Design Science

Three decades of machine learning with neural networks in computer-aided architectural design (1990–2021)

Jinmo Rhee[1], Pedro Veloso[2] and Ramesh Krishnamurti[1]

[1]School of Architecture, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Fay Jones School of Architecture and Design, University of Arkansas, Fayetteville, AR, USA

Abstract

Over the past years, computational methods based on deep learning—that is, machine learning with multilayered neural networks—have become state-of-the-art in main research areas in computer-aided architectural design (CAAD). To understand current trends of CAAD with deep learning, to situate them in a broader historical context, and to identify future research challenges, this article presents a systematic review of publications that apply neural networks to CAAD problems. Research papers employing neural networks were collected, in particular, from CumInCad a major open-access repository of the CAAD community and categorized into different types of research problems. Upon analyzing the distribution of the papers in these categories, namely, the composition of research subjects, data types, and neural network models, this article suggests and discusses several historical and technical trends. Moreover, it identifies that the publications analyzed typically provide limited access to important research components used as part of their deep learning methods. The article points out the importance of sharing training experiments and data, of describing the dataset, dataset parameters, dataset samples, model, learning parameters, and learning results to support reproducibility. It proposes a guideline that aims at increasing the quality and availability of CAAD research with machine learning.

Keywords: machine learning, deep learning, computer-aided architectural design, review, reproducibility

Received 30 July 2022
Revised 05 July 2023
Accepted 11 July 2023

Corresponding author
Jinmo Rhee
jinmor@cmu.edu

Des. Sci., vol. 9, e25
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2023.21

1. Introduction

The new wave of artificial intelligence (AI) and machine learning based on deep neural network models—so-called deep learning models—has been supported, increasingly, by the availability of data, efficient algorithms, and models, greater compute power, and access to machine learning libraries and frameworks. Developments in deep learning rely on a history of advances in AI and related fields, some of which have been explored in computer-aided architectural design (CAAD) (Zhang 2019; Rodríguez *et al.* 2020; Rhee & Veloso 2021). In this respect, recent conferences on CAAD, such as CAAD Futures (2021) and CAADRIA (2021), have established thematic sections based on AI and machine learning. Online communities, such as DigitalFUTURES (2020), have organized diverse workshops and public discussions based on deep learning. There are emerging groups in academia and industry with a focus on combining creativity, architecture, and deep learning, as evidenced by their scholarly production and practical activities (Newton 2018; Del Campo 2019; CRAIDL 2020).

In this study, we provide a broad investigation of research publications in CAAD that are based on deep learning approaches. We use data from research repositories to quantitatively evaluate historical and recent trends of deep learning in CAAD and discuss the current state and future challenges in the field. Specifically, we propose a guideline to overcome some of these challenges in a bid to improve the quality of future research. Our investigation covers three decades of research between 1990 and 2021. The year 2021 was only partially explored as the data archives at the time of the analysis described in this paper is based was still in the process of being updated.

2. Background

2.1. From shallow to deep neural networks

Deep learning is the subset of AI that relies on computational models to make accurate data-driven decisions by representing and learning complex concepts modeled as hierarchical representations of simpler concepts (Goodfellow, Yoshua, & Aaron 2016). Typically, these models are feedforward neural networks composed of a sequence of layers with differentiable mathematical parameters, operations, and functions. These layers act as a hierarchical representation for identifying and extracting patterns from large datasets that accurately map from sets of complex inputs to good decision outcomes (Kelleher 2019).

Deep learning emerged after the second AI winter (1987–1993), the period when the previous dominant paradigm, expert systems, had reached its limits and research funding radically shrank. AI researchers re-invented the back-propagation algorithm in the 1980s, which revived the interest in the connectionist models of AI (Russell & Norvig 2021). Around the same time, there was an interest in establishing more scientific approaches to AI research. Many subfields, such as computer vision, robotics, speech recognition, multiagent systems, and natural language processing were gradually reunified with the "newfound appreciation for data, statistical model, optimization, and machine learning" (Russell & Norvig 2021).

In the 2000s, advances in computing power and the creation of the worldwide web supported the exponential growth and availability of training data and the shift toward ML. This has lightened the key burden of statistical estimation, leading to the dramatic progress of statistical generalization and applicability of deep learning (Goodfellow *et al.* 2016).

Developments in computer infrastructure, such as larger memory capacity and denser computing units, enabled faster and heavier computation and have also contributed to deep learning. While neural networks have been developed since the 1940s, the recent spread of computing environments enabled stacking more layers in a neural network to perform 'deep' computations with high accuracy to process larger datasets (Goodfellow *et al.* 2016). Furthermore, neural network models have not only increased in their number of processing units or layers but also in the configuration types used to address different forms of representations. From the general use neural networks, such as multilayer perceptron and artificial neural networks, to more specialized models, such as convolutional neural networks and graph convolutional networks. These terms are well known; their definitions can be found in Goodfellow *et al.* (2016).

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

Overall, in the last decade, a wave of deep learning has emerged with the successful development and application of neural network-based algorithms and models to problems from different domains. Two examples of such breakthroughs are as follows: AlexNet (Krizhevsky, Sutskever, & Hinton 2012), a learning system for image classification proposed by Geoffrey Hinton and his team; and AlphaGo (Russell & Norvig 2021; Silver *et al.* 2016), a reinforcement learning model able to defeat human champions in the game of Go.

2.2. Mapping AI technologies in CAAD

The deep learning wave has radically impacted and changed many fields and industries, and CAAD is no exception. In the face of growing deep learning research in the field, it is important to both understand the trends and to map predecessors.

In this regard, Mateusz Zwierzycki, a Polish architect and researcher in digital design published a quantitative review of AI in CAAD. He considers that "AI research efforts kept intensifying, with a growth peak around 1995–2005" (Zwierzycki 2020) and "in the next decade (2010–2020) there seems to have been another growth period in the field" (Zwierzycki 2020). He suggests three factors to explain the growth of AI-related design research: the maturity of Internet applications and data-oriented services, the popularity of parametric modeling, and a general increase in architectural research. These factors offer an interesting hypothesis to explain the rapid growth of AI research in CAAD.

Zwierzycki's analysis relies on large domains of AI and application, which limits the capacity to infer specific trends and details about machine learning. For example, his analysis of AI tools used in design research over time comprises a wide range of computational concepts, such as genetic algorithms and CNN.

Abraham Noah Wu and his colleagues reviewed articles that used a specific deep learning model, generative adversarial networks (GAN), for solving challenging tasks in the built environment (Wu, Stouffs, & Biljecki 2022). They found that GAN is a cutting-edge technology with a wide range of applications, from improving performance in existing problems to opening new frontiers in previously overlooked areas. They also claimed that GAN can be applied at various scales in the built environment and is being used in several unique application domains including data augmentation, privacy protection, and building design generation. They also pointed out that a common challenge with GAN currently is the lack of high-quality datasets curated specifically for problems in the built environment.

The study conducted by Wu, Stouffs, and Biljecki diverged from the investigation carried out by Zwierzycki in its focus on the examination of the utilization and trends of a particular deep learning model. This research attempts to bridge the gap between these two different scales by exploring the general trend of deep learning in CAAD research. We narrow the scope of our review to machine learning with neural networks in CAAD research over the past 30 years. This refined scope enables a more accurate understanding and analysis of trends and precedents related to deep learning in CAAD.

2.3. Taxonomy and terms

To systematically collect and analyze CAAD research papers that employ machine learning with neural networks, we established a clear terminology related to deep

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

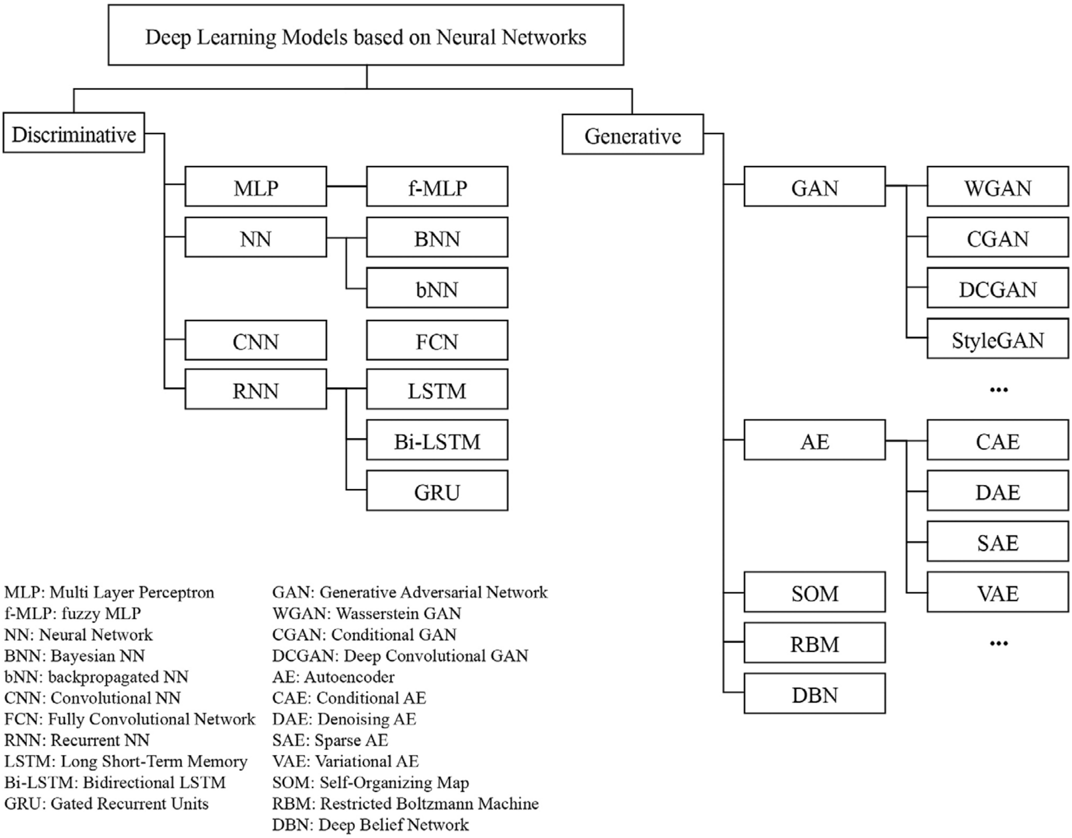

Figure 1. Taxonomy of deep learning models based on neural networks.

learning based on Liu & Lang (2019) and Sarker (2021). For this study, deep learning technology is classified into two categories based on the role of the model, namely, whether it is discriminative or generative. A discriminative model uses conditional probability to make predictions on unseen data and can be applied to classification or regression problems. In contrast, a generative model centers on the distribution of a dataset, returning a probability for a specific instance (Figure 1).

Various discriminative and generative models were considered, and the following were selected based on the presence of neural networks: multi-layer perceptron (MLP), simple neural network (NN), convolutional neural network (CNN), and recurrent neural network (RNN) for discriminative models; and GAN, autoencoder (AE), self-organizing map (SOM), restricted Boltzmann machine (RBM), and deep belief network (DBN) for generative models.

It is important to state that some of the categories are not completely exclusive. For example, generative models, such as GAN and AE, can incorporate convolutional layers or entire CNNs as part of their architecture. In this study, cases of non-generative use of CNNs are classified under the CNN category, while cases of generative use of CNNs are classified based on the type of model that consists of multiple CNN layers. Despite the widespread use of the terms and concepts used in the proposed taxonomy, it is important to acknowledge that slight variations in usage exist.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

Table 1. Six main- and 15 sub-categories of research problem types

Categories	Research problems		
evaluation	performance evaluation	construction engineering	simulation
recognition	drawing recognition	object recognition	
theory	design theory	design pedagogy	
fabrication	robotic fabrication	digital fabrication	
generation	design linguistics	design tool making	design synthesis
analysis	urban analysis	environment analysis	design analysis

3. Data collection

3.1. Scraping the literature

The first step in analyzing trends and characteristics of research in CAAD that employ machine learning with neural networks is collecting data on research publications. The source for our data is the Cumulative Index about publications in Computer Aided Architectural Design, CumInCAD (1998), the earliest open access database initiated by Martens and Turk in 1998 (Martens, B., & Turk, Z. 2003). CumInCAD is the dedicated digital archiving platform supported by several conferences and journals in the CAAD research community. A site-specific crawler was developed to collect information on publications. The crawler and statistical data used in this research can be accessed through here.[1]

At this juncture, it is worth noting that there are limitations to our study which should be acknowledged. The data source, being limited to CumInCAD, likely excludes relevant research found in publications relating to CAAD in less open/public access repositories. Moreover, as the study aims to capture the domain of machine learning and deep learning, specific models from related fields such as deep reinforcement learning, and natural language processing may also be omitted.

Prior to crawling, a list of keywords belonging to the field of deep learning is set: ["deep learning", "deep neural network", "artificial neural network", "neural network", "multi-layer perceptron", "convolutional neural network", "recurrent neural network", "generative adversarial network", "autoencoder", "self-organizing map", "Boltzmann machine"]. The keywords are established by using a minimal matching algorithm where, for example, terms like "deep learning," "deep-learning" and "deep-learning-based" would all be equivalently matched. Note that terms relating to either "deep learning" or to "neural network" are included in the list, since the term neural network also pertains to research specific to deep learning with deep neural networks. Data collection is restricted to publications between 1990 and 2021.

The crawler accessed all 16,182 publications in CumInCAD and extracted research information such as research id, title, year, authors, source, abstract, references, and so forth. It then checked whether the information has at least

[1] https://github.com/leeuack/Three-Decades-of-Machine-Learning-with-Neural-Networks-in-Computer-aided-Architectural-Design.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

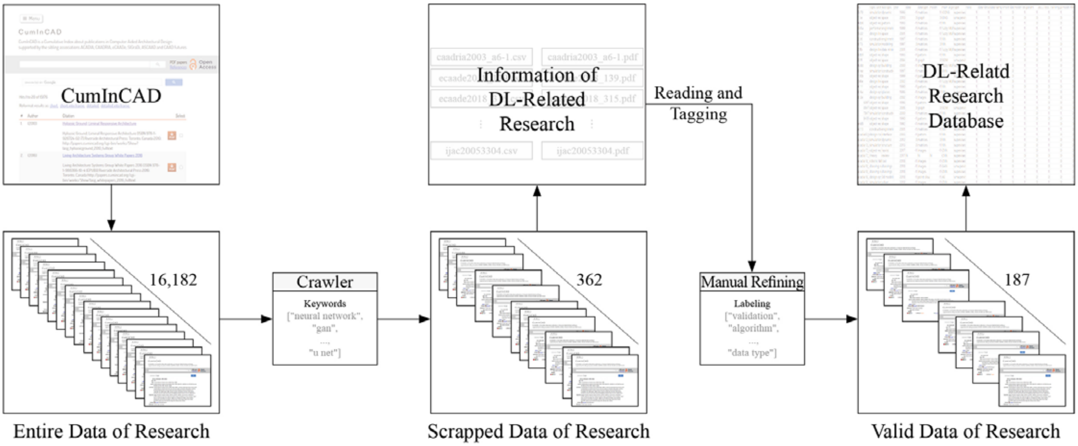

Figure 2. Pipeline for establishing a neural network-related research database from CumInCAD.

one of the keywords. If the publication information includes a keyword, the full-text pdf file, and its respective repository information as a csv file are saved. A total of 362 research papers and journal articles were collected (Figure 2).

3.2. Refinement: Filtering out papers that exclude learning

We manually filtered out papers that are not closely related to deep learning. For this, each of the collected research papers was reviewed and tagged for identifying their statistical features. The tags are as follows: validity, data, model, and content.

Validity is binary valued: 0, if the publication does not include any learning process or only includes keywords in its references; and 1, otherwise. By this refinement, 175 of the original research papers and journal articles were excluded, and the remaining 187 publications were set as the final pool for statistical analysis.

There are tags that indicate the main neural network model and data type identified in the research samples, which are further analyzed in the following sections. Besides, there are three distinct data tags and five distinct model tags related to reproducibility. The tags represent reader access author's data and models: data description, data sample, entire data set, model description or structure, model parameters, model accuracy after training, training process, and model file. These tags are used to estimate the level of reproducibility of the papers in the next sections.

Tags also comprise two characteristics of the content of the research: the research problem and the research subject. There are 15 different types of research problems that are categorized into six larger clusters: analysis, generation, fabrication, theory, recognition, and evaluation. These are used in the evaluation of trends in the Analysis section.

We likewise tagged research subjects with as much detail as possible, referencing the keywords list in CumInCAD. The resulting number of subjects is 60 and is used in analysis and described in Section. Research subject.

In the sequel, the term "neural network-related research" is used to denote the selected publications in the final pool that either apply or review deep learning

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

techniques in CAAD. This terminology is employed to clearly distinguish the nature of the studies included.

4. Analysis

4.1. Trend and research problems

General trend analysis

Figure 3 has three axes: year, problem categories, and the number of publications. The axis of year starts in 1990 and ends in 2021. Six different types of problems constitute the axis of problem categories. The bars depicts the number of publications by problem type. The evaluation category encompasses issues related to assessing the efficacy and efficiency of models developed through simulations or computational calculations using neural networks. The recognition category pertains to the challenge of recognizing drawings or objects using neural networks. The theory category encompasses the development of design theories, education, and teaching methodologies utilizing neural networks. The fabrication category encompasses digital fabrication techniques such as robotics or three-dimensional (3D) printing, which are enabled by neural networks. The generation category encompasses the creation of design vocabularies, the development of design tools, and the synthesis of design outcomes through generative models enabled by neural networks. The analysis category encompasses the challenge of transforming elements associated not only with cities and environments but also with the design process, into data for examination using neural networks.

One notable inference that can be drawn from Figure 3 is that integrating neural networks into CAAD research is not an entirely new phenomenon. From the early 1990s to 2015, the number of papers on neural network-related research has been relatively steady with small fluctuations year by year. Because of the archiving limitations, there is no clear indication of machine learning in design prior to 1990 in CumInCAD. However, Figure 3 also suggests that efforts to employ neural networks in design research is an extension of earlier research initiatives within the CAAD realm. Specifically, as described in Section. Model trend, neural networks have been consistently used.

Another notable indication from Figure 3 is that there are two distinct periods when neural network-related research was active: 1995 to 2005, and 2015 to 2021. The first period follows the second AI winter. This period had a small number of, but continuing, attempts to implement neural networks in CAAD. However, by 2005, research using neural networks declined until 2015. During this period, shallow neural networks were still predominantly used.

Post-2015, the second period illustrates a steep rise in the number of papers on neural network-related research. Prior to 2016, the number of publications related to neural network-related research within the CAAD research community was fewer than 5 or more less per year. Afterward, it has been mostly exponentially increasing with 3 publications in 2016, 5 in 2017, 15 in 2018, 35 in 2019, 43 in 2020, and 29 in 2021.

Problem type

Analyzing the number of publications by problem type shows the trend in active research by period. Using color-coded histograms, Figure 3 illustrates changes in

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

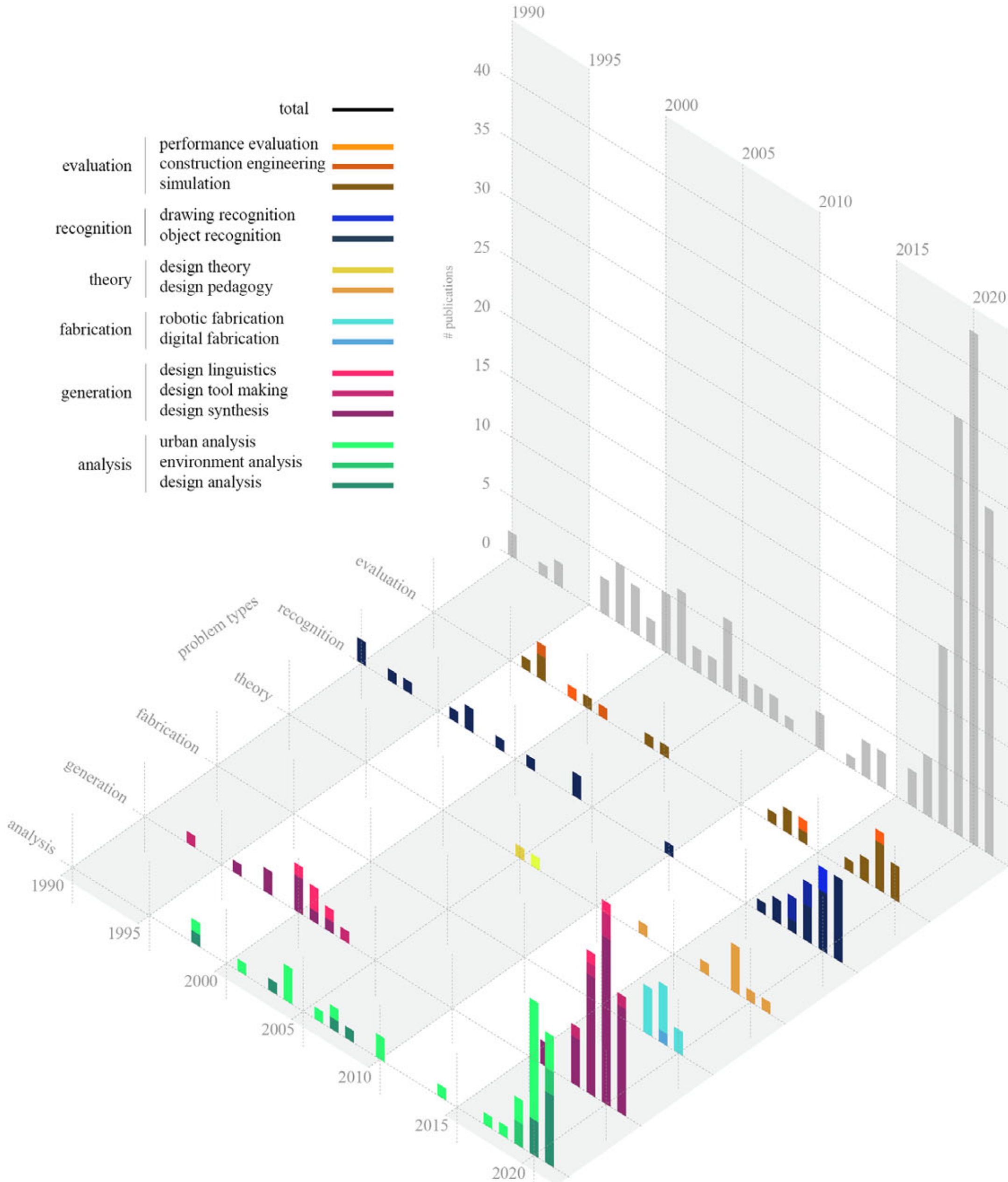

Figure 3. Number of neural network-related research in CAAD by time and research problem.

the number of neural network-related research papers by problem types and changes in the total number of neural network-related research papers.

- The growth rate of publications depends on the problem type; a common feature is that they have grown rapidly since 2015.
- There are three distinct trends indicated by problem type:

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

– Problem types that have persisted for three decades or more and grew recently after experiencing a severe recession in 2005–2015: generation, recognition, and evaluation.
– Problem types that have persisted for three decades or more and grew recently after experiencing a mild recession in 2005–2015: analysis.
– Problem types that have emerged recently and have not experienced a recession or boom in the past, for example, fabrication and theory.

- Rapid increase in the number of publications for each problem type and the emergence of several problem types after the recession is suggestive of the influence of external factors, such as the impact of deep learning and the increasing access to libraries and frameworks.

Design generation is a problem type that presents the greatest change, especially, in design synthesis research. In the late 1990s and early 2000s, there were attempts to implement machine learning for design generation problems. The trend in the number of publications during those years is noticeably discernible compared to other types. However, from 2005 to 2015, this trend is almost nonexistent. After 2015, the number of publications increases rapidly, following the general trend in the total number of neural network-related publications.

Recognition and evaluation problems follow similar trends except for the fact that these problem types have not experienced the rapid rise of design generation problems. Pattern recognition is the earliest research domain for machine learning in CAAD. It typically comprises methods for object recognition, such as shape, furniture, and recognition of space, color, and tone. Drawing recognition emerged around 2015. Design evaluation using machine learning has not been studied as steadily as recognition problems; it has had a shorter 'recession' between 2005 and 2015 than the problems of generation or recognition.

Design analysis is another emerging problem type with the shortest recession in 2005–2015. Despite being in recession, design analysis problems with deep learning techniques have been steadily studied. Over the last 5 years, it experienced an increase in publications like the other problem types.

Fabrication and theory follow another notable research trend. Both have recently emerged and have experienced neither a recession nor boom in the past three decades. Specifically, neural network-related robotic fabrication shows rapid growth when it appeared after 2015, and research on design theory using deep learning techniques also re-emerged in this period.

4.2. Research subject

The multifaceted nature of CAAD research often requires a multidisciplinary approach that encompasses multiple themes. In order to effectively categorize and analyze these complex subject areas, we have chosen to focus on the specific domain to which the neural network is being applied. By doing so, we aim to provide a clearer examination of the various applications of neural networks in CAAD research.

Figure 4 presents through a pie chart the composition of the subjects in neural network-related research in CAAD. The area of the pies is proportional to the number of publications. The darker shades also indicate more publications. The

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

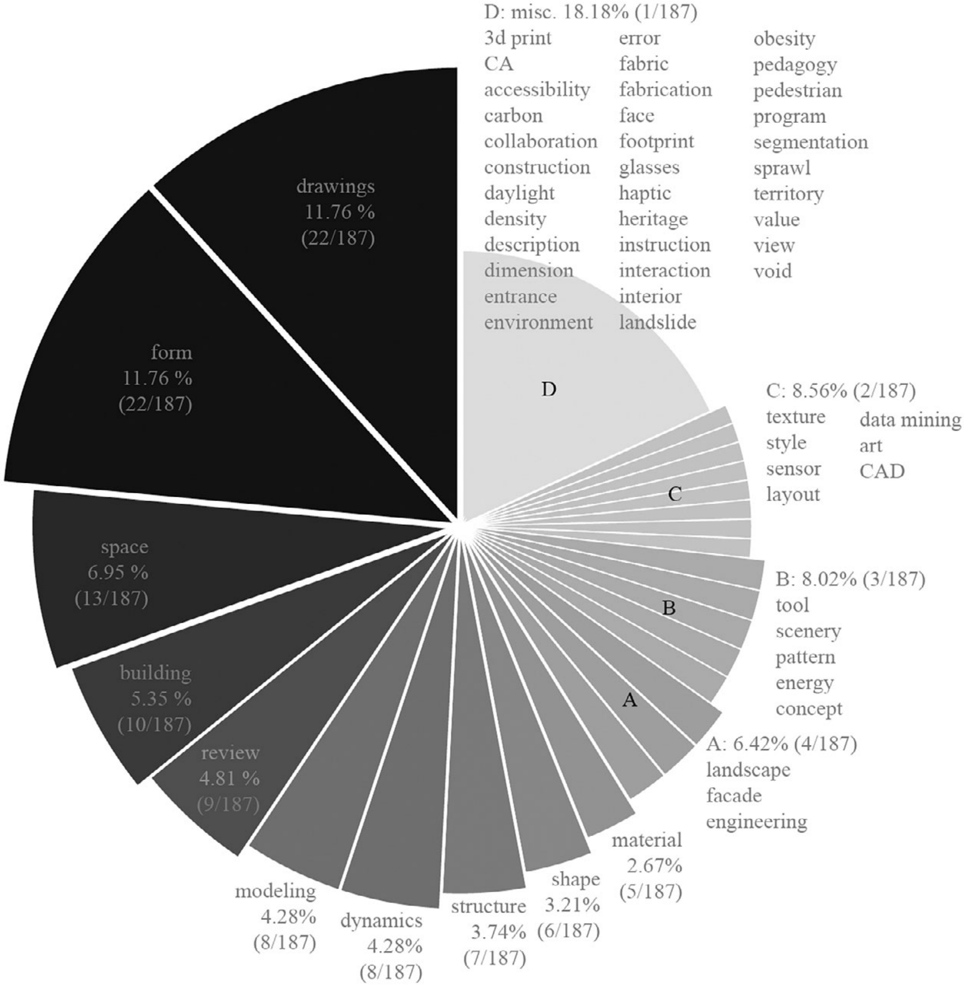

Figure 4. The composition of the subjects of neural network-related research in CAAD.

chart shows that there are various subjects in neural network-related research and a distinctive trend of dominance in subjects.

Drawing and form are the two slightly dominant research subjects in the dataset—both at 11.76% of the total—followed by other subjects with considerably smaller slices. On the other extreme, there are 34 subjects with only one publication, which comprises approximately 18% of the total.

4.3. Model trend

Figure 5 shows trends in the usage of different machine learning models in CAAD by year. Eight publications were excluded from the model analysis as they did not involve the use of model. The rows contain models that were used in the publications and the columns represent years. There are three main findings.

- NNs have been used over the whole period reviewed.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

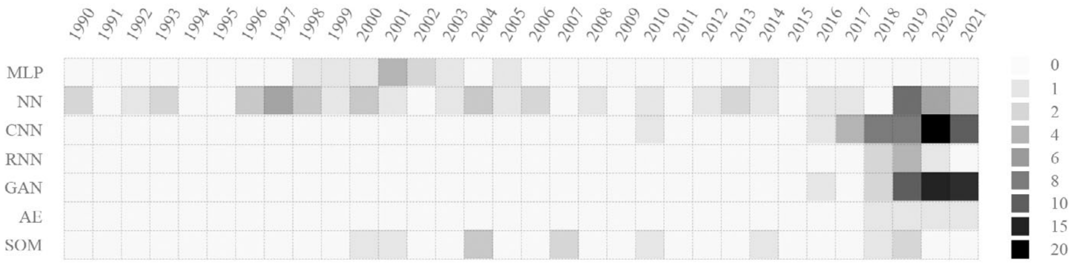

Figure 5. Changes in different models of neural network-related research in CAAD.

- CNNs are emerging models in CAAD research.
- A transition in the employment of models from MLP or SOM to more diversified neural network-based models since 2015.

The chart in Figure 5 shows that simple NNs, including ANN, are models that have been in use over the past 30 years in neural network-related research. The darker-colored cells in the row neural network in the period of 2016–2021 show that even the standard neural network models have been widely used in the framework of deep learning.

Another finding from Figure 5 is that CNN are the emerging models in neural network-related research. Considering that GAN models usually rely on CNN, the latter can be considered the most popular model over the past 5 years. We postulate that the dominance of CNN over the last 5 years is related to the strong visual culture in architecture, in comparison to other models such as MLP or SOM. In the period between 2010 and 2021 there were fewer occurrences of models other than CNN. Specifically, Figure 5 shows that the occurrence of MLP was interrupted in 2015. After 2015, SOM was sporadically adopted and newer neural network-based models emerged in research. This transition suggests that the introduction of CNN-based models has increased the possibilities of machine learning research in CAAD.

4.4. Data type

Figure 6 illustrates the occurrence of the different data types used in the publications, arranged by year, showing the dominance and trends in CAAD. Eight publications were excluded from the model analysis as they did not involve the use of data to be processed in such research subject of theory and pedagogy. We classify the data types into five categories: graph, pattern, matrix, point cloud, and voxel. A graph is an abstract structure that represents entities as vertices and their potential relations as connections. An image is a raster graphic that represents drawings and photographs as matrices of pixels. A pattern is a raster graphic that represents simplified visual entities as an arrangement of binary values. A matrix is an arrangement of numbers into rows and columns. A point cloud is a collection of points representing 3D forms. A voxel is a regular and volumetric representation of forms in a 3D grid based on discrete entities called voxels. Based on these categories, there are four main findings:

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

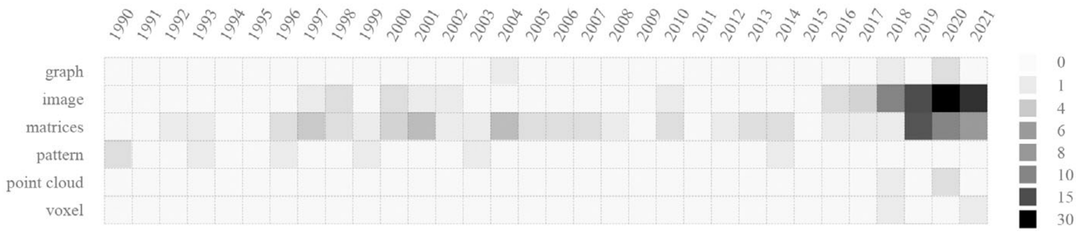

Figure 6. Changes in different data format of neural network-related CAAD research.

- Matrices have been used steadily over the last 30 years; and its use recently is rapidly increasing.
- Image data have been extensively used recently, which may be correlated to the large adoption of CNN indicated in Section Model trend.
- Newer 3D data types, such as point clouds and voxels, have emerged.
- The utilization of pattern data, which is a technique for representing shapes through occupancy in a 2D grid, has experienced a marked decrease.

Overall, image and matrix have replaced pattern data, and newer data types that directly represent 3D space and forms are starting to be adopted in machine learning for CAAD.

4.5. Reproducibility

An important criterion in evaluating neural network-related research is reproducibility, an important principle of the scientific method and a key aspect of computer science research. For the findings of any study to be reproducible it should contain sufficient information for an experienced practitioner to repeat the experiment with (nearly) identical results (Kajiya 1993). In this respect, machine learning and deep learning conferences such as ICML (International Conference of Machine Learning) and neurIPS (Neural Information Processing Systems) now require submission of code, model, and data for the purpose of reproducibility (Pineau 2019; NeurIPS 2020). Although in CAAD publications neural network-related research may neither be fully technical or scientific, nor presented in a similar structure or format, deep learning implementations should be properly presented within a scientific frame. In our analysis, we did not take into consideration the availability of code repositories such as Github, which was released in 2008, and results from earlier publications would not have been so readily reproducible.

Nine publications were excluded from the reproducibility analysis as they did not involve the use of data and were focused on design theory and physical toolmaking instead. Three data tags (red-schemed in the rows in Figure 7) and five model tags (blue-schemed in the rows in Figure 7) were used to investigate how neural network-related research publications address reproducibility. Figure 7 shows the frequency of combinations of the eight tags; the more on the left, the more frequently appeared.

Most publications include data description, data sample, and model structure, but do not include data files, learning process, and model (hyper)parameters. So, while these publications provide important information about the data, the

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

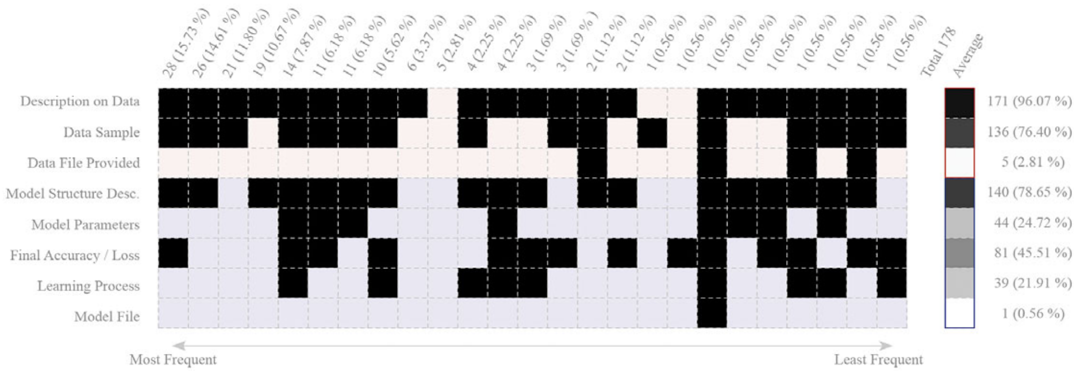

Figure 7. Reproducibility of neural network-related research in CAAD.

information on the model is relatively insufficient. Considering both data and model information is critical for reproducing deep learning-implemented research, for most publications, the experiments were impossible to reproduce in terms of data and model.

5. Discussion

5.1. Findings

Table 2 recaps the findings from the analyses of the characteristics of neural network-related research in CAAD. Trends in neural network-related research can be divided into pre- or post-2015. Neural networks have been used for at least 30 years for design recognition, analysis, synthesis, and evaluation. After 2015, the number of publications increased in all six types of problems. While the design synthesis problem is actively being studied, new research problems have emerged. Moreover, CNNs have actively been implemented for problems using image data, and diverse deep learning models and research subjects have been explored.

However, regardless of these two periods, most of the neural network-related publications do not provide access to the data, learning process, and model (hyper) parameters used in the research. Therefore, they are not reproducible by other researchers and practitioners.

5.2. Impact of CNN to CAAD research

To understand the relationship between technology and research trends in depth it is necessary to consider multiple factors. For instance, research on robotics for architecture based on neural networks is related both to the development and accessibility of robots and of neural networks. Nevertheless, in this article, we focus on reflections directly related to neural network technology. For example, the time-period when there was a significant change in neural network-related research using conventional neural network coincides with when CNN was introduced into CAAD. It is therefore reasonable to assume and investigate how the introduction of CNN affects deep learning-implemented research.

A speculative interpretation of this rapid change in a short period may be related to the progress in image processing (Deng *et al.* 2009) during the 2010s.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

Table 2. Findings from characteristics analysis of neural network-implemented research in computational design

Analysis type	Results
General	Neural networks have been implemented in CAAD for the past three decades.
	The number of neural network-related publications soared post-2015.
	There is a recession in neural network-related research between 2005 and 2015.
Research problem	Research on all six problem types has grown rapidly since 2015.
	The earliest research problem in CAAD using neural networks is pattern recognition.
	Design synthesis is the research problem that has grown most rapidly.
	Design analysis has gone through the shortest recession in 2005–2015.
	Fabrication and theory have emerged recently.
Research subject	Subjects in neural network-related research vary with two slight dominant subjects: drawing and form.
Algorithm trend	Neural networks have been steadily used in CAAD for the past three decades.
	CNN emerged in CAAD research after 2015 and became one of the most dominant trends.
	A transition in the use of models from MLP or SOM to more diversified neural network-based models since 2015.
Data type	Pattern data is being replaced with image or matrix data.
	New data types that can directly represent 3D space have recently been adopted.
Reproducibility	Most publications share data description, data sample, and model structure.
	Most publications do not include data file, learning process, and model (hyper-) parameters.

ImageNet is an open source to provide "a critical resource for developing advanced, large-scale content-based image search and image understanding algorithms, as well as for providing critical training and benchmarking data for such algorithms" (Deng *et al.* 2009).

The organizers of ImageNet hold an annual competition using the dataset for image classification, called ImageNet Large Scale Visual Recognition Challenge (ILSVRC). ILSVRC shows a noticeable improvement of object classification error rate from the 2010s. By 2011, the best entries' error rates were approximately 25% (Robbins 2016). The adoption of CNN (AlexNet) decreased the error rate to approximately 16% (Alom *et al.* 2018). On attaining this achievement, deep neural networks became the image classification technique of choice, and within a couple of years error rates were down to a few percent (Robbins 2016). Nowadays, CNNs are one of the more popular techniques in deep learning; they are widespread across a variety of academics and industrial fields.

The dramatic progress of ILSVRC using CNN affected neural network-related research in CAAD. Figure 5 presents and confirms this emerging trend of CNN in

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

neural network-related research. Pre-2016, there were no publications with CNN as the main model. However, post-2016, the use of CNN has rapidly increased. We can also assume that most GAN models are built with CNN; that is, CNN is the most popular method and model in the current research trend in neural network-related research.

Furthermore, for neural network-related research in CAAD, CNN not only served as a new deep learning model, but it was also the catalyst that brought it out of the recession (2005–2015). In this period, neural networks were used only for design analysis, implemented research continued to a small extent, and there was no expansion of research diversity (Figure 3).

The introduction of CNN revitalized the interest in neural network-related research in CAAD. It led to a drastic increase in the number of neural network-related research in a variety of research problems. With CNN, spatial data and drawings become the major research data types and research subjects respectively. Specifically, CNN provoked renewed research into design recognition, analysis and generation problems in CAAD through the use of drawings and photographs (See Figures 3, 4, 5, and 6).

However, the positive impacts of CNN contrast with the main shortcoming of neural network-related research in CAAD, namely, the low level of reproducibility. CNN requires larger datasets and more complex models, which foregrounds the importance of access to data and models. However, most publications do not provide sufficient information about them or how to access them. Based on our observations, we may expect that without radical changes to research practice, the more CAAD studies use CNN, the more pronounced this problem will become.

5.3. Reproducibility for technical evaluation, educational purpose, accelerated development

The lack of reproducibility leads to three main challenges:

- Poor technical evaluation of neural network-related publications in the review process.
- Research inefficiency within the CAAD community by repeating similar studies.
- Restricted access of methods and findings for architectural education and research.

Without sufficient information on the data and model in neural network-related research, other researchers cannot review and validate the technical progress and value in the publication during the review process. For example, when researchers use CNN, the way to split a dataset into three parts for training, validation, and test is one of the factors that has a profound influence on training outcomes. By changing the ratio between these three parts, learning might fail. Therefore, information on pre-processing must be provided, such as how the data is mixed and split, as well as how the data is collected and structured.

Another important factor in technically validating an experiment is the inclusion of graphs of the learning process. Most papers mention the final error rate or accuracy with the sample of the predicted values or images from trained CNN models. However, given only that information, reviewers cannot identify undesirable training behaviors such as overfitting. Overfitting is a fatal issue "which prevents us from perfectly generalizing the models to well fit observed data on

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

training data, as well as unseen data on testing set" (Ying 2019). Overfitting can "vary significantly in different regions of the model" (Caruana, Lawrence, & Giles 2001) and there are many cases of different overfitting depending on the relationships of the progress between four criteria: training error, training accuracy, validation error, and validation accuracy. By scrutinizing how these criteria have changed during learning progresses, reviewers can decide whether the model is overfitting. In particular, publications that mention 99.99% accuracy using CNN are more likely to be using a model that is overfitted.

If overfitting and data splitting are not identified, a more serious problem arises with regard to the reliability of neural network-related research: "cherry-picking" the results. Here is a scenario for the perfect fabrication of NN-related research results. After a training model is overfitted, if a researcher provides the trained data as test data, the prediction value or image will seem 'good'. In this case, without demonstrating or duplicating the learning experiment, no one can validate whether the experiment results are fabricated or not. Failure to filter out even these basic errors through the review process can reduce the credibility not only of publications but also of CAAD conferences and journals.

The lack of reproducibility also leads to inefficiency in CAAD research by allowing the repetition of similar or identical studies. This repetition hinders the continuous development of neural network-related research based on the successful and inadequate analysis of previous studies in design disciplines. On similar subjects, studies with only slight differences using similar models and data will continue to appear. For example, research on automatic floorplan generation using deep learning is a current trend in CAAD. However, on the training side, it consists of selecting and using plan drawings as dataset for well-known CNN-based generative models. Without investigating the technical capacities of the pre-existing model imported from computer science research, this approach is restricted to the adoption of the model as a design tool. Thereby, it is difficult to produce and accumulate research knowledge on the unique characteristics of deep learning required in design disciplines. If sufficient reproducibility is achieved, researchers can find and overcome technical limitations and challenges of space layout problems using deep learning.

Lastly, the lack of reproducibility inhibits the expansion of neural network-related research by restricting the access to previous deep learning experiments. With the current neural network-related design research environment, even for educational purposes, implementation is almost impossible due to the lack of data and model accessibility. After all, it means that someone must develop new educational materials to teach deep learning in the field of design. By providing the relevant information about the data and models, research publications can be reproduced, which will enable researchers, practitioners, and students to learn how deep learning is used in design research. Access to practical and high-quality materials will facilitate the spread of neural network-related research methods and help to discover new perspectives and knowledge in CAAD research.

5.4. Formatting deep learning papers in CAAD

Despite the sudden increase in CAAD research based on neural networks over the last 5 years and the increasing use of large models that are data-hungry and expensive, there are still no guidelines available for CAAD researchers. Until

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

now, it can be understood that the absence of this guideline has not been discussed in the research community because it has been neither prominent nor ripe. For potentially increasing the quality of discussion and of technical advancements in CAAD, guidelines for formatting neural network-related research publications should be clearly presented to avoid problems caused by the lack of reproducibility. Therefore, following some good practices in the machine learning community, this study proposes a publication format guideline for CAAD research with neural network. These suggestions are summarized in Table 3:

Table 3. Formatting guidelines for research publications for neural network-related research in CAAD

Item	Description
Dataset description	Provides information about the source of the raw data, method of data collection, and the scope and license of data usage.
Dataset parameter	Easily overlooked. Allows researchers and practitioners to accurately reproduce neural network-related research. It includes all parameters used in the pre-processing of the data according to the research purpose – for example, filtering keywords, key for sorting, ratio between training, validation, and test data.
Dataset sample	Samples of the collected and preprocessed data that characteristically represent the entire data. This visually details the data utilized and the methodology employed for its processing.
Data	Disclosing the data itself as well as data information is a way to ensure the reproducibility and reliability of the research. Further, this is particularly important for CAAD research, because it is still deficient in producing and sharing problem-specific datasets. If there is a concern about the intellectual property and copyrights of the data, the researcher can only disclose the data in the peer review process for evaluation. Ideally, learning data related to CAAD research would be archived in an integrated platform such as CumInCAD. However, if there are practical challenges, data can be disclosed by linking existing platforms that are active, such as GitHub or Kaggle.
Model description	The model is an important element for learning. Neural network-related research should describe the architecture of the model and the initialization of the parameters.
Learning parameter	Learning parameters are indispensable for reproducing research. It should include learning rate and its schedule, epochs, hidden layers, activation functions, batch size, etc.
Learning result	This includes not only training, validation, and test accuracy and error rates but also their changes during training. These results are typically presented as a graph.
Model	Disclosing the trained model used by the researcher allows other researchers and practitioners to accurately reproduce the research. As with data, it is possible to set a minimum scope of the disclosure for peer review.

The field of CAAD comprises a vast range of themes and research methods, from creative design experiments to the development of computational methods for problem-solving. In this sense, guidelines for CAAD research require a balance in ensuring scientific accuracy without compromising imagination and flexibility

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

of creative research. The eight items of the guidelines are important for reproducibility and scientific accuracy of neural network-related research, so it is highly recommended that they are available in a repository to support reproducibility and diffusion of knowledge. With that said, depending on the scope of the research, some of the items might be more important than others. For example, a publication that proposes a new learning method for design should provide a proper description of model, parameters, training, and so forth and share them in a repository. However, another research that uses a pre-existing model for a pedagogical experience may only need to share the model and training data in a repository. Through a more comprehensive discussion, the above items could be refined and adapted to the needs of the CAAD research community.

6. Conclusion

In this study, we analyzed the characteristics of neural network-related publications in CAAD. While neural networks are not new in CAAD research, in the past 5 years, the use of CNN has rapidly increased. Considering the temporal coincidence with CNN's dramatic progress, this increase can be interpreted as a phenomenon triggered by the advancements in deep learning, rather than internal development in CAAD. Despite this boom of deep learning in CAAD research, most publications analyzed here do not provide sufficient data and model information to reproduce their experiments. Insufficient reproducibility impairs the reliability of neural network-related research in CAAD, weakens research efficiency due to repetition of similar research, and can eventually inhibit the expansion of future neural network-related research.

Therefore, this paper points out the necessity of formatting guidelines for neural network-related publications in CAAD research and posits eight criteria for improved reproducibility of the publication results. Since these eight criteria address a minimum range of information for reproducibility, they should be further specified through discussion and agreement within the CAAD research community. Due to the data-hungry and data-sensitive characteristics of learning models, the suggested guidelines for neural network-related research are critical for scientific advancements and diffusion of knowledge in CAAD. However, it is important that these guidelines do not function as constraints to suppress research culture and creative experimentation.

For future studies, a larger scope of deep learning with a more diverse set of CAAD publications will be investigated. Terminology from emerging fields such as natural language processing, multimodal learning, and deep reinforcement learning will extend the set of keywords and tags for a more refined investigation of the trends in neural network-related research. Moreover, the future analysis of the trends in the application of neural networks within the field of CAAD could greatly benefit from a more nuanced and granular categorization of the various approaches and techniques utilized. Extending this research, a critical examination of current practices and methodologies, through a lens of differentiated subtopics such as those proposed by Zwierzycki and Wu, will provide a valuable foundation for further understanding and advancing the role of AI in CAAD research.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

Acknowledgment

The authors thank Chloe Hong at Carnegie Mellon University for her help and assistance in organizing the data. The authors are grateful to the anonymous reviewers for their helpful and constructive comments.

References

Alom, M. Z., **Taha, T. M.**, **Yakopcic, C.**, **Westberg, S.**, **Sidike, P.**, **Nasrin, M. S.**, **Van Esesn, B. C.**, **Awwal, A. A. S.**, & **Asari, V. K.** 2018. The history began from AlexNet: A comprehensive survey on deep learning approaches. arXiv:1803.01164 [Cs].

CAAD Futures. 2021. Computer-aided architectural design futures. https://www.caadfutures2021.org/call.html

CAADRIA. 2021. International conference of the association for computer-aided architectural design research in Asia. CAADRIA 2021. https://caadria2021.org/papers-all

Caruana, R., **Lawrence, S.**, & **Giles, C. L.** 2001. Overfitting in neural nets: Backpropagation, Conjugate Gradient, and Early Stopping. In *Advances in Neural Information Processing Systems 13* (ed. T. K. Leen, T. G. Dietterich & V. Tresp), pp. 402–408. MIT Press. http://papers.nips.cc/paper/1895-overfitting-in-neural-nets-backpropagation-conjugate-gradient-and-early-stopping.pdf

CRAIDL. 2020. Creative AI and design launchpad. CRAIDL. https://craidl.group/

CumInCAD. 1998. Research repository. Cumulative index about publications in computer aided architectural design. 1998. http://papers.cumincad.org/

Del Campo, M. 2019. SPAN. SPAN. https://span-arch.org/

Deng, J., **Dong, W.**, **Socher, R.**, **Li, L.-J.**, **Li, K.**, & **Fei-Fei, L.** 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE. https://doi.org/10.1109/CVPR.2009.5206848

DigitalFUTURES. 2020. DigitalFUTURES 2020. https://digitalfutures.international/

Goodfellow, I., **Yoshua, B.**, & **Aaron, C.** 2016. *Introduction to Deep Learning*. MIT Press. http://www.deeplearningbook.org

Kajiya, J. 1993. How to get your SIGGRAPH paper rejected. 4. https://www.siggraph.org/sites/default/files/kajiya.pdf

Kelleher, J. D. 2019. *Deep Learning*. MIT Press.

Krizhevsky, A., **Sutskever, I.**, & **Hinton, G. E.** 2012. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Vol. **25**. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Liu, H., & **Lang, B.** 2019. Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences* **9**, 4396. https://doi.org/10.3390/app9204396

Martens, B., & **Turk, Z.** 2003. Cumulative index of CAAD: Current status and future directions. *International Journal of Architectural Computing* **1** (2), 219–231. https://doi.org/10.1260/147807703771799193

NeurIPS. 2020. NeurIPS 2020 Call for Papers. https://nips.cc/Conferences/2020/CallForPapers.

Newton, D. 2018. Computational architecture research lab. https://www.computationalarchitecturelab.org

Pineau, J. 2019. ReproducibilityChecklist-v2.0.2.

https://doi.org/10.1017/dsj.2023.21 Published online by Cambridge University Press

Rhee, J., & **Veloso, P.** 2021. Type investigation in the form of high-rise building using deep neural network. In *Computer-Aided Architectural Design. Design Imperatives: The Future is Now* (ed. D. Gerber, E. Pantazis, B. Bogosian, A. Nahmad, & M. Constantinos), pp. 175–189. Springer.

Robbins, M. 2016. Does an AI need to make love to Rembrandt's girlfriend to make art? [News Paper]. The Guardian. http://www.theguardian.com/science/2016/may/06/does-an-ai-need-to-make-love-to-rembrandts-girlfriend-to-make-art.

Rodríguez, J., **Villafañe, M.**, **Piškorec, L.**, & **Caparrini, F.** 2020. Generation of geometric interpolations of building types with deep variational autoencoders. *Design Science* **6**, e34. https://doi.org/10.1017/dsj.2020.31

Russell, S. J., & **Norvig, P.** 2021. *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sarker, I. H. 2021. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science* **2** (6), 420. https://doi.org/10.1007/s42979-021-00815-1

Silver, D., **Huang, A.**, **Maddison, C. J.**, **Guez, A.**, **Sifre, L.**, **van den Driessche, G.**, **Schrittwieser, J.**, **Antonoglou, I.**, **Panneershelvam, V.**, **Lanctot, M.**, **Dieleman, S.**, **Grewe, D.**, **Nham, J.**, **Kalchbrenner, N.**, **Sutskever, I.**, **Lillicrap, T.**, **Leach, M.**, **Kavukcuoglu, K.**, **Graepel, T.**, & **Hassabis, D.** 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* **529** (7587), 484–489. https://doi.org/10.1038/nature16961

Wu, A., **Stouffs, R.**, & **Biljecki, F.** 2022. Generative adversarial networks in the built environment: A comprehensive review of the application of GANs across data types and scales. *Building and Environment* **223**, 109477. https://doi.org/10.1016/j.buildenv.2022.109477

Ying, X. 2019. An overview of overfitting and its solutions. *Journal of Physics: Conference Series* **1168**, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

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

Zwierzycki, M. 2020. On AI adoption issues in architectural design–identifying the issues based on an extensive literature review. In *Anthropologic: Architecture and Fabrication in the Cognitive Age - Proceedings of the 38th ECAADe Conference - Volume 1, TU Berlin, Berlin, Germany, 16–18 September 2020* (ed. Werner, L. & Koering, D.), pp. 515–524. eCAADe (Education and research in Computer Aided Architectural Design in Europe). http://papers.cumincad.org/cgi-bin/works/paper/ecaade2020_139