Application of Artificial Intelligence in the Acquisition of Architectural Forms.

Raúl Felipe Buzó da Silveira, Ángel Armagno Gentile

Facultad de Arquitectura, Diseño y Urbanismo (UdelaR), Montevideo, Uruguay
raulbuzo@fadu.edu.uy; angelarmagno@fadu.edu.uy

Abstract. The primary objective of this research was to explore the effectiveness of Neural Radiance Fields (NeRF) in acquiring architectural forms and compare them with traditional photogrammetry results. The study began with a comprehensive literature review on AI in architecture and NeRF. Afterwards, a single case study applicable to both NeRF and photogrammetry was selected for comparison. The NeRF model showed the ability to accurately represent details and light effects, adapting reflections and transparencies to real-world conditions, as well as handling occlusions, and inferring three-dimensional information. In similar situations, Photogrammetry generated less coherent volumetrics or failed to interpret objects. Additionally, tests with a reduced number of images showed that the NeRF model maintained its characteristics, while photogrammetry suffered a decrease in quality and completeness. However, NeRF's performance was influenced by data collection quality. Insufficient data led to lower-quality volumetrics with imperfections, highlighting the importance of careful data collection, even with technologies like NeRF.

Keywords: Neural Radiance Fields (NeRF), Photogrammetry, Artificial intelligence, Design, Architecture.

1 Introduction

In the current technological landscape, there is an observed merger between digital, physical, and biological technologies, a phenomenon which Klaus Schwab (2017) has termed the Fourth Industrial Revolution. Among the disciplines affected by this revolution, architecture has seen significant transformations due to the adoption of new technologies. As noted by Pena et al. (2021), Artificial Intelligence (AI) emerges not just as a tool to automate processes, but also as a powerful collaborator that enhances capabilities and stimulates creativity in the way we conceptualize, design, and build within the discipline.

In architecture, a specific aspect of these changes has been the representation of forms. Historically, it has evolved from manual methods to advanced digital instruments. AI, with its ability to analyze and process large volumes of data efficiently, has proven to be an essential tool in various stages of architecture. Photogrammetry has traditionally been used to transform 2D images into 3D models (Almagro, 2000). However, recent innovations in AI, particularly Neural Radiance Fields (NeRF), which are neural network-based techniques designed to model three-dimensional scenes from images, as proposed by Mildenhall et al. (2021), emerge as promising alternatives to enhance the accuracy and functionality of three-dimensional modeling.

Given this backdrop, the question arises: How can AI, through the use of NeRF, refine the acquisition of architectural forms in contrast to traditional photogrammetry? This paper proposes a comparative analysis between NeRF and photogrammetry, focusing on the advantages and challenges of both approaches. The aim is to highlight the role that AI is taking in architectural representation and to anticipate the implications that these advancements might have on the discipline in the near future.

The core purpose of this study is to guide professionals and academics looking to integrate AI into architecture, highlighting the opportunities that tools like NeRF can provide. In an era of constant technological evolution, AI stands out as a pivotal transformative element in contemporary architecture, promising to further revolutionize the field in the coming decades.

2 Methodology

Within the context of this research, the aim was to explore the efficacy of the NeRF neural networks in the acquisition of architectural forms and contrast them with traditional photogrammetry. To comprehensively address the research question, a mixed-methodology was adopted, combining a literature review with a practical analysis of case studies.

The technical approach focused on the key differences between the interpretation of images through NeRF and photogrammetry. According to Mildenhall et al. (2021), NeRF utilizes a combination of neural networks and optimization algorithms to infer the 3D geometry and appearance of a scene from 2D images, in contrast to photogrammetry that employs triangulation techniques to generate 3D models.

The choice of a mixed-methodology was grounded in the combination of theoretical and practical aspects of NeRF usage in architecture. The literature review offered an overview of the theoretical framework and current trends related to NeRF and photogrammetry. Additionally, the analysis of case studies provided an empirical context to assess the practical applications of these technologies and discern their advantages and limitations.

To ensure compatibility between the NeRF and photogrammetry case studies, a case applicable to both techniques was chosen. This case originates from the FRAC/Arq program (Close-Range Photogrammetry for Architecture), an elective course from the Faculty of Architecture, Design, and Urbanism (Udelar). This decision aimed to ensure that variables such as building characteristics and capture conditions were consistent, thus avoiding distortions in the comparative results.

Through this methodology, the intent was to offer a broad understanding of NeRF's efficacy in acquiring architectural forms and how it compares to traditional photogrammetry. This approach also served as a starting point for deeper discussions about the future applications and implications of NeRF in architecture.

2.1 Literature Review

To understand the functionality of NeRFs, it is necessary to have a general knowledge of real-time computer graphics and rendering techniques. A relevant parallel would be three-dimensional video games, in which user interactions continuously modify the game's internal state, requiring the software to constantly recalculate aspects such as position, lighting, shading, and texture, in a process known as rendering.

"Ray tracing" has emerged as one of the most prominent rendering techniques in recent times. Its approach focuses on replicating the interaction of photons with the environment to emulate visual perception in a digital context. Its methodology involves projecting rays of light from each pixel towards light sources to determine that pixel's color (Whitted, 1980). Despite its accuracy, it has faced challenges in terms of computational cost, with rendering times that can extend from hours to days. However, recent innovations in hardware and software have facilitated significant acceleration in this process, positively impacting areas like cinematography, video games, and architectural visualization.

While ray tracing holds a dominant position in computer graphics, it is not the only approach to interpret three-dimensional data. This research focuses on "volumetric ray marching," a volumetric rendering technique. Contrary to ray tracing, which focuses on reflections on surfaces, this method involves rays that traverse and integrate with objects in the scene (Sitzmann et al., 2019).

When applying this technique to a three-dimensional space, the concept of a voxel is introduced, the three-dimensional equivalent of a pixel. Each voxel contains information used to determine the color and opacity of a corresponding pixel in the rendered image. The information of each voxel is translated into color and opacity through a transfer function, a process adaptable to visualize different aspects of a volume of data (Mildenhall et al., 2021).

In this context, Neuronal Radiance Fields (NeRF), through the use of artificial intelligence and machine learning, propose an innovative approach to rendering based on volumetric ray marching, specifically aiming to optimize those aspects where this technique faces greater challenges.

2.1.1 Volumetric Rendering with NeRF

The primary objective of NeRF is, through training a system based on neural networks, to acquire the ability to generate any new perspective requested, starting from a pre-existing set of images: photographs of the object intended to be represented.

According to Mildenhall et al. (2021), volumetric rendering involves casting a ray, recording its path, and assigning a color using a transfer function. However, the challenge of the study lies in working with pre-existing images. When selecting one of these images, the desired color for each pixel is known and can be used to supervise the training of a neural network tasked with learning to encode the scene.

As previously emphasized, although the transfer function is crucial, it can be replaced by a neural network that predicts the radiated color of each voxel based on its position and orientation (Mildenhall et al., 2021). Initially, the network produces random colors, but with the knowledge of the actual image, it is possible to correct the error between the prediction and the real image. This process is repeated, refining the predictions, until the network can accurately predict the light emitted at each point, generating synthetic yet realistic three-dimensional scenes.

Following this line, Sitzmann et al. (2019) mention that neural networks represent the scene as a continuous function of depth and color. This representation facilitates the generation of detailed images from any perspective, encapsulating the scene as a function that maps a 3D coordinate to specific scene properties.

Despite its apparent complexity, the representation based on the NeRF technique seeks to offer an innovative solution to the traditional limitations of architectural form acquisition. As Mildenhall et al. (2021) state, the ability to generate highly detailed representations from 2D images has the potential to change the conventional focus in computer graphics, impacting how architecture is presented and communicated.

2.2 Case Study

This study is framed within the elective module of Close-Range Photogrammetry (CRP) from the Faculty of Architecture, Design, and Urbanism (FADU) at the University of the Republic (UdelaR). The selection of this context arises from the alignment between the objectives of the CRP module and the current investigative approach.

The CRP module aims at introducing and exploring new technologies that facilitate the interpretation of the geometric, morphological, and visual characteristics of architectural environments. In this regard, the integration of

AI, through NeRF, aligns with the module's purpose and provides a conducive academic context to experiment with the technology and assess its advantages and distinctions compared to photogrammetry.

To ensure the validity of comparisons between the case studies using NeRF and photogrammetry, as previously mentioned, a methodology was adopted involving the selection of a unique case study applicable to both techniques. The purpose of this strategy is to control and keep uniform the factors that could influence the results, such as inherent features of the building and its surroundings, and the specific conditions under which image captures were conducted.

In line with this methodology, the final delivery of a subgroup of CRP module students was chosen as the case study. This group consisted of Mariano Cabeza, Rodrigo Gómez, Santiago Irureta, Romina Martínez, and Ulises Morín, during the first semester of 2023. The study elements included in this case, encompassing both captured images and generated data, were published on the website: https://ramblahistorica.wixsite.com/inicio.

The primary goal of the work carried out by the students was to investigate the relationship between the city of Montevideo and its waterfront throughout various historical stages. To achieve this goal, multiple records were taken at points of interest located along the city's waterfront. Specifically, two historically relevant architectural sites were selected: the English Temple and the South Cube, both located near Plaza España, in the Old City district.

For capturing images of the study area, a planned flight was conducted in DroneDeploy with the DJI Mavic Pro 3 drone. Concurrently, students took ground-level photographs of the studied site to enhance the final survey outcome, resulting from a combination of aerial and ground photos. For processing the obtained information using the photogrammetry technique, the Agisoft Metashape software was utilized. Initially, a point cloud was created from the collected photographs, and subsequently, a textured mesh was embedded on the website.

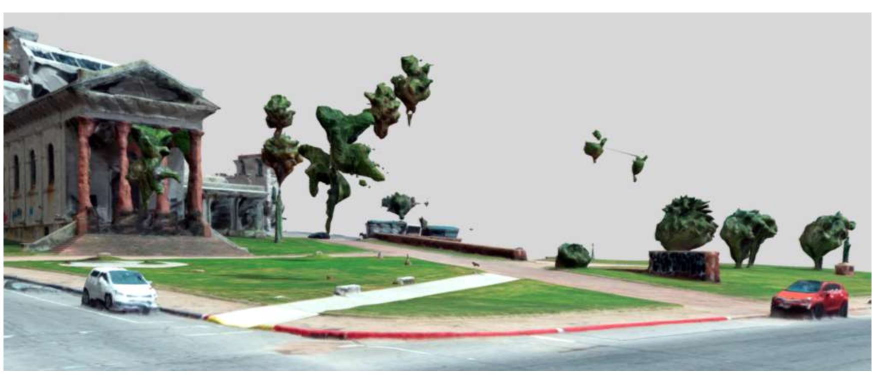

Figure 1. Photogrammetry of the English Temple. Source: https://ramblahistorica.wixsite.com/inicio/general.6

In the next phase, the techniques of photogrammetry and NeRF applied to the case study will be compared. The goal is to determine whether Artificial Intelligence, particularly NeRF, enhances the capture and representation of architectural forms compared to traditional photogrammetric methods.

3 Results

The software used for training the NeRF model was based on an open-source project developed by Nvidia, named Instant NGP. The source code is available on GitHub (an online platform that allows hosting, sharing, and collaborating on open-source projects) for users to download, compile, and use in their own projects and research.

Upon setting the dataset and the software, the NeRF implementation in the study context was initiated. The training process evolved through multiple phases, starting with a preprocessing period where minor adjustments were made to the images to ensure proper interpretation by the neural network.

During the initial training phase of the neural network, it was observed that within a five-second span, the basic volumetry was generated. Even though this volumetry allowed for identification of the general structure, it was marked by significant volumetric noise. By the end of the first minute, the representation of the scene, spanning approximately 300 m^2, had achieved a level of detail that enabled contextual interpretation. However, some regions with insufficient information, especially those far or out of photographic range, need to be highlighted. Subsequently, after 30,000 cycles and around eight minutes of training, the object showed no significant improvements, leading to the decision to terminate the training.

Upon concluding the training, a set of convergence and divergence points between both technologies (NeRF and photogrammetry) was identified. The first notable distinction emerges from the underlying conception of each system. Regarding NeRF, it doesn't infer the polygonal mesh of an object, but its entire three-dimensional volume. In this manner, the result obtained with NeRF encompassed the luminous information emitted by the scene in its entirety, adopting a volumetric rendering approach, which contrasts the surface rendering characteristic of photogrammetry.

In this context, the NeRF technology achieves a faithful representation of details and lighting effects, closely aligned with reality. Thus, depending on the viewpoint from which the resulting volumetry is observed, one can appreciate the precise adaptation of reflections and transparencies to the real scene conditions.

When analyzing specific elements, such as reflections and gleams produced by water, as well as glasses placed in the apertures of the English Temple or other buildings, NeRF's ability to interpret them appropriately is evident, producing coherent volumetry. This differs from the results obtained through photogrammetry, which, under similar situations, led to the generation of amorphous volumetry or a lack of object interpretation.

Moreover, NeRF’s capability to handle occlusions in the study scene stands out. In a three-dimensional environment, occlusions are areas hidden or blocked by other objects in the scene. For photogrammetry, these occluded zones posed a challenge due to the absence of direct visual information. In contrast, results from Instant NGP showcase the ability to infer three-dimensional information from multiple angles, even in occlusion situations, as occurred with the bottom faces of the AEBU building.

When analyzing the modeling of complex surfaces, another distinction between photogrammetry and NeRF technology arises. Photogrammetry, conditioned by its dependence on interpreting two-dimensional visual information, faced challenges in generating highly complex volumes, such as trees, vehicles, and the rocky coastline profile. Conversely, NeRF’s volumetric approach proved more effective in handling these surfaces. This capability contributed to the generation of a more accurate and coherent volumetry, even in the presence of surfaces and textures of high complexity.

Figure 2. Volumetry generation using NeRF. Source: author’s photo

In the presented case study, it was not possible to effectively assess the changing light conditions or the management of moving objects. Data collection was conducted on a cloudy day, resulting in consistent light conditions throughout the 18-minute flight period, and the only moving objects were vehicles, which were not a crucial component of the model focused on architecture and the environment.

Furthermore, although both models were generated using the same set of photographs, additional tests were carried out with fewer images. In all evaluated scenarios, the NeRF model demonstrated more efficient volumetric handling, even when using fewer photographs.

In line with the above, it is worth noting that the technology underpinning NeRF and the broader field of AI is characterized by its dynamism. This quality implies the need for constant updating and vigilance against potential improvements, developments, and modifications that can directly influence its effectiveness in architectural visualization. While recent advancements have overcome some inherent limitations of these technologies, as evidenced in initiatives like "NeRF in the Wild" by Martin-Brualla et al. (2021) and "NeRF++" by Zhang, Riegler, Snavely & Koltun, (2020), continuous research and adaptation to emerging changes are essential to maintain their relevance and effectiveness.

The next chapter will delve deeper into these aspects, highlighting key considerations for the adoption and effective application of NeRF and other AI technologies in the field of architecture and visualization.

4 Discussion

This research has delved into the environment of new digital media and its relationship with architecture, positioning it within the conceptual context of emerging architectural practices. Starting from the notion that architecture, as a discipline, is not immune to the challenges and advances posed by digital media, this work has analyzed the convergence of both domains from an academic and practical perspective.

In this regard, the study revealed that although the application of NeRF in the field of architecture offered significant potential in the representation and visualization of forms, its successful implementation required a vast range of technical and practical factors. The primary hurdle was found to be the high computational requirements and the need for specialized knowledge in computer science and programming, limiting its applicability in technologically restricted settings. While there are alternatives such as cloud execution through services like LUMA AI, they do not offer the same quality and accuracy in representations.

In addition to the considerations mentioned above, it was observed that the scalability and generalization of NeRF in various contexts and environments posed specific challenges. Transferring a NeRF-trained model to a new context, especially to different software for further work, required substantial adjustments. This revealed a limitation in the technology's adaptability and was identified as a significant challenge in integration with existing tools and workflows in architectural practice.

The challenges and limitations identified in the adoption of NeRF in architectural practice highlight the inherent complexity of this emerging technology. Transferring trained models to new contexts, the demand for specialized programming skills, and compatibility with conventional tools are obstacles reflecting a limitation in its adaptability and efficacy. Additionally, the infrastructure needed to implement NeRF is complex and demands powerful, updated equipment, limiting its efficient application to a broader audience.

Despite the complexity, NeRF has been recognized as a promising technique, albeit still in development. The possibility of a future technological transformation in architectural form acquisition has been identified, but further research and improvement in areas where NeRF has limitations are still required. The contrast between photogrammetry and NeRF hasn't been drastic; however, it has been suggested that AI represents an evolution of a technique rather than a definitive solution. The exploration of NeRF is seen as an advancement in technique that still requires refinement and adaptation.

The research has emphasized the importance of carefully assessing available capabilities and resources, as well as the necessary training to effectively implement these technologies, despite their current limitations in implementation and usage. Ongoing research and development in these areas will be crucial to overcome these obstacles and fully leverage NeRF's potential in contemporary practice.

In this context, the study has revealed that the adoption and application of NeRF in architectural form acquisition are in an early stage and require further research and developments. While currently not showing a significant difference from photogrammetry, NeRF has the potential to become, under certain conditions in the future, the standard methodology for form acquisition. This finding provides a valuable contribution to the architectural field, establishing a framework for the future of architectural representation, where NeRF could play a pivotal role in the digital era.

References

Almagro, A. (2000). Photogrammetry for Architects. The State of Affairs.

Cubo sur | Rambla Histórica. (n.d.). Rambla Histórica. https://ramblahistorica.wixsite.com/inicio/general-6

Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7210-7219).

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

Pena, M. L. C., Carballal, A., Rodríguez-Fernández, N., Santos, I., & Romero, J. (2021). Artificial intelligence applied to conceptual design. A review of its use in architecture. Automation in Construction, 124, 103550.

Schwab, K. (2017). The fourth industrial revolution. Currency.

Sitzmann, V., Zollhöfer, M., & Wetzstein, G. (2019). Scene representation networks: Continuous 3d-structure-aware neural scene representations. Advances in Neural Information Processing Systems, 32.

Whitted, T. (1980). An improved illumination model for shaded display. Communications of the ACM, 23(6), 343-349.

Zhang, K., Riegler, G., Snavely, N., & Koltun, V. (2020). Nerf++: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010.07492.