Archi-Base: Automated Dataset Construction of Architectural Imagery for Deep Neural Networks

Michael Hasey [1], Sage Elliott [2] and Jinmo Rhee [3]

[1] Carnegie Mellon University, Computational Design Laboratory, Pittsburgh PA 15213, USA
[2] Sixgill, San Francisco CA 94104, USA
[3] Carnegie Mellon University, Computational Design Laboratory, Pittsburgh PA 15213, USA

mhasey@andrew.cmu.edu

Abstract. This paper proposes "Archi-Base", a supportive tool for the quick and automated generation of large custom datasets of sorted and labelled architectural imagery for deep neural network (DNN) training within architectural research. Despite DNN's potential to advance architectural design and knowledge, their use has been limited due to the inordinate amount of time and labor needed to construct their large training datasets. Architectural imagery databases must be manually created due to a lack of support tools and source data being spread widely across multiple platforms and may take several days to properly compile, label and prepare. Archi-Base autonomously constructs these custom large datasets in a fraction of the time, thus removing a significant barrier towards DNN's wider adoption within architectural research. To do this, Archi-Base uses a three-step pipeline that mimics the manual process typically used when creating custom image databases: subject and size identification, image scraping, and image labeling. In the first step, the user identifies the dataset subject and size desired (e.g. Zaha Hadid, 20,000 images). In the second step, Archi-Base autonomously searches for and aggregates images matching the identified subject from multiple online databases of publicly available information. In the final step, a pre-trained image classifier model sorts and labels the images according to their content type. For example, interior images of buildings, exterior images of buildings, aerial images of buildings, parts of buildings, or building textures. Unrelated images are discarded. The result is a sorted and labelled dataset that matches the subject and size specified by the user. Experiments were then conducted to validate the quality and robustness of an Archi-Base dataset with deep neural networks. For example, a 50,000 images dataset of Brutalist style architecture compiled by the Archi-Base tool was used to train an IntroVAE Convolutional Neural Network using both traditional VAE frameworks and Generative Adversarial Neural Networks [14]. By analyzing the results, we developed and employed a criteria matrix to qualitatively evaluate the performance of the dataset. Based on this evaluation, the quality of the dataset was sufficient for deep neural networks to effectively capture the defining architectural features and image composition of the subject. However, two bottlenecks were identified: errors in synthesis and a lack of sufficient dataset diversity to fully represent the architectural style inherent in the specified subject. To address these issues, we might try to include a wider range of image types (plans, sections, elevations, etc.) and test against several other models to more precisely evaluate the qualitative performance of the datasets in future studies. Despite these challenges, developing Archi-

Base is an effort to systematically collect, sort and label architectural image data for DNN training, drastically reducing datasets creation time (e.g. about 18.97 times faster) and increasing the use of deep neural networks within architectural research.

Keywords: Training Dataset, Deep Neural Networks, Machine Learning Datasets, Deep Learning Datasets, Model Training, Data Aggregation, Architectural Image, Web Crawler

1 Background

1.1 The Systemization of Architectural Design

Within the last century, architectural design has slowly moved away from its handcrafted and manual origins towards an increasingly systemized and computational approach. As part of a four step process, this transition first began with modular design methods in the 1930's where strict grammars and mathematical logic drove resultant form , to the advent of computer aided design in the 1960's, to the complex forms generated by parametric design in the early 2000's, and most recently, to the seemingly autonomous and post-human abilities of machine learning and artificial intelligent systems [1]. In conjunction, these four steps represent a slow march towards an increasingly systematized and computational architectural design practice.

Beyond design tools and methodologies, technological progress has also shifted design agency away from the individual and towards the algorithmic tool and design system. These changes became most prominent during the height of the parametric design stage when complex rule-based design systems and logics were used to create designs far more complex than could have been achieved manually. Such inhuman ability began to challenge the agency of the single designer against the computational design system itself. Though many of these designs remained as digital representations due to their sheer construction complexity and scale, many did escape into the real world as manifested, and physical architectural works. Many of Zaha Hadid's, Morphosis's, and Coop Himmelb(l)au's work represents this stage of parametric and systemized computational design. Though, earlier experiments in parametricism extend back to the 1930's with Frei Otto's work with tensile structures, the full manifestation of the rule-based concepts and driving foundations wouldn't emerge until decades later.

The recent integration of A.I. into architectural design, however, has introduced a major shift where rule-setting is no longer the responsibility of the individual but rather of the artificially "intelligent" system. Using data as a resource, intelligent systems are now capable of autonomously setting their own parameters (as opposed to a parametricism, where they are defined by humans) based on "*information either collected from data or transmitted by the user*" [1]. As a result of this increasing independence, AI has broken through the previous human-dependent systems via its ability to autonomously digest large amounts of information and create solution forms through the exploitation of its own landscape of autonomously defined parameters and logics [1].

This new paradigm has recently caused an exponential shift in the way we might now view, analyze and solve architectural design problems; seemingly away from the individual and all-knowing genius, towards the wisdom of the intelligent system. As a result, a great deal of research has now been devoted towards AI, and how it might alter the way in which buildings are designed, analyzed, and optimized. Already, various companies, institutions, research groups, and individuals are using A.I. methods to either supplant previously human-dominated work or to augment and enhance human ability to carry out certain tasks through increased efficiency, design capability or optimization.

1.2 Data Collection Problem

A common assumption made when considering A.I. technology is that it consistently outperforms human cognitive ability. While true in many cases, A.I. performance is entirely dependent on the quality of training data it "learns" and extracts "knowledge" from. Therefore, dataset quality and robustness are key indicators of an A.I. models potential performance. However, dataset creation has become one of the most challenging and time-consuming tasks in the implementation of AI methods, leading to major slowdowns and problems for both researchers and practitioners alike.

The cause of this challenge is complex and varies depending on the type of training data needed, the particular algorithmic model used, and the particular project goal or intent defined. For the majority of cases however, large, targeted, and often prelabelled datasets (if conducting supervised learning) are typically required to feed into a model. But as these datasets do not exist and often need to be constructed manually, an inordinate amount time must be spent searching for, amassing and preparing massive datasets for deep learning projects.

Though a universal challenge, there have been surprisingly few studies undertaken or resources made available to help position researchers and practitioners within the complex web of data collection methods and provide solutions for this problem. One recent attempt however was proposed in the paper "A Survey on Data Collection for Machine Learning" [2] which lays out the landscape of existing data collection and dataset construction methods and provided guidelines to help clarify how one can efficiently and quickly build their own dataset depending on individual project needs and parameters. However, this guideline is not industry specific, and to the best of the authors knowledge, no practical architectural-oriented resource exists to help architects build image datasets quickly or efficiently for image-based deep learning research and work. Within this landscape of uncertainty, Archi-Base aims to provide a simple and straightforward method of dataset building for those seeking out large, labelled and sorted sets of architectural imagery.

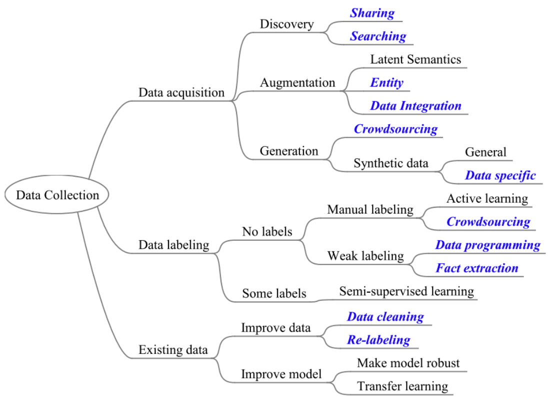

Fig. 1. A high-level research landscape of data collection methods for machine learning as proposed in "A Survey on Data Collection for Machine Learning" [2]

1.3 Architectural Data: The Image

Within the realm of architecture, image data has become one of the most exploited resources for training data within DNN architectural research projects. Within this context, there are two primary architectural research streams that use imagery as a main resource; architectural design research, and architectural analysis research.

Within the architectural design research stream, recent studies have explored how particular AI algorithms such as GANs, IntroVAE, Style Transfer systems, or other image synthesis models can assist with, augment, or supplant the traditional design process with new designs. Projects like "AI & Architecture: Towards a New Approach" [3] and "House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation" [4] have revealed how AI deep learning methods such as GANs and IntroVAEs can autonomously learn patterns inherent in images of floor plans and then recreate thousands of novel and feasibly constructable floor plans in any given style in extremely short periods of time. Other projects such as "Machine Hallucination: NYC" by Refik Anadol Studio operate in the more artistic realm of AI generated design and uses AI to analyze thousands of images of city scenes in order to recreate wonderous and dreamlike recreations in similar yet eerily unique aesthetic compositions. Furthermore, companies like XKool Technologies are developing new sketch-assist programs that use large databases of architectural photography imagery to transform simple user-generated sketches into realistic architectural renderings in real time [4]. Together, these projects leverage AI's ability to mine knowledge from existing architectural imagery as means to augment and enhance our creative abilities through systematized AI assisted processes.

Fig. 2. XKool Technology uses Deep Learning to translate sketches into realistic architectural renderings. Large datasets of housing image data were likely used for model training. [4]

Within the analysis stream, recent projects like "Distant Reading" [5] and "Architecture as a Graph" [6] attempt to extract deep patterns and relationships of design based architectural information embedded within technical drawing imagery such as floor plans and sections in hopes of codifying and revealing new insight into the deeply embedded design logics that drives architectural form and function. Though computational approaches have been used for decades as means to derive the logic behind seemingly subjective architectural form, past studies have typically been limited by restrictive manual methods, limited computational power, or the inability to scale beyond a limited number of building test cases. However, with A.I. tools and very large image datasets, more recent approaches are able to absorb and extract exceedingly greater insight by analyzing thousands of buildings at a time through the digestion of large amounts of image data, therefore deriving new knowledge that could only be acquired at such a scale.

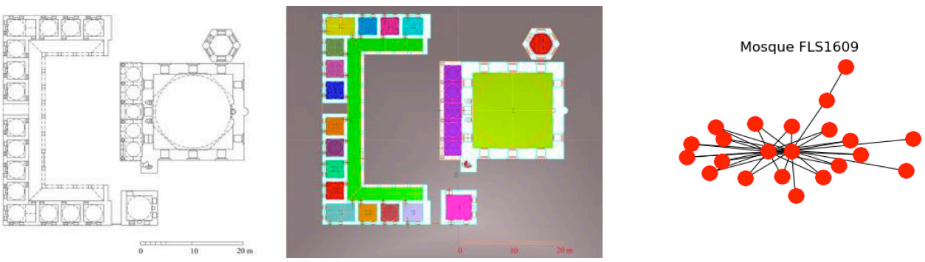

Fig. 3. In the paper "Architectural Distant Reading" use image databases of religious building plans and machine learning methods to explore how machine learning techniques can autonomously identify their typological and functional traits [5]

Though opposite in intent, both the design and analytical research streams share the common theme of using the architectural image as a primary data source. Whether it be a rendering, photograph, architectural plan, or sketch, all three methods require large amounts of targeted imagery to properly train their models and achieve their goals. As a result, as AI research within these realms increases, so does the demand for large volumes of targeted, categorical digital architectural imagery.

1.4 Thinly Spread Data Across Multiple Platforms

Currently, digital architecture image data is decentralized and spread amongst various public and private platforms, catalogues and databases. Similar to other fields like medicine, where health related data in the United States might be spread thinly across multiple platforms [7], such decentralization leads to partial understandings, ineffective solutions, and potential messy mistakes. As such, the full spectrum of a building's image record might be spread across multiple platforms or digital catalogues, with no single source containing all relevant information. As a result, obtaining a full and comprehensive image record or training image dataset of a specific building, architectural style or body of work is currently a very challenging and time-consuming task.

As more emphasis is placed on the importance of data collection for deep learning and as AI use within architectural research becomes more widespread, the urgency to locate and aggregate related yet dispersed imagery data becomes increasingly important. Ironically however, architecture's natural emphasis on image representation positions it well for deep learning as a great deal of imagery has already been documented, digitized and deposited within various online databases, ready to be collected. Below is a brief list of some of the more popular online repositories for architectural image data.

Open-Source Databases of Training Datasets

- *CKAN:* A popular and searchable open source data portal platform where datasets can be made accessible, accessed, shared, viewed by anyone. Built on Python backend and JavaScript front end.
- Kaggle: Another dataset repository, often holds competitions. Currently 63,748 datasets [8]. Already contains a number of interesting datasets related to architecture. For example, the Architectural Styles Dataset which contains 10113 images of 25 different architectural styles. Or Street View House Numbers, which contains 600, 000 images of houses with house numbers from Google Street View.

Digitized Library Collections

- *Artstor:* 300 collections composed of 2.5 million images related to art and design for educational and scholarly use. All images include high quality metadata.

Independent Architectural Databases

- *Archinform:* Claims to be the largest resource of architectural data online, with information on 83,000 built and unbuilt architectural works spanning the globe.
- *Nextroom:* A Europe-centric Austrian platform for the mediation of high quality realized architectural projects.
- *SAH Archipedia:* an online encyclopedia of the U.S. built environment that contains histories, photographs, and maps for over 20,000 structures and places. Architecture can be searched by architect, material, type or style.

Architectural Blogs

- *ArchDaily:* Worlds most visited architecture website containing thousands of well documented global projects.
- *Dezeen:* A digital collection of architecture, design and interior, boasting as "the world's most popular and influential architecture and design magazine".

Search Engines

- *Google Street View:* As of 2019, 10 million miles of streets and associated buildings have been photographed and are available online, making it possibly the most valuable source of architectural imagery currently available.
- *Google Images*
- *Bing Images*

2 Method

2.1 Framework & Pipeline

Archi-Base exists as a Python based Jupyter Notebooks application hosted on Google's Colab environment. As an end-to-end application, its pipeline is built upon three main components: a user input terminal, an image scraper and an image classification tool.

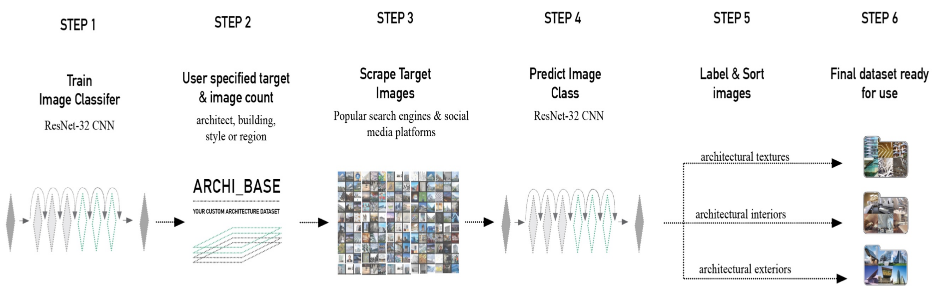

Fig. 4. Archi-Base workflow diagram

User Input Terminal

The user input terminal prompts the user to define the subject or "target" of the required dataset as well as the desired volume of images needed. The target may include any architectural style, typology, office, architect, or building. Multiple targets may also be specified as a means to gather even more or refined image data of a desired subject. For example, one can specify an architect as well as the names of the buildings they designed. However, due to the type of databases used to obtain images, searching for terms that may return larger volumes of images such as architectural styles (ex. "Postmodern Architecture"), typologies (ex. "Gothic cathedrals") or firms (ex. "Bjarke Ingels Group") typically return more results compared to more niche terms (ex. "Dome in the Desert House" by "Paolo Soleri"). This limitation, however, prompts the need for further investigation and integration of image sources that may contain more niche image data to draw from. Finally, users can specify the size of the dataset required. However, the final size is dictated by image availability and may not always meet targeted size as a result. However, Archi-Base consistently built datasets of 10,000 + images for various search terms tested such as "Jean Nouvel", "Gothic Architecture", "Zaha Hadid", and "Brutalist Architecture" and has been able to build even larger datasets of 50,000 + images.

Image Scraper

Python based open source image scraping tools are integrated and used to crawl various image databases and platforms and scrape all imagery associated with the user's target input terms. Scraping is typically accomplished by accessing the platform or databases open-source API, identifying all related images that have descriptions (ex. hashtags) that match the user specified target search term(s), and then downloading en-masse. For Archi-Base, images are directly downloaded and stored on Google Drive's cloud platform for later analysis, labelling and sorting.

Pre-Trained Image Classification Model

A deep learning image classification model is then used to analyze, classify, label and sort new scraped and downloaded images into one of 13 categories based on image content. Categories include images of street views, closeup, interior or aerial views, city skylines, images with strong horizon lines, night images, people dominant images, images of sketches, technical drawings, images featuring dominant gaps between buildings, and images of print material (ex. Books, magazines, posters, etc.). The classification model is built upon Fastai's v2 vision platform and uses a resNet-32 convolutional neural network [9] in combination with Pytorches Pyimage computer vision library [10]. Once the model has classified a new image (ex. "interior image"), additional functions label the image according to its class (ex. "interior-image-1") and then places it within a correspondingly named folder (ex. A folder named "interior-images"). When all images have been classified, labelled and sorted, each individual sorted folder is now a complete, targeted and labelled dataset and is ready to be used for DNN research projects. For example, a 50,000-image dataset of "Jean Nouvel" architecture may yield a 9,000-image dataset of Jean Nouvel architectural "street" images, a 12,000-image dataset of Jean Nouvel architectural "interior" images, a 5,000-image dataset of Jean Nouvel architectural "texture" images, and so on. However, the sum of all images in all 13 image class datasets equals the original volume specified by the user, and in this case, a total of 50,000 images.

2.2 Data Collection Method

Archi-Base has focused its data collection method on publicly available data acquired from a combination of social media websites and large, popular search engines.

Social Media

Social media applications contain an inordinate amount of image data collected and shared publicly by users online. Due to their inherent social nature, the majority of these images originate from casual non-expert users who tend to focus their imagery on the external style and outward appearance of architecture or on people within the scene rather than perhaps on the finer construction details, documents, or technical aspects of buildings. In addition, attention and image volume tends to gravitate towards more "popular" buildings, cities, architects, or styles, therefore creating a certain level of subject bias within the content. Finally, the social media websites targeted are Western or American in origin and may likely contain an unbalanced emphasis towards Western generated content. Non-western social media platforms were not used in this study, though their integration would be very beneficial in future work.

Popular Search Engines

Popular search engines are another resource used by Archi-Base to source architectural imagery. Like many images however, the vast majority originate from websites, blogs, online encyclopedias and so on, rather than from digitized books, magazines, and pdf documents, papers or reports. As a result, many of the acquired images are often refined and representative of finished products, or originate from marketing material, architectural firm websites, reviews, articles, and so on. Like images sourced from social media platforms, technical drawings, diagrams, sketches, plans, sections, and so on are few and unequally represented in the acquired data. Finally, Archi-Bases search queries were carried out in English, thus limiting our datasets to imagery that originated from English language dominant online resources, hence reinforcing our dataset's bias towards English language-oriented content. Nonetheless, multi-language translation can be integrated in future work, thus broadening the range of non-western and non-English language sources Archi-Base searches and gathers image data from.

3 Results

3.1 Image Classifier Performance

Archi-Base's image classification model was trained on 6,750 hand-picked architectural images that were organized into 13 different content categories and labelled appropriately. After completing 19 training epochs, our model achieved a minimum training loss of 0.151 and a maximum accuracy score of 89%. Per the confusion matrix below (fig. 5), the majority of classification errors occur when the model attempted to differentiate between "building closeup" imagery and "street" imagery (images of buildings taken from the street or roadway), between "technical drawings" and "sketches" and between images "horizon" and "night" images. As apparent in the prediction analysis images below (fig. 6), many of the incorrectly predicted images contain two or more defining classification category elements. For example, the image of the city skyline during the evening contains elements of both the "city skyline" class and the "night" class, thus confusing the model and leading to a classification error. This issue of images with multi-class content questions whether the 89% accuracy score truly reflects the model's real performance. Rather, these errors help identify how some of the original labels manually applied to the training set were not sufficient descriptors, or too broad for the content, thus revealing original labelling errors, or highlighting the need for multiple labels to provide increased accuracy.

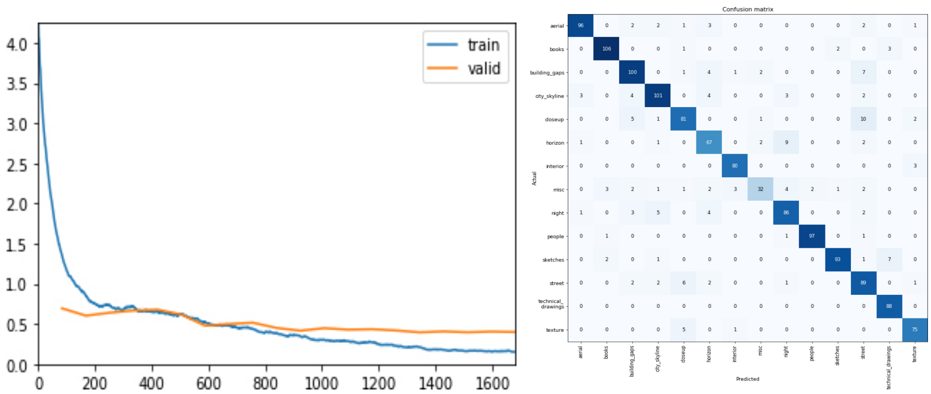

Fig. 5. Image classifier model training & validation loss (left). Image classifier model confusion matrix (right).

Fig. 6. Model classification errors due to unanticipated multi-class image content

3.2 Archi-Base Speed Benchmarks

The below speed performance benchmarks (fig. 7) are based on Archi-Bases GPU and internet speeds at the time of the experiment. Archi-Base uses Google Colab Pro's GPU capacity, which includes Nvidia's T4 or Tesla P100 GPU, though access was not always guaranteed. For internet, Archi-Base ran on Verizon's fiber optic Fios service which provided a 700 mbps download speed and a 125 mbps upload speed. Archi-Base operates at approximately 6.32 times faster than the traditional manual image data base creation method. Manual database construction method assumes that images are searched for online, identified, opened, downloaded, labelled and sorted into content categories by hand. Three manual image preparation tests were conducted, and it was determined that on average, it takes about 11 seconds to prepare one image manually by hand. Archi-Base can complete this same task in 1.74 seconds. As a result, a 10,000-image database that would normally take 30.55 hours to create manually would take Archi-Base only 4.83 hours; a non-negligible and significant reduction. However, if taking into account normal time delays in manual collection such as distractions, break-taking, and time spent eating and sleeping, it can be assumed that a realistic collection time would be much higher. Assuming a human can maintain a cumulative average of 8 hours of total uninterrupted maximum-performance dataset creation time per day, it would take at minimum, 3.82 days to build a 10,000-image dataset. As a result, Archi-Base is not 6.32x faster, but rather 18.97x faster. With even larger datasets of 50,000 images, which is quite common, Archi-Base can accomplish this task in a single day (24.15 hours) instead of taking 19.1 days to compile by hand if taking into account time delays. By implementing Archi-Base, large datasets can now be built quickly and efficiently in an afternoon, overnight or in a day, therefore freeing up time to devote to more important aspects of research or practice while simultaneously making CNNs more accessible to a wider spectrum of the architectural research community.

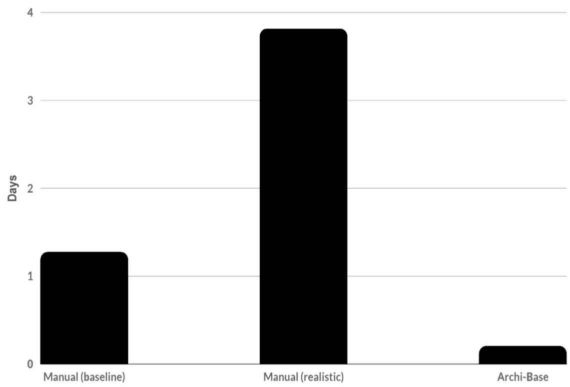

Fig. 7. Average Archi-Base vs. manual build time per days for a 10,000 -image dataset. "Manual (baseline)" assumes non-stop dataset creation which is not possible. "Manual (realistic)" accounts for breaks, time spent sleeping / eating, etc.

3.3 Dataset Generation Test: “Brutalist Architecture”

To test Archi-Base, we used it to build a 50,000-image dataset of “brutalist style” architectural images. Archi-Base built this dataset in just over one day (24.15 hours) with all images labelled and sorted into 13 different image categories. To compare, this task would have taken approximately 19.1 days to complete if carried out manually. Only a single search term “brutalist architecture” was used as user-input for image scraping. As for the dataset focus, this particular style was chosen for two reasons. First, brutalist architecture has experienced a resurgence in interest among designers and design enthusiasts alike. As a result, an extremely high volume of brutalist architecture images exists online, therefore ensuring that a 50,000-image dataset could be created. Secondly, a 50,000-image brutalist dataset lends itself well to our evaluation method, which uses it to train an IntroVAE CNN to generate new images in a similar style. As brutalism is a very distinct architectural style characterized by minimal monochromatic construction that showcases bare building materials (typically concrete) and heavy structural elements over decorative design, determining whether the IntroVAE’s new synthesized images embody these features becomes very apparent and easy to do.

Fig. 8. An example of Brutalist architecture; The National Theatre in London designed by Denys Lasdun [11]

4 Evaluation

4.1 IntroVAE Test

In order to evaluate the quality and robustness of Archi-Base datasets for use in deep CNN research, we trained an Introspective Variational Autoencoder (IntroVAE) on the Archi-Base generated brutalist street image dataset (7,818 images) and then qualitatively measured how accurately the synthesized images matched the content and style of the original training images. IntroVAE was chosen over other DNN models such as GAN or WGAN, due to its ability to produce much higher resolution synthesized output images compared to other competing models. To briefly summarize, IntroVAE achieves this high level of image quality by integrating both traditional VAE and GAN generative frameworks, while preserving the advantages of both, such as stable training and latent manifold as well as the classic internal adversary / critic architecture [12]. After training the IntroVAE on the Brutalist street image dataset for 75 epochs over the course of 13.76 days, we studied the final synthesized output images and determined that they sufficiently captured both the defining features of the basic brutalist style as well as the general composition and content within the street image class category.

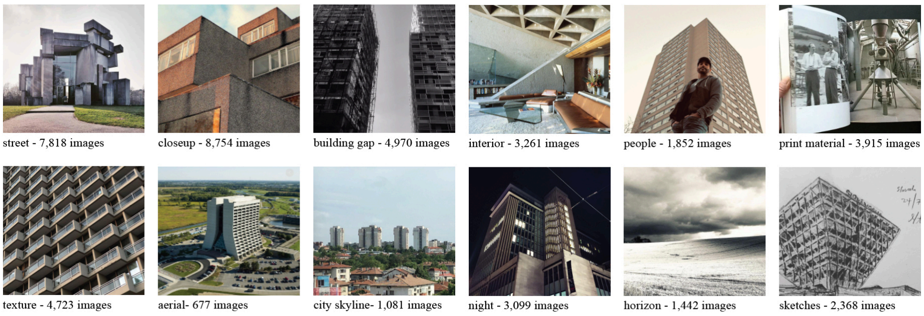

Fig. 9. The final Brutalist dataset labelled and sorted into 13 distinct classes (misc. class not shown)

Brutalist Style

As apparent in the synthesized images below (fig. 10), the Archi-Base dataset was robust enough to ensure that the distinctly raw, minimal, and heavy-weighted character of brutalist architecture was maintained, carried forward and expressed in new IntroVAE synthesized images. To begin, the geometry includes both the signature rectilinear and angular design language of brutalist styles. In addition, the theme of tectonic expression and rhythm is captured and clearly expressed in nearly all new synthesized generations. This is very apparent in the window grids and heavy linear masses of banded concrete that wrap some of the building designs at floor changes. Meanwhile, the characteristic concrete color, texture and even decay is clearly incorporated in the synthesized images. As for composition, synthesized building images maintain a certain degree of logic and cohesive organization, mimicking the data absorbed and learned from the Archi-Base dataset. Finally, architectural atmosphere is maintained through the inclusion of appropriately placed shadows, highlights, lighting conditions, change in tone and "mood", which again was learned through the recognition of the deeply embedded patterns of similar qualities within the Archi-Base "brutalist" street image dataset.

Fig. 10. An example of IntroVAE synthesized images that effectively capture the defining architectural features of the Brutalist style present within the Archi-Base dataset.

Street Class

Beyond brutalist aesthetic, the large and sorted Archi-Base dataset also allowed the IntroVAE to capture and express the general composition of building photographs taken from the street. As a result, synthesized images often contain the main building composition in the center of the frame and at an angle that reflects the position of a photograph taken from the street, therefore matching the original training images. Further, synthesized images often contain other primary features such as a clearly defined skies and ground planes, as well as secondary features such as trees, individual clouds and distant or surrounding buildings. Tertiary features such as sidewalks, landscaping and vehicles are also included at times, further reinforcing the level of image cohesion and consistency within the autonomously generated Archi-Base Brutalist dataset.

Real image from Brutalist street view training set | IntroVAE synthesized image | Real image from Brutalist street view training set | IntroVAE synthesized image

Fig. 11. An example of IntroVAE synthesized images that effectively capture the defining features of the street view composition and content

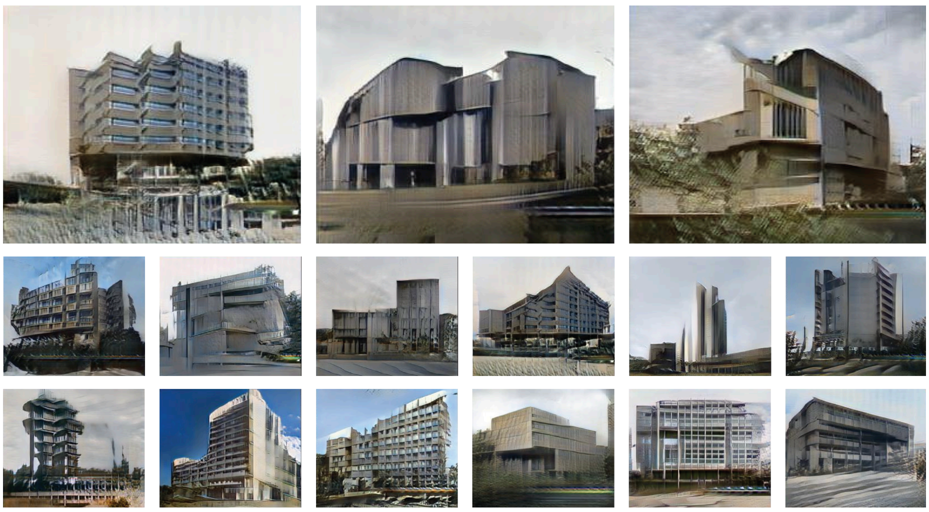

Fig. 12. IntroVAE synthesized images in Brutalist street style

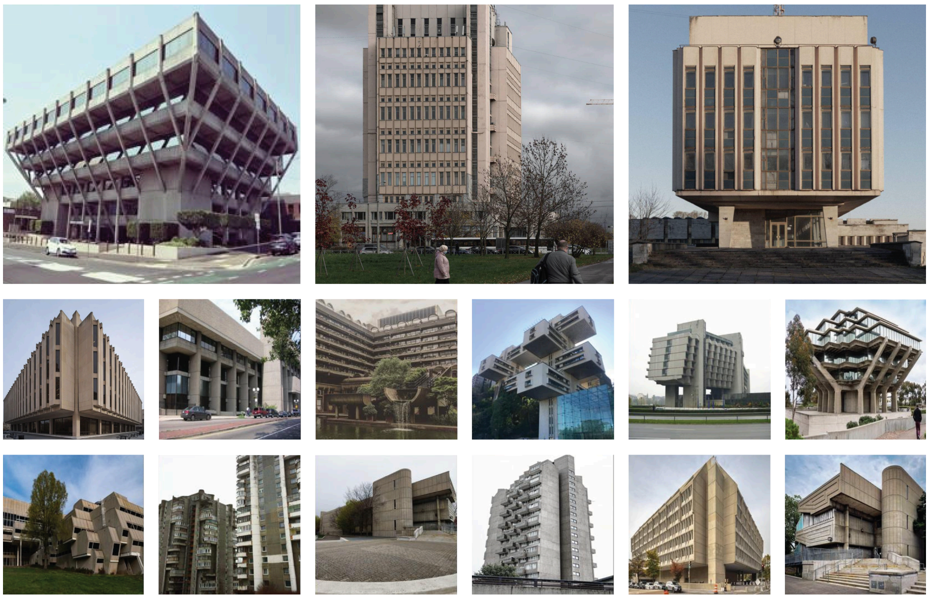

Fig. 13. Original images from the Archi-Base 50,000 image Brutalist dataset

5 Limitations

5.1 Dataset Diversity

As discussed in section "2.2 Data Collection Method" Archi-Base image data is collected from two platform types; popular image search engines and social media applications. As a result, final datasets are primarily comprised of non-technical, aesthetic forward images such as building photographs and renderings. Though these platforms do contain more technical images such as plans and sections, they make up an extremely limited percentage of the final Archi-Bases generated image datasets (under 0.1%). By broadening the number and type of resources, platforms, and databases used for image search and collection, or including a wider range of search terms, Archi-Base may be able to expand the diversity of its datasets by including more technical drawings or professional quality photographs. Possible resources to incorporate might include online architectural databases such as Archinform or SAH Archipedia, architectural blogs such as ArchDaily or Dezeen, or government planning websites that might include publicly available technical drawings. By expanding dataset diversity, Archi-Base could become the go-to resource for those needing to build large datasets of architectural floor plans or other technical drawings. As research with large datasets of architectural floor plans is already underway (see section 1.3 Architectural Data: The image"), such a feature would benefit a broader range of researchers and practitioners. Finally, Google Street View provides an enormous database of available architectural street imagery. Mining this database for architectural imagery in an autonomous and targeted fashion would provide an enormous means of aggregating huge databases of imagery, therefore providing researchers with a means to study architecture and related urban phenomenon at a scale far beyond anything previously imagined and potentially revealing new insight into and knowledge of the complexity and interconnectedness of our built world within a broader social, economic, political, environmental and cultural context.

5.2 Evaluation Models

This paper offers only a single method of evaluating dataset robustness and consistency. As training an IntroVAE on an Archi-Base dataset is a good way to qualitatively measure dataset robustness through image synthesis analysis, it does not provide a quantitative means of evaluation or scoring. As a solution, a manual check may be required to check for classification errors. Although a visual inspection of each image one by one is possible, doing so for a 50,000-image dataset is not realistic nor time efficient. In fact, it would take approx. 13.85 hours, or nearly 2 full working days to do this if at least 1 second was spent checking each image. Thus, a manual check goes against the aim of this research, which is to drastically reduce dataset build time and human intervention. Instead, a more productive means would be to focus on improving image classification performance. Beyond this, other models can be used to test dataset robustness such as BIG GAN, or WGAN. StyleGAN can also be used to generate larger and more detailed images, thus allowing for increased image scrutiny and dataset evaluation.

5.3 Computational Efficiency & Speed

Computationally, the performance of Archi-Base relies heavily on GPU and internet access speed. GPU speed is essential for Archi-Base as it directly dictates the speed of its image-classification model. Internet download and upload speed also affects image scraping and is dependent on the user's internet connection speed. By increasing the GPU speed with better hardware, or internet speed by upgrading or changing providers, or both, Archi-Bases benchmark speeds would likely increase significantly, therefore pushing performance beyond current performance (see section 3.2 Archi-Base Speed Benchmarks).

Finally, Archi-Base downloads and sorts images belonging to all 13 image categories en masse, regardless of whether the user needs images from all categories or not. As a result, a good portion of time spent gathering and sorting images may be reduced in future iterations of Archi-Base by providing a more targeted means of collecting and downloading images as requested by the user. One possible solution is to integrate the image classifier midway through the pipeline and prior to image download phase. By enabling the user to input required classes at the start, (ex. a user only wants "aerial images") the image classifier would act as a gateway, identifying image class, and then only allowing download if the identified class matches the class category or categories specified by the user. This method would further speed up the scraping process and avoid unnecessary time and computation power spent gathering and sorting images that are not needed.

6 Reflection

6.1 Dataset Culture

During a 2020 Digital Futures panel discussion regarding AI & Architecture, Architect/Artist Guvenc Ozel suggested that "*GANS have a bias towards western culture*" [4]. Though generally correct, it is not GANs that are biased, but rather the datasets which they are trained upon. This issue of dataset bias has found its way into discussions beyond architecture and has become a major issue among all industries that use artificial intelligence to make both hypothetical and real-world predictions and decisions. In the case of GANs, a deep learning algorithm capable of creating seemingly "novel" designs, it can only create "newness" and "novelty" to the extent to which its training data allows. In other words, GANs can only reproduce new images that match the visual patterns inherent within its image training dataset. And as Ozel highlighted, a great deal of popular and publicized Architectural GAN research seems to be using Western oriented data, and thus, recreating western ideas and designs in the seemingly "new" and "novel" synthesized designs. Within this framework, one could postulate a number of reasons why increased Western bias might exist. First, Western architecture publications may be less aware of the wider range of non-Western research being conducted globally, especially where papers are being publicized in languages other than English. Secondly, the major image datasets or databases used to train DNN's such as the "Open Images Dataset" or "Google Images" may contain disproportionately more American, Western and English-language oriented content, thus leading to increasingly Western biased datasets and models. This hypothesis was in fact confirmed by Google, thus prompting their 2018 "Inclusive Images Competition", which aimed to expand image oriented DNN model's cultural sensitivity and awareness through a wider and more culturally inclusive image database [13]. Thirdly, deep learning research is expensive as it often requires the latest and most powerful computing hardware. As a result, DNN research might limit the full spectrum of potential global research that might have otherwise been possible if more affordable hardware were available. Finally, the perceived ability to access cross-culture data may be more difficult and limited than had previously been imagined, thus preventing Western researchers and practitioners from easily accessing non-Western content, hence reinforcing the sense that Western bias dominates GAN generated research. This last suggestion was supported by another panelist who argued against Orzel's statement by highlighting that their own GAN research was in fact not using any Western data at all and was rather entirely composed of Chinese real estate image data [4], something that may be difficult for non-Chinese speaking researchers to access. With this in mind, datasets, when using localized resources, can be seen as a type of cultural artifact, that embodies not only the values and biases of its creators, but also the inherent image cultures of the databases from which they originate, and in turn, the inherent values and biases of those who created the data in the first place. Such a deeply embedded digital culture, though perhaps disproportionately dominated by Western content, acknowledges the wider range of ethnographic data variation that exists online, and how it may emerge in GANs or other image synthesis algorithms through "richly embedded" regional image datasets, especially when created by those who exist within those specific cultural ecologies. Given this position, we might start to see artificial intelligence as something more related to the inherited values, biases, local knowledge, and culture in which it was produced rather than the globalized neutrality that it has commonly been associated with. Furthermore, by acknowledging the specificities of place, we can now position datasets and their creators within a limited local environment of specific cultural values and social conditions, thus reinforcing dataset creation as a means to generate culturally appropriate constructs, designs, and increasingly localized solutions.

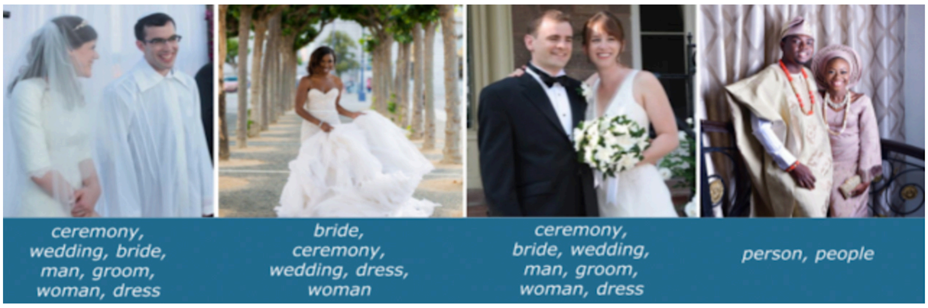

Fig. 14. Wedding photos labelled by an image classifier trained on the Western biased Open Images Dataset. Notice, the far-right image is not labelled as a wedding due to the unaccounted-for wedding dress of non-Western cultures. [13]

Fig. 15. Examples of labelled Google Inclusive Competition dataset that incorporates far more non-western images than the Open Image Dataset. [13]

7 Conclusion

Archi-Base at its core, is a response to the very simple problem of the inordinate amount of time and effort needed to build large, labelled and sorted architectural image-based datasets for architecture-based DNN research. Though a universal issue amongst researchers and practitioners, surprisingly few tools or solutions have been offered to the best of the authors knowledge. For those who may need unique datasets where images are spread thinly across multiple platforms, the time and effort required is compounded, resulting in multiple days or weeks devoted to dataset building. As architectural related DNN research within these realms increases, so does the demand for large volumes of high quality, targeted, categorical image databases.

Archi-Base is only one solution out of a realm of possibilities but attempts to remedy this issue through the use of deep learning tools, and a rapid algorithmic pipeline that autonomously searches for, aggregates, labels and sorts architectural imagery into meaningful datasets for use in deep learning projects and research. Though its current form exists within a Western and English oriented framework and returns datasets of primarily non-technical architectural imagery, future versions may easily become more culturally and image-content inclusive and sensitive through various algorithmic additions, re-arrangements and modifications. In addition, users might be able to dictate where, and what kind of data is scraped to ensure that algorithmically dictated predictions are sensitive and responsive to the user's cultural, value based, social and overall environmental specificities. Nonetheless, Archi-Base begins to open the doors towards exploring methods for autonomous and targeted dataset creation at an extremely rapid pace, with the intent of expediting research projects through increased productivity while simultaneously making deep learning research and projects more accessible, and responsive to a wider and more diverse spectrum of the architectural research and design community.

References

1. Chaillou, S.: A Tour of AI in Architecture. Towards Data Science (2020). Last accessed 2021/01/28. www.towardsdatascience.com/a-tour-of-ai-in-architecture-ef8a6bf33fa8
2. Roh, Y., Heo, G., Whang, E. S.: A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. arXiv:1811.03402v2 [cs.LG] (2019)
3. Chaillou, S. AI + Architecture: Towards a New Approach. Harvard University (2019).
4. DigitalFUTURES Talks: Artificial Intelligence. YouTube, uploaded by DigitalFUTURES world, 25 July 2020. Last accessed 2021/01/28. www.youtube.com/watch?v=2YO0loKIzBg&feature=youtu.be&ab_channel=DigitalFUTURES world
5. Ferrando, C., Dalmasso, N., Mai, J., Llach, D. C.: Architectural Distant Reading: Using Machine Learning to Identify Typological Traits Across Multiple Buildings.: Carnegie Mellon University, Computational Design Laboratory (2019)
6. Landes, J., Dissen, H., Fure, H., Chaillou, S.: Architecture as a Graph: A Computational Approach (2020). Last accessed 2021/01/28. www.towardsdatascience.com/architecture-as-a-graph-6a835d46f918

7. Glaser, J. It's Time for a New Kind of Electronic Health Record. Harvard Business Review (2020). Last accessed 2021/01/28. https://hbr.org/2020/06/its-time-for-a-new-kind-of-electronic-health-record.
8. Kaggle. A customizable Jupyter Notebooks environment with access to a huge repository of community published data & code. Available at www.kaggle.com/datasets
9. Howard, J., Thomas, R.: Fastai: A Layered API for Deep Learning. Information 2020, 11(2), 108. Available at https://doi.org/10.3390/info11020108
10. Torchvision. A computer vision library that is part of the PyTorch Project, an open source machine learning framework. Available at https://pytorch.org/docs/stable/torchvision/index.html
11. Lasdun, Denis.: Royal National Theatre. Ignant. Last accessed 2021/01/28. https://www.ignant.com/2019/05/31/the-royal-national-theatre-london-uk/
12. IntroVAE. A pytorch implementation of Paper "IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis". Last accessed 2021/01/28. www.github.com/hhb072/IntroVAE?fbclid=IwAR1b-Y3nO7bMhUz2Njd9DTpUXdYbG1CEUvZIUEaIXfzxRDGtfZYmH4RwEyc
13. Doshi, T. Introducing the Inclusive Images Competition. Google AI Blog (2018). Last accessed 2021/01/28. https://www.technologyreview.com/2018/12/02/138843/ai-has-a-culturally-biased-worldview-that-google-has-a-plan-to-change/
14. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. ArXiv:1701.07875 [Cs, Stat].http://arxiv.org/abs/1701.07875