Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

The 16th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2025)
October 28-30, 2025, Istanbul, Türkiye

Leveraging Machine Learning Techniques in Converting 2D Floorplans Images to 3D Models: Applications, Challenges, and Future Directions

Ibsen G. Bazie[a,b], Boaz N. Nzazi[a,b,c], Jirince K. Biaba[a,b,c], Tasho Tashev[d], Witesyavwirwa V. Kambale[e], Kyandoghere Kyamakya[f], Nathanaël M. Kasoro[a,c], Selain K. Kasereka[a,c,d,*]

[a] *ABIL Research Center, Kinshasa, Democratic Republic of the Congo*
[b] *Institut Francophone International (IFI), Vietnam National University, Hanoi, Vietnam*
[c] *Department of Mathematics, Statistics and Computer Science, University of Kinshasa, Kinshasa, Democratic Republic of the Congo*
[d] *Department of Information Measurement Systems, Technical University of Sofia, Sofia, Bulgaria*
[e] *Faculty of Information and Communication Technology, Tshwane University of Technology, Pretoria, South Africa*
[f] *Institute of Smart Systems Technologies, University of Klagenfurt, Klagenfurt, Austria*

Abstract

With the growing demand for automation in fields such as architecture, real estate, and digital twin technologies, the ability to efficiently convert 2D floorplan images into accurate 3D structural models has become increasingly critical. Traditional CAD (Computer-Aided Design)-based approaches, while precise, often lack scalability and adaptability in dynamic or large-scale environments. In response, recent advancements in machine learning have opened new possibilities for intelligent 3D reconstruction. This short review explores these developments, surveying key machine learning pipelines, benchmark datasets, evaluation metrics, and real-world applications. It also addresses persistent challenges including generalization, occlusion, and dataset limitations. The paper highlights promising directions such as diffusion models and foundation models, that aim to overcome current barriers and shape the future of automated 3D modeling from 2D sources. This work contributes to a clearer understanding of the current landscape, identifies gaps in existing approaches, and outlines strategic pathways for future research in data-driven architectural modeling.

Keywords: 2D to 3D reconstruction; Floorplan analysis; Deep learning; Neural rendering; Real estate automation; Indoor modeling

[*] Corresponding author. Tel.: +243-821-828-964.
E-mail address: selain.kasereka@unikin.ac.cd

1. Introduction

The conversion of two-dimensional (2D) floorplan images into three-dimensional (3D) structural models has emerged as a critical area of research and development in recent years. This process underpins various domains, including real estate visualization [1, 2], indoor navigation for robotics, virtual and augmented reality (VR/AR) [3, 4], and the creation of digital twins for smart cities. Traditional approaches, such as computer-aided design (CAD) tools and manual reconstruction, are often labor-intensive, time-consuming, and lack scalability across diverse architectural styles and datasets [5].

With the rapid advancements in artificial intelligence, machine learning (ML) techniques have emerged as promising alternatives for automating 2D-to-3D conversion. Convolutional neural networks (CNNs) [6], generative adversarial networks (GANs) [7], recurrent neural networks (RNNs) [8], and transformer-based architectures [9] have enabled researchers to develop systems capable of reconstructing 3D geometry and textures directly from 2D floorplans [10, 11]. These systems support diverse applications from property marketing with interactive 3D tours to indoor scene understanding for autonomous systems and immersive VR/AR environments.

Despite these advancements, significant challenges remain. Floorplans often lack depth information and may exhibit occlusions or ambiguous elements, complicating the reconstruction process [12]. Generalizing ML models to handle diverse architectural conventions and noisy or incomplete input data remains an open research problem. Furthermore, computational constraints and the demand for real-time performance pose additional barriers to widespread deployment [13].

The objective of this study is to provide a comprehensive review of recent machine learning approaches for converting two-dimensional (2D) floorplan images into three-dimensional (3D) structural models. Specifically, the paper aims to:

- Synthesize academic and applied research contributions on machine learning approaches for 2D-3D floorplan conversion;
- Identify core machine learning techniques used in real-world applications such as real estate visualization, robotics, and digital twins;
- Evaluate commonly used benchmark datasets and metrics for performance assessment;
- Discuss major challenges including input ambiguity, architectural diversity, and computational limitations;
- Propose future research directions including emerging techniques like diffusion models and foundation models.

The remainder of this paper is organized as follows: Section 2 details the research methodology used for identifying and analyzing relevant studies. Section 3 discusses core machine learning techniques for 2D-to-3D conversion. Section 4 explores major application domains, while Section 5 presents current challenges and limitations. Section 6 provides an overview of the explored techniques and models explained in the paper. Section 7 outlines future research directions, and Section 8 concludes the study.

2. Research Methodology

This literature review follows the PRISMA 2020 framework [14] to investigate machine learning techniques for converting 2D floorplan images into 3D structural models. The search strategy queried six primary databases (Google Scholar, IEEE Xplore, ScienceDirect, Springer, MDPI, arXiv) and two additional repositories (ResearchGate, TheCVF Open Access) using Boolean operators and keywords including ”2D floorplan to 3D model”, ”floorplan 3D reconstruction”, ”generative models for architectural modeling”, and ”neural rendering”. The search covered peer-reviewed articles and preprints published between January 2020 and July 2025. An initial corpus of 89 records was retrieved. After removing duplicates (n = 8) and applying automated filters (n = 15), 66 records underwent title and abstract screening. Studies focusing solely on hardware implementation or traditional CAD workflows were excluded (n = 18). The remaining 48 articles were assessed for full-text eligibility, with 15 excluded for insufficient relevance to ML-based 2D-to-3D conversion. Ultimately, 33 studies were selected: 22 presenting novel ML architectures for floorplan-based 3D modeling, 8 providing comparative evaluations, and 3 contributing foundational techniques or datasets. Figure 1 shows the PRISMA flowchart.

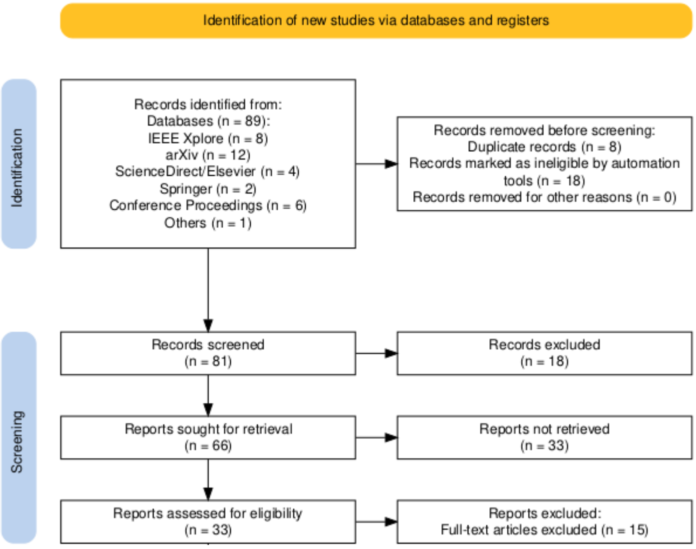
Identification of new studies via databases and registers
Identification
Records identified from:
Databases (n = 89):
IEEE Xplore (n = 8)
arXiv (n = 12)
ScienceDirect/Elsevier (n = 4)
Springer (n = 2)
Conference Proceedings (n = 6)
Others (n = 1)
Records removed before screening:
Duplicate records (n = 8)
Records marked as ineligible by automation tools (n = 18)
Records removed for other reasons (n = 0)
Screening
Records screened (n = 81)
Records excluded (n = 18)
Reports sought for retrieval (n = 66)
Reports not retrieved (n = 33)
Reports assessed for eligibility (n = 33)
Reports excluded:
Full-text articles excluded (n = 15)

Fig. 1. PRISMA flow diagram for systematic literature review process

3. Machine Learning Techniques for 2D-to-3D Floorplan Conversion

3.1. CNN-Based Approaches

Convolutional Neural Networks (CNNs) have been widely adopted for 2D-to-3D floorplan conversion due to their strong spatial feature extraction capabilities, as shown in Eq. (1)–(3).

$$(f * x)(i, j) = \sum_{m=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \sum_{n=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} f(m, n) \cdot x(i - m, j - n), \tag{1}$$

where $f(m, n)$ represents filter weights, $x(i, j)$ is the input feature map, k is the kernel size, and $\lfloor \cdot \rfloor$ denotes the floor function.

$$y_{i,j}^{(\ell+1)} = \sigma\Big(\big(f^{(\ell)} * x^{(\ell)}\big)_{i,j} + b^{(\ell)}\Big), \tag{2}$$

where $y_{i,j}^{(\ell+1)}$ is the output feature, σ is the activation function, $f^{(\ell)}$ represents learned filters, and $b^{(\ell)}$ is the bias term.

$$y_{i,j}^{\text{pool}} = \max_{(m,n)\in\mathcal{W}} x_{i+m,j+n}, \tag{3}$$

where $y_{i,j}^{\text{pool}}$ is the pooled output, $\mathcal{W}$ represents the pooling window, and $x_{i+m,j+n}$ are input values within the window.

3DPlanNet achieves 95.3% wall detection accuracy [2], while Cambeiro Barreiro et al. achieved IoU scores of 0.81 on CubiCasa5k [10]. Fu and Makino employed U^2-Net for 15-second 3D mesh reconstruction [3], 3DPlanNet extends standard CNNs through ensemble learning with rule-based heuristics using only 30 training images.

3.2. GAN-Based Approaches

Generative Adversarial Networks enable realistic 3D synthesis from 2D floorplans, as shown in Eq. (4)–(5).

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))], \tag{4}$$

where $V(D, G)$ is the minimax value function, G is the generator, D is the discriminator, x denotes real data, and z is the latent vector.

$$\mathcal{L}_G = -\mathbb{E}_{z\sim p_z}[\log D(G(z))], \tag{5}$$

where $\mathcal{L}_G$ is the non-saturating generator loss and $D(G(z))$ is the discriminator's probability for generated samples.

Plan2Scene integrates texture generation with Graph Neural Networks [1], while Pix2Vox++ achieves IoU scores of 0.84 on ShapeNet [15]. Plan2Scene modifies standard GANs by implementing Graph Neural Network propagation for texture synthesis on unobserved surfaces.

3.3. Transformer-Based Approaches

Transformers model long-range dependencies in floorplan conversion, as shown in Eq. (6)–(7).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{6}$$

where Q, K, V are query, key, value matrices, d_k is the key dimension, and $\text{softmax}(\cdot)$ normalizes attention weights.

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad \text{MultiHead}(Q, K, V) = [\text{head}_1; \ldots; \text{head}_h]W^O \tag{7}$$

where head_i is the i-th attention head, W_i^Q, W_i^K, W_i^V are projection matrices, and W^O combines information from all heads.

Zheng et al. achieved 68.5–80% geometry win rates using rectified flow Transformers [11], while Para et al. outperformed StyleGAN in perceptual studies [16]. Zheng et al. enhance transformers through masked rectified flow by treating partially completed scene latents as generation constraints.

3.4. Hybrid Methods

Hybrid approaches combine multiple paradigms, as shown in Eq. (8)–(9).

$$h_t = \phi(W_x x_t + W_h h_{t-1} + b), \tag{8}$$

where h_t is the hidden state, x_t is the input, W_x, W_h are weight matrices, and ϕ is the activation function.

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}), \quad r_t = \sigma(W_r x_t + U_r h_{t-1}), \\ \tilde{h}_t &= \tanh(W_h x_t + U_h(r_t \odot h_{t-1})), \quad h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \tag{9}$$

where z_t, r_t are update and reset gates, W_z, W_r, W_h are input weights, U_z, U_r, U_h are recurrent weights, and $\tilde{h}_t$ is the candidate state.

3DPlanNet created over 110,000 3D models [2], while I-Design integrated GPT-4 agents with CLIP embeddings [17]. Hybrid methods offer flexibility but inherit individual component challenges.

4. Applications of 2D-to-3D Floorplan Conversion

The conversion of 2D floor plan images into 3D structural models has found significant adoption across multiple domains, with ML techniques enhancing automation, scalability, and user experiences.

4.1. Real Estate Visualization and Marketing

The real estate industries increasingly leverage ML-based 2D-to-3D conversion for interactive property showcases. Plan2Scene [1] transformed residential floor plans into realistic textured 3D meshes using graph neural networks, achieving superior realism in user studies despite limitations in semantic segmentation and room type assumptions. Similarly, 3DPlanNet [2] combined deep learning with rule-based heuristics, achieved 95.3% wall recognition accuracy while scaling to generate over 110,000 3D models, though requiring manual corrections for complex cases.

4.2. Building Information Modeling (BIM) and Digital Twins

ML approaches enabled automated creation of semantic 3D models for BIM integration. Barreiro et al. [10] proposed a framework converting legacy 2D plans into IFC-compatible 3D BIMs using ResNet and Feature Pyramid Networks, achieving state-of-the-art IoU scores of 0.81 for masks on the CubiCasa5k dataset. However, dataset inconsistencies and reproducibility challenges persist across the field.

4.3. Virtual and Augmented Reality Applications

VR/AR applications benefited significantly from 2D-to-3D conversion for immersive experiences. Fu and Makino [3] developed a VR system using U2-Net for semantic segmentation, reconstructing virtual houses in approximately 15 seconds per floorplan. While effective for interactive exploration, current approaches showed reduced accuracy (76.7%) for complex layouts and require high-quality input images.

4.4. Urban Modeling and Entertainment

At larger scales, ML techniques reconstructed urban layouts and generated virtual worlds. Zheng et al. [11] synthesized coherent 3D towns from single top-down images using rectified flow transformers, achieved 92% geometry win rates in user studies. In entertainment, Persistent Nature created unbounded 3D natural scenes from single-view photos [13], though computational intensity remains a limitation for real-time applications.

5. Challenges and Limitations

Despite significant advancements in leveraging machine learning for 2D-to-3D floorplan conversion, several challenges and limitations persist across current methodologies.

1. **Data-Related Challenges**: Existing datasets often lack diversity in architectural styles, making it difficult for models to adapt to non-Western floorplans with fundamental structural differences. Traditional Japanese designs feature flexible tatami-mat room divisions with sliding panels that create fluid spaces contradicting wall-detection assumptions. Chilean buildings show non-uniform wall systems with complex cross-sections, while traditional African compound structures feature circular or polygonal rooms connected by outdoor pathways, challenging rectangular room assumptions inherent in current datasets like RPLAN [18] and LIFULL [19].
2. **Model Limitations**: CNN-based methods struggle to capture long-range spatial dependencies, leading to semantic inconsistencies in complex layouts [20]. GAN-based approaches suffer from training instabilities and mode collapse, resulting in artifacts and the loss of fine structural details [1, 13]. Transformer architectures introduce high computational overhead and memory requirements, restricting their scalability despite promising global context modeling capabilities [11].
3. **Generalization and Semantic Consistency**: Several methods assume fixed room types and standard architectural elements, limiting applicability to diverse building types. Plan2Scene requires predefined room-image correspondences and handles only limited surface types per room [1]. Hybrid approaches combining deep learning with heuristic rules may fail in corner cases where handcrafted constraints conflict with learned representations.
4. **Computational and Deployment Challenges**: High-resolution volumetric reconstruction demands substantial GPU resources, as observed in Pix2Vox++ and Persistent Nature, which require multi-GPU setups for training and inference [15, 13]. This computational burden poses barriers to real-time applications such as interactive VR/AR systems, while integrating ML pipelines with rendering engines remains an open engineering challenge.

6. Comparative Analysis of State of the Art Methods

Table 1 provides a comprehensive overview of recent advances in 2D to 3D scene reconstruction and generation, comparing methodologies, results, and limitations across different approaches.

Table 1. State-of-the-art in 2D to 3D Scene Reconstruction and Generation

Ref.	Purpose	Methodology	Results	Limitations	Hardware
[1]	Convert residential floorplans and interior photos into textured 3D mesh models	6-step pipeline: Floorplan vectorization, 3D geometry construction, object placement, photo rectification, texture generation, GNN propagation	In Plan2Scene's case, User study shows superior realism and accuracy over baselines	Semantic segmentation errors, lighting issues, limited to 3 surface types	PyTorch + PyTorch Geometric, GPU (V100/A100)
[2]	Generate 3D vector models from 2D floor plan images using ensemble methods	Hybrid approach: Pattern recognition, object detection (TensorFlow API), node/edge generation, plan scaling	Wall accuracy: 95.3%, Junction accuracy: 92.2%, Generated 110k+ models	Manual correction needed, weak object detection with minimal training data	3DPlanNet experiment's hardware : i7-875H 2.20GHz notebook with GPU
[10]	Automate 2D floor plan digitalization into semantic 3D BIM models	ResNet backbone, FPN segmentation, Faster-RCNN detection, IFC-compatible output	IoU: 0.81 (mask), 0.8 (vectorized), SOTA on CubiCasa5k	Limited room annotations, assumed height estimation, lack of public code	Not specified
[3]	Generate interactive 3D virtual house from single floorplan for VR experience	U2-Net segmentation, 3D mesh construction, VR framework with furniture library	15s generation time, 70k iterations training (55h), interactive VR immersion	Limited furniture types, manual labeling, estimated window properties	i7-10750H, 16GB RAM, RTX 2070, HTC Vive Pro
[4]	Convert 2D floor plans to interactive 3D models using image processing and AR	Image processing for wall detection, Unity3D + Vuforia for AR rendering, Android GUI	76.70% accuracy in wall recognition and reconstruction	Accuracy depends on plan clarity, limited to plan complexity	Android (Flutter 3.3.7), Node.js on Amazon EC2, Python 3.10
[11]	Synthesize realistic 3D towns from single top-down image	Modular decomposition, pretrained rectified flow transformer, landmark initialization, masked latent inpainting	Geometry win rate: 68.5-80% (human), 82-92% (GPT-4o), Texture win rate: up to 92.3%	Duplicated façades, coarse depth maps, lacks scene-level fine-tuning	NVIDIA RTX A5000 (24GB), Trellis, Florence2, SAM2
[17]	Personalized LLM interior designer for 3D room layouts from natural language	I-Design uses Multi-agent reasoning (AutoGen), procedural scene graph, VLM evaluation, Objaverse retrieval	0% out-of-bound objects vs 57.6% baseline, GPT-4V rating: 5.7/10 vs 4.8, 3x more objects placed	Dense scene failures, asset mismatches, lacks retexturing, slow backtracking	GPT-4, AutoGen, OpenShape, CLIP, Blender
[21]	Unsupervised single-view 3D scene extrapolation with diffusion models	Conditional diffusion with corrupted/ground-truth RGBD, anchored conditioning, lookahead conditioning	PSNR: 23.56, SSIM: 0.68, 2x COLMAP points (3124 vs 1476), consistent fly-through videos	DiffDreamer has faced Slow inference, limited content diversity, struggles with complex terrains	2× NVIDIA RTX 8000, PyTorch3D, Palette, 1 week training
[20]	Reconstruct outdoor buildings as planar graphs from single RGB image	Faster-RCNN corner detection, graph formulation, convolutional message passing, edge classification	Region F1-score: 54.2, outperforms PolyRNN++, handles non-Manhattan geometries	Memory intensive, limited iterations (T=3), fails on large buildings	Conv-MPN experiment was built on 2× NVIDIA TitanX (24GB), PyTorch, DRN, 20-40h training
[15]	Reconstruct 3D voxel objects from single/multiple uncalibrated images	Parallel processing, weighted multi-scale fusion, sigmoid voxel occupancy, U-Net refiner	Pix2Vox++ scored ShapeNet IoU: 0.670 (1 view), 0.843 (8 views), 7x faster than 3D-R2N2	Memory intensive at high resolution, no camera parameters, poor generalization	NVIDIA GTX 1080 Ti, PyTorch, 15 servers for training

Real-world deployment reveals significant barriers limiting widespread adoption. Computational requirements present substantial obstacles, with methods like DiffDreamer [21] experienced ”slow inference” and demanding high-end GPU configurations shown in Table 1, significantly increasing deployment costs. Manual intervention requirements further constrain deployment, as 3DPlanNet's [2] commercial use required manual correction of ”insufficient parts” despite 95% wall accuracy, while Plan2Scene [1] faced semantic segmentation errors requiring human intervention for production use.

7. Future Research Directions

As machine learning continues to evolve, emerging techniques present promising avenues to overcome current limitations in 2D-to-3D floorplan conversion.

7.1. Diffusion Models for 3D Scene Generation

Diffusion models demonstrate remarkable capabilities in generating consistent 3D scenes from single-view inputs. DiffDreamer showed how conditional diffusion models enable unsupervised 3D scene extrapolation with improved neural field reconstructions [21]. However, computational intensity remains a barrier for real-time architectural applications, necessitating lightweight architectures for practical deployment.

7.2. Explainable AI for Architectural Design Transparency

Explainable AI (XAI) techniques can significantly enhance architect-designer workflows by providing visual explanations of spatial reasoning decisions in 2D-to-3D conversion. XAI methods could generate attention maps highlighting critical floorplan features, visualize reconstruction confidence scores, and explain why specific 3D interpretations were chosen over alternatives [22]. This transparency would enable architects to validate model decisions, identify potential errors, and iteratively refine designs with greater confidence in automated reconstruction tools.

7.3. Foundation Models and Vision-Language Integration

Foundation models such as Florence-2 and GPT-4V enabled integration of spatial understanding with semantic reasoning for 3D reconstruction [23, 24]. The I-Design framework demonstrated multi-agent LLM systems for procedural 3D room layout generation from natural language inputs [17]. This approach could revolutionize architectural workflows by enabling conversational design interfaces and supporting diverse architectural styles through pretrained knowledge.

7.4. Few-Shot Learning for Data-Efficient Reconstruction

Few-shot learning techniques reduce reliance on large annotated datasets by enabling adaptation to new floorplan styles with minimal supervision [25]. Zheng et al.'s modular approach to 3D town synthesis illustrated benefits of pretraining transformer models on limited domain-specific data [11]. This strategy is particularly valuable for regions with scarce architectural datasets and non-standard building conventions.

7.5. Neuro-Symbolic Approaches for Constraint Satisfaction

Neuro-symbolic systems can integrate building code compliance checking with generative reconstruction pipelines [26]. Recent research demonstrates AI systems that analyze generated layouts for regulatory violations, such as excessive Maximum Travel Distances (MTD), and trigger automatic redesign when non-compliance is detected. This post-generation validation approach combines neural flexibility in 3D generation with symbolic reasoning for code compliance, ensuring reconstructed models meet both aesthetic and regulatory requirements.

8. Concluding Remarks

This paper has presented a comprehensive review of machine learning techniques for converting 2D floorplan images into 3D models, highlighting their applications, strengths, and limitations. CNNs have proven effective for structural feature extraction and semantic segmentation, while GANs excelled in producing high-fidelity textures and volumetric reconstructions. Transformer-based architectures and hybrid methods further enhanced global context modeling and flexibility in handling diverse layouts. Despite these advancements, significant challenges remain, including limited availability of large-scale annotated datasets, computational complexity, and difficulties in ensuring semantic consistency across varied architectural styles. Emerging techniques such as diffusion models, foundation models, and neuro-symbolic AI offer promising directions to address these challenges and enable more robust, scalable, and user-centric 3D reconstruction pipelines. Future work should focus on developing data-efficient and lightweight architectures that support real-time applications, facilitating broader adoption in real estate visualization, virtual reality, and smart urban planning.

Acknowledgments

This work has been realized with financial support by the European Regional Development Fund within the Operational Programme “Bulgarian national recovery and resilience plan”, Procedure for direct provision of grants “Establishing of a network of research higher education institutions in Bulgaria”, under the Project BG-RRP-2.004-0005 “Improving the research capacity and quality to achieve international recognition and resilience of TU-Sofia”.

References

[1] M. Vidanapathirana, Q. Wu, Y. Furukawa, A. X. Chang, and M. Savva. Plan2scene: Converting floorplans to 3d scenes. In *CVPR*, pages 10733–10742. IEEE, 2021. DOI: 10.1109/CVPR46437.2021.01059.

[2] Sungsoo Park and Hyeoncheol Kim. 3dplannet: Generating 3d models from 2d floor plan images using ensemble methods. *Electronics*, 10(22), 2021. DOI: 10.3390/electronics10222729.

[3] H. Fu and M. Makino. A vr-based indoor visualization system from floorplan images with deep learning. In *IWAIT 2022*, page 72. SPIE, 2022. DOI: 10.1117/12.2625978.

[4] P. Deshmukh, S. Kulkarni, D. Samuel, et al. 2D to 3D Floor plan Modeling using Image Processing and Augmented Reality, may 2023. DOI: 10.1109/REEDCON57544.2023.10151165.

[5] M. A. Elhamy Kamel and M. W. Ibrahim Khalil. The impact of using computer-aided design (cad) on the creativity of architecture students, 7 2023. DOI: 10.47750/jett.2023.14.04.021.

[6] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9:611–629, 2018. DOI: 10.1007/s13244-018-0639-9.

[7] D. Kim, S. H. Kang, S.-Y. Lim, et al. Review on generative adversarial networks: Focusing on computer vision and its applications. *Electronics*, 10(10):1216, 2021. DOI: 10.3390/electronics10101216.

[8] I. D. Mienye, T. G. Swart, and G. Obaido. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9):517, 2024. DOI: 10.3390/info15090517.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017. DOI: 10.5555/3295222.3295349.

[10] A. C. Barreiro, M. Trzeciakiewicz, A. Hilsmann, and P. Eisert. Automatic reconstruction of semantic 3d models from 2d floor plans. In *MVA 2023*, pages 1–5. IEEE, 2023. DOI: 10.23919/MVA57639.2023.10215746.

[11] K. Zheng, R. Zhang, J. Gu, J. Yang, and X. E. Wang. Constructing a 3D Town from a Single Image, May 2025. DOI: 10.48550/arXiv.2505.15765.

[12] A. Kalervo, J. Ylioinas, M. Häikiö, A. Karhu, and J. Kannala. Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis. In *SCIA 2019*, pages 28–40. Springer, 2019.

[13] L. Chai, R. Tucker, Z. Li, P. Isola, and N. Snavely. Persistent nature: A generative model of unbounded 3d worlds. In *CVPR*, pages 20863–20874. IEEE, 2023. DOI: 10.1109/CVPR52729.2023.01999.

[14] Seung Won Lee and Min Ji Koo. PRISMA 2020 statement and guidelines for systematic review and meta-analysis articles, and their underlying mathematics: Life Cycle Committee Recommendations. *Life Cycle*, 2, 2022. Publisher: Life Cycle.

[15] H. Xie, H. Yao, S. Zhang, S. Zhou, and W. Sun. Pix2Vox++: Multi-scale Context-aware 3D Object Reconstruction from Single and Multiple Images. *International Journal of Computer Vision*, 128(12):2919–2935, dec 2020. DOI: 10.1007/s11263-020-01347-6.

[16] W. Para, P. Guerrero, T. Kelly, L. Guibas, and P. Wonka. Generative layout modeling using constraint graphs. In *ICCV*, pages 6670–6680. IEEE, 2021. DOI: 10.1109/iccv48922.2021.00662.

[17] A. Çelen, G. Han, K. Schindler, L. Van Gool, I. Armeni, A. Obukhov, and X. Wang. I-design: Personalized llm interior designer, 2024. DOI: 10.48550/arXiv.2404.02838.

[18] W. Wu, X.-M. Fu, R. Tang, Y. Wang, Y.-H. Qi, and L. Liu. Data-driven interior plan generation for residential buildings. *ACM Transactions on Graphics*, 38(6):118:1–118:12, 2019. DOI: 10.1145/3355089.3356556.

[19] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, and Y. Furukawa. House-gan: Relational generative adversarial networks for graph-constrained house layout generation. In *ECCV 2020*, pages 162–177. Springer, 2020. DOI: 10.1007/978-3-030-58452-8_10.

[20] F. Zhang, N. Nauata, and Y. Furukawa. Conv-mpn: Convolutional message passing neural network for structured outdoor architecture reconstruction. In *CVPR*, pages 2795–2804. IEEE, 2020. DOI: 10.1109/cvpr42600.2020.00287.

[21] S. Cai, E. R. Chan, S. Peng, M. Shahbazi, A. Obukhov, L. Van Gool, and G. Wetzstein. Diffdreamer: Towards consistent unsupervised single-view scene extrapolation with conditional diffusion models. In *ICCV*, pages 2139–2150. IEEE, 2023. DOI: 10.1109/ICCV51070.2023.00204.

[22] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, et al. Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020. DOI: 10.1016/j.inffus.2019.12.012.

[23] B. Xiao, H. Wu, W. Xu, X. Dai, H. Hu, Y. Lu, M. Zeng, C. Liu, and L. Yuan. Florence-2: Advancing a unified representation for a variety of vision tasks, 2023. DOI: 10.48550/arXiv.2311.06242.

[24] OpenAI, J. Achiam, S. Adler, S. Agarwal, et al. Gpt-4 technical report, 2024. DOI: 10.48550/arXiv.2303.08774.

[25] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3):1–34, 2020. DOI: 10.1145/3386252.

[26] F. Yang and J. Zhang. Prompt-based automation of building code information transformation for compliance checking, 2024. DOI: 10.1016/j.autcon.2024.105817.