EPFL in 2014. He is Co-founder of Baraki Architecture and Engineering and Scientific Assistant in the Media x Design Lab at EPFL.