



Integration of convolutional and adversarial networks into building design: A review

Jean Parente^{a,b,*}, Eugénio Rodrigues^a, Bárbara Rangel^{b,c}, João Poças Martins^{c,d}

^a Univ Coimbra, ADAI, Department of Mechanical Engineering, Rua Luís Reis Santos, Pólo II, 3030-788 Coimbra, Portugal

^b CEAU/T2P, University of Porto, Via Panorâmica Edgar Cardoso 215, 4150-564 Porto, Portugal

^c CONSTRUCT/GEQUALTEC, Department of Civil Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

^d BUILT CoLAB – Collaborative Laboratory for the Future Built Environment, 4150-003 Porto, Portugal

ARTICLE INFO

Keywords:

CNN
GAN
Deep learning
Generative models
Building design

ABSTRACT

Convolutional and adversarial networks are found in various fields of knowledge and activities. One such field is building design, a multi-disciplinary and multi-task process involving many different requirements and preferences. Although showing several advantages over traditional computational methods, they are still far from being part of the daily design practice. Nevertheless, if fully integrated, these methods are expected to accelerate design and automate procedures. This paper reviews these methods' latest advances and applications to identify current barriers and suggests future developments. For that, a systematic literature review extended with forward and backward snowball methods was carried out. The focus was on the first design phases, including site layout, floor planning, furniture arrangement, and facade design. The network models show great potential in exploring novel design paths, comparing alternative solutions, and reducing task-associated time and cost. In addition, newer approaches may benefit from combining convolutional and adversarial networks in decision-making since they may complement analysis and synthesis. However, the lack of a smooth integration into the design process and the need for a high-level mastery limit their widespread use. Furthermore, ethical issues arise, such as models being trained with biased datasets, ignoring the intellectual property of the data creators, potential violation of privacy, and the models limiting human creativity.

1. Introduction

Researchers have strived to develop generative design methods for several decades to create candidate solutions from the sometimes conflicting project requirements and physical constraints of the building [1]. These methods require the user to specify numerous requirements and constraints while demanding user experience in fine-tuning the method's parameters to produce feasible solutions [2]. In contrast, current learning methods find new solutions with minimal user input, relying on past training data to learn implicit patterns [3]. Once the model learns the probability distribution, the method synthesizes new examples with a high degree of realism [4]- an efficient supervised approach in terms of time and cost [5].

Building design is a creative process that cannot be fully described mathematically and may express personal preferences and judgments [6]. Although several technical, functional, and performance aspects may be objectively formulated [7], its associated subjectivity poses an obstacle to any automated generation or optimization procedure [8]. Nonetheless, researchers developed many

* Corresponding author. Univ Coimbra, ADAI, Department of Mechanical Engineering, Rua Luís Reis Santos, Pólo II, 3030-788 Coimbra, Portugal.
E-mail address: jmnparente@gmail.com (J. Parente).

efforts to use heuristic models. For example, the shape of a building may result from the interaction between project requirements and the physical constraints of the building, such as automated floor planning [9,10]. However, such an approach requires the user to specify many design inputs and to choose adequate algorithm parameters to produce feasible solutions [2]. Other examples of automated procedures are facade design [11,12], generation of indoor 3D scenes [13], and building energy and thermal optimization [14,15].

The literature divides the generative methods used in building design into image analysis (or computer vision) and image synthesis (or computer graphics) [16]. While computer vision can analyze an image and create a model, computer graphics provides tools that enable the analysis model to create images [17]. Image analysis and synthesis are foundation procedures in image processing [18].

Image analysis has achieved relevant results in visual recognition tasks [19]. In general, image analysis methods can be divided into four steps [20]: (i) image acquisition, (ii) image input and pre-processing, (iii) feature extraction and segmentation, and (iv) recognition and application. In image acquisition, digital images are collected or produced from satellites, cameras in aerial vehicles, photographic cameras, or 3D scanners. In pre-processing, image denoising or enhancements are carried out to improve the quality of the images. In feature extraction, specific image information, such as edges, corners, and contours, is obtained, and images are segmented into homogeneous groups without overlapping regions. Lastly, in image recognition, images are classified (categorizing and labeling pixels, or regions in an image, into several groups), objects are detected (bounding and classifying objects), registered (overlapping images into a single image), retrieved (searching for images in a database), and reconstructed (creating a 3D model from images).

Unlike traditional generative design methods, deep learning methods discover implicit patterns from the training data [3]. Once the probability distribution is learned, these models synthesize new examples with high realism [4]. Deep learning is an efficient supervised approach in terms of time and cost, which can be identified from the increasing accuracy rates in various technology areas [5] as in the following building design tasks: image classification [21], semantic segmentation [22], object detection [23].

Two deep learning methods are predominant in literature: Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN). CNN is the most widely used deep learning approach in computer vision applications [24], and its architecture involves three main components: the visible, hidden, and output layers. In the visible layer, the image's descriptor is the input, and the hidden layers (convolutional, pooling, and fully connected) learn high-level representation by gradually extracting low-level features [25,26]. In the case of convolutional layers, these layers extract features from the convolution operation developed for a filter (usually a kernel) composed of weights and biases, which generate feature maps [26]. Then, an activation function is applied, making it possible to model non-linear functions in the network—this process repeats on every convolutional layer. Next, the pooling layers replace a small neighborhood of feature maps with an invariant function, such as max or mean operations. This replacement reduces the dimension of the feature image and network parameters, thus invariant to shift and distortion [26]. Lastly, fully connected layers recognize and classify the image by adjusting the weights, which determine the degree of influence of the input layer on the output layer.

Different algorithms or networks derived from CNN can classify, detect, or segment images [27]. In image classification, the most popular architectures are AlexNet [28], Visual Geometry Group (VGG) [29], and GoogLeNet [30]. In image detection, the most used are the region-based CNN (R-CNN) [31], Fast R-CNN [32], and You Only Look Once (YOLO) [33]. Lastly, in image segmentation, U-shaped encoder-decoder (U-Net) [34] and RefineNet [35] are the most common.

A model generates a new image by learning the distribution of features from past information [16,18]. For example, a house layout may be created using a model that learns the general rules of possible arrangements from a dataset with a specific building typology. The Boltzmann machine, deep belief network, Variational AutoEncoder (VAE), and GAN are examples of methods that can be used. GANs comprise two competing networks, a generator and a discriminator [36]. These networks behave like a min-max game in an iterative optimization process. While a generator produces synthetic data from random noise, the discriminator evaluates results as real or fake images. As the two models compete, the generator learns to deceive the discriminator, culminating in an undistinguishable synthetic image from a real one.

The number of studies using deep learning in building design has been increasing steadily. The applications are vast, including object detection in floor plans [23], indoor scene segmentation [22], and indoor image classification [21]. Other building design-related applications are the design of an indoor scene [37], facade layout [38], and floor plan arrangement [39]. However, a critical literature review on applying these methods in architectural design needs to be included. The reviews which might be considered closest to the topic cover artificial intelligence in general [40], real estate [41], and construction [42].

This paper attempts to fill the gap mentioned above and answers the following research questions:

- Q1 – ‘Which are the latest advances in CNN and GAN that are oriented towards architectural design?’
- Q2 – ‘What are the challenges and barriers to integrating these methods into architectural design?’

The present paper is divided into six sections. After the current introduction, the systematic literature review method is described in the material and methods section. Then, the literature analysis is carried out by covering different building design-related applications and benchmarking sections. Next, the research questions are answered, and the implications of the findings are debated in the discussion section. Finally, the main conclusions of this study are presented.

2. Material and methods

In order to answer the stated questions, a systematic literature review on the latest advancements in CNN and GAN in architectural design was carried out in a three-step methodology.

In the first step, ‘Literature Survey,’ title words, such as “architectural design,” “building footprint,” “deep learning,” “CNN,” “GAN,” “networks,” and “image” were combined to search for publications in the Scopus and Google Scholar databases. A period filter was applied from January 2016 to December 2022 to limit the analysis to the last six years. An initial sample of 1201 publications was obtained (Fig. 1). After eliminating review articles, thesis, and publications dealing with different topics but having the same nomenclature, 155 relevant publications were selected and used as a gold set. Lastly, the snowballing method was used, an iterative process with two complementary parts: *backward* and *forward* [43]. *Backward snowballing* starts by selecting other articles from the reference list of the 155 publications. The relevance of the identified publication is then analyzed, and the reference list is once again screened. When the new publications match the search goal, these are included in the overall sample—this iterative process ends when new papers can no longer be found. A similar case occurs in *forward* snowballing, but other papers that cite the initial set of documents are selected. After completing both parts, the final sample included 394 publications, divided into 109 conference proceedings, 23 preprints (ArXiv), and 262 journal articles. Foundational publications in deep learning are also included in this sample, despite some being published before 2016. As the initial search had produced a number of false positives due to taxonomy similarities among the different topic areas, a method that allowed us to find and screen each publication was required. This explains the reason for the use of snowballing methods, resulting in a significant increase in publications from the initial set.

During the second stage, ‘Data Extraction and Tabulation,’ information from the retrieved publications was extracted according to four applications of deep computer vision in building design: building footprint (perimeter of a building measured on a horizontal plane), building floor plan (visual representation of a home layout on a horizontal plane), 3D indoor scene (3D representation of interior room), and building facade (representation of external faces of a building on an elevation). The information is then grouped in each category according to topics in computer vision, such as segmentation, retrieval, generation, vectorization, fusion, detection, classification, and reconstruction. Finally, references, method architecture, input and output variables, and publication conclusions are tabulated.

In the third and last step, ‘Synthesis of Findings,’ the tabulated information is analyzed and synthesized to answer the research

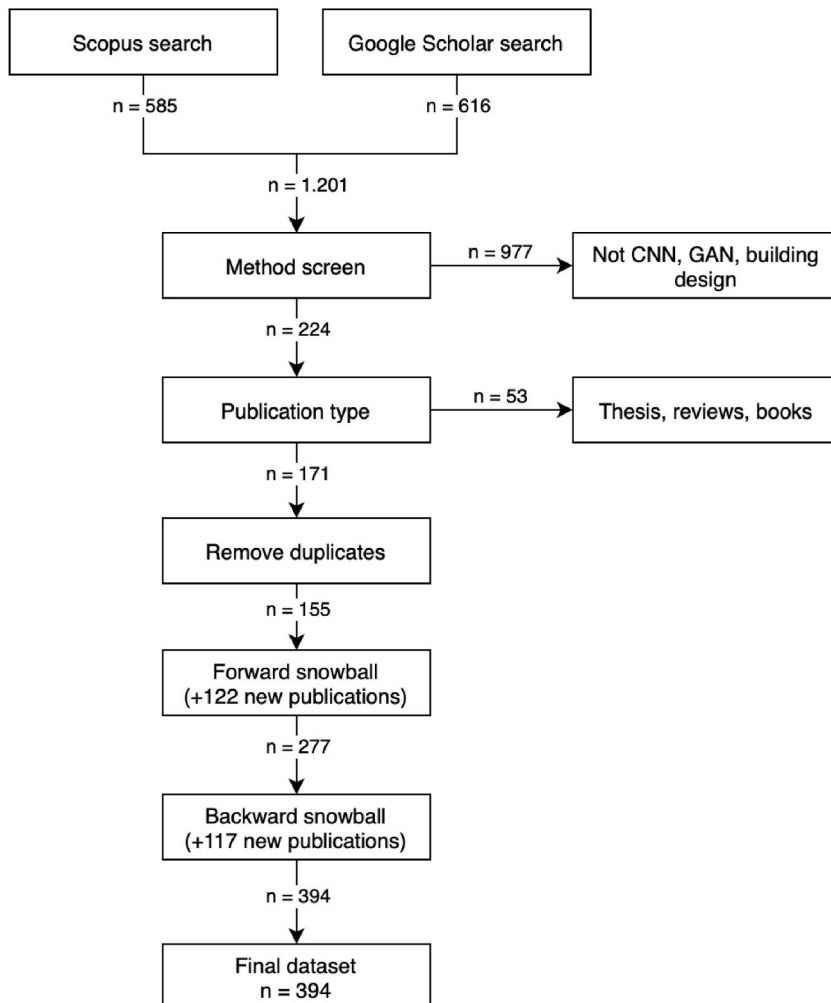


Fig. 1. Literature survey flowchart.

questions. Lastly, information about network architecture and dataset sources is compiled to develop deep learning approaches.

3. CNN and GAN-based applications

The initial stage of this study entailed the analysis of publications on CNN and GAN in building design. The statistical analysis included: (i) quantification of publications that used CNN or GAN-based methods in image analysis or image synthesis procedures and (ii) distribution of publications per computer vision tasks into four building design categories: building footprint, building floor plan, building facade, and 3D indoor scene.

This initial assessment indicated that the CNN model was the most used technique in image analysis (Fig. 2). The model is particularly effective in building footprint and 3D indoor scene tasks due to its ability to extract features accurately and rapidly without any human supervision from remote sensing data or other imagery. Specific examples of applications are to identify roofs, facades, furniture, and other building system components. Several derived architectures were implemented, including VGGNet, FCN, Faster-RCNN, Mask R-CNN, U-Net, and residual connections (ResNet).

Rather than extracting information, the GAN model creates new synthetic images. An example is creating building floor plans from a simple building boundary or the location of the openings in the perimeter. Another example is recreating a building facade or reconstructing a 3D indoor scene. This capability to produce or complete imagery is particularly helpful when the data sample is insufficient or the image resolution needs to be enhanced.

As shown in Fig. 3, building extraction is the most common application in the building footprint category, having the same number as all other applications combined (segmentation, detection, classification, and generation). However, CNN and GAN-based techniques generating new building footprints have nine cases.

The generation process stands out in the building floor plan category with 34 studies (nine CNN and 25 GAN), while segmentation has 14 (13 CNN and one GAN). In this case, the number of applications matches one of the most creative yet most common tasks among designers: the spatial organization of a building layout. A segmentation task in a layout design involves recognizing building floor plan elements from learning semantic information, such as walls, doors, windows, furniture, and their relationships. Other applications (two GAN and nine CNN in total) include the detection of spaces, retrieval of information, reconstruction of design, and classification of spaces.

In the building facade task, segmentation is dominant with 21 studies (19 CNN and two GAN). It fulfills similar functions as in the floor plan category but is applied to the exterior surfaces of the building and its components. This is followed by generation, CNN with seven and GAN with one publication, as this plays an essential role in facade composition studies. Finally, detection, reconstruction, and retrieval complete the other ten studies (12 CNN).

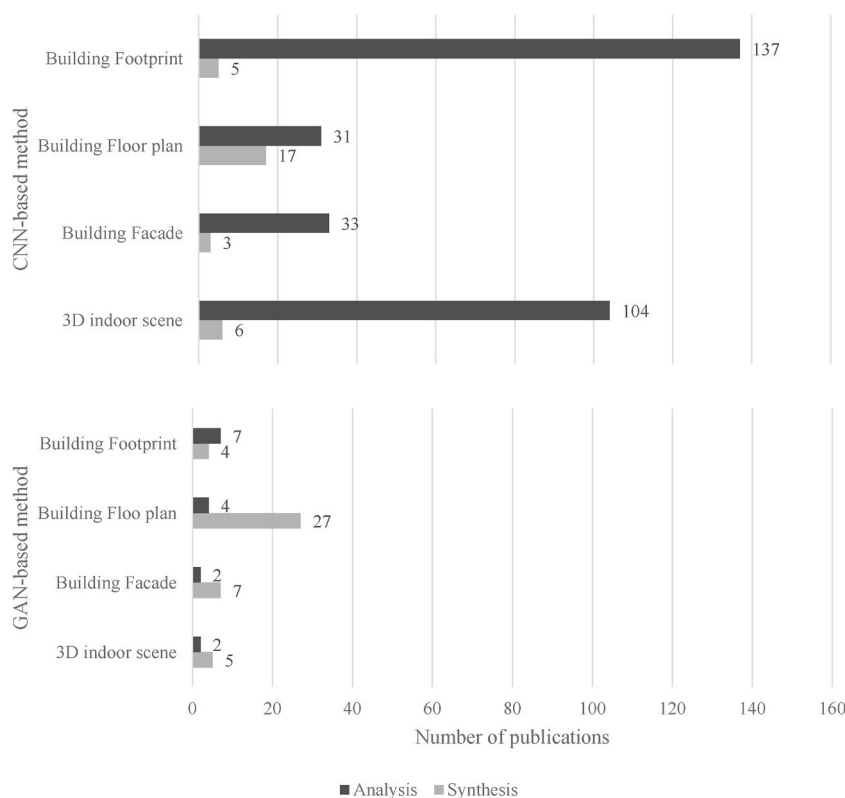


Fig. 2. Distribution of CNN and GAN-based methods in image analysis or synthesis procedures.

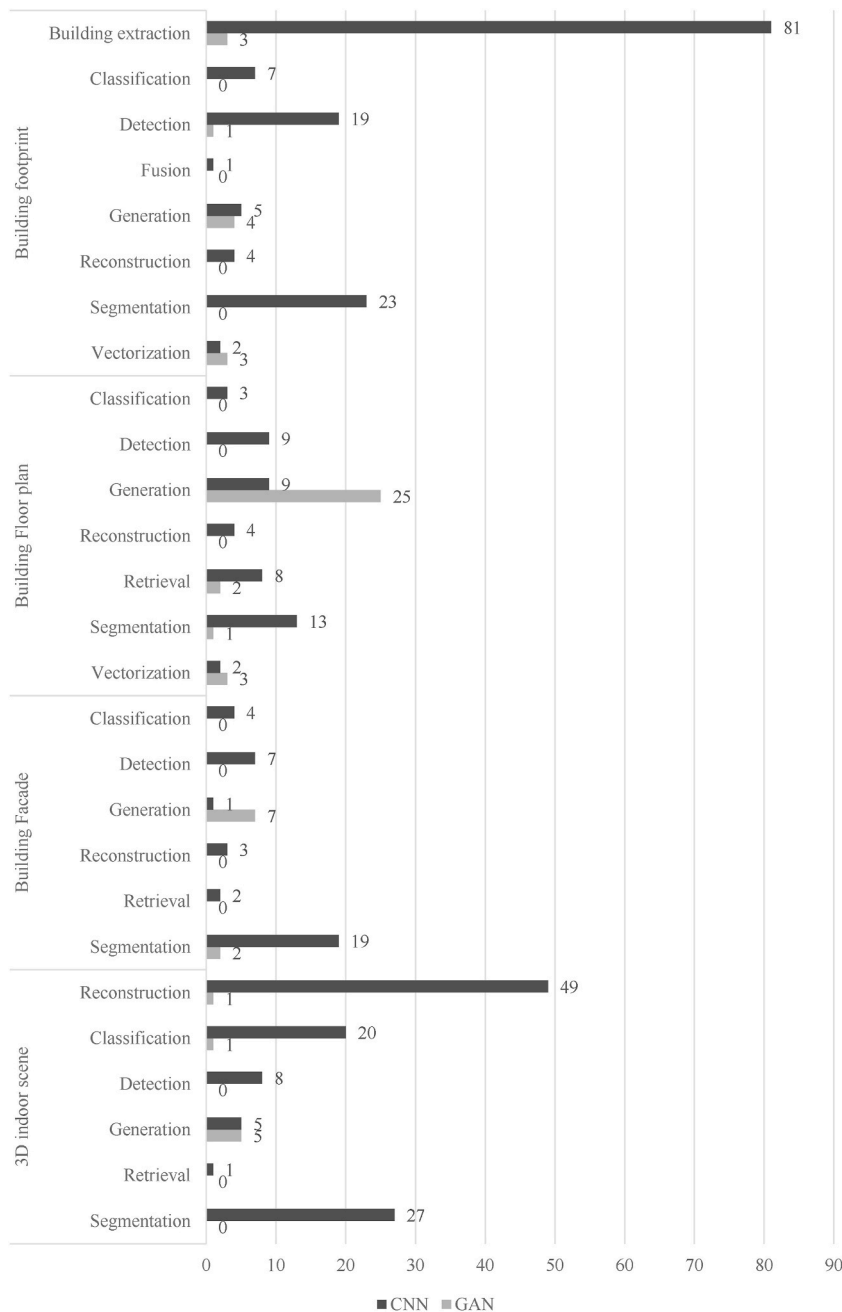


Fig. 3. Distribution of publications per computer vision task.

In the 3D indoor scene category, reconstruction (49 studies in CNN), which estimates the 3D arrangement of a room from 2D images gathered using robots and virtual or augmented reality, is followed by segmentation (27 studies in CNN) and classification (20 and one in CNN and GAN, respectively). Generation also stands out with ten publications (five studies in both methods) where studies show creating 3D layouts for indoor scenes or changing a layout from a 3D image. Finally, detection and retrieval add up to the nine CNN cases.

The following subsections compare the last CNN and GAN methods with traditional ones for each main building design task. Segmentation (CNN) and generation (GAN) stand out among all construction categories. However, other relevant tasks in their specific categories are shown, such as building extraction (footprint), retrieval and detection (floor plan), classification, and reconstruction (facade and 3D indoor scene). Furthermore, these comparisons show the evolution of the analysis (CNN) and synthesis (GAN) processes in different building categories. Thus, it is also possible to identify integration paths between computer graphics and computer vision in building design.

3.1. Building footprint

The building footprint may be retrieved from remote sensing techniques, such as through digital sensors aboard satellites and airplanes [44] or other sources, including unmanned aerial vehicles (such as light detection and ranging), crowdsourcing (phone imagery), and advanced driver-assistance systems.

Thanks to CNN, mainly Fully Convolutional Networks (FCN), the interest in image segmentation in remote sensing has grown recently [45]. Some examples of CNN architectures in building footprint include FCN, Faster/Mask R-CNN [46], and Res-U-Net, which are based on ResNet and SegNet. The main advantage of CNN is the capacity to accumulate contextual information about large objects over vast receptive fields. However, the process becomes challenging with low spatial resolution and blurry object boundaries [47]. Nonetheless, several strategies were employed in many recent studies to improve the semantic segmentation process in remote sensing [46].

For decades, non-deep learning methods such as Random Forest [48], Adaboost, and Support Vector Machine [49] were used in building extraction. However, these were not applied to complex, high-diversity urban regions [50]. Moreover, these methods were based on handcrafted features, making it a problem for non-experts as experience is required in order to define specific design features [51], particularly due to the contour features of buildings, their complex morphology, and occlusions from shadows. The CNN-based methods presented higher accuracy [51–55].

Automated building extraction reduces time and cost when used to produce large-scale maps. In addition to FCN, other CNN approaches are widely explored in remote sensing tasks, such as ResNet [56–58] and U-Net [59,60]. Nevertheless, these methods still demanded efforts to reduce the prediction error to the ground truth pixels [61]. To address this issue, GAN has a specific architecture for classifying output images as real or fake. GAN to segmentation map [62] improves accuracy with an F1-score of 96.8%, compared with 95.2% from an FCN model. To help with image completion, Wasserstein GAN architecture (WGAN) [63] creates patches to fill gaps in an unfinished image. The network uses residual learning procedures to improve the produced picture. In a remote sensing application, WGAN and CGAN make up the CWGAN, which improves the quality of building footprint generation. When exploring the advantages of WGAN and CGAN, the CWGAN-GP method [61] improved the quality of building footprint generation compared to methods such as CGAN and U-Net.

Res2-UNet [64] is a new and prominent study of building footprint detection from high-resolution remotely sensed images. The purpose of this algorithm is to deal with intricate details of background objects and problems, such as small buildings, which can be omitted in an urban context. Res2-UNet achieved state-of-the-art performance over three public datasets: WHU Aerial building, Satellite II Dataset, and Massachusetts building. However, this method's limitation is distinguishing some confusing bare roads with buildings, which causes background objects to be classified erroneously as buildings.

3.2. Building floor plan

Image recognition of floor plans is an effective procedure in automated document analysis, part of the computer vision tasks [6], from which techniques such as floor plan retrieval [65], floor plan vectorization [66], and automatic floor plan generation [39] can be developed.

Architectural floor plan analysis is a specific application domain of graphics recognition whose automation was incorporated into computer-aided design platforms [67]. In recent decades, several works explored techniques for floor plan recognition, such as room and wall segmentation [67–72]. Other steps follow floor plan analysis in deep computer vision, such as information segmentation, structural analysis, and semantic information extraction and alignment [70]. These steps made it possible for graphic segmentation VGG [73], graphic and text segmentation using FCN architecture [6], wall extraction and detection using U-Net [74], and Faster R-CNN [23]. Beyond the traditional orthogonal forms of floor plan representation, irregular shapes may also be recognized from kernels and the GAN training process [75].

The floor plan retrieval procedure is (i) an image-based recommendation that includes the preferences of a home purchaser or (ii) a creative source of information to help the designer in the initial stage of the design process. Traditionally, researchers use these techniques for symbol recognition [76,77] or segmentation of the floor plan [69,78]. For example, the a.Scatch system is a sketch-based application that automatically extracts semantic structure from old projects or an architect's sketch to retrieve a similar floor plan in a repository [78]. In the last years, modelers used a modified AlexNet [65], FRCNN [79], and Cyclic GAN [80,81] to capture semantic features from floor plans. The latter is more efficient in matching the sketch and the image query [80]. While the Conditional GAN needs an aligned image pair to be trained, the Cyclic GAN does not.

An architectural drawing, such as a floor plan or cross-section, is often created in a vector format and later converted into a raster image for publication in a digital medium. This procedure limits the post-process, such as analysis, synthesis, or modifications. Moreover, this procedure results in the loss of structured geometry and semantic information [82]. Recovering this is a challenging process (known as vectorization), which converts raster data into vector data from the transformation of pixels into geometric primitives, such as lines, polylines, and curves. This problem may be solved using several methods, including a skeleton-based approach [68] and coupling of Hough Transforms with vectorization [69]. For example, a CNN algorithm developed for low-level rasterized semantic and geometric information achieved 90% accuracy in a vectorization process [66]. In another example, an application of vectorization used a multi-task ResNet to create a large-scale floor plan with 5000 samples arranged into 80 categories from the dataset known as CubiCasa5K [82].

Recently, researchers have been developing generative models to create novel floor plans. The primary method used in layout and furniture generation is GAN, which can be adopted singly [39,83,84] or linked to some complementary source, such as a graph of the connections between the building spaces (each node represents a room, and each edge represents an adjacency). For example, because

GAN only has the noise signal to control the generation process, InfoGAN [85] improves the method by using a given graph to produce and control a variety of design solutions. Using this method in floor plan generation allows the creation of early conceptual designs through latent code learning to incorporate topology features or functional configuration spaces obtained from various designs [86]. Another example of this application is the House-GAN [87]. House-GAN was developed specifically to generate a house floor plan which combines GAN and a graph constraint. The output quality is measured according to realism, diversity, and compatibility with the input graph constraint. Similarly, even though Graph2graph is a CNN-based approach, it is also graph-constrained and allows the user to generate a variety of floors for the same input boundary plans [3].

Conditional GAN (cGAN) [88] is a notable generate method that creates synthetic layouts satisfying the geometric constraint from bubbles diagrams and space allocation heat maps to achieve topological preferences. In addition, this method allows the designers greater control over the result, which is often a challenge for many unconditional GAN methods [89]. However, the main limitation of this work is that there is no application for typologies with more than one floor. When addressing the multi-story generation, vertical circulation spaces and their relationship with the remaining rooms must be addressed in order to have coherent designs.

3.3. 3D indoor scene

The 3D scene understanding of the indoor environment has grown in recent years, mainly in applications such as augmented reality, virtual reality, robotics, games, and interior design. The artificial intelligence agent must recognize the scene's functional attributes and semantic labels and understand apparent and hidden relationships between these components [90]. Therefore, these applications consider the complex geometric and semantic context of all parts of the analyzed space and their relationship.

Regarding human visual scene interaction, modelers should consider human perception beyond the analysis of image features and include the user's memory, language application, and constraints of the visual system [91]. For example, measuring a person's feelings in interior design is promising research on human behavior and decision-making. In this case, the emotional responses are determined using an electroencephalography-based deep learning model [92].

The reconstruction of 3D indoor scenes, which consists in building a three-dimensional shape from two-dimensional images, was frequently studied in the last decade. Manhattan world assumption [93], Bayesian network recognition [94], and geometric context [95] were some of the methods used. Recent methods, such as CNN, improve 3D reconstruction, which can be accomplished with a single input image [96–102]. However, incomplete images, low-depth resolutions, missing data, and sensor noise still pose a challenge [90]. For example, LayoutNet [98] works directly on the panoramic image, unlike other recent studies that decompose it into perspective images. Compared to previous methods, this method is similar to RoomNet [102] but has better accuracy with perspective images. The improvement of the LayoutNet occurs due to image alignment from the vanishing points, predicting multiple layout elements, using the Manhattan constraints. Another example is HorizonNet [101], which also uses panoramic images but outperforms LayoutNet by representing room layout as three 1D vectors that encode the edges between walls-floors, ceiling-wall, and wall-wall.

A relevant way to understand a 3D scene is semantic segmentation from a 3D point cloud. A volumetric 3D point cloud representation can be achieved with a voxel grid, which consists of a regular grid in the 3D space. Several investigations have adopted the combination of voxel grid and convolutional frameworks, such as DGCNN [103], 3D-FCN [104], SCSS-Net [105], and CNN with conditional random fields (CRFs) [106]. The CRFs combine the advantage of classification, graphical modeling, and efficient parameter optimization. Therefore, it is an expected resource for 3D point semantic segmentation [107–111]. In addition, 3D fully convolutional networks (3D-FCN), combined with CRFs, are potential deep learning methods for the classifier stage in 3D point cloud segmentation [112].

There are some approaches to 3D indoor scene modeling – the arrangement of a room from the relationship of a given set of furniture and elements– such as human activities centric [113], example-based data-driven scene synthesis [114], and action-driven 3D indoor scene evolution [115]. The data-driven scene modeling may also be based on other types of resources, such as text [116, 117], sketches [118–120], and images [121]. Recently, researchers used a graph conditional prediction network (GC-LPN) to estimate the layout of a room and a language conditional texture GAN (LCT-GAN) to generate an interior texture [122]. The generative recursive autoencoders [123] and hybrid generative models (GAN and VAE) [124] are also alternatives in indoor scene synthesis.

To achieve high accuracy in 3D indoor scenes, parsing methods, such as PGDENet [125], brought substantial improvements. In the RGB-D image case, the method works by having multimodal information (color and depth) and fusing it into discriminable features that allows a correct scene classification. Trained with NYUv2 and SUN datasets, the model achieved superior performance than previous methods. However, despite the availability of structured models which integrate various data types to enhance scene understanding, there are still challenges in effectively retrieving and fully utilizing the different modalities to ensure accurate and efficient results. Additionally, this method often suffers from reduced performance in the presence of changes in illumination, resulting in unclear object boundaries and difficulties in detecting small objects [126].

3.4. Building facade

One of the applications in building facade design is classification according to the facades' historical style. Modelers used a multimodal latent logistic regression [127], Support Vector Machines [128], and k-means [129–132] to group images according to a building's historical period in Mexico [133]. In addition, these techniques, among others, allow the grouping of architectural elements into categories, such as doors, windows, and columns.

The parsing of building facades has evolved over the last years in applications such as 3D city modeling. The challenge of this task is defined by the variations of scenarios, the changes in illumination, visual perspective, and occlusions [134]. One of the most common techniques for 3D city modeling is the reconstruction of the building facade, which, together with the scene understanding, is of interest to architectural design, digital scenarios in movies, and virtual environments in games. For many of these applications in deep

learning, facade segmentation has been a fundamental vision task for detecting components of the facade from architectures such as CNN [135], ConvNets [38], Mask R-CNN [136], Faster R-CNN [137], RPN [138], SegNet [135,139], FCN [140], and Pyramid ALKNet [141]. Pyramid ALKNet has performed better with occluded regions, ambiguities, and other image issues.

The 3D reconstruction from a 2D facade image or point cloud is another application of the methods. Researchers applied computer vision methods in location-based services, architectural restoration, and urban planning. Automated facade reconstruction may be employed using a grammar-based algorithm describing the visual structure [142–146]. Compared to the global methods, this approach is flexible for accounting facade variation [145]. In recent years, researchers reconstructed a 3D building from point clouds of facades using CNN [147] or Faster R-CNN in complex scenarios that included texture [148] and GAN-based FCRN-Depth in cases with a higher number of segmentation classes [149].

The generation of facades is a complex challenge because its design must relate to environmental, functional, structural, and aesthetical parameters. For this reason, a multi-objective genetic algorithm is a helpful optimization procedure to find the best performance solution from a combined set of two or more decision variables. For example, in daylight optimization, building shape results from the relationship between parameters such as the window-to-wall ratio, number of windows, window distribution, length of shading devices, and others [150]. The split grammar, a variety of shape grammar [151], is also a shape control approach for generating a facade which offers an application to create different building styles [152,153]. This approach's main limitation is its ability to learn, although it is suitable for automating design creation from rules defined by the designer [154]. The pix-to-pix, a variety of cGAN, is a framework that learns semantic facade images from network training and creates a new 2D design [155], City-GAN [156], and DCGAN [154].

Deep learning with prior knowledge of facades [157] can detect their elements automatically in pixel-wise complex environments. The framework outperforms other state-of-the-art models on two public datasets: Ecole Centrale Paris (ECP) and ArtDeco. One of the limitations of this study is the laborious labeling of building facade elements. Transfer learning can be used as an alternative to address this issue, yielding improved accuracy with less labeling data.

3.5. Recognition and other applications

The architectural design process is often divided into analysis, synthesis, and appraisal, as demonstrated in Fig. 4. Complementarily, the decision made in an early design stage sets the starting point for a designer to elaborate on the next stage or design process. The CNN and GAN methods, on the other hand, can be divided into recognition (e.g., image segmentation, object detection, and image classification) and other applications (e.g., automation in vectorization, image retrieval, reconstruction, and generation). The integration between CNN-GAN methods and early architectural design can be represented by Fig. 5, which summarizes the main steps of this approach in this section.

CNN and GAN allow automation in different steps of the design process, which can be identified as automation in recognition, automation in the query process, automation in the conversion process, and automation in the generation process.

Automation in recognition is the basis of applications in building design since all applications use at least object detection, such as chair, door, and bed [23], facade segmentation [158], land use [159], and land cover [160] classification, or 3D scene recognition [161] in their computer vision process. Most of the methods used in these applications are CNN. On the other hand, the role of GAN in this process can be to improve the accuracy of the segmentation map [62], floor plan [75], and image completion [63].

Through the different CNN and GAN methods for recognition, it is possible to transform image representation into data, which is required for the analysis stage to make decisions in building design. Then, in sequence, the application for the synthesis stage is started.

Automation in the query process is the comparative part of the design and can count on the support of several references in the analysis process. Sketch-based image retrieval (SBIR) applied to floor plans [65,80,162,163] and facades [164,165], for example, is the task that explores the intuition humans express through the relationship between drawing and technical image results.

The use of sketch as an aid in the creative stage is usual in developing an idea in architectural design. Therefore, the automation in the query process can be seen as a methodological approximation between the traditional design model and the CNN-GAN method. With this result, it is possible to automate the conversion process, a step to follow to make the result manageable in CAD.

Automation in the conversion process can be divided into vectorization and reconstruction. While vectorization converts raster format into a vector file, reconstruction converts a real object into a 3D model through sensors or a set of 2D images obtaining space depth estimation from a 2D image. For example, a 3D layout estimation from a 2D image helps build design analysis and create new datasets [166]. In addition, these samples can be used to train and test generative networks for new building designs.

In architectural design, the analysis phase usually requires building recognition as indoor layout information [167], which may be

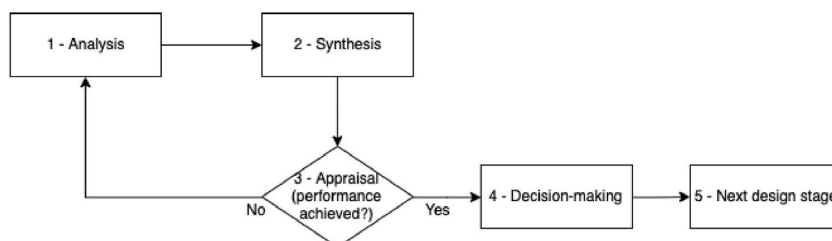


Fig. 4. Early architectural design method.

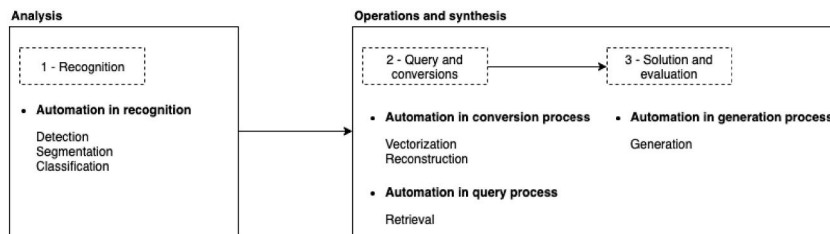


Fig. 5. Integration between CNN-GAN methods and early architectural design.

implemented in complex drawings with better scores in a one-to-one match compared to existing methods. Therefore, a vectorization function [66,167,168] is required to enable CAD manipulation. These results are then used in the automation for 3D modeling.

When the CNN-GAN method is used for building design, **automation in the generation process** occurs between the synthesis and decision-making stages. For example, with the GAN method, feasible layouts [169], facades [170], or different 3D indoor scene styles [171] are created according to training image input and network architecture. The generative process is also fundamental when using synthetic datasets and increased dataset size to train and evaluate the classification model [92].

The generation process is an iterative process that needs to be evaluated. If the result is not suitable, the process should return to the analysis stage to resume the process until a new evaluation is carried out.

In the following, from Figs. 6–9, the main structure identified throughout the paper when analyzing the relationship between CNN-GAN methods and building design is presented. In this structure, a pattern composed of data input, analysis, and synthesis tasks was found. Furthermore, these results have two flow categories: required and possible. Required flow is the necessary path to completely fulfill a task or a set of tasks, while the possible flow is only an alternative.

Modelers use image analysis for urban planning, diligence for law, real estate for land development, landscape architecture, environmental remediation in engineering, and building design for architectural concepts. In addition, the result of possible flows between the analysis and generation processes (Fig. 6) allows decision-making in building design, such as in the dimensional aspects of the site: boundaries, location, zoning classification, and dimensional implications.

With land size, zone, and constraints, the information will flow through the retrieval, conversion, and analysis to make the floor plan components recognizable and manipulable, such as walls and doors (Fig. 7).

The analysis of the 3D indoor scene (Fig. 8) is composed of 3D reconstruction (considering geometry, semantic, and topological modeling) and recognition (including semantic, detection, and classification) of indoor scene components. The possible flow between the analysis and synthesis processes, or other inputs such as 3D images and natural language text, allows the generation of a new 3D indoor scene and its furniture arrangement.

Reconstruction creates a facade model from a 2D image and a 3D point cloud (Fig. 9). This process results in facade segmentation, components detection, such as a frame for doors and windows, and architectural style classification. Finally, the last possible flow follows from the analysis to the synthesis process, when the generator is able to create a new building facade or 3D outdoor scene, completing the analysis and synthesis stages of the conception of form in architectural design. In summary, the whole process consists in determining the location of the building, arrangement of the rooms, distributing equipment and fixtures in indoor scenarios, and designing the facade of a building by integrating deep learning into architectural design conception.

3.6. Advantages, challenges, and barriers

The main advantages of CNN are: being free from hand-crafted featurings and having better generalization capability than other classification-based approaches, both in accuracy and efficiency [172]. CNN frequently performs better than other machine learning techniques for sketch-based image retrieval in building design [173]. Other sketch recognition techniques, such as histograms of oriented gradients and scale-invariant feature transform, cannot capture the abstract nature of sketches as CNN [80]. Also, those other recognition techniques are hand-crafted image classification approaches. CNN-based applied in building design has in its favor the ability to deal with complicated background details and detection of small buildings in building footprint dataset [64], in addition to detecting elements in complex environments on building facades dataset [157] or achieving high accuracy in 3D scene recognition [125]. As a hypothetical example, CNN may reconstruct a 3D scene using meshes, textures [174], and complex shapes [175] from a floor plan image.

In building footprint, the advantages of GAN can be found in improving the image quality [61] or complete images due to occlusions [62]. In floor plans [86] and 3D indoor scenes [122], the combination between GAN methods and graphs or bubble graphs and space allocation heat maps [176] allows greater control over the outcome of the design process. Together with CNN, for example, GAN-based can convert raster-to-vector with combinatorial optimization in junction units of the floor plan with complex drawing [177] and, combined with U-Net for vectorization of the building, it can generate 3D models directly [168]. The GAN method is frequently used for the creative process in the early design phase, especially in facades [178] and floor plans [3].

However, integrating these methods also faces several challenges and barriers, such as the lack of a ready-to-use tool for most users and the lack of interpretability in the deep model, also known as a black box issue [179]. Other limitations are found when the method is unable to distinguish bare roads and buildings from the building footprint [64], tackle light changes, detect small objects, and retrieve RGB-D data from different 3D modalities of an indoor scene [125]. Although CNN's key ingredient is data training [180], a

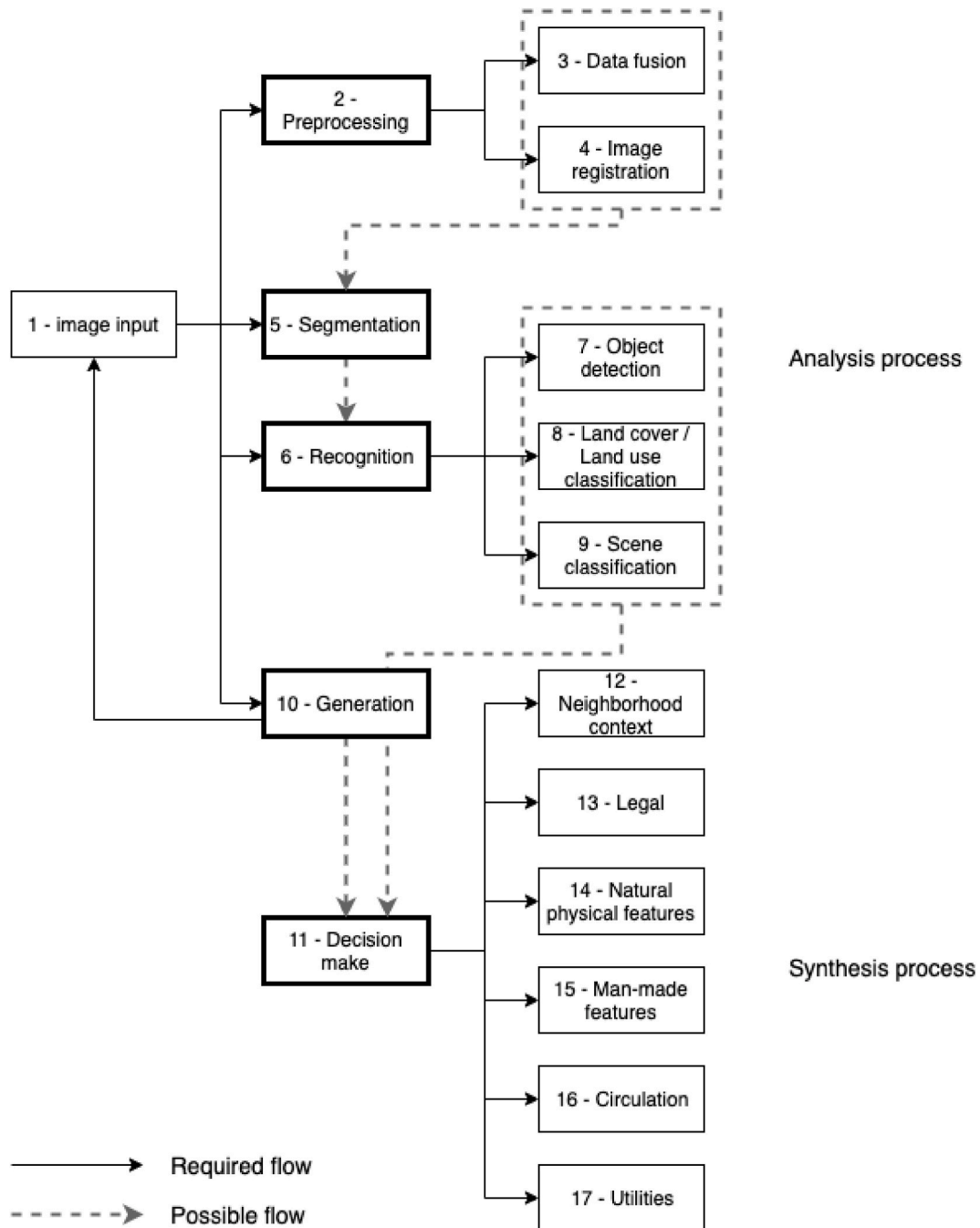


Fig. 6. Analysis and decision-making from the relationship between deep learning in remote sensing.

technical challenge is the limitation of size and quality of publicly available building footprint datasets [46]; lack of floor plan representative datasets [181]; insufficient data for indoor scenes, if compared to the outdoor scene [182]; and the challenge to collect large-scale annotated dataset for segmentation [183]. High-quality datasets still need to be more common despite the rise in free datasets [180] – several present scenes with occlusions or low-quality images, resulting in the need for pretreating data. Data annotation, on the other hand, is an arduous and manual task requiring precision, thus prone to error [184]. Furthermore, having human experts label and annotate the building design data about architectural elements, styles, and spatial relationships is indispensable. Supervised methods may require human labeling [181,185], which is time-consuming and costly.

It is also important to highlight the ethical challenges associated with creating models from datasets. First, datasets may be biased [186]. For example, these may favor a particular architectural style, architect, or cultural aspect that will be prevalent in the generated

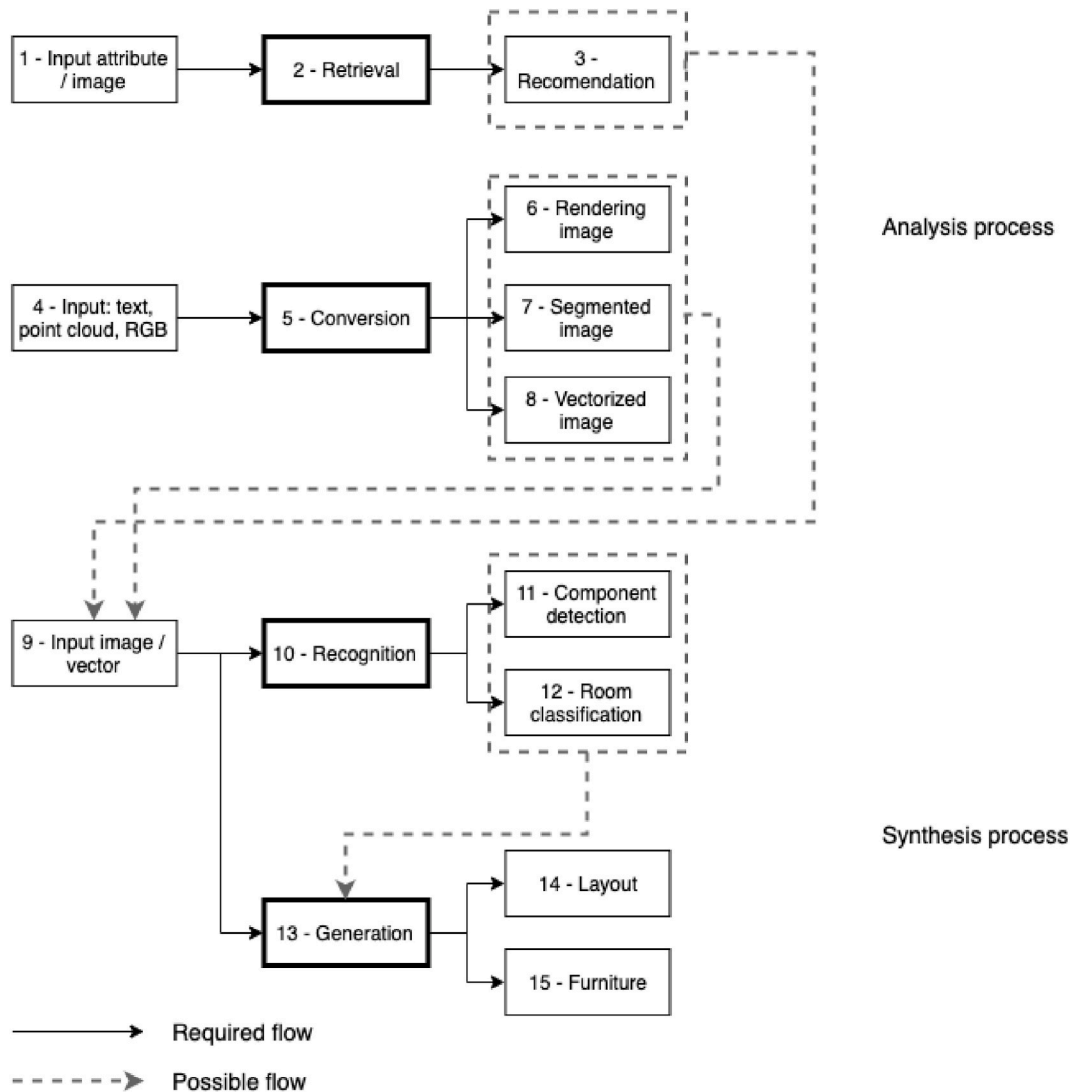


Fig. 7. Layout generation and furniture from the relationship between deep learning and 2D layout plan.

designs. If how the model was trained is opaque, in the sense that the source data is unknown, the user may be unaware that the produced result may segregate or have negative cultural impacts. Second, these models may limit the role of humans in the creative design process [187]. Human creativity in building design involves applying judgment, intuition, and subjective decision-making based on experience, values, and ethical considerations that vary among societies and users. And lastly, it is important to comprehend how these models will respect the intellectual property and privacy rights of the data creators and data subjects [188]. The data creators have authorship and intellectual rights that are not being credited for or monetarily compensated if the model is used commercially. In addition, the source data may contain imagery subject to privacy rights, which may vary depending on each nation's legislation. Thus, transparency, explainability, and mechanism for human supervision [189] are crucial to address some of these ethical issues.

4. Methods and benchmarks

In addition to mastering network architectures, deep learning requires large datasets of images when applied to architectural design. Although many papers use proprietary data, there are several open-access datasets available. Table A 1 (Appendix A) lists these datasets according to different task categories: footprint image, facade image, 3D indoor scene image, 3D point cloud outdoor scene, 3D point cloud indoor scene, and floor plan image.

Table A 2 (Appendix A) presents the most common CNN and GAN methods found in the literature review. Several network architectures are used, such as ResNet-based methods in category building footprint [56,57,190–195], GAN-based in building floor plan [80,81,83,84,86,167,169,196], AlexNet-based in building facade [38,197–200], and SegNet-based methods in a 3D indoor scene [98,

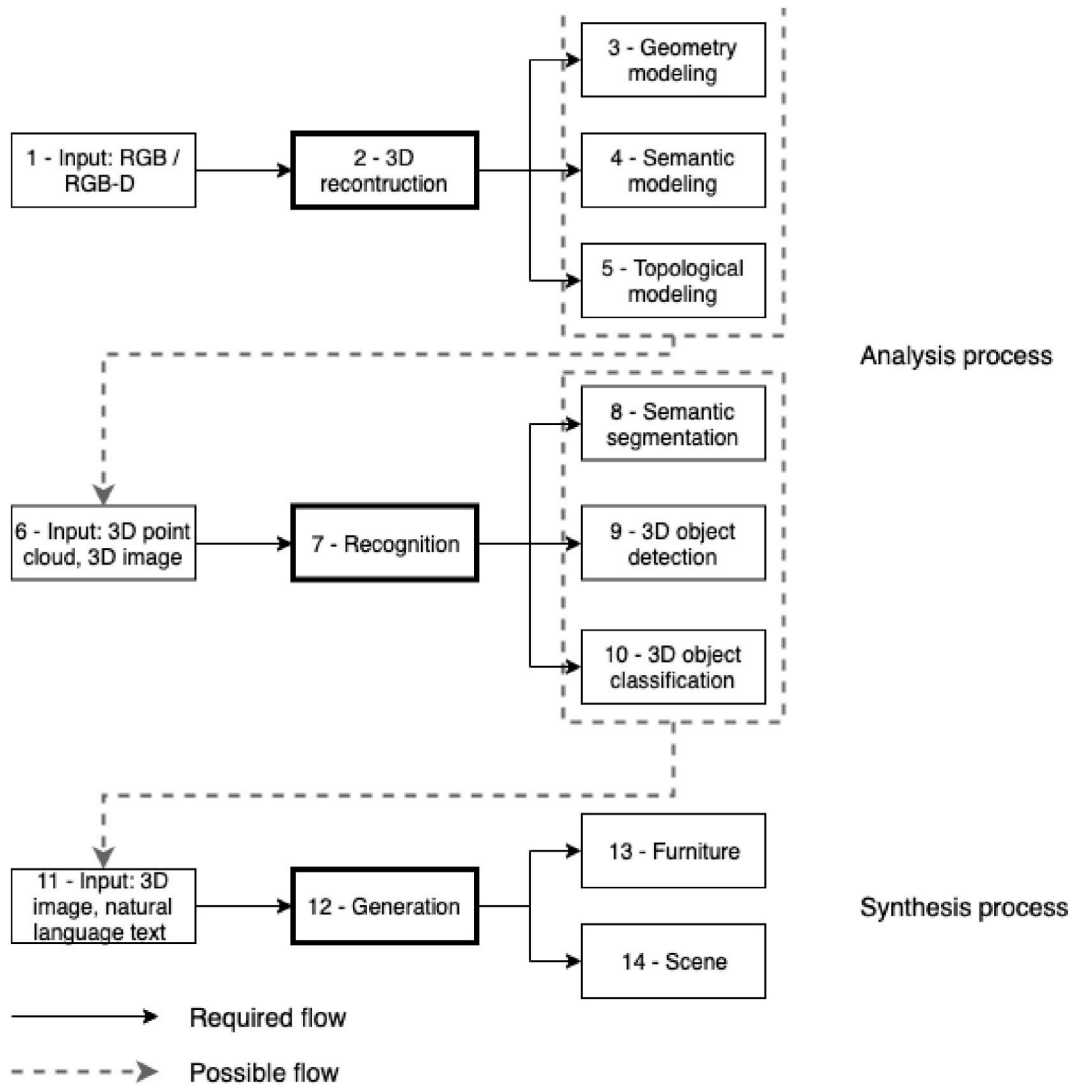


Fig. 8. 3D indoor scene generation from the relationship between deep learning and 3D indoor scene plan.

201,202].

Benchmarking compares different algorithms or methods from free and high-quality training data. Computer vision has a long tradition in using datasets with RGB, RGB-D, and HSV images, panoramic images, 3D point clouds, 3D mesh segmentation, and vector images in robotics and remote sensing. Modelers use these data in computer vision tasks such as image and point cloud segmentation, scene classification, object detection, 3D indoor or outdoor scene reconstruction, image retrieval, and object generation. For this purpose, they test diverse network architectures to find the best results, as each method's accuracy depends on the dataset's size and the data quality. In addition, they may also split the data to guarantee the diversity of objects in training.

Benchmarking can present the state-of-the-art based on quantitative or qualitative criteria. Among the analyzed works, the CNN method has the most metrics in Overall Accuracy (OA), Precision (P), Recall (R), F1-Score (F1), Intersection over Union (IoU), and others, which are defined as follows:

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

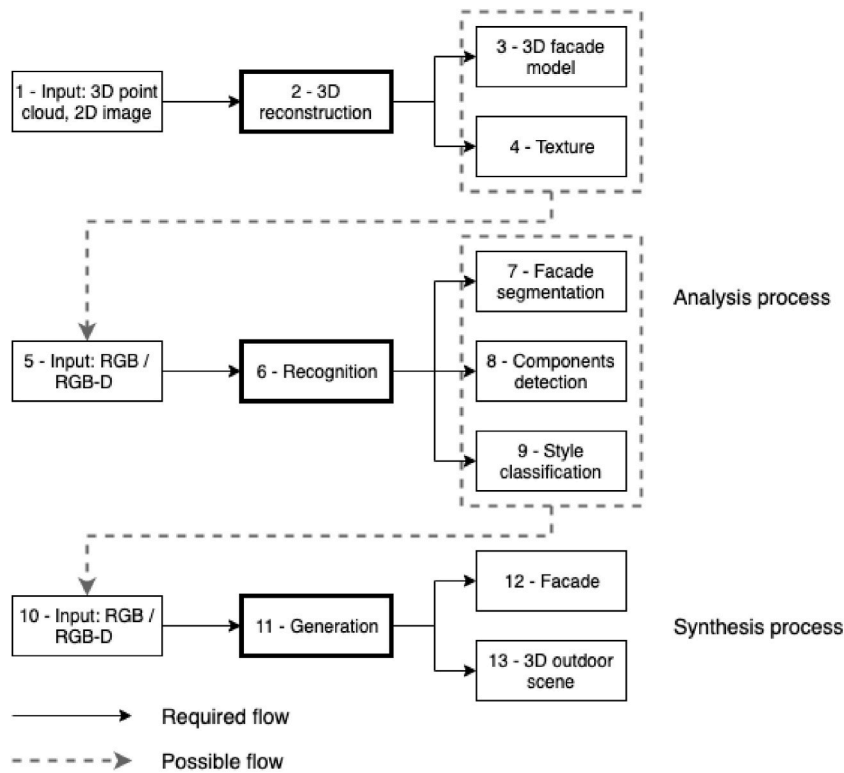


Fig. 9. Facade generation from the relationship between deep learning and 3D indoor scene plan.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{5}$$

FN, FP, TP, and TN are the number of false negatives, false positives, true positives, and true negatives, respectively. Other available error metrics are the mean squared error (MSE) and mean absolute error (MAE).

Depending on the purpose of the method (classification or regression), different evaluation metrics may be used. For example, when a user inputs a facade image, the result may be a value indicating the level or intensity of the attribute being predicted—e.g., indoor daylighting. Therefore, metrics are usually MSE and MAE. Conversely, in classification, the same facade image may be used to categorize its architectural style—e.g., modernism, classic architecture, postmodernism, and so on. In this case, the metrics are OA, P, R, and F1.

The main methods used in different building categories in recent years are compared in Table 1. The OA, P, R, F1, and IoU metrics are the most common in CNN, while FID is the most frequent among GAN applications. In some cases, more than one dataset is used to compare the same methods and verify how they behave in different data conditions. The state-of-the-art results follow the order of metrics in the table, whereas the bold font represents the better result compared to other benchmark methods.

In the building footprint category, the extraction task has high accuracy in small buildings and complex roof configurations. However, it has a low R and does not use land cover classification in its applications [203]. In the detection task, the shadow extraction methods may falsely detect the dark land covers as dense vegetation [205]. Also, it has no better P and R.

In the building floor plan category, the detection task has demonstrated the advantages of graph-based representations for analysis from quantitative results [207]. However, the point drop in performance for the baseline method in their applied dataset demonstrates the difficulty in constructing a repository that supports graph and raster-based approaches. Furthermore, recognition fails in villas when images have different floors in a single figure. Also, it does not perform well in semantic segmentation for basic elements such as walls, windows, rooms, and icons such as sink or fire [75]. Finally, in the retrieval task, the multi-oriented framework handles scan time rotations and is also invariant to scale [79]. However, performance is reduced when severe noise is present. In this case, new descriptors are needed with specific resources to annotate the global figure shape automatically.

In the building facade category, misclassifications may occur when there are similar or indistinguishable classes of houses [210], such as “villa” and “detached house,” suggesting that a better class definition is needed. Some approaches overcome misclassification by adding several pre-processing features to extract different channel information. However, the procedure uses more memory and a longer training time. Furthermore, in the generation task, there are limitations in the functionality of the assignment, such as location,

Table 1
Benchmark between main methods of different tasks and building categories.

Category	Task	Author	Method	Metric						Dataset	Accuracy
				OA	P	R	F1	IoU	Other		
Footprint (CNN)	Building extraction	[203]	SegNet, U-Net, DeepLabV3		x	x	x	x		Massachusetts Building	0.8857, 0.8238, 0.8536, 0.7446
					x	x	x	x			WHU Aerial Building
	Classification	[204]	2DCNN, ECNN, GCNN, 3DCNN, SSRN, MSDNSA	x				x	Kappa	Houston 2018	72.57, 66.80, 64.89, 58.02
Footprint (GAN)	Detection	[205]	FC-EF, FC-Siam-conc, FC-Siam-diff, DTCDSN, STANet		x	x	x			LEVIR-CD WHU-CD	0.916, 0.865, 0.89
					x	x	x				0.914, 0.869, 0.891
	Generation	[206]	Pix2Pix						mPA, mA, mIoU, mFWIU	Orthofoto Tirol	0.96, 0.89, 0.83, 0.90
Floor plan (CNN)	Detection	[207]	[6,66,73], FCNs	x	x	x		x		Lifull Home's	60.8, 75.2, 75.9 (IoU > 0.95); 94.5, 96.7, 97.6 (IoU > 0.5)
	Segmentation	[75]	PSPNet, DeepLabV3, Deepfloorplan	x					mIoU	R2V	0.92, 0.81
Floor plan (GAN)	Retrieval	[79]	CNN [208,209]	x					mIoU	R3D	0.94, 0.75
	Generation	[87]	GAN, CNN, GCN						mAP	ROBIN	0.7446
Facade (CNN)	Classification	[210]	[133], ResNet 18 and ResNet 34	x	x	x	x			Particular	13.6, 9.4, 14.4, 11.6, 20.1
	Detection	[211]	[212]	x						ECP eTrims Graz 50	96.4%, 96.9%, 96.1%, 95.9%
				x					95		
	Segmentation	[141]	[134,138,213], and others	x			x		Class Average, mIoU	ECP	84 90 88.40, 84.09, 83.86, 73.42
Facade (GAN)	Generation	[178]	WGAN, SAGAN						mIoU	RueMonge2014	53.66, 79.58, 46.92, 60.05
										Class Average, mIoU	ECP-Occcluded and RueMonge2014
	3D indoor scene (CNN)	3D reconstruction	[215]	[102,216], and others	x				PE, CE	Hedau LSUN	7.36 81.22, 6.94, 5.16
3D indoor scene (GAN)	Feature selection	[217]	[218,219], and others	x						SUN RGB-D NYUD v2	56.2%
				x							68.1%
	Fusion	[220]	DF ² Net, MAPNet, TRecNet						mA	SUN RGB-D	56.4, 44.1, 58.6
									mA	NYUD-Depth v2	67.8, 59.2, 75.1
Segmentation	[101]	[98,175,221], and others					x	CE, PE (SS), PE (CS)	SUN360	78.87, 0.75, 2.49, 3.03	
							x	Cerr, PE (SS), PE (CS)	Stanford 2D-3D	65.23, 1.44, 4.75, 6.05	
3D indoor scene (GAN)	Generation	[171]	OASIS, SPADE, CC-FPSE, SESAME, LGGAN						mIoU, FID, VC	SceneNet	0.622, 94.258, 62.694

neighborhood, culture, and climate, to customize the architectural design [178].

The generation task in the 3D indoor scene category has two noticeable limitations [171]. First, the network cannot model the view-dependent light effect because it assigns one color for each 3D coordinate. In addition, the model does not recognize transparency, thus considering all colors opaque. Second, large scenes take a long time to train, reaching up to 15 h for one scene of the SceneNet dataset.

5. Conclusion

This paper reviews the application of CNN and GAN algorithms in several tasks of automated building design, such as building footprint, floor plan, 3D indoor scene, and facade. The increasing number of publications reflects the growing interest in deep learning in building design, particularly computer vision and computer graphics techniques. In addition, the increasing availability of public datasets and source code will encourage more professionals to adopt these techniques, thus opening a path for radical and innovative

solutions in building design. Nonetheless, their use is still in its infancy and is uncommon among building engineering and architectural design professionals.

CNN and GAN-based approaches have great potential for integration into building design. They can be easily employed after training without experience and expertise, thus overcoming the downfalls of traditional artificial intelligence methods. Also, these approaches have a high degree of automation and produce good-quality results (still highly dependent on the type of task). Although these methods display more advantages in the analyzed studies than in other methods, the workflow between the different stages of the building design is neither smooth nor accessible to many designers. In conclusion, CNN-based networks provide the necessary understanding, manipulation, and results when adequately applied. Meanwhile, GANs were found to help complete images, generate building layouts, and create facades for different architectural styles.

Furthermore, CNN and GAN improve performance compared to traditional methods, reducing classification errors and improving results. In addition, the automation between CNN and GAN can overcome common barriers in image processing by facilitating the participation of non-specialists in specific steps such as remote sensing or floor plan vectorization to reconstruct these in 3D.

As observed in the previous section, the number of deep-learning approaches used in building design is growing, indicating an increase in the automation flow and integration between different building representations of the design process. Therefore, several advances are described in the literature, particularly in building extraction, integration between CNN and GAN in vectorization, reconstruction of 3D models from 2D imagery, and broader and more reliable retrieval of models from sketches. In addition, the quality and accuracy of the results incentivize non-experts and designers to use them.

However, these methods face several challenges and barriers, such as low quality and occlusions in datasets, and present obstacles for some users due to the lack of a user-friendly interface. In addition, several ethical issues are associated with the datasets used to create these methods that will eventually be regulated by societies and may pose a challenge to integrating these models in building design.

Future research paths may seek further integration between these networks and building design, going beyond computer vision to explore different types of data that incorporate different building performance criteria, such as thermal and visual comfort. For example, these studies may consider building energy simulation from input data such as weather conditions and building characteristics, daylight optimization from training GAN on building orientation, window size, and shading design.

Another research path may be the customization of architectural design according to the location, culture, and neighborhood context. In other words, the model may be able to produce design solutions that consider the surroundings by integrating the new building stylistically. Other aspects may be included as well. For example, it may generate building facades that consider alignments and floor heights of the adjacent buildings. These features will be particularly helpful in sensitive situations involving heritage sites and historical architecture. In floor plan design, the generation process could consider multiple floors, structural alignments, and the relationship between vertical circulation and the functional program of the building.

For this, the CNN-GAN tools must assume multi-task functionalities, and multiscale datasets must be available for different categories of building design representations. Furthermore, another path may integrate both CNN and GAN in different categories since analysis, evaluation, and synthesis are requirements in the decision-making process. Lastly, as the number of experts and non-experts continues to grow in this field, it is essential to understand how these methods may be smoothly integrated into building design methodology without disrupting the creative process.

Funding

This work was supported by the Foundation for Science and Technology (FCT) [grant number PTDC/EME-REN/3460/2021]. This work was also financially supported by: Base Funding – UIDB/04708/2020 and Programmatic Funding – UIDP/04708/2020 of the CONSTRUCT – Instituto de I&D em Estruturas e Construções and the Base Funding – UIDB/00145/2020 of the CEAU – Center for Studies in Architecture and Urbanism, both funded by national funds through the FCT/MCTES (PIDDAC). This work is also co-funded by the European Social Fund (ESF) through the Northern Regional Operational Programme (Norte 2020) [Funding Reference: NORTE-06-3559-FSE-000176].

Eugénio Rodrigues is funded by the Portuguese Foundation for Science and Technology (FCT) through a researcher contract [2021.00230.CEECIND].

The authors are grateful to Anabela Reis for proofreading the document.

CRediT authorship contribution statement

Jean Parente: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Eugénio Rodrigues:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Bárbara Rangel:** Writing – review & editing, Supervision. **João Poças Martins:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The presented work is framed under the *Energy for Sustainability Initiative* of the University of Coimbra (UC).

Appendix A

Table A 1
Datasets.

Design building category	Dataset	Applications	Link to dataset
Building Footprint	AI Crowd Mapping Challenge	Segmentation	https://www.crowdai.org/challenges/mapping-challenge
	AIRS	Segmentation	https://www.airs-dataset.com/
	Bing Map	Classification	https://www.bing.com/maps
	CDNet	Detection	http://changedetection.net/
	Deepglobe workshop (CVPR)	Building extraction	http://deepglobe.org/
	GeoEye	Building extraction	https://apollomapping.com/geoeye-1-satellite-imagery
	Google Earth	Building extraction, Classification, Detection	https://earth.google.com/web/
	Google Street View	Classification, Detection	https://www.google.com/maps/
	HyRANK	Classification	https://www2.isprs.org/commissions/comm3/wg4/hyrank/
	IEEE GRSS DFC	Building extraction	http://dase.grss-ieee.org/
	INRIA	Building extraction, Classification	https://project.inria.fr/aerialimagelabeling/
	Kaggle Open Source	Generation	https://www.kaggle.com/c/draper-satellite-image-chronology/data
	LINZ data service	Segmentation	https://data.linz.govt.nz/
	Massachusetts Building	Building extraction, Classification, Detection	https://www.kaggle.com/balraj98/massachusetts-buildings-dataset
	NWPU VHR	Detection	https://github.com/chaozhong2010/VHR-10_dataset_coco
	Orthofoto Tirol	Segmentation	https://www.arcgis.com/apps/webappviewer/
	PlanetScope Satellite Imagery	Generation, Segmentation	https://apollomapping.com/planetoscope-satellite-imagery
	RIT - 18	Segmentation	https://github.com/rmkemker/RIT-18
	SpaceNet	Building extraction, Detection, Segmentation	https://spacenet.ai/datasets/
	Building Floor plan	Urban3D	Building extraction, Segmentation
USGS		Building extraction	https://earthexplorer.usgs.gov/
WHU Aerial Building		Building extraction	http://gpcv.whu.edu.cn/data/building_dataset.html
Worldview-2		Detection	https://apollomapping.com/worldview-2-satellite-imagery
Arcbazar		Generation	https://www.arcbazar.com
COCO		Detection, Reconstruction	https://cocodataset.org/#home
CODE SUBMIT/dataset2		Generation	https://github.com/CODE-SUBMIT/dataset2
Cubicasa5k		Segmentation	https://github.com/CubiCasa/CubiCasa5k
CVC-FP		Segmentation	http://dag.cvc.uab.es/resources/floorplans/
EAIS		Segmentation	https://open.eais.go.kr/main/main.do
Google's image		Classification	https://www.google.com/imghp
Lifull Home's		Classification, Generation, Segmentation	https://www.nii.ac.jp/dsc/idr/en/lifull/
LSUN		Reconstruction	https://www.yf.io/p/lsun
R2V and R3D		Segmentation	https://github.com/zleng/DeepFloorplan
ROBIN		Retrieval	https://github.com/gesstalt/ROBIN
RPLAN		Generation, Retrieval	http://staff.ustc.edu.cn/~fuxm/projects/DeepLayout/index.html
ScanNet		Reconstruction, Segmentation	http://www.scan-net.org/
SESYD		Detection, Generation	http://mathieu.delalandre.free.fr/projects/sesyd/
Building Facade	CMP	Detection, Generation, Segmentation	https://cmp.felk.cvut.cz/~tylecr1/facade/
	e-Trims	Detection, Segmentation, Reconstruction	http://www.ipb.uni-bonn.de/projects/etrims_db/
	ECP-Monge	Detection, Segmentation, Generation	http://vision.mas.ecp.fr/Personnel/teboul/data.php
	Google Street View	Segmentation	https://www.google.com/maps/
	Graz	Segmentation	https://people.ee.ethz.ch/~rhayko/paper/cvpr2012_riemenschneider_lattice/
	ImageNet	Classification, Detection	https://image-net.org/
Make3D	Segmentation	http://make3d.cs.cornell.edu/data.html	

(continued on next page)

Table A 1 (continued)

Design building category	Dataset	Applications	Link to dataset
3D indoor scene	Mapillary	Segmentation	https://www.mapillary.com/
	Oxford Buildings	Classification	https://www.robots.ox.ac.uk/~vgg/data/oxbuildings/
	RueMonge2014	Segmentation	https://people.ee.ethz.ch/~rhayko/paper/eccv2014_riemenschneider_multiviewsemseg/
	SUN	Generation	https://vision.cs.princeton.edu/projects/2010/SUN/
	TSG-20	Detection	http://dib.joanneum.at/cape/TSG-20/
	TSG-60	Detection	http://dib.joanneum.at/cape/TSG-60/
	ZuBuD	Detection	https://icu.ee.ethz.ch/research/datasets.html
	ADE20k	Segmentation	https://groups.csail.mit.edu/vision/datasets/ADE20K/
	COCO	3D reconstruction	https://cocodataset.org
	Hedau	3D reconstruction	http://vhedau-uiuc.cs.uiuc.edu/%7Evarsha/Research/data/groundtruth.zip
	ImageNet	Detection	https://www.image-net.org/
	LSUN	3D reconstruction	https://www.yf.io/p/lsun
	Matterport3D	3D reconstruction	https://niessner.github.io/Matterport/
	MIT67	Classification, Detection	http://web.mit.edu/torralba/www/indoor.html
	NYU Depth v1	3D reconstruction, Segmentation	https://cs.nyu.edu/~silberman/datasets/nyu_depth_v1.html
	NYU Depth v2	Segmentation, Classification, fusion	https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
	PanoContext	Segmentation	https://panocontext.cs.princeton.edu/
	Places 365	Classification	http://places2.csail.mit.edu/download.html
	Realtor 360	3D reconstruction	N/A
	S3DIS	Segmentation	http://buildingparser.stanford.edu/dataset.html#Download
	ScanNet	3D reconstruction, Segmentation	http://www.scan-net.org/
	Scene 15	Classification	https://figshare.com/articles/dataset/15-Scene_Image_Dataset/7007177
	SceneNet	3D reconstruction	http://103.24.77.34/scenenn/home/
	ShapeNet	Segmentation	https://shapenet.org/
	Stanford 2D3D	3D reconstruction, Segmentation	http://buildingparser.stanford.edu/dataset.html#Download
	SUN 360	3D reconstruction, Detection, Segmentation	https://vision.cs.princeton.edu/projects/2012/SUN360/data/
	SUN RGB-D	Segmentation, Fusion, Classification, Detection	https://rgbd.cs.princeton.edu/
	XAUT	Segmentation	N/A

Table A 2

Code sources.

Category	Network	Applications	Link to codebase
Building Footprint	ASLNet	Analysis for building extraction	https://github.com/ggsDing/ASLNet
	DenseNet (based)	Analysis for image segmentation	https://github.com/liangjian2018/MFRN
	Faster R-CNN	Analysis for image segmentation	https://github.com/liangjian2018/MFRN
	FCN	Analysis for image segmentation, object detection, and object classification	https://github.com/lixiang-ucas/Building-A-Nets
	GAN	Analysis for image segmentation and synthesis for building footprint generation	https://github.com/SimJeg/FC-DenseNet
	GCN	Analysis for building extraction and image segmentation	https://github.com/rbgirshick/py-faster-rcnn
	Mask R-CNN	Analysis for building extraction and image segmentation	https://github.com/divamgupta/image-segmentation-keras
	RefineNet	Analysis for building extraction and image segmentation	https://github.com/shelhamer/fcn.berkeleyvision.org
	ResNet	Analysis for Building extraction, object detection, and image segmentation	https://github.com/martinarjovsky/WassersteinGAN
	ResNext	Analysis for building extraction	https://phillipi.github.io/pix2pix/
	RetinaNet	Segmentation	https://github.com/tkipf/gcn
	SegNet	Analysis for building extraction, detection, and image segmentation	https://github.com/matterport/Mask_RCNN
	U-Net	Analysis for building extraction, classification, detection, and image segmentation	https://github.com/guosheng/refinenet
Building floor plan	AlexNet (modified)	Analysis for image retrieval	https://github.com/KaimingHe/deep-residual-networks
	Faster R-CNN	Analysis for object detection	https://github.com/facebookresearch/ResNeXt
	FloorNet	Analysis for image segmentation	https://github.com/facebookresearch/Detectron
	GAN	Analysis for image segmentation and synthesis for building footprint generation	https://github.com/nshaud/DeepNetsForEO

(continued on next page)

Table A 2 (continued)

Category	Network	Applications	Link to codebase
Building facade	LeNet (Modified)	Analysis for room classification	https://github.com/JiananLi2016/LayoutGAN-Tensorflow https://colab.research.google.com/github/d2l-ai/d2l-en-colab/blob/master/chapter_convolutional-neural-networks/lenet.ipynb
	Mask R-CNN	Analysis for object detection	https://github.com/matterport/Mask_RCNN
	ResNet	Analysis for image segmentation	https://github.com/KaimingHe/deep-residual-networks
	PixelDCL/ PixelTCN	Analysis for image segmentation	https://github.com/divelab/PixelTCN
	YOLO/YOLO v2	Analysis for object detection	https://pjreddie.com/darknet/yolo/
	AlexNet	Analysis for classification and segmentation, and synthesis for building facade generation	https://github.com/deep-diver/AlexNet
	ALKNet	Analysis for segmentation	https://github.com/qsworkers/wohaiyo/Pyramid_ALKNet
	DenseNet	Analysis for building facade classification	https://github.com/liuzhuang13/DenseNet
	Faster R-CNN	Analysis for object detection	https://github.com/rbgirshick/py-faster-rcnn
	FCN	Analysis for image segmentation	https://github.com/shelhamer/fcn.berkeleyvision.org
	FCRN	Analysis for building facade reconstruction	https://github.com/iro-cp/FCRN-DepthPrediction
	GAN	Synthesis for building facade generation	https://phillipi.github.io/pix2pix/ https://github.com/Newmu/dcgan_code
	GoogLeNet	Analysis for building facade classification	https://paperswithcode.com/model/googlenet
	Mask R-CNN	Analysis for object detection	https://github.com/matterport/Mask_RCNN
	NASNet	Analysis for building facade classification	https://github.com/johannesu/NASNet-keras
	ResNet	Analysis for object detection, image segmentation, and building facade classification	https://github.com/KaimingHe/deep-residual-networks
	YOLO	Analysis for object detection	https://pjreddie.com/darknet/yolo/
	EfficientNet	Analysis for 3D indoor scene classification	https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet
	Faster R-CNN	Analysis for 3D reconstruction and object detection	https://github.com/ShaoqingRen/faster_rcnn
	FCN	Analysis for 3D reconstruction and image/point cloud segmentation	https://github.com/iro-cp/FCRN-DepthPrediction
GAN	Synthesis for 3D indoor scene generation	https://github.com/chenhsuanlin/spatial-transformer-GAN	
GCN	Analysis for 3D indoor scene classification and synthesis for 3D indoor scene generation	https://github.com/tkipf/gcn	
MobileNet	Analysis for image segmentation	https://github.com/tensorflow/models/blob/master/research/slim/nets/mobilenet_v1.md	
ResNet	Analysis for 3D reconstruction, 3D indoor scene classification, object detection, image retrieval, image/point cloud segmentation, and synthesis for 3D indoor scene generation	https://github.com/TUI-NICR/ESANet https://github.com/GitBoSun/roomnet https://github.com/ownstyledu/Translate-to-Recognize-Networks https://github.com/sunset1995/HorizonNet https://github.com/brownvc/planit https://github.com/JindongJiang/RedNet	
SegNet	Analysis for 3D reconstruction and image segmentation	https://github.com/zouchuhang/LayoutNet	
U-Net	Analysis for image segmentation	https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/	
YOLO	Analysis for 3D indoor scene classification	https://pjreddie.com/darknet/yolo/	

References

- [1] M. Zawadzki, K. Tateyama, I. Nishikawa, The constraints satisfaction problem approach in the design of an architectural functional layout, *Eng. Optim.* 43 (2011) 943–966, <https://doi.org/10.1080/0305215X.2010.527005>.
- [2] T. Chugh, K. Sindhya, J. Hakanen, K. Miettinen, A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms, *Soft Comput.* 23 (2019) 3137–3166, <https://doi.org/10.1007/s00500-017-2965-0>.
- [3] R. Hu, Z. Huang, Y. Tang, O. Van Kaick, H. Zhang, H. Huang, Graph2Plan: learning floorplan generation from layout graphs, *ACM Trans. Graph.* 39 (2020) 1–14, <https://doi.org/10.1145/3386569.3392391>.
- [4] Z. Wang, Q. She, T.E. Ward, Generative adversarial networks: a survey and taxonomy, *ArXiv* (2019) 1–16.
- [5] S. Dargan, M. Kumar, M.R. Ayyagari, G. Kumar, A survey of deep learning and its applications: a new paradigm to machine learning, *Arch. Comput. Methods Eng.* 27 (2020) 1071–1092, <https://doi.org/10.1007/s11831-019-09344-w>.
- [6] S. Dodge, J. Xu, B. Stenger, Parsing floor plan images, in: *Proc. 15th IAPR Int. Conf. Mach. Vis. Appl. MVA 2017*, 2017, pp. 358–361, <https://doi.org/10.23919/MVA.2017.7986875>.
- [7] E. Rodrigues, M.S. Fernandes, Á. Gomes, A.R. Gaspar, J.J. Costa, Performance-based design of multi-story buildings for a sustainable urban environment: a case study, *Renew. Sustain. Energy Rev.* 113 (2019), 109243, <https://doi.org/10.1016/j.rser.2019.109243>.
- [8] E. Rodrigues, A.R. Gaspar, Á. Gomes, An evolutionary strategy enhanced with a local search technique for the space allocation problem in architecture, Part 1: methodology, *Comput. Des.* 45 (2013) 887–897, <https://doi.org/10.1016/j.cad.2013.01.001>.

- [9] J.J. Michalek, P.Y. Papalambros, Interactive design optimization of architectural layouts, *Eng. Optim.* 34 (2002) 485–501, <https://doi.org/10.1080/03052150214021>.
- [10] E. Rodrigues, A.R. Gaspar, Á. Gomes, An approach to the multi-level space allocation problem in architecture using a hybrid evolutionary technique, *Autom. Construct.* 35 (2013) 482–498, <https://doi.org/10.1016/j.autcon.2013.06.005>.
- [11] W. Li, M.Y. Yang, Efficient semantic segmentation of man-made scenes using fully-connected conditional random field, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 41 (2016) 633–640, <https://doi.org/10.5194/isprsarchives-XLI-B3-633-2016>.
- [12] D. Dai, H. Riemenschneider, G. Schmitt, L. Van, Example-based facade texture synthesis, *Proc. IEEE Int. Conf. Comput. Vis.* (2013) 1065–1072, <https://doi.org/10.1109/ICCV.2013.136>.
- [13] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2004) 91–110, <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [14] S. Yigit, B. Ozorhon, A simulation-based optimization method for designing energy efficient buildings, *Energy Build.* 178 (2018) 216–227, <https://doi.org/10.1016/j.enbuild.2018.08.045>.
- [15] E. Rodrigues, A.R. Gaspar, Á. Gomes, Improving thermal performance of automatically generated floor plans using a geometric variable sequential optimization procedure, *Appl. Energy* 132 (2014) 200–215, <https://doi.org/10.1016/j.apenergy.2014.06.068>.
- [16] R. Szeliski, *Computer Vision: Algorithms and Applications*, second ed., Springer, 2020. <https://szeliski.org/Book/>.
- [17] L. McMillan, S. Gortler, Image-based rendering, *ACM SIGGRAPH Comput. Graph.* 33 (1999) 61–64, <https://doi.org/10.1145/345370.345415>.
- [18] J. Gomes, L. Velho, *Image Processing for Computer Graphics*, Springer New York, New York, NY, 1997, <https://doi.org/10.1007/978-1-4757-2745-6>.
- [19] S. Khan, H. Rahmani, S.A.A. Shah, M. Bennamoun, A guide to convolutional neural networks for computer vision, *Synth. Lect. Comput. Vis.* 8 (2018) 1–207, <https://doi.org/10.2200/s00822ed1v01y201712cov015>.
- [20] R. Szeliski, *Computer Vision : Algorithms and Applications*, second ed., Springer, 2021, p. 1196.
- [21] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, M. Atri, Indoor Image Recognition and Classification via Deep Convolutional Neural Network, 2020, pp. 364–371, https://doi.org/10.1007/978-3-030-21005-2_35.
- [22] Z. Zhu, D. Li, Y. Hu, J. Li, D. Liu, J. Li, Indoor scene segmentation algorithm based on full convolutional neural network, *Neural Comput. Appl.* 0 (2020), <https://doi.org/10.1007/s00521-020-04961-0>.
- [23] Z. Ziran, S. Marinali, Object Detection in Floor Plan Images 11081 (2018) 383–394, https://doi.org/10.1007/978-3-319-99978-4_30.
- [24] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, H. Ghayvat, Cnn variants for computer vision: history, architecture, application, challenges and future scope, *Electron J* (2021) 1–28, <https://doi.org/10.3390/electronics10202470>.
- [25] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016. <http://www.deeplearningbook.org/>.
- [26] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444, <https://doi.org/10.1038/nature14539>.
- [27] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, *Neurocomputing* 187 (2016) 27–48, <https://doi.org/10.1016/j.neucom.2015.09.116>.
- [28] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM* 60 (2017) 84–90, <https://doi.org/10.1145/3065386>.
- [29] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. (2015) 1–14.
- [30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015.
- [31] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.* (2014) 580–587, <https://doi.org/10.1109/CVPR.2014.81>.
- [32] S. Ren, K. He, R. Girshick, J. Sun, R.-C.N.N. Faster, Towards real-time object detection with region proposal networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), *Adv. Neural Inf. Process. Syst.* 28, Curran Associates, Inc., 2015, pp. 91–99. <http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>.
- [33] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-Decem (2016) 779–788, <https://doi.org/10.1109/CVPR.2016.91>.
- [34] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, *Lect. Notes Comput. Sci.* 9351 (2015) 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
- [35] G. Lin, A. Milan, C. Shen, I. Reid, RefineNet, Multi-path refinement networks for high-resolution semantic segmentation, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 2017, pp. 5168–5177, <https://doi.org/10.1109/CVPR.2017.549>, 2017-Janua.
- [36] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, *Adv. Neural Inf. Process. Syst.* 3 (2014) 2672–2680.
- [37] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, 2016, pp. 1–16.
- [38] M. Schmitz, H. Mayer, A convolutional network for semantic facade segmentation and interpretation, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* (2016) 709–715, <https://doi.org/10.5194/isprsarchives-XLI-B3-709-2016>.
- [39] M. Rahbar, M. Mahdavejnad, M. Bemanian, A.H. Davaei Markazi, L. Hovestadt, Generating synthetic space allocation probability layouts based on trained conditional-GANs, *Appl. Artif. Intell.* 33 (2019) 689–705, <https://doi.org/10.1080/08839514.2019.1592919>.
- [40] M.L. Castro Pena, A. Carballal, N. Rodríguez-Fernández, I. Santos, J. Romero, Artificial intelligence applied to conceptual design. A review of its use in architecture, *Autom. Construct.* 124 (2021), <https://doi.org/10.1016/j.autcon.2021.103550>.
- [41] D. Koch, M. Despotovic, S. Leiber, M. Sakeena, M. Döller, M. Zeppezauer, Real estate image analysis: a literature review, *J. R. Estate Lit.* 27 (2019) 271–300, <https://doi.org/10.22300/0927-7544.27.2.269>.
- [42] S. Xu, J. Wang, W. Shou, T. Ngo, A.M. Sadick, X. Wang, Computer vision techniques in construction: a critical review, *Arch. Comput. Methods Eng.* 28 (2021) 3383–3397, <https://doi.org/10.1007/s11831-020-09504-3>.
- [43] C. Wohlin, Guidelines for snowballing in systematic literature studies and a replication in software engineering, *ACM Int. Conf. Proceeding Ser.* (2014), <https://doi.org/10.1145/2601248.2601268>.
- [44] J.A. LaGro, *Site Analysis : a Contextual Approach to Sustainable Land Planning and Site Design*, second ed., 2007. New Jersey.
- [45] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B.A. Johnson, Deep learning in remote sensing applications: a meta-analysis and review, *ISPRS J. Photogrammetry Remote Sens.* 152 (2019) 166–177, <https://doi.org/10.1016/j.isprsjrs.2019.04.015>.
- [46] D. Griffiths, J. Boehm, Improving public data for building segmentation from Convolutional Neural Networks (CNNs) for fused airborne lidar and image data using active contours, *ISPRS J. Photogrammetry Remote Sens.* 154 (2019) 70–83, <https://doi.org/10.1016/j.isprsjrs.2019.05.013>.
- [47] D. Marmanis, K. Schindler, J.D. Wegner, S. Galliani, M. Dacu, U. Stilla, Classification with an edge: improving semantic image segmentation with boundary detection, *ISPRS J. Photogrammetry Remote Sens.* 135 (2018) 158–172, <https://doi.org/10.1016/j.isprsjrs.2017.11.009>.
- [48] M. Belgiu, L. Drăguț, Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery, *ISPRS J. Photogrammetry Remote Sens.* 96 (2014) 67–75, <https://doi.org/10.1016/j.isprsjrs.2014.07.002>.
- [49] R. Chen, X. Li, J. Li, Object-based features for house detection from RGB high-resolution images, *Rem. Sens.* 10 (2018) 451, <https://doi.org/10.3390/rs10030451>.
- [50] W. Li, C. He, J. Fang, J. Zheng, H. Fu, L. Yu, Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data, *Rem. Sens.* 11 (2019), <https://doi.org/10.3390/rs11040403>.
- [51] P. Liu, X. Liu, M. Liu, Q. Shi, J. Yang, X. Xu, Y. Zhang, Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network, *Rem. Sens.* 11 (2019), <https://doi.org/10.3390/rs11070830>.

- [52] G. Wu, Z. Guo, X. Shi, Q. Chen, Y. Xu, R. Shibasaki, X. Shao, A boundary regulated network for accurate roof segmentation and outline extraction, *Rem. Sens.* 10 (2018) 1–19, <https://doi.org/10.3390/rs10081195>.
- [53] P. Schuegraf, K. Bittner, Automatic building footprint extraction from multi-resolution remote sensing images using a hybrid FCN, *ISPRS Int. J. Geo-Inf.* 8 (2019) 1–16, <https://doi.org/10.3390/ijgi8040191>.
- [54] Z. Ye, Y. Fu, M. Gan, J. Deng, A. Comber, K. Wang, Building extraction from very high resolution aerial imagery using joint attention deep neural network, *Rem. Sens.* 11 (2019) 1–21, <https://doi.org/10.3390/rs11242970>.
- [55] Y. Liu, J. Zhou, W. Qi, X. Li, L. Gross, Q. Shao, Z. Zhao, L. Ni, X. Fan, Z. Li, ARC-net: an efficient network for building extraction from high-resolution aerial images, *IEEE Access* 8 (2020) 154997–155010, <https://doi.org/10.1109/ACCESS.2020.3015701>.
- [56] W. Deng, Q. Shi, J. Li, Attention-gate-based encoder-decoder network for automatic building extraction, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 14 (2021) 2611–2620, <https://doi.org/10.1109/JSTARS.2021.3058097>.
- [57] D. Zhou, G. Wang, G. He, T. Long, R. Yin, Z. Zhang, S. Chen, B. Luo, Robust building extraction for high spatial resolution remote sensing images with self-attention network, *Sensors* 20 (2020) 1–19, <https://doi.org/10.3390/s20247241>.
- [58] A. Milosavljević, Automated processing of remote sensing imagery using deep semantic segmentation: a building footprint extraction case, *ISPRS Int. J. Geo-Inf.* 9 (2020), <https://doi.org/10.3390/ijgi9080486>.
- [59] G. Pasquali, G.C. Iannelli, F. Dell'Acqua, Building footprint extraction from multispectral, spaceborne earth observation datasets using a structurally optimized U-Net convolutional neural network, *Rem. Sens.* 11 (2019) 1–20, <https://doi.org/10.3390/rs11232803>.
- [60] X. Pan, F. Yang, L. Gao, Z. Chen, B. Zhang, H. Fan, J. Ren, Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms, *Rem. Sens.* 11 (2019) 917, <https://doi.org/10.3390/rs11080917>.
- [61] Y. Shi, Q. Li, X.X. Zhu, BFGAN-building footprint extraction from satellite images, in: 2019 Jt. Urban Remote Sens. Event, JURSE 2019, IEEE, 2019, pp. 1–4, <https://doi.org/10.1109/JURSE.2019.8809048>.
- [62] A. Abdollahi, B. Pradhan, S. Gite, A. Alamri, Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture, *IEEE Access* (2020), <https://doi.org/10.1109/ACCESS.2020.3038225>, 1–1.
- [63] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein Gan. <http://arxiv.org/abs/1701.07875>, 2017.
- [64] F. Chen, N. Wang, B. Yu, L. Wang, Res2-Unet, a new deep architecture for building detection from high spatial resolution images, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 15 (2022) 1494–1501, <https://doi.org/10.1109/JSTARS.2022.3146430>.
- [65] D. Sharma, N. Gupta, C. Chattopadhyay, S. Mehta, DANIEL: a deep architecture for automatic analysis and retrieval of building floor plans, in: Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, 2017, pp. 420–425, <https://doi.org/10.1109/ICDAR.2017.76>.
- [66] C. Liu, J. Wu, P. Kohli, Y. Furukawa, Raster-to-Vector: revisiting floorplan transformation, in: Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2214–2222, <https://doi.org/10.1109/ICCV.2017.241>.
- [67] L.-P. de las Heras, D. Fernandez, E. Valveny, J. Lladós, G. Sanchez, Unsupervised wall detector in architectural floor plans, in: 2013 12th Int. Conf. Doc. Anal. Recognit., IEEE, 2013, pp. 1245–1249, <https://doi.org/10.1109/ICDAR.2013.252>.
- [68] P. Dosch, K. Tombe, C. Ah-Soon, G. Masini, A complete system for the analysis of architectural drawings, *Int. J. Doc. Anal. Recogn.* 3 (2000) 102–116, <https://doi.org/10.1007/PL00010901>.
- [69] S. Macé, H. Locteau, E. Valveny, S. Tabbone, A system to detect rooms in architectural floor plan images, in: Proc. 8th IAPR Int. Work. Doc. Anal. Syst. - DAS'10, ACM Press, New York, New York, USA, 2010, pp. 167–174, <https://doi.org/10.1145/1815330.1815352>.
- [70] S. Ahmed, M. Liwicki, M. Weber, A. Dengel, Improved automatic analysis of architectural floor plans, *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR.* (2011) 864–869, <https://doi.org/10.1109/ICDAR.2011.177>.
- [71] S. Ahmed, M. Liwicki, M. Weber, A. Dengel, Automatic room detection and room labeling from architectural floor plans, in: 2012 10th IAPR Int. Work. Doc. Anal. Syst., IEEE, 2012, pp. 339–343, <https://doi.org/10.1109/DAS.2012.22>.
- [72] L.-P. de las Heras, S. Ahmed, M. Liwicki, E. Valveny, G. Sánchez, Statistical segmentation and structural recognition for floor plan interpretation, *Int. J. Doc. Anal. Recogn.* 17 (2014) 221–237, <https://doi.org/10.1007/s10032-013-0215-2>.
- [73] Z. Zeng, X. Li, Y.K. Yu, C.W. Fu, Deep floor plan recognition using a multi-task network with room-boundary-guided attention, in: Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 9095–9103, <https://doi.org/10.1109/ICCV.2019.00919>.
- [74] H. Jang, J.H. Yang, Y. Kiyun, Automatic wall detection and building topology and property of 2D floor plan, in: Leibniz Int. Proc. Informatics, LIPIcs, 2018, pp. 2–6, <https://doi.org/10.4230/LIPIcs.GIScience.2018.33>.
- [75] Y. Zhang, Y. He, S. Zhu, X. Di, The Direction-Aware, Learnable, Additive Kernels and the Adversarial Network for Deep Floor Plan Recognition, 2020. <http://arxiv.org/abs/2001.11194>.
- [76] N. Navef, T.M. Breuel, On the use of geometric matching for both: isolated symbol recognition and symbol spotting, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2013, pp. 36–48, https://doi.org/10.1007/978-3-642-36824-0_4.
- [77] K.C. Santosh, B. Lamiro, L. Wendling, Integrating vocabulary clustering with spatial relations for symbol recognition, *Int. J. Doc. Anal. Recogn.* 17 (2014) 61–78, <https://doi.org/10.1007/s10032-013-0205-4>.
- [78] S. Ahmed, M. Weber, M. Liwicki, C. Langenhan, A. Dengel, F. Petzold, Automatic analysis and sketch-based retrieval of architectural floor plans, *Pattern Recogn. Lett.* 35 (2014) 91–100, <https://doi.org/10.1016/j.patrec.2013.04.005>.
- [79] R. Khade, K. Jariwala, C. Chattopadhyay, U. Pal, A rotation and scale invariant approach for multi-oriented floor plan image retrieval, *Pattern Recogn. Lett.* 145 (2021) 1–7, <https://doi.org/10.1016/j.patrec.2021.01.020>.
- [80] D. Sharma, N. Gupta, C. Chattopadhyay, S. Mehta, REXplore: a sketch based interactive explorer for real estates using building floor plan images, in: Proc. - 2018 IEEE Int. Symp. Multimedia, ISM 2018, IEEE, 2019, pp. 61–64, <https://doi.org/10.1109/ISM.2018.00018>.
- [81] D. Sharma, N. Gupta, C. Chattopadhyay, S. Mehta, A novel feature transform framework using deep neural network for multimodal floor plan retrieval, *Int. J. Doc. Anal. Recogn.* 22 (2019) 417–429, <https://doi.org/10.1007/s10032-019-00340-1>.
- [82] A. Kalervo, J. Ylioinas, Markus Häikiö, A. Karhu, J. Kannala, 1, CubiCasa5K: A Dataset and an Improved Multi-Task Model for Floorplan Image Analysis, 2019, pp. 427–438, <https://doi.org/10.1007/978-3-030-20205-7>.
- [83] C. Zhao, J. Yang, J.-T. Li, Generation of hospital emergency department layouts based on generative adversarial networks, *J. Build. Eng.* (2021), 102539, <https://doi.org/10.1016/j.jobe.2021.102539>.
- [84] C. Zhao, J. Yang, W. Xiong, J. Li, Two generative design methods of hospital operating department layouts based on healthcare systematic layout planning and generative adversarial network, *J. Shanghai Jiao Tong Univ. (Sci.)* 26 (2021) 103–115, <https://doi.org/10.1007/s12204-021-2265-9>.
- [85] X. Chen, Y. Duan, R. Houthoof, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: interpretable representation learning, in: 30th Conf. Neural Inf. Process. Syst. (NIPS 2016), 2016, pp. 2172–2180, <https://doi.org/10.1007/978-3-319-16817-3>.
- [86] I. As, S. Pal, P. Basu, Artificial intelligence in architecture: generating conceptual design via deep learning, *Int. J. Architect. Comput.* 16 (2018) 306–327, <https://doi.org/10.1177/1478077118800982>.
- [87] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, Y. Furukawa, House-Gan, Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation, 2020, pp. 162–177, https://doi.org/10.1007/978-3-030-58452-8_10.
- [88] M. Rahbar, M. Mahdavejad, A.H.D. Markazi, M. Bemanian, Architectural layout design through deep learning and agent-based modeling: a hybrid approach, *J. Build. Eng.* 47 (2022), 103822, <https://doi.org/10.1016/j.jobe.2021.103822>.
- [89] A.N. Wu, R. Stouffs, F. Biljecki, Generative Adversarial Networks in the built environment: a comprehensive review of the application of GANs across data types and scales, *Build. Environ.* 223 (2022), 109477, <https://doi.org/10.1016/j.buildenv.2022.109477>.
- [90] M. Naseer, S. Khan, F. Porikli, Indoor scene understanding in 2.5/3D for autonomous agents: a survey, *IEEE Access* 7 (2019) 1859–1887, <https://doi.org/10.1109/ACCESS.2018.2886133>.
- [91] D. Rose, P. Bex, The linguistic analysis of scene semantics: lass, *Behav. Res. Methods* 2020 (2020), <https://doi.org/10.3758/s13428-020-01390-8>.

- [92] S. Chang, H. Jun, Hybrid deep-learning model to recognise emotional responses of users towards architectural design alternatives, *J. Asian Architect. Build Eng.* 18 (2019) 381–391, <https://doi.org/10.1080/13467581.2019.1660663>.
- [93] J.M. Coughlan, A.L. Yuille, Manhattan World: compass direction from a single image by Bayesian inference, in: *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, vol. 2, IEEE, 1999, pp. 941–947, <https://doi.org/10.1109/ICCV.1999.790349>.
- [94] E. Delage, Honglak Lee, A.Y. Ng, A Dynamic Bayesian Network Model for Autonomous 3D Reconstruction from a Single Indoor Image, in: *2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. vol. 2*, IEEE, n.d.: pp. 2418–2428. <https://doi.org/10.1109/CVPR.2006.23>.
- [95] V. Hedau, D. Hoiem, D. Forsyth, Recovering the spatial layout of cluttered rooms, in: *2009 IEEE 12th Int. Conf. Comput. Vis.*, IEEE, 2009, pp. 1849–1856, <https://doi.org/10.1109/ICCV.2009.5459411>.
- [96] H. Izadinia, Q. Shan, S.M. Seitz, IM2CAD, in: *2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2017, pp. 2422–2431, <https://doi.org/10.1109/CVPR.2017.260>.
- [97] S. Tulsiani, S. Gupta, D. Fouhey, A.A. Efros, J. Malik, Factoring shape, pose, and layout from the 2D image of a 3D scene, in: *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2018, pp. 302–310, <https://doi.org/10.1109/CVPR.2018.00039>.
- [98] C. Zou, A. Colburn, Q. Shan, D. Hoiem, LayoutNet: reconstructing the 3D room layout from a single RGB image, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2051–2059, <https://doi.org/10.1109/CVPR.2018.00219>.
- [99] P.M. Chu, Y. Sung, K. Cho, Generative adversarial network-based method for transforming single RGB image into 3D point cloud, *IEEE Access* 7 (2019) 1021–1029, <https://doi.org/10.1109/ACCESS.2018.2886213>.
- [100] H. Kim, J. Moon, B. Lee, RGB-to-TSDF: direct TSDF prediction from a single RGB image for dense 3D reconstruction, *IEEE Int. Conf. Intell. Robot. Syst.* (2019) 6714–6720, <https://doi.org/10.1109/IRROS40897.2019.8968566>.
- [101] C. Sun, C.-W. Hsiao, M. Sun, H.-T. Chen, HorizonNet: learning room layout with 1D representation and pano stretch data augmentation, in: *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2019, pp. 1047–1056, <https://doi.org/10.1109/CVPR.2019.00114>.
- [102] C.Y. Lee, V. Badrinarayanan, T. Malisiewicz, A. Rabinovich, RoomNet: end-to-end room layout estimation, in: *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4875–4884, <https://doi.org/10.1109/ICCV.2017.521>.
- [103] F. Engelmann, T. Kontogianni, J. Schult, B. Leibe, Know what your neighbors do: 3D semantic segmentation of point clouds. https://doi.org/10.1007/978-3-030-11015-4_29, 2019.
- [104] S. Wang, S. Suo, W.C. Ma, A. Pokrovsky, R. Urtasun, Deep parametric continuous convolutional neural networks, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2589–2597, <https://doi.org/10.1109/CVPR.2018.00274>.
- [105] S. Deng, Q. Dong, B. Liu, SCSS-net: Superpoint Constrained Semi-supervised Segmentation Network for 3D Indoor Scenes, 2021. <http://arxiv.org/abs/2107.03601>.
- [106] N. Zhao, End2End semantic segmentation for 3D indoor scenes, in: *MM 2018 - Proc. 2018 ACM Multimed. Conf.*, 2018, pp. 810–814, <https://doi.org/10.1145/3240508.3243933>.
- [107] A. Martinovic, J. Knopp, H. Riemenschneider, L. Van Gool, 3D all the way: semantic segmentation of urban scenes from start to end in 3D, in: *2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2015, pp. 4456–4465, <https://doi.org/10.1109/CVPR.2015.7299075>.
- [108] T. Wang, J. Li, X. An, An efficient scene semantic labeling approach for 3D point cloud, in: *2015 IEEE 18th Int. Conf. Intell. Transp. Syst.*, IEEE, 2015, pp. 2115–2120, <https://doi.org/10.1109/ITSC.2015.342>.
- [109] D. Wolf, J. Prankl, M. Vincze, Fast semantic segmentation of 3D point clouds using a dense CRF with learned parameters, in: *2015 IEEE Int. Conf. Robot. Autom.*, IEEE, 2015, pp. 4867–4873, <https://doi.org/10.1109/ICRA.2015.7139875>.
- [110] D. Wolf, J. Prankl, M. Vincze, Enhancing semantic segmentation for robotics: the power of 3-D entangled forests, *IEEE Rob. Autom. Lett.* 1 (2016) 49–56, <https://doi.org/10.1109/LRA.2015.2506118>.
- [111] X. Liu, H. Li, W. Meng, S. Xiang, X. Zhang, 3D Point Cloud Classification Based on Discrete Conditional Random Field, 2017, pp. 115–137, https://doi.org/10.1007/978-3-319-65849-0_14.
- [112] L. Tchepami, C. Choy, I. Armeni, J. Gwak, S. Savarese, SEGCloud: semantic segmentation of 3D point clouds, in: *Proc. - 2017 Int. Conf. 3D Vision*, vol. 2018, 3DV, 2017, pp. 537–547, <https://doi.org/10.1109/3DV.2017.00067>.
- [113] M. Fisher, M. Savva, Y. Li, P. Hanrahan, M. Nießner, Activity-centric scene synthesis for functional 3D scene modeling, *ACM Trans. Graph.* 34 (2015), <https://doi.org/10.1145/2816795.2818057>.
- [114] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, P. Hanrahan, Example-based synthesis of 3D object arrangements, *ACM Trans. Graph.* 31 (2012), <https://doi.org/10.1145/2366145.2366154>.
- [115] R. Ma, H. Li, C. Zou, Z. Liao, X. Tong, H. Zhang, Action-driven 3D indoor scene evolution, *ACM Trans. Graph.* 35 (2016) 1–13, <https://doi.org/10.1145/2980179.2980223>.
- [116] A.X. Chang, M. Savva, C.D. Manning, Learning spatial knowledge for text to 3D scene generation, in: *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, 2014, pp. 2028–2038, <https://doi.org/10.3115/v1/d14-1217>.
- [117] A.X. Chang, M. Eric, M. Savva, C.D. Manning, SceneSeer: 3D Scene Design with Natural Language, 2017. <http://arxiv.org/abs/1703.00050>.
- [118] J. Lee, T. Funkhouser, Sketch-based search and composition of 3D models, in: *Proc. Fifth Eurographics Conf. Sketch-Based Interfaces Model*, 2008, pp. 97–104, <https://doi.org/10.2312/SBM/SBM08/097-104>.
- [119] R. Schmidt, B. Wyvill, M.C. Sousa, J.A. Jorge, ShapeShop: sketch-based solid modeling with BlobTrees, *SIGGRAPH 2006 - ACM SIGGRAPH 2006 courses*. <https://doi.org/10.1145/1185657.1185775>, 2006.
- [120] K. Xu, K. Chen, H. Fu, W.-L. Sun, S.-M. Hu, Sketch2Scene, *ACM Trans. Graph.* 32 (2013) 1–15, <https://doi.org/10.1145/2461912.2461968>.
- [121] T. Liu, A. Hertzmann, W. Li, T. Funkhouser, Style compatibility for 3D furniture models, *ACM Trans. Graph.* 34 (2015), <https://doi.org/10.1145/2766898>.
- [122] Q. Chen, Q. Wu, R. Tang, Y. Wang, S. Wang, M. Tan, Intelligent home 3D: automatic 3D-house design from linguistic descriptions only, in: *2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, IEEE, 2020, pp. 12622–12631, <https://doi.org/10.1109/CVPR42600.2020.01264>.
- [123] M. Li, A.G. Patil, K. Xu, S. Chaudhuri, O. Khan, A. Shamir, C. Tu, B. Chen, D. Cohen-Or, H. Zhang, Grains: generative recursive autoencoders for indoor scenes, *ACM Trans. Graph.* 38 (2019), <https://doi.org/10.1145/3303766>.
- [124] D. Ritchie, K. Wang, Y.A. Lin, Fast and flexible indoor scene synthesis via deep convolutional generative models, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6175–6183, <https://doi.org/10.1109/CVPR.2019.00634>.
- [125] W. Zhou, E. Yang, J. Lei, J. Wan, L. Yu, PGDENet: progressive guided fusion and depth enhancement network for RGB-D indoor scene parsing, *IEEE Trans. Multimed.* 9210 (2022), <https://doi.org/10.1109/TMM.2022.3161852>, 1–1.
- [126] T. Gong, W. Zhou, X. Qian, J. Lei, L. Yu, Global contextually guided lightweight network for RGB-thermal urban scene understanding, *Eng. Appl. Artif. Intell.* 117 (2023), 105510, <https://doi.org/10.1016/j.engappai.2022.105510>.
- [127] Z. Xu, D. Tao, Y. Zhang, J. Wu, A.C. Tsoi, Architectural Style Classification Using Multinomial Latent Logistic Regression, 2014, pp. 600–615, https://doi.org/10.1007/978-3-319-10590-1_39.
- [128] B. Zhang, Y. Song, S. Guan, Y. Zhang, Historic Chinese architectures image retrieval by SVM and Pyramid histogram of oriented gradients features, *Int. J. Soft Comput.* 5 (2010) 19–28, <https://doi.org/10.3923/ijscmp.2010.19.28>.
- [129] G. Shalunts, Y. Haxhimusa, R. Sablatnig, Architectural style classification of building facade windows, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2011, pp. 280–289, https://doi.org/10.1007/978-3-642-24031-7_28.
- [130] G. Shalunts, Y. Haxhimusa, R. Sablatnig, Architectural Style Classification of Domes, 2012, pp. 420–429, https://doi.org/10.1007/978-3-642-33191-6_41.
- [131] A. Goel, M. Juneja, C.V. Jawahar, Are buildings only instances? Exploration in architectural style categories, *ACM Int. Conf. Proceeding Ser.* (2012), <https://doi.org/10.1145/2425333.2425334>.
- [132] G. Shalunts, Architectural Style Classification of Building Facade Towers, 2015, pp. 285–294, https://doi.org/10.1007/978-3-319-27857-5_26.
- [133] A.M. Obeso, J. Benois-Pineau, A.Á.R. Acosta, M.S.G. Vázquez, Architectural style classification of Mexican historical buildings using deep convolutional neural networks and sparse features, *J. Electron. Imag.* 26 (2016), 011016, <https://doi.org/10.1117/1.jei.26.1.011016>.

- [134] R. Fathalla, G. Vogiatis, A deep learning pipeline for semantic facade segmentation, in: *Br. Mach. Vis. Conf. 2017*, vol. 2017, BMVC, 2017, pp. 1–13, <https://doi.org/10.5244/c.31.120>.
- [135] A. Fond, M.-O. Berger, G. Simon, Facade proposals for urban augmented reality, in: *2017 IEEE Int. Symp. Mix. Augment. Real.*, IEEE, 2017, pp. 32–41, <https://doi.org/10.1109/ISMAR.2017.20>.
- [136] F. Lippoldt, Window detection in facades for aerial texture files of 3D CityGML models. https://openaccess.thcvf.com/content_CVPRW_2019/papers/DOAI/Lippoldt_Window_Detection_in_Facades_for_Aerial_Texture_Files_of_3D_CVPRW_2019_paper.pdf, 2019.
- [137] H. Liu, J. Zhang, J. Zhu, S.C.H. Hoi, Deepfacade: a deep learning approach to facade parsing, in: *IJCAI Int. Jt. Conf. Artif. Intell.*, 2017, pp. 2301–2307.
- [138] K. Rahmani, H. Mayer, High quality facade segmentation based on structured random forest, region proposal network and rectangular fitting, in: *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, 2018, pp. 223–230, <https://doi.org/10.5194/isprs-annals-IV-2-223-2018>.
- [139] A. Fond, M.-O. Berger, G. Simon, Model-image registration of a building's facade based on dense semantic segmentation, *Comput. Vis. Image Underst.* 206 (2021), 103185, <https://doi.org/10.1016/j.cviu.2021.103185>.
- [140] X. Zhuo, M. Monks, T. Esch, P. Reinartz, Facade segmentation from oblique UAV imagery, in: *2019 Jt. Urban Remote Sens. Event*, vol. 2019, JURSE, 2019, pp. 2–5, <https://doi.org/10.1109/JURSE.2019.8809024>.
- [141] W. Ma, W. Ma, S. Xu, H. Zha, Pyramid ALKNet for semantic parsing of building facade image, *Geosci. Rem. Sens. Lett. IEEE* (2020) 1–5, <https://doi.org/10.1109/IGRS.2020.2993451>.
- [142] N. Ripperda, C. Brenner, Reconstruction of façade structures using a formal grammar and RjMCMC, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 4174 LNCS (2006) 750–759, https://doi.org/10.1007/11861898_75.
- [143] N. Ripperda, Determination of facade attributes for facade reconstruction, in: *ISPRS Congr. Beijing 2008 Proc. Comm. III*, 2008, pp. 285–290.
- [144] N. Ripperda, C. Brenner, Application of a formal grammar to facade reconstruction in semiautomatic and automatic environments, *12th agil, Int. Conf. Geogr. Inf. Sci.* (2009) 1–12. <http://www.ikg.uni-hannover.de/agile/fileadmin/agile/paper/44.pdf>.
- [145] G. Wan, A. Sharf, Grammar-based 3D facade segmentation and reconstruction, *Comput. Graph.* 36 (2012) 216–223, <https://doi.org/10.1016/j.cag.2012.02.003>.
- [146] F. Sadeghi, H. Arefi, A. Fallah, M. Hahn, 3D building facade reconstruction using handheld laser scanning data, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* - ISPRS Arch. 40 (2015) 625–630, <https://doi.org/10.5194/isprsarchives-XL-1-W5-625-2015>.
- [147] R.G. Lotte, N. Haala, M. Karpina, L.E.O.e.C. de Aragão, Y.E. Shimabukuro, 3D façade labeling over complex scenarios: a case study using convolutional neural network and structure-from-motion, *Rem. Sens.* 10 (2018), <https://doi.org/10.3390/rs10091435>.
- [148] S. Hensel, S. Goebels, M. Kada, Facade reconstruction for textured LOD2 citygml models based on deep learning and mixed integer linear programming, in: *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, 2019, pp. 37–44, <https://doi.org/10.5194/isprs-annals-IV-2-W5-37-2019>.
- [149] K. Bacharidis, F. Sarri, L. Ragia, 3D building façade reconstruction using deep learning, *ISPRS Int. J. Geo-Inf.* 9 (2020) 1–24, <https://doi.org/10.3390/ijgi9050322>.
- [150] J. Gagne, M. Andersen, A generative facade design method based on daylighting performance goals, *J. Build. Perform. Simul.* 5 (2012) 141–154, <https://doi.org/10.1080/19401493.2010.549572>.
- [151] G. Stiny, Introduction to shape and shape grammars, *Environ. Plann. Plann. Des.* 7 (1980) 343–351, <https://doi.org/10.1068/b070343>.
- [152] P. Wonka, M. Wimmer, F. Sillion, W. Ribarsky, Instant Architecture, *ACM SIGGRAPH 2003 Pap, SIGGRAPH '03*, 2003, pp. 669–677, <https://doi.org/10.1145/1201775.882324>.
- [153] F. Wu, D.M. Yan, W. Dong, X. Zhang, P. Wonka, Inverse procedural modeling of facade layouts, *ACM Trans. Graph.* 33 (2014), <https://doi.org/10.1145/2601097.2601162>.
- [154] D. Newton, Generative deep learning in architectural design, *Technol. Archit. Des.* 3 (2019) 176–189, <https://doi.org/10.1080/24751448.2019.1640536>.
- [155] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 2017, pp. 5967–5976, <https://doi.org/10.1109/CVPR.2017.632>.
- [156] M. Bachl, D.C. Ferreira, City-GAN: Learning Architectural Styles Using a Custom Conditional GAN Architecture, 2019, pp. 1–10. <http://arxiv.org/abs/1907.05280>.
- [157] G. Zhang, Y. Pan, L. Zhang, Deep learning for detecting building façade elements from images considering prior knowledge, *Autom. Construct.* 133 (2022), 104016, <https://doi.org/10.1016/j.autcon.2021.104016>.
- [158] M. Dai, G.M. Meyers, D.O. Densley Tingley, M. Mayfield, Initial investigations into using an ensemble of deep neural networks for building façade image semantic segmentation, 2019, 1115708, <https://doi.org/10.1117/12.2532828>.
- [159] B. Huang, B. Zhao, Y. Song, Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, *Remote Sens. Environ.* 214 (2018) 73–86, <https://doi.org/10.1016/j.rse.2018.04.050>.
- [160] W. Li, R. Dong, H. Fu, J. Wang, L. Yu, P. Gong, Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping, *Remote Sens. Environ.* 237 (2020), <https://doi.org/10.1016/j.rse.2019.111563>.
- [161] A. Caglayan, N. Imamoglu, A.B. Can, R. Nakamura, When CNNs meet random RNNs: towards multi-level analysis for RGB-D object and scene recognition, *ArXiv* (2020) 1–16.
- [162] D. Sharma, C. Chattopadhyay, ROBIN and S-ROBIN dataset. <https://github.com/gesstalt/ROBIN.git>, 2017. (Accessed 9 November 2021).
- [163] J. Jin, Z. Xue, B. Leng, SHRAG: semantic hierarchical graph for floorplan representation, in: *2022 Int. Conf. 3D Vis.*, IEEE, 2022, pp. 271–279, <https://doi.org/10.1109/3DV57658.2022.00039>.
- [164] S. Zhao, L. Wang, X. Qian, J. Chen, Enhancing performance-based generative architectural design with sketch-based image retrieval: a pilot study on designing building facade fenestrations, *Vis. Comput.* 37 (2021). <https://link.springer.com/article/10.1007%2Fs00371-021-02170-x>.
- [165] K. Ma, B. Wang, Y. Li, J. Zhang, Image retrieval for local architectural heritage recommendation based on deep hashing, *Buildings* 12 (2022) 1–16, <https://doi.org/10.3390/buildings12060809>.
- [166] J. Xu, B. Stenger, T. Kerola, T. Tung, Pano2CAD: room layout from a single panorama image, in: *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, 2017, pp. 354–362, <https://doi.org/10.1109/WACV.2017.46>.
- [167] H. Kim, Evaluation of deep learning-based automatic floor plan analysis technology: an ahp-based assessment, *Appl. Sci.* 11 (2021), <https://doi.org/10.3390/app11114727>.
- [168] S. Dong, W. Wang, W. Li, K. Zou, Vectorization of floor plans based on EdgeGAN, *OR Inf.* 12 (2021), <https://doi.org/10.3390/info12050206>.
- [169] X. Di, P. Yu, D. Yang, H. Zhu, C. Sun, Y. Liu, End-to-end Generative Floor-Plan and Layout with Attributes and Relation Graph, 2020, pp. 1–8. <http://arxiv.org/abs/2012.08514>.
- [170] S. Meng, Exploring in the latent space of design: a method of plausible building facades images generation, properties control and model explanation Base on StyleGAN2, in: *Proc. 2021 Digit.*, Springer Singapore, Singapore, 2021, pp. 55–68, https://doi.org/10.1007/978-981-16-5983-6_6.
- [171] J. Jeong, J. Jo, J. Wang, S. Cho, J. Park, Realistic image synthesis with configurable 3D scene layouts. <http://arxiv.org/abs/2108.10031>, 2021.
- [172] Q. Li, Y. Shi, X. Huang, X.X. Zhu, Building footprint generation by integrating convolution neural network with feature pairwise conditional random field (FPCRF), *IEEE Trans. Geosci. Rem. Sens.* (2020) 1–18, <https://doi.org/10.1109/tgrs.2020.2973720>.
- [173] S. Zhao, L. Wang, X. Qian, J. Chen, Enhancing performance-based generative architectural design with sketch-based image retrieval: a pilot study on designing building facade fenestrations, *Vis. Comput.* 38 (2022) 2981–2997, <https://doi.org/10.1007/s00371-021-02170-x>.
- [174] M. Vidanapathirana, Q. Wu, Y. Furukawa, A.X. Chang, M. Savva, Plan2Scene, Converting floorplans to 3D scenes, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10728–10737, <https://doi.org/10.1109/CVPR46437.2021.01059>.
- [175] C. Fernandez-Labrador, J.M. Facil, A. Perez-Yus, C. Demonceaux, J. Civera, J.J. Guerrero, Corners for layout: end-to-end layout recovery from 360 images, *ArXiv* 5 (2019) 1255–1262.
- [176] M. Rahbar, M. Mahdavinjad, A.H.D. Markazi, M. Bemanian, Architectural layout design through deep learning and agent-based modeling: a hybrid approach, *J. Build. Eng.* 47 (2022), 103822, <https://doi.org/10.1016/j.jobbe.2021.103822>.

- [177] S. Kim, S. Park, H. Kim, K. Yu, Deep floor plan analysis for complicated drawings based on style transfer, *J. Comput. Civ. Eng.* 35 (2021), 04020066, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000942](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000942).
- [178] W. Qian, Y. Xu, H. Li, A self-sparse generative adversarial network for autonomous early-stage design of architectural sketches, *Comput. Civ. Infrastruct. Eng.* (2021) 1–17, <https://doi.org/10.1111/mice.12759>.
- [179] Z. Luo, W. Huang, FloorplanGAN: vector residential floorplan adversarial generation, *Autom. Construct.* 142 (2022), 104470, <https://doi.org/10.1016/j.autcon.2022.104470>.
- [180] Q. Li, Y. Shi, S. Auer, R. Roschlaub, K. Möst, M. Schmitt, C. Glock, X.X. Zhu, Detection of undocumented building constructions from official geodata using a convolutional neural network, *Rem. Sens.* 12 (2020) 1–21, <https://doi.org/10.3390/rs12213537>.
- [181] A. Kalervo, J. Ylioinas, M. Häikiö, A. Karhu, J. Kannala, CubiCasa5K, A dataset and an improved multi-task model for floorplan image analysis, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2019, pp. 28–40, https://doi.org/10.1007/978-3-030-20205-7_3.
- [182] A. Basu, L. Petropoulakis, G. Di Caterina, J. Soraghan, Indoor home scene recognition using capsule neural networks, in: *Procedia Comput. Sci.*, Elsevier B.V., 2020, pp. 440–448, <https://doi.org/10.1016/j.procs.2020.03.253>.
- [183] M. Dai, W.O.C. Ward, G. Meyers, D. Densley Tingley, M. Mayfield, Residential building facade segmentation in the urban environment, *Build. Environ.* 199 (2021), 107921, <https://doi.org/10.1016/j.buildenv.2021.107921>.
- [184] J. Zhang, T. Fukuda, N. Yabuki, Automatic generation of synthetic datasets from a city digital twin for use in the instance segmentation of building facades, *J. Comput. Des. Eng.* 9 (2022) 1737–1755, <https://doi.org/10.1093/jcde/qwac086>.
- [185] L. Zhang, L. Zheng, Y. Chen, L. Huang, S. Zhou, CGAN-assisted renovation of the styles and features of street facades—a case study of the wuyi area in fujian, China, *Sustain.* Times 14 (2022), <https://doi.org/10.3390/su142416575>.
- [186] P. Arroyo, A. Schöttle, R. Christensen, The Ethical and Social Dilemma of AI Uses in the Construction Industry, 2021, pp. 227–236, <https://doi.org/10.24928/2021/0188>.
- [187] J. Bieser, How Can AI Support Human Creativity? Here's what a New Study Found, *World Econ. Forum.*, 2023. <https://www.weforum.org/agenda/2023/02/ai-can-catalyze-and-inhibit-your-creativity-here-is-how/>. (Accessed 30 May 2023).
- [188] A. Paullada, I.D. Raji, E.M. Bender, E. Denton, A. Hanna, Data and its (dis)contents: a survey of dataset development and use in machine learning research, *Patterns 2* (2021), 100336, <https://doi.org/10.1016/j.patter.2021.100336>.
- [189] U. Ehsan, Q.V. Liao, M. Muller, M.O. Riedl, J.D. Weisz, Expanding explainability: towards social transparency in AI systems, in: *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.*, ACM, New York, NY, USA, 2021, pp. 1–19, <https://doi.org/10.1145/3411764.3445188>.
- [190] R. Davari Majd, M. Momeni, P. Moallem, Transferable object-based framework based on deep convolutional neural networks for building extraction, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 12 (2019) 2627–2635, <https://doi.org/10.1109/JSTARS.2019.2924582>.
- [191] H. Hosseinpour, F. Samadzadegan, Convolutional neural network for building extraction from high-resolution remote sensing images, in: *2020 Int. Conf. Mach. Vis. Image Process.*, IEEE, 2020, pp. 1–5, <https://doi.org/10.1109/MVIP49855.2020.9187483>.
- [192] D. Stilller, T. Stark, M. Wurm, S. Dech, H. Taubenbock, Large-scale building extraction in very high-resolution aerial imagery using Mask R-CNN, in: *2019 Jt. Urban Remote Sens. Event, IEEE*, 2019, pp. 1–4, <https://doi.org/10.1109/JURSE.2019.8808977>.
- [193] C. Ayala, R. Sesma, C. Aranda, M. Galar, A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery, *Rem. Sens.* 13 (2021) 3135, <https://doi.org/10.3390/rs13163135>.
- [194] N. Audebert, B. Le Saux, S. Lefèvre, Beyond RGB: very high resolution urban remote sensing with multimodal deep networks, *ISPRS J. Photogrammetry Remote Sens.* 140 (2018) 20–32, <https://doi.org/10.1016/j.isprsjprs.2017.11.011>.
- [195] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, S.L. Waslander, Aerial imagery for roof segmentation: a large-scale dataset towards automatic mapping of buildings, *ArXiv* (2018).
- [196] N. Nauata, K.H. Chang, C.Y. Cheng, G. Mori, Y. Furukawa, House-GAN: relational generative adversarial networks for graph-constrained house layout generation, *ArXiv* (2020), https://doi.org/10.1007/978-3-030-58452-8_10.
- [197] Y. Li, Y. Chen, A. Rajabifard, K. Khoshelham, M. Aleksandrov, Estimating building age from google street view images using deep learning, in: *Leibniz Int. Proc. Informatics, LIPICs*, 2018, pp. 1–7, <https://doi.org/10.4230/LIPICs.GIScience.2018.40>.
- [198] A.M. Obeso, M.S.G. Vázquez, A.A.R. Acosta, J. Benois-Pineau, Connoisseur, Classification of styles of Mexican architectural heritage with deep learning and visual attention prediction, *ACM Int. Conf. Proceeding Ser. Part F1301* (2017), <https://doi.org/10.1145/3095713.3095730>.
- [199] A.M. Obeso, J. Benois-Pineau, M.S.G. Vázquez, A.A.R. Acosta, Saliency-based selection of visual content for deep convolutional neural networks: application to architectural style classification, *Multimed. Tool. Appl.* 78 (2019) 9553–9576, <https://doi.org/10.1007/s11042-018-6515-2>.
- [200] G. Nishida, A. Bousseau, D.G. Aliaga, Procedural modeling of a building from a single image, *Comput. Graph. Forum* 37 (2018) 415–429, <https://doi.org/10.1111/cgf.13372>.
- [201] C. Zou, J.W. Su, C.H. Peng, A. Colburn, Q. Shan, P. Wonka, H.K. Chu, D. Hoiem, 3D Manhattan Room Layout Reconstruction from a Single 360° Image, *ArXiv*, 2019.
- [202] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, in: *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, pp. 2481–2495, <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [203] J. Cai, Y. Chen, MHA-net: multipath hybrid attention network for building footprint extraction from high-resolution remote sensing imagery, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 14 (2021) 5807–5817, <https://doi.org/10.1109/JSTARS.2021.3084805>.
- [204] R. Hang, Z. Li, Q. Liu, P. Ghamisi, S.S. Bhattacharyya, Hyperspectral image classification with attention-aided CNNs, *IEEE Trans. Geosci. Rem. Sens.* 59 (2021) 2281–2293, <https://doi.org/10.1109/TGRS.2020.3007921>.
- [205] H. Chen, W. Li, Z. Shi, Adversarial instance augmentation for building change detection in remote sensing images, *IEEE Trans. Geosci. Rem. Sens.* (2021) 1–16, <https://doi.org/10.1109/TGRS.2021.3066802>.
- [206] A. Ziaee, R. Dehbozorgi, M. Döllner, A novel adaptive deep network for building footprint segmentation. <http://arxiv.org/abs/2103.00286>, 2021.
- [207] C.P. Simonsen, F.M. Thiesson, M.P. Philipsen, T.B. Moeslund, Generalizing floor plans using graph neural networks, in: *2021 IEEE Int. Conf. Image Process.*, IEEE, 2021, pp. 654–658, <https://doi.org/10.1109/ICIP42928.2021.9506514>.
- [208] D. Sharma, C. Chattopadhyay, G. Harit, A unified framework for semantic matching of architectural floorplans, *Proc. - Int. Conf. Pattern Recognit.* 0 (2016) 2422–2427, <https://doi.org/10.1109/ICPR.2016.7899999>.
- [209] D. Sharma, C. Chattopadhyay, High-level feature aggregation for fine-grained architectural floor plan retrieval, *IET Comput. Vis.* 12 (2018) 702–709, <https://doi.org/10.1049/iet-cvi.2017.0581>.
- [210] V. Tavakkoli, K. Mohsenzadegan, K. Kyamakya, A visual sensing concept for robustly classifying house types through a convolutional neural network architecture involving a multi-channel features extraction, *Sensors* 20 (2020) 1–16, <https://doi.org/10.3390/s20195672>.
- [211] C.K. Li, H.X. Zhang, J.X. Liu, Y.Q. Zhang, S.C. Zou, Y.T. Fang, Window detection in facades using heatmap fusion, *J. Comput. Sci. Technol.* 35 (2020) 900–912, <https://doi.org/10.1007/s11390-020-0253-4>.
- [212] H. Liu, Y. Xu, J. Zhang, J. Zhu, Y. Li, S.C.H. Hoi, DeepFacade: a deep learning approach to facade parsing with symmetric loss, in: *IEEE Trans. Multimed.*, 2020, pp. 3153–3165, <https://doi.org/10.1109/TMM.2020.2971431>.
- [213] J. Femiani, W.R. Para, N. Mitra, P. Wonka, facade segmentation in the wild. <http://arxiv.org/abs/1805.08634>, 2018.
- [214] W. Ma, S. Xu, W. Ma, H. Zha, Multiview feature aggregation for facade parsing, *Geosci. Rem. Sens. Lett. IEEE* (2020) 1–5, <https://doi.org/10.1109/lgrs.2020.3035721>.
- [215] W. Zhang, W. Zhang, J. Gu, Edge-semantic learning strategy for layout estimation in indoor environment, *IEEE Trans. Cybern.* 50 (2020) 2730–2739, <https://doi.org/10.1109/TCYB.2019.2895837>.
- [216] S. Dasgupta, K. Fang, K. Chen, S. Savarese, DeLay: robust spatial layout estimation for cluttered indoor scenes, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 616–624, <https://doi.org/10.1109/CVPR.2016.73>.

- [217] Z. Xiong, Y. Yuan, Q. Wang, MSN: modality separation networks for RGB-D scene recognition, *Neurocomputing* 373 (2020) 81–89, <https://doi.org/10.1016/j.neucom.2019.09.066>.
- [218] X. Song, L. Herranz, S. Jiang, Depth CNNs for RGB-D scene recognition: learning from scratch better than transferring from RGB-CNNs, in: *31st AAAI Conf. Artif. Intell. AAAI 2017, 2017*, pp. 4271–4277.
- [219] Y. Li, J. Zhang, Y. Cheng, K. Huang, T. Tan, Df 2 Net: a discriminative feature learning and fusion network for RgB-D indoor scene classification, in: *32nd AAAI Conf. Artif. Intell. AAAI 2018, 2018*, pp. 7041–7048.
- [220] A. Mosella-Montoro, J. Ruiz-Hidalgo, 2D–3D Geometric Fusion network using Multi-Neighbourhood Graph Convolution for RGB-D indoor scene classification, *Inf. Fusion* 76 (2021) 46–54, <https://doi.org/10.1016/j.inffus.2021.05.002>.
- [221] S.T. Yang, F.E. Wang, C.H. Peng, P. Wonka, M. Sun, H.K. Chu, DuLa-Net, A dual-projection network for estimating room layouts from a single RGB panorama, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3358–3367, <https://doi.org/10.1109/CVPR.2019.00348>.