International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

A Literature Survey on Room Analysis and Renovation Visualization Using AI

Daisi M[1], Divya C[2], Kavyashri L H[3], Nagashri Narayan Upadhya[4], Dr. Kavita K. Patil[5]

[1,2,3,4]Students, [5]Professor, Department of Information Science and Engineering, Global Academy of Technology, Bengaluru, India

***Abstract*—Maintaining an indoor space involves more than improving its appearance. Cracks, moisture marks, peeling paint, mold, and other surface problems can gradually affect the condition of a room and may require timely attention. Conventional inspection generally depends on visual checking by a person, which can be slow and may produce different observations for the same room.**

This survey examines how Artificial Intelligence (AI) can support different stages of room inspection and renovation planning. The reviewed studies cover image classification and damage recognition using convolutional neural networks, object localization using YOLO-based methods, semantic segmentation, transformer-based image understanding, and generative models for producing interior design alternatives. Approaches such as Stable Diffusion and ControlNet are considered particularly relevant because they can create modified room views while using information from the original image to maintain its overall structure. Recommendation techniques are also reviewed for their potential role in selecting materials and interior styles.

The comparison shows that these technologies are generally developed for separate purposes. Damage detection systems identify defects, generative models create visual alternatives, and recommendation systems assist with selection. Only limited work connects these stages into a continuous room analysis and renovation workflow. This observation points toward the need for an integrated approach in which inspection results can directly support renovation visualization and subsequent recommendations.

I. Introduction

The condition of an indoor space can change considerably over time. Everyday use, moisture, poor ventilation, aging materials, and environmental exposure may result in problems such as cracks, peeling paint, damp patches, mold, and other forms of surface deterioration. Identifying these issues early is useful because a small visible defect can sometimes indicate a larger maintenance requirement.

In many practical situations, inspection is still performed by visually examining the room. Such an approach depends strongly on the experience of the person carrying out the inspection. It can also become inconvenient when several rooms have to be examined or when the defects are small and difficult to distinguish from normal surface variations.

Image-based Artificial Intelligence provides an alternative by allowing a photograph to be processed automatically and examined for visual patterns associated with different types of damage.

Convolutional Neural Networks (CNNs) established an important foundation for automated image analysis by learning useful visual representations directly from image data [1], [2]. Later architectures such as VGGNet and ResNet demonstrated how deeper networks could improve feature representation and classification performance [3], [4]. For room inspection, these developments are relevant because defects may differ considerably in shape, size, texture, and appearance.

Classification alone, however, does not indicate where a defect occurs in an image. Object detection methods such as YOLO and Faster R-CNN address this limitation by locating visual objects within an image [5], [6]. YOLO-based approaches are particularly attractive for applications where detection speed is important, while segmentation methods can provide more detailed information about the affected pixels [7], [8].

Another branch of recent research has changed the way renovation concepts can be presented to users. Instead of preparing a design manually, generative models can create a possible appearance of the same space after changes have been made. Generative adversarial networks introduced an important direction for image synthesis [9], while conditional image-to-image translation demonstrated how an input image could guide the generation of a related output [10]. Diffusionbased approaches subsequently achieved strong results in highquality image synthesis. In particular, latent diffusion models provide an efficient basis for generating detailed images [11], and ControlNet adds structural conditioning that can help retain information from the source image [12].

Transformer-based models provide another perspective on visual analysis. The original Transformer introduced an attention-based architecture [13], and the Vision Transformer demonstrated that a similar principle can be applied directly to image recognition [14].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

These models are relevant to room analysis because understanding an indoor scene often requires relationships between different regions rather than examining every local feature independently.

Recommendation methods can complement these visual techniques. Traditional recommender systems have used collaborative and content-based approaches to select items according to user preferences and available information [15], [16]. In a renovation setting, similar ideas could be used to connect detected room conditions with appropriate paint types, flooring choices, furniture arrangements, or interior styles.

Although the individual technologies have progressed considerably, the literature reveals a separation between inspection and renovation visualization. A system may identify a crack without showing how the room could look after repair, while an image-generation system may create a new interior without first considering the actual defects in the input image. Recommendation methods are also commonly treated as a separate component. This separation motivates the study of an integrated AI-based approach to room analysis and renovation visualization.

The purpose of this literature survey is therefore to examine the major techniques used across these related areas, compare their capabilities and limitations, and determine where existing research leaves room for an integrated solution.

II. Literature Review

Research relevant to AI-assisted room renovation can be grouped into four closely related areas: visual damage detection, defect localization and segmentation, image generation, and recommendation. Studying these areas separately helps explain both the progress already achieved and the difficulty of combining them into one workflow.

A. CNN-Based Damage Analysis

CNNs have played a central role in computer vision because they can learn visual patterns directly from training images. The early work of LeCun et al. established the use of convolution-based learning for image recognition [1]. AlexNet later demonstrated the practical advantages of deep convolutional models on large image datasets [2]. VGGNet increased network depth to obtain richer representations [3], while ResNet introduced residual connections that made the training of deeper networks more practical [4].

These developments are relevant to structural inspection because damage patterns are not uniform. A crack may appear as a thin line, whereas dampness or peeling paint may occupy a larger irregular region.

Consequently, a useful inspection model must learn differences in texture, shape, and spatial appearance rather than relying on a fixed visual rule.

Park et al. investigated deep learning for concrete crack detection and quantification, showing the usefulness of learned image features for structural defect analysis [17]. However, concrete-focused datasets do not necessarily represent the appearance of defects in residential rooms. Indoor images introduce additional variation from lighting, furniture, wall finishes, shadows, and camera position.

B. Object Detection and Segmentation

Classification identifies the category associated with an image or region, but practical inspection often requires the location of the problem as well. R-CNN introduced a regionbased strategy for object detection [18], followed by Faster R-CNN, which incorporated region proposal generation into a unified detection framework [6]. YOLO approached the problem differently by predicting objects in a single detection pipeline, making real-time applications more practical [5]. Later versions such as YOLOv3 and YOLOv4 improved the balance between detection quality and computational speed [19], [20].

Segmentation provides still finer information because it assigns labels at the pixel level. U-Net demonstrated the effectiveness of encoder-decoder segmentation for detailed region identification [7], while Fully Convolutional Networks established an end-to-end approach to semantic segmentation [8]. Such techniques could be useful when the exact extent of peeling paint, dampness, or a crack needs to be estimated rather than simply enclosed by a bounding box.

C. Generative Models for Renovation Visualization

Image generation has developed from adversarial approaches toward diffusion-based models. Generative Adversarial Networks demonstrated that two competing networks could be used to synthesize realistic images [9]. Pix2Pix extended conditional image generation to image-to-image translation using paired examples [10]. These ideas are relevant to renovation because the input room photograph can provide information that guides the desired output.

Diffusion models introduced another powerful route to image synthesis. Denoising Diffusion Probabilistic Models showed how an image could be generated through a gradual denoising process [21]. Latent Diffusion Models moved this process into a lower-dimensional representation, making highresolution synthesis more practical [11]. ControlNet further demonstrated how structural conditions can be incorporated into diffusion-based generation [12].

For room renovation, structural conditioning is important because an attractive generated image is not sufficient if walls, doors, windows, or major furniture positions are changed unintentionally.

D. Transformers and Visual Understanding

Attention-based architectures provide a different way of modelling visual information. The Transformer architecture introduced self-attention as a central mechanism for learning relationships between elements in a sequence [13]. The Vision Transformer subsequently applied this idea to image patches and showed that transformer-based image recognition can achieve strong performance [14].

For indoor scenes, contextual relationships are important. For example, a wall, window, floor, and furniture item cannot always be interpreted independently. Transformer-based representations may therefore complement CNN-based approaches when broader scene context is required.

E. Explainable AI and Recommendation

A prediction is more useful in an inspection application when the user can understand why a model reached that decision. Grad-CAM provides a visual explanation by identifying image regions that contribute strongly to a CNN prediction [22]. Such explanations may be useful when displaying detected damage to a homeowner or maintenance professional.

Recommendation systems provide a separate but complementary capability. Matrix factorization became a widely used approach for collaborative recommendation [16], while Adomavicius and Tuzhilin reviewed several recommendation paradigms and their possible extensions [15]. Within renovation planning, these techniques could be adapted to relate user preferences and room conditions to suitable materials or design alternatives.

F. Overall Observation

The literature shows strong progress in each individual component. CNNs and related architectures support visual classification, detection models provide spatial localization, segmentation gives detailed region information, and generative models make virtual design alternatives possible. Transformers contribute broader visual representation, while explainability and recommendation methods can improve the usability of AI-assisted systems.

The main limitation is that these capabilities are usually studied separately. The reviewed work provides a strong technical foundation, but there is still scope for connecting room inspection results directly with renovation generation and material recommendations.

III. Comparative Analysis

This section compares the major research papers reviewed in this survey. The comparison is based on the methodology used, the advantages of each approach, its limitations, and the research gap identified.

The comparison in Table I shows that no single method addresses every requirement of an AI-assisted renovation workflow. CNN and ResNet-based approaches are useful when the primary task is recognizing damage from images. Detection models add spatial information, which is important when a user needs to know where a defect is located. Segmentation methods go a step further by describing the affected area at pixel level.

For visualization, GAN-based and diffusion-based methods offer a different set of capabilities. They can transform the appearance of an input scene, but the quality of the final result depends strongly on the conditioning information and the generation process. ControlNet is particularly relevant to room renovation because structural guidance can reduce unwanted changes to the original scene.

Recommendation approaches address yet another part of the problem. They can help users choose among possible materials or styles, but conventional recommendation models do not normally receive structural damage information from an image-analysis module.

Taken together, the studies suggest that the most promising direction is not simply selecting one existing algorithm. Instead, an effective room renovation system would need to connect complementary techniques so that the output of inspection can influence visualization and recommendations.

IV. Research Gap

The reviewed studies provide useful solutions for individual stages of room analysis, but they leave a clear separation between inspection, visualization, and recommendation.

First, damage-oriented approaches generally concentrate on recognizing or locating defects. Their output is useful for inspection, but it does not normally answer the next question faced by a user: how could the room look after the problem has been addressed?

Second, image-generation research focuses primarily on producing visually convincing results. Although conditional generation and structural guidance can preserve important information from an input image, these models are not generally connected to an automated damage-assessment stage.

Third, recommendation systems are usually designed around user preferences, historical choices, or item relationships. In a renovation scenario, useful recommendations should ideally consider both what the user wants and what the room actually requires.

The literature therefore points to an integration gap. A practical system would need to pass information between these stages: an uploaded room image should first provide evidence about its condition, the detected information should influence the renovation visualization, and the resulting context should support recommendations for materials and design choices.

Another challenge is maintaining the identity of the original room during image generation. A useful renovation preview should modify appearance without unnecessarily changing important structural elements such as walls, doors, windows, and major spatial relationships.

Hence, the research opportunity lies in developing a connected workflow rather than another isolated detection or generation model. The combination of visual damage assessment, spatial localization, structure-aware renovation generation, and context-sensitive recommendation represents the main direction identified from this survey.

V. Conclusion

This survey examined how AI techniques are being applied to problems that arise during room inspection and renovation planning. The reviewed work covers convolutional models for image analysis, detection and segmentation methods for locating defects, transformer-based visual representations, generative models for producing interior alternatives, and recommendation techniques for supporting user choices.

The comparison suggests that these technologies are individually capable, but they are rarely connected into one continuous workflow. Damage recognition can describe the condition of a room, while generative models can illustrate a possible redesign; however, the link between these two stages is still limited.

A useful direction for future research is therefore an integrated system that uses the detected condition of a room to guide renovation visualization and subsequent recommendations. Such a system could make AI-assisted renovation planning more practical by allowing users to move from inspection to visual exploration and material selection within the same workflow.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

TABLE I COMPARATIVE ANALYSIS OF EXISTING RESEARCH

Ref.	Paper	Method	Advantages	Limitations	Research Gap
[1]	Gradient-Based Learning Applied to Document Recognition	CNN	Introduced convolutional neural networks for image recognition.	Limited to simple image datasets.	Does not address building damage detection or renovation visualization.
[2]	ImageNet Classification with Deep CNN	AlexNet	Improved classification accuracy using deep CNNs.	Requires high computational resources.	Not designed for indoor room analysis.
[3]	Very Deep Convolutional Networks for Large-Scale Image Recognition	VGGNet	Learns detailed image features using deeper networks.	Large number of parameters.	No localization of damaged regions.
[4]	Deep Residual Learning for Image Recognition	ResNet	Solves degradation problem in deep networks.	Computationally expensive.	Does not provide renovation suggestions.
[5]	Concrete Crack Detection and Quantification Using Deep Learning	CNN	Accurate crack detection under different conditions.	Mainly focuses on concrete structures.	No visualization of repaired output.
[6]	You Only Look Once	YOLO	Performs real-time object detection.	Small objects may be missed.	Does not generate renovation previews.
[7]	YOLOv3: An Incremental Improvement	YOLOv3	Better multi-scale object detection.	Less accurate than recent versions.	Focuses only on detection.
[8]	YOLOv4: Optimal Speed and Accuracy of Object Detection	YOLOv4	High speed with improved detection accuracy.	Requires GPU for best performance.	No recommendation module.
[9]	Faster R-CNN	Faster R-CNN	Accurate object localization.	Slower than YOLO.	Not suitable for real-time mobile systems.
[10]	U-Net: Convolutional Networks for Biomedical Image Segmentation	U-Net	Precise pixel-level segmentation.	Developed mainly for medical images.	Needs adaptation for structural damage detection.
[11]	High-Resolution Image Synthesis with Latent Diffusion Models	Stable Diffusion	Generates realistic highresolution images.	Computationally intensive.	Does not analyze room damage automatically.
[12]	Adding Conditional Control to Text-toImage Diffusion Models	ControlNet	Preserves structural layout while generating images.	Depends on pretrained diffusion models.	No automatic damage identification.
[13]	Attention Is All You Need	Transformer	Captures long-range dependencies effectively.	Requires large datasets and computation.	General architecture; not developed for room analysis.
[14]	An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale	Vision Transformer (ViT)	Achieves strong image recognition using selfattention.	Requires large-scale datasets for training.	Not specialized for structural damage analysis.
[15]	Denoising Diffusion Probabilistic Models	Diffusion Model	Produces high-quality image synthesis with stable training.	High computational cost during inference.	Does not preserve room layout during renovation.
[16]	Matrix Factorization Techniques for Recommender Systems	Collaborative Filtering	Provides accurate personalized recommendations.	Suffers from cold-start problem.	Not integrated with computer vision applications.
[17]	Toward the Next Generation of Recommender Systems	Hybrid Recommendation	Combines multiple recommendation techniques.	Increased system complexity.	Does not supportrenovation planning.
[18]	Generative Adversarial Nets	GAN	Generates realistic synthetic images.	Difficult to train due to instability.	No room damage detection capability.
[19]	Image-to-Image Translation with Conditional GANs	Pix2Pix	Converts input images into realistic output images.	Requires paired training datasets.	Limited renovation customization.
[20]	Grad-CAM: Visual Explanations from Deep Networks	Grad-CAM	Improves interpretability of CNN predictions.	Provides explanation only after prediction.	Does not perform damage localization independently.
[21]	Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation	R-CNN	Introduced deep-learningbased object detection.	Slow inference speed.	Unsuitable for real-time inspection.
[22]	Fully Convolutional Networks for Semantic Segmentation	FCN	Enables end-to-end semantic segmentation.	Lower boundary precision than recent methods.	Needs improvement for complex indoor scenes.
[23]	Pixel Recurrent Neural Networks	PixelRNN	Learns pixel-level image distributions effectively.	Very slow image generation.	Not applicable to renovation visualization directly.
[24]	Densely Connected Convolutional Networks	DenseNet	Encourages feature reuse and reduces parameters.	High memory consumption.	Not specifically designed for room analysis.
[25]	The Cityscapes Dataset for Semantic Urban Scene Understanding	Cityscapes Dataset	Provides benchmark dataset for segmentation research.	Focused on outdoor urban scenes.	Cannot directly support indoor room damage analysis.

REFERENCES

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, 2012.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779– 788.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems, vol. 28, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431– 3440.

[9] I. Goodfellow et al., "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014.

[10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1125–1134.

[11] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "Highresolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684–10695.

[12] L. Zhang and M. Agrawala, "Adding conditional control to text-toimage diffusion models," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 3836–3847.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in International Conference on Learning Representations (ICLR), 2021.

[15] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[16] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30–37, 2009.

[17] S. E. Park, S. H. Eem, and H. Jeon, "Concrete crack detection and quantification using deep learning and structured light," Construction and Building Materials, vol. 252, p. 119096, 2020.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.

[19] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[21] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in Neural Information Processing Systems, vol. 33, 2020.

[22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.