A Case Study on Architectural Sketch Recognition Utilizing Deep Learning Networks for Exterior and Interior Datasets

Veli Mustafa Yönder[1], Hasan Burak Çavka[2], Fehmi Doğan[2]

[1]Kütahya Dumlupınar University, Kütahya, Turkey
[2]İzmir Institute of Technology, İzmir, Turkey
velimustafa.yonder@dpu.edu.tr; fehmidogan@iyte.edu.tr
hasancavka@iyte.edu.tr

Abstract. Sketching is a pivotal component in facilitating the effective conveyance of ideas and the actualization of architectural design concepts. The potential applications of machine learning and computer vision algorithms in the fields of technical drawing and architectural graphic communication are substantial, presenting a diverse array of possibilities. This research investigates the effectiveness of deep learning-based classification techniques in analyzing both indoor and outdoor freehand architectural perspective drawings. Furthermore, the transfer learning approach was employed in this binary classification problem. The primary aim of this study is to train deep neural networks to recognize and interpret freehand architectural perspective drawings effectively and precisely. In this context, pre-trained models such as GoogLeNet, ResNet-50, AlexNet, ResNet-101, Places365-GoogLeNet, and DarkNet-53 were fine-tuned. The findings indicate that the ResNet-101 architecture has significant levels of validation accuracy, yet the validation accuracy of the Places365-GoogLeNet and AlexNet pretrained models is comparatively lower.

Keywords: Machine Learning, Transfer Learning, Drawing Recognition, Deep Neural Nets, Image Classification

1 Introduction

The utilization of sketching as a means of conceptualizing and resolving design challenges is a commonly employed approach within the field of design activity. Sketching serves as a multifaceted instrument for not only effectively communicating ideas, but also as an indispensable process in the materialization of design concepts. Researchers in the domain of design have conducted investigations on the effectiveness of sketches as a method for developing and visualizing concepts (Bilda et al., 2006). To solve an extant design problem or to visualize an existing space, architects or designer may sketch. Besides, sketching holds a significant role in the curriculum of architecture students' technical drawing and graphic communication courses, as it can be executed through various drawing techniques and styles. Multiple styles and technical approaches are employed to enhance the expression of ideas or to visually represent existing concepts in a more technically proficient manner. Hence, it is imperative to adhere to specific drawing conventions such as proportion, scale, material selection, and others, as well as observe established rules including line types, proximity-distance, and line thickness, among others. These standards are also applicable to the practice of freehand drawing. Nevertheless, freehand drawing offers a greater degree of flexibility and tolerance. This phenomenon can present certain benefits in particular contexts, while also posing specific difficulties in other scenarios.

Machine learning (ML) and computer vision (CV) algorithms have the potential to be utilized in technical drawing and architectural graphic communication for a multitude of purposes. Research has been conducted on the application of advanced algorithms for the recognition of architectural plans (Gimenez et al. 2016; Zeng et al. 2019; Jang et al. 2020; Huang and Zheng 2018; Kim et al. 2021; Goyal et al. 2021). In the field of computer vision, computational methods play a crucial role in enabling the recognition of contours and shapes of objects within an image by means of employing diverse filters and geometric operations. The advancement of intricate artificial intelligence (AI) models has led to the emergence of artificial neural networks, including Capsule Networks (CapsNet), Graph Neural Nets (GNNs), and Convolutional Neural Networks (CNNs), which have potential usage in the field of object recognition and scene classification problems. In this context, convolutional neural networks (ConvNet or CNNs), one of the deep networks, have many different layers (O'Shea and Nash, 2015). CNN's most significant layer is convolution, which requires the most time, and network performance is dependent on the number of layers (Albawi et al., 2017). Deep learning models possess the capability to effectively address challenges, such as those encountered in gradient problems. These models have the capacity to be utilized not only in the context of classification and clustering problems, but also in various image and video generating applications. The task of image classification poses significant challenges and necessitates a substantial amount of data. The extraction of various features from images can be facilitated by the utilization of the training data. A range of pretrained AI models can be employed for the purposes of image classification and clustering operations. With the assistance of pretrained models, classification procedures can attain high levels of accuracy. The models went thru training using extensive datasets, such as ImageNet, and were optimized to achieve high success scores. In addition, the utilization of these networks necessitates significant processing resources (i.e., extensive input data) and encompassing hardware specifications, owing to their intricate nature. These networks are customized for specific problems through the process of fine-tuning and hyper-parameterization, which involves adjusting various mathematical constraints. “Transfer learning” approach has also potential efficacy in this context. Transfer learning is a technique that enables the utilization of models that have been trained on extensive datasets for the purpose of addressing different classification problems.

Artificial intelligence methodologies and models have a potential utility in addressing challenges related to image classification and drawing classification tasks. However, it is important to note that these approaches are characterized by their inherent complexity. Training the models holds significant importance, particularly in the domain of drawing classification. Hence, this research paper presents a field study into the classification of freehand architectural perspective drawings. Classification methods based on deep learning are examined in this study to see how well they perform on both interior and exterior sketches. The fundamental goal of this research is to understand the capacity of deep neural networks to recognize and correctly interpret freehand architectural perspective drawings.

1.1 Research Questions and Scope of this Study

This study examines the efficacy of deep learning-based classification methods on both indoor and outdoor sketches. The initial objective involves the optimization and fine-tuning of deep neural networks for the purpose of accurately identifying and interpreting freehand architectural perspective drawings. Within this context, thereby it is compulsory to carry out a comparison examination of the scores and performance demonstrated by the models. The main questions of this research are as follows:

i. Is it possible for a deep learning model to accurately predict whether freehand architectural perspective drawings depict indoor or outdoor scenes?
ii. What are the training parameters (e.g., epoch, learning rate, optimization algorithm, training time, and validation frequencies etc.) of the deep learning network models if the deep networks accurately predict the outcome?

2 Theoretical Background

The computer scientists and neurobiologists trying to understand the relationship of cognitive actions such as learning, decision-making, and problem-solving with the brain have contributed to the modeling and development of artificial neural networks. Biological learning systems exhibit a sophisticated network of interconnected neurons, which has served as a source of inspiration for the advancement of artificial neural networks (ANNs) (Mitchell, 1997). An artificial neural network that can learn a mathematical function that maps points from one set to another set is trained using backpropagation and gradient descent algorithms (Zhou, 2021). Shallow artificial neural network structure basically consists of input, output, bias, weights, and hidden layers (Figure 1). The configuration of interrelated layers and the structure of the network are determined in accordance with the research question (e.g., ill-defined, or well-defined). Nevertheless, the operational mechanism of these networks can be compared to a black box due to the presence of hidden layers. Still, they possess the capability to effectively address numerous engineering challenges (Tayfur, 2014). The field of modern machine learning (e.g., supervised, unsupervised, reinforcement, and semi-supervised) incorporates a wide range of methodologies that are characterized by their multidimensional nature, consisting of various constituent components and models. AI tools usage areas have diversified with the centralizing of data (i.e., data-centric approach) and the development of reference models (i.e., model-centric approach) over time.

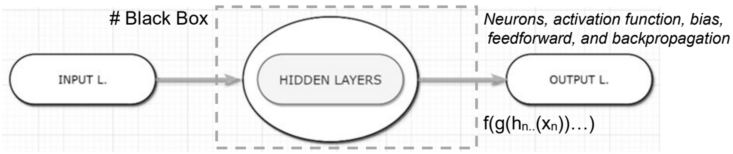

Figure 1. Artificial Neural Networks' (ANNs) Diagrammatic Structure.
Source: Authors, 2023.

Modern machine learning and deep learning algorithms undergo testing in various sub-disciplines, including computational design, architectural digital fabrication, building information modeling (BIM), building energy analysis, sustainable design protocols, architectural heritage, and project management (Figure 2). Furthermore, the influence of generative artificial intelligence (AI) has been extensive in various academic disciplines, with particular attention given to its implications for the design industry. Generative artificial intelligence (generative AI) encompasses various cutting-edge methodologies, including generative adversarial neural networks (GANs), variational autoencoders (VAEs), diffusion models, and transformer architecture (NVIDIA, n.d.). The utilization of generative models in design exhibits significant potential across various stages, ranging from visualization to concept design. Furthermore, textto-image diffusion models could have a potential use in the field of generative models, especially in generating new design alternatives and combining several variations.

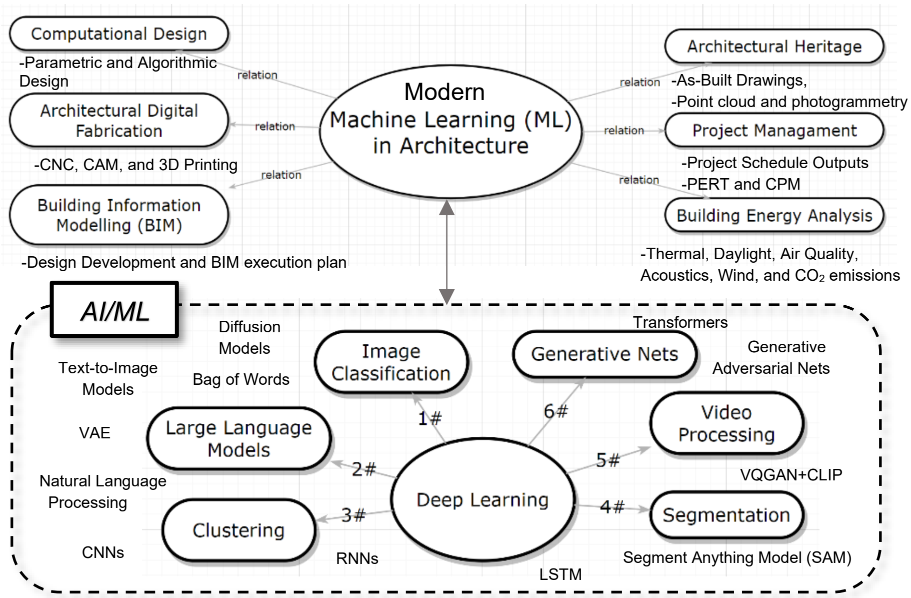

Figure 2. Interaction Areas of Modern AI and Architecture.
Source: Authors, 2023.

This study employs deep neural networks as opposed to shallow networks, given the effectiveness of deep neural networks in image classification. Consequently, a deep learning approach is adopted. “Autoencoder (AE), Deep Belief Network (DBN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Recursive Neural Network and Direct Deep Reinforcement Learning” are all examples of common deep learning methods (Shinde and Shah 2018, p.4). The domains of interest in artificial intelligence (AI) encompass the utilization of big data for learning purposes, as well as the concepts of “transfer learning” for developing solution to new problems. Transfer learning plays a vital role in machine learning when addressing the fundamental challenge of inadequate training data (Tan et al., 2018). Pan and Yang (2009, p. 1356) examine the concept of transfer learning, categorizing it into three distinct types: “inductive transfer learning, transductive transfer learning, and unsupervised transfer learning transfer learning.” Consequently, the transfer learning method was employed in this study for deep learning pipeline.

Architectural technical drawing entails various subcategories. The sub-headings encompassed within the topic include perspective drawing, orthographic projection, 3D representation methods, and sketching (Figure 3).

Several studies have been conducted employing artificial intelligence methods and computer vision methodologies within the context of drawing/sketch classification. Within the scope of these studies, the process of generating the dataset and the subsequent training of machine learning algorithms are separate efforts. In the field of sketch classification, the TU Berlin sketch dataset (Eitz et al., 2012) is an example of a comprehensive dataset (composed of 20,000 human sketches) that has been used in numerous studies. Another example dataset is Sketchy database. There are 12,500 original images and 75,471 hand-drawn drawings of the same objects included in this database, which extends across 125 groups (Sangkloy et al., 2016, p. 119:3). The process of generating the drawing dataset entails the utilization of novel procedures, and novel methodologies can also be employed for the purpose of training neural networks. CNN architectures have the potential to be employed for the purpose of sketch classification. Ballester and Araujo (2016) studied the effectiveness of GoogLeNet and AlexNet networks in classifying drawings. In the study of Hayat et al. (2019, p.447), they suggested "a deep CNN-based framework for sketch recognition via transfer learning with global average pooling strategy."

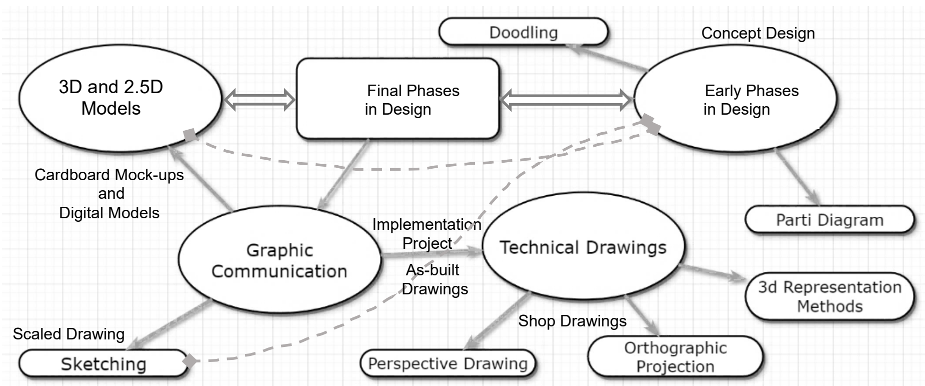

Figure 3. Architectural Graphic Communication Stages.
Source: Authors, 2023.

The disparity in the amount of detail observed in the drawings necessitates the CNN architecture to have the ability to discern and comprehend varying levels of information. According to Zhang et al., (2020), the main focus in the field of sketch recognition pertains to acquiring robust and distinguishing features. In cases where the input dataset differs from huge datasets like ImageNet, which are commonly used to train traditional convolutional neural network (CNN) designs, it may be advantageous to develop specific problem centric CNNs. Yu et al. (2015, p.2), introduced a novel deep neural network (DNN) called Sketch-a-Net, which is designed for the purpose of recognizing freehand drawings.

3 Methodology

Classification of drawings can be solved using either traditional computer vision techniques or deep learning models. Deep neural networks can be used in the domain of drawing classification (Yu et al., 2015; Ballester and Araujo, 2016). The methodology employed in this study comprises several components (Figure 4). Every primary division comprises subordinate headings. The sequential process consists of the following steps: data acquisition, data labeling, data augmentation, data splitting, selection of a convolutional neural network, fine tuning, and benchmarking. The procedures were executed for both outdoor and indoor freehand drawings. The raw dataset has a total of forty-five images, with an equal distribution of fifty percent indoor and fifty percent outdoor drawings. The drawings are predominantly rendered in black and white, with a limited number being depicted in color. Data collection was conducted using handheld image scanner devices that had a resolution of 900 DPI (i.e., high resolution level was selected). Some freehand drawings were discarded due to their failure to meet the desired quality standards (e.g., scanning errors).

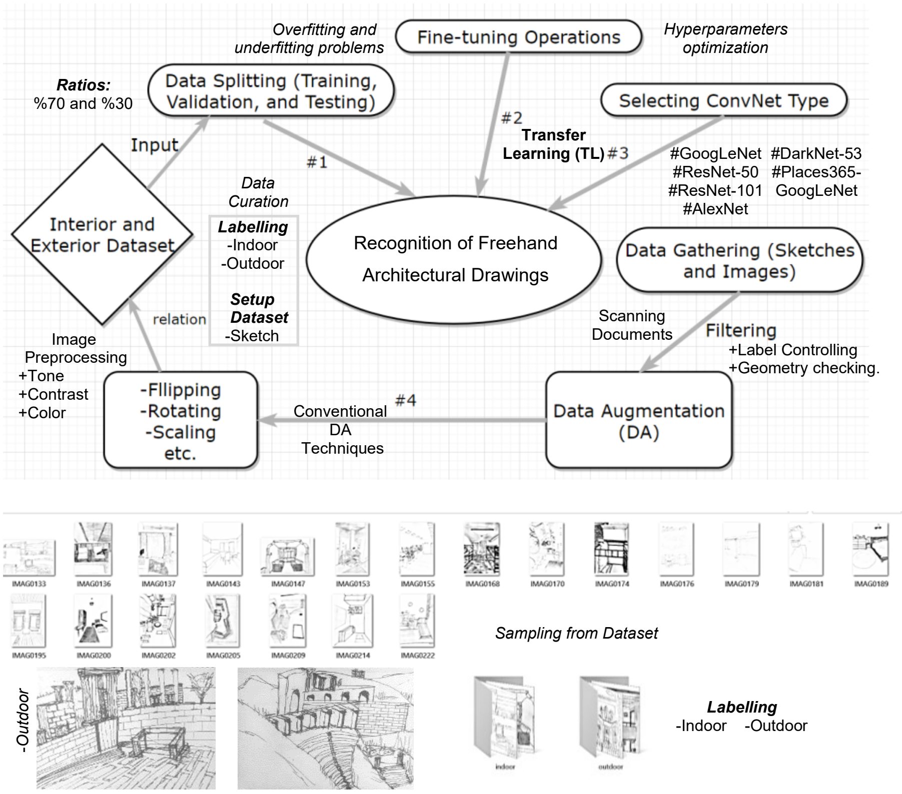

Figure 4. Schematic Workflow of the Study
Source: Authors, 2023

Data augmentation (DA) techniques commonly employed in classical approaches include random rotation, random rescaling, and translation. The entire dataset was partitioned into two distinct categories: indoor and outdoor. The dataset was partitioned into two distinct subsets, namely the training set (%70) and the validation set (%30), to facilitate the process of model training (Figure 5). The pretrained models, namely AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2014), DarkNet-53 (Redmon and Farhadi, 2018), ResNet-50 (He et al., 2016), ResNet-101, and Places365-GoogLeNet (GoogLeNet was trained on Places365 dataset) were subsequently trained on drawings. The complete dataset has been labeled as indoor and outdoor. The training phases of the models were conducted within the MathWorks Matlab® R2023b environment. Besides, the hardware specifications employed consist of an NVIDIA RTX3080TI graphics card and an Intel i7 12700K 3.60 GHz processor.

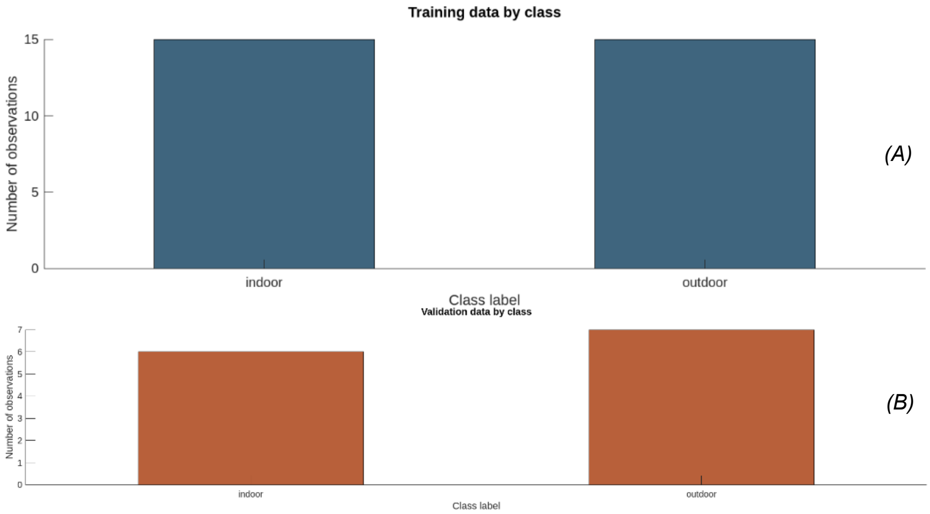

Figure 5. Distribution of Training (A) and Validation (B) Datasets by Indoor and Outdoor Class'. Source: Authors, 2023.

4 Results

The collection of architectural perspective drawings was trained using different pretrained network models. The training of the dataset model, both indoors and outdoors, involved the utilization of 30 epochs and 50 iterations. Conversely, during the validation process, the frequency of iterations was set at 50. Besides, learning rate was set as 0,01. The pretrained network models utilized in the study were GoogLeNet, ResNet-50, Places365-GoogLeNet, ResNet-101, AlexNet, and DarkNet-53. Pretrained network models consist of different depth, parameter, and size metrics. The models were ranked based on their validation accuracy, with ResNet-101 achieving the highest accuracy, followed by ResNet-50, GoogLeNet, DarkNet-53, Places365-GoogLeNet, and AlexNet (Table 1).

Table 1. Pretrained Network Models' Scores

Network Type	*Validation Accuracy*	*Training Time (sec)*	*Epoch*	*Validation Frequency (iterations)*	*Learning Rate*	Pretrained Network Model		
						Depth	*Parameters*	*Size*
GoogLeNet	69%	94	30	50	0.01	22	7.00 M	27
ResNet-50	77%	163	30	50	0.01	50	25.6 M	96
Places365-GoogLeNet	46%	95	30	50	0.01	22	7.00 M	27
ResNet-101	84.62%	203	30	50	0.01	101	44.6 M	167
AlexNet	46%	25	8	50	0.01	8	61 M	227
DarkNet-53	62%	229	30	50	0.01	53	41.6 M	155

Source: Authors, 2023

The ResNet-101 model was selected for its superior validation accuracy. Validation accuracy and loss values are visualized using plotted graphs (Figure 6).

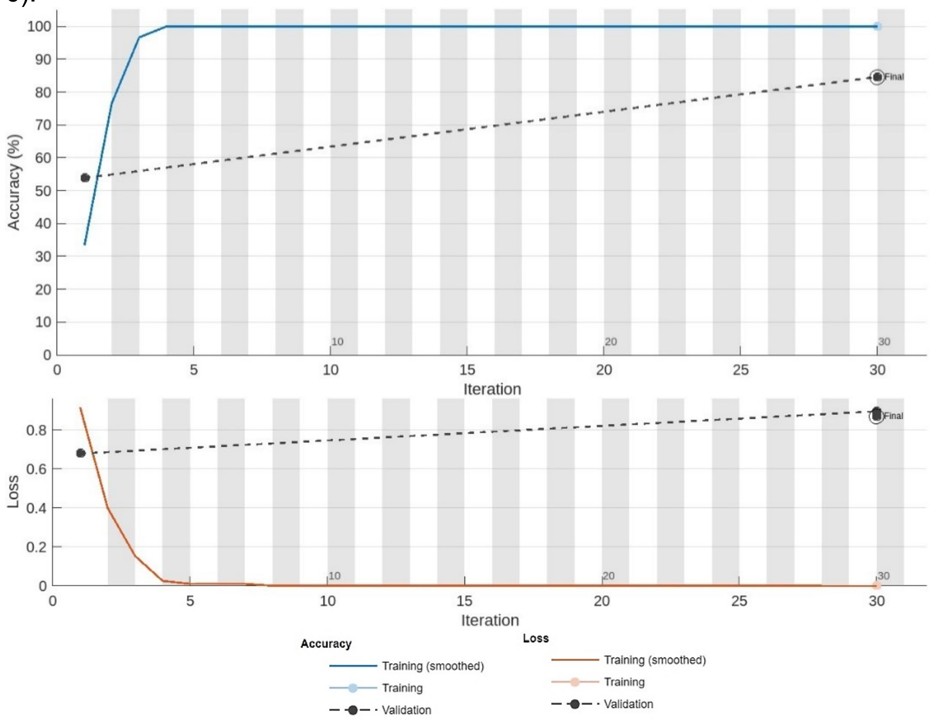

Figure 6. Validation Accuracy (Upper) and Loss (Lower) Plots
Source: Authors, 2023

5 Discussion

Within the field of architecture, sketching is a commonly employed technique for the purpose of ideation and problem-solving. In addition to its efficacy in communicating concepts, sketching plays a crucial role in the manifestation of design ideas. Architects and designers frequently employ sketching as a means to address design challenges or gain a sense of a prospective space. The classification of sketches holds significant relevance for computational designers and computer scientists across various domains. Hence, this work employs a limited dataset of architectural drawings to categorize them based on the binary classification principle. The architectural drawings database comprises pre-rendered perspective drawings. This study employs deep learning methods in lieu of traditional computer vision and machine learning technologies. Deep learning, which is a specialized branch within the field of machine learning, exhibits a multi-layered architecture and demonstrates commendable efficacy in the domains of image classification and object identification. The impact of deep learning approaches is steadily increasing across various domains, ranging from generative AI to computer vision (CV). Modern deep learning and data-centric algorithms have the potential to be utilized in architectural field for a multitude of purposes. Innovative applications of deep learning algorithms, scripting libraries, and apps can be explored by computational designers and architectural theorists.

This study involved training deep learning networks using the transfer learning approach. The training results indicate that the ResNet-101 model achieves a high level of validation accuracy. This network has a greater number of parameters and a deeper architecture compared to previous networks. The dataset used in this study is somewhat restricted, however it will be expanded in subsequent investigations. Furthermore, the quantity of tested deep neural networks will be augmented.

References

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)* (pp. 1-6). Ieee.

Ballester, P., & Araujo, R. (2016, February). On the performance of GoogLeNet and AlexNet applied to sketches. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, No. 1).

Bilda, Z., Gero, J. S., & Purcell, T. (2006). To sketch or not to sketch? That is the question. *Design studies*, *27*(5), 587-613.

Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects?. *ACM Transactions on graphics (TOG)*, *31*(4), 1-10.

Gimenez, L., Robert, S., Suard, F., & Zreik, K. (2016). Automatic reconstruction of 3D building models from scanned 2D floor plans. *Automation in Construction*, *63*, 48-56.

Goyal, S., Chattopadhyay, C., & Bhatnagar, G. (2021). Knowledge-driven description synthesis for floor plan interpretation. *International Journal on Document Analysis and Recognition (IJDAR)*, *24*(1-2), 19-32.

Hayat, S., She, K., Yu, Y., & Mateen, M. (2019). Deep cnn-based features for hand-drawn sketch recognition via transfer learning approach. *Editorial Preface From the Desk of Managing Editor…*, *10*(9).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Huang, W., & Zheng, H. (2018, October). Architectural drawings recognition and generation through machine learning. In *Proceedings of the 38th annual conference of the association for computer aided design in architecture, Mexico City, Mexico* (pp. 18-20).

Jang, H., Yu, K., & Yang, J. (2020). Indoor reconstruction from floorplan images with a deep learning approach. *ISPRS International Journal of Geo-Information*, *9*(2), 65.

Kim, S., Park, S., Kim, H., & Yu, K. (2021). Deep floor plan analysis for complicated drawings based on style transfer. *Journal of Computing in Civil Engineering*, *35*(2), 04020066.

Mitchell, T. M. (1997). Machine learning.

NVIDIA. (n.d.). What is Generative AI? Retrieved from NVIDIA website: https://www.nvidia.com/en-us/glossary/data-science/generative-ai/

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359. (p. 1356)

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Sangkloy, P., Burnell, N., Ham, C., & Hays, J. (2016). The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)*, *35*(4), 1-12.

Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE. (p. 4)

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2014. Going deeper with convolutions. CoRR abs/1409.4842.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece,* *October 4-7, 2018, Proceedings, Part III 27* (pp. 270-279). Springer International Publishing.

Tayfur, G. (2014). *Soft computing in water resources engineering: Artificial neural networks, fuzzy logic and genetic algorithms*. WIT Press.

Yu, Q., Yang, Y., Song, Y. Z., Xiang, T., & Hospedales, T. (2015). Sketch-a-net that beats humans. *arXiv preprint arXiv:1501.07873*.

Zeng, Z., Li, X., Yu, Y. K., & Fu, C. W. (2019). Deep floor plan recognition using a multi-task network with room-boundary-guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9096-9104).

Zhang, X., Huang, Y., Zou, Q., Pei, Y., Zhang, R., & Wang, S. (2020). A hybrid convolutional neural network for sketch recognition. *Pattern Recognition Letters*, *130*, 73-82.

Zhou, Z. H. (2021). *Machine learning*. Springer Nature.