An AI-based system offering automatic DR-enhanced AR for indoor scenes

Georgios, G. A, Albanis

Information Technologies Institute / Centre for Research and Technology, galbanis@iti.gr

Vasileios, V. G, Gkitsas

Information Technologies Institute / Centre for Research and Technology, gkitsasv@iti.gr

Nikolaos, N. Z, Zioulis

Information Technologies Institute / Centre for Research and Technology, nzioulis@iti.gr

Stefanie, S. O, Onsori-Wechtitsch

JOANNEUM RESEARCH, Stefanie.Onsori-Wechtitsch@joanneum.at

Richard, R. W, Whitehand

Usability Partners, richard.whitehand@usabilitypartners.se

Per, P. S, Ström

Usability Partners, per.strom@usabilitypartners.se

Dimitrios, D. Z, Zarpalas

Information Technologies Institute / Centre for Research and Technology, zarpalas@iti.gr

Figure 1: Imagine that you want to redesign your living space and replace existing furniture with new ones. Most of us are not so creative, and conceiving how a place would look is challenging. AR technology allows for inserting virtual objects in real environments, and can thus, assist in better understanding how new furniture would fit in our place.

However, this won't work in already furnished environments. These existing objects will disrupt the AR experience, and in most cases will have the opposite of the intended effect, leaving the users confused. We propose a system consisting of various AI services for enabling next-generation AR indoor (re-)planning and design experiences. Our system strives for user-friendliness and bypasses error-prone and cumbersome user scanning processes. Instead, users only require a single 360° camera capture that produces a spherical panorama of their indoor space. Then, our AI-based system automatically generates a high-level understanding of the scene, both semantically and structurally, enabling automatic selection of objects to be removed or replaced. This is driven by employing DR technology that incorporates the inferred scene structural prior to generate plausible hallucinations, eventually offering a compelling and effective AR experience. Top row shows the overall concept and higher level component connections, while the bottom row shows an actual example from the Structured3D dataset, where a bed is replaced within a room. It is worth noticing that apart from the virtual bed placement, all other steps are automatic, that the diminished room with the original bed removed respects the room's layout, and also that a significantly large object was replaced

In this work, we present an AI-based Augmented Reality (AR) system for indoor planning and refurbishing applications. AR can be an important medium for such applications, as it facilitates more effective concept conveyance, and additionally acts as an efficient and immediate designer-to-client communication channel. However, since AR only overlays, and cannot replace or remove, our system relies on Diminished Reality (DR) to support deployment to real-world already furnished indoor scenes. Further, and contrary to the traditional mobile AR application approach, our system offers on demand Virtual Reality (VR) viewing, relying on spherical (360°) panoramas, capitalizing on their user-friendliness for indoor scene capturing. Given that our system is an integration of different AI services, we analyze its performance differentials concerning the components comprising it. This analysis is both quantitative and qualitative, with the latter realized through user surveys, and provides a complete systemic assessment of an attempt for a user-facing, automatic AR/DR system.

CCS CONCEPTS • **Human-centered computing ~ Human-Computer Interaction (HCI) ~ Interaction paradigms ~ Mixed / augmented reality**

Additional Keywords and Phrases: Augmented Reality, indoor AR, Diminished Reality, Artificial intelligence, User Study, 360º panoramas, Image Inpainting, Computer Vision

1 INTRODUCTION

Interior design can be a challenging and stressing process, requiring bidirectional communication between users and experts. Experts usually express their ideas in traditional 2D drawings produced by Computer Aided Design (CAD) software, making it difficult for the end-users to comprehend them. AR is an emerging technology that allows users to superimpose Computer-Generated (CG) elements over the real world. In the particular case of interior design, AR can be used for placing virtual 3D objects within the real environment bridging the communication gap between experts (designers) and non-experts (clients). In this way, AR serves as a medium between digitized concepts and the real scene, facilitating effective and efficient communication and feedback between its users, improving the iterative design process. Indicatively, an AR system for rearranging a furniture layout was proposed in [15], while in [12] a system employing a dynamic user interface for placing 3D virtual furniture models was developed. However, both aforementioned systems required multiple QR markers to allow users to physically position the virtual furniture.

Even though AR enables the interaction with virtual objects inside real environments, its nature is pure of additive nature, with a practical problem befalling when working in occupied and filled indoor scenes as is the case for AR home design applications [20]. Concepts like redecoration cannot be delivered solely through AR technology, as users would only be capable of superimposing CG elements on top of existing real-world objects,

hindering understanding due to a conflicting mental response. To overcome this, AR needs to be supported by DR which can diminish existing objects prior to overlaying new virtual ones, and provide users with an enhanced view to assess furniture fit into their spaces. DR is an intriguing technology that can enable novel concepts. One example is intercar see-through vision, which aims at preventing accidents [14] and diminishes (i.e. ”removes”) the front car. In this particular case, DR is driven by multi-view observations and view synthesis. There are cases though, where no view behind the removed object is available, and then DR needs to hallucinate content, typically referred to as infilling or inpainting [10]. Pioneering work in the DR domain was presented by [5], where a patch-based image inpainting method was developed. Follow up work [8] moved beyond image based diminishing and transitioned towards respecting scene geometry by exploiting SLAM-based localisation. More recently, an inpainting method for non-planar scenes was developed [16] that considered both color and depth information. Still, in both cases, manual selection of the region to be removed in the image domain was required. To allow for easier selection of the object to be diminished in indoor scenes for interior design, [17] used a manually positioned and scaled volume to enclose the object of interest. In addition, the floor plane was identified by inserting a marker into the scene. Real-time 6 degrees-of-freedom DR without manual object selection is challenging [13], and requires a 3D reconstruction of the scene without the object of interest but with the diminishing area annotated, limiting its flexibility. When considering AR interior home refurnishing, where quickly prototyping ideas is very important, minimizing interactions is very important, as users will also need to position the new elements into the scene as well [7].

Still, all of the aforementioned studies work in a narrow field of view inputs, limiting the amount of information of each scene and thus degrading their performance on big objects (e.g.furniture), while at the same time they do not strictly respect the structure of the environment. To overcome this, moving cameras are employed relying on SLAM [16] or wider field of view captures [7], but they limit user friendliness and are more error-prone. In this work, we present a system that addresses the challenges of cumbersome user diminished area selection and user scanning, delivering DR-enhanced, AR for indoor scene planning and design. To achieve that, our system is AI-based, operating on a single monocular image capture, exploiting recent advances in data-driven inpainting methods [6]. In addition, albeit image-based, it takes the scene’s structure into account, an important cue for the targeted application domain. Our main contributions are summarised below:

- A novel AI-based DR-enhanced AR system with various data-driven components connected in parallel and cascade structure using only monocular 360° images as input.
- A holistic system evaluation including a systemic point of view analysis to identify the weakest link in the system, and a user-study focused on the importance and relevance of DR in indoor planning applications.

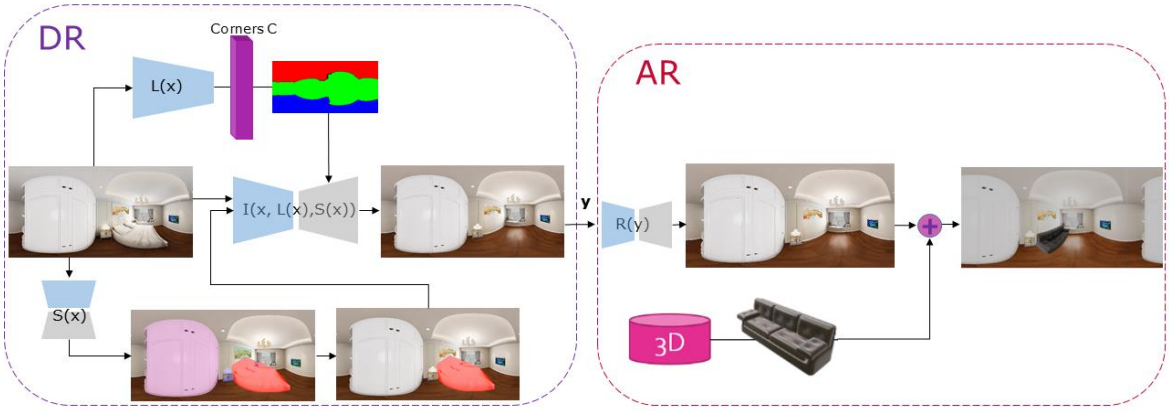

*Figure 2: Overview of the proposed automatic DR-enhanced 360o AR system. The system can be dissected into two high-level sub-systems, the **DR** one the left and the **AR** one the right, operating in cascade. The former is responsible for the automatic diminishing of the scene and the latter for user-driven augmentation. For a given scene, captured by a 360o camera, the input panorama x is fed to the system. The scene's junctions L(x) and objects' masks S(x) are first estimated in parallel by the corresponding data-driven components, with L and S being the layout and segmentation AI models respectively. Then, for each separate object mask segmented in the scene, the data-driven inpainting component is invoked I(x,L(x),S(x)), with I being the respective AI model. Diminishing is achieved by inpainting the object's mask in a structure-aware manner using the dense layout map that facilitates the preservation of the wall-ceiling/floor boundaries. The result y is a diminished scene for that particular object. Subsequently, the result is up-sampled by invoking R(y), where R is a super-resolution model, and the higher resolution panorama is ready to be augmented by the virtual 3D object that the user positions into the scene, producing the DR-enhanced AR 360o image z.*

2 SYSTEM OVERVIEW

Figure 2 shows a high-level overview of our system comprising two main sub-systems, and the nominal data flow among the various components. Each component is an AI model, trained on the Structured3D dataset [23].

As presented in Figure 2 the two sub-systems operate in cascade, while the DR sub-system also includes a parallel component connection. The DR sub-system first processes the input panoramic image by estimating the scene's layout and segmenting the distinct objects inside the scene. Then, for each segmented object in the scene, the inpainting component is invoked to diminish the object and prepare the input for the AR superimposition. Since data-driven models typically operate in lower resolutions than required for panorama viewing, the AR sub-system first invokes a super-resolution component to rescale the diminished area back to 360° viewing resolution. AR is user-driven by positioning elements into the scene that interact with the masked regions depending on their projection to select the appropriate diminished panorama. Still, users may simply require to remove an object from the scene which is straightforwardly supported. In the following subsections, the different AI building blocks comprising our automatic DR-enhanced AR system are presented.

2.1 Object Segmentation Component

In order to diminish an object from a residential indoor scene, the object's pixel-aligned area within the image must be available. For this purpose, we employ a semantic segmentation network to infer objects mask for a set of a priori selected classes, commonly present in residential scenes. We use the DeepLabv3 architecture [3] with a ResNet50 [4] backbone, which has shown reliable and robust results in segmentation tasks, offering a great compromise between accuracy and speed. The network was supervised using cross-entropy and trained for 133 epochs using the Adam optimization algorithm [9], with default parameters, a learning rate of 0.0002, and a scheduler halving it every 20 epochs.

2.2 Layout Estimation Component

Another prerequisite of the inpainting component is the scene's dense layout segmentation (i.e. the per-pixel classification into the ceiling, wall, or floor classes). This is required to preserve the scene's structure during diminishing which is a very important cue for the downstream applications (i.e. planning or designing). We use the HorizonNet model [19] to estimate the locations of the scene's junctions.

2.3 Inpainting Component

The core of our AI-based DR sub-system is the inpainting AI model which is responsible for object diminishing. Apart from the input panorama, it additionally requires an object mask and the scene's layout segmentation map, as depicted in Figure 2. The latter provides the structure of the scene as corner positions, which are subsequently reconstructed as the dense layout, while the former is a requisite for specifying the object to be diminished. We adopt a structure-aware 360° inpainting model [2] that uses SEAN residual blocks [24] to aid in hallucinate plausible content with semantic coherency in the diminished region. SEAN blocks leverage the structural information provided by the input semantic maps (the layout segmentation in our case), and uses it as structural guidance.

2.4 Super-resolution Component

For alleviating the aforementioned issue concerning the low-resolution of the panoramas to be processed, we resort to a lightweight super-resolution model [22], to upscale the diminished result up to (×4) times. That way, we offer results appropriate for panorama viewers, without degrading their visual quality.

2.5 Implementation & Orchestration

Our models are trained with PyTorch [11] and delivered as services using TorchServe [1]. Our components share a common communication interface that is built around callback URLs, with all inputs and outputs delivered as endpoints to either retrieve (GET) or submit (POST) data. This interface makes our system highly modular since the communication interface is decoupled from the back-end functionality of each component.

The system orchestration is realised as a web server, where each upload triggers a chain of events as follows. At first, the object segmentation and layout estimation models are invoked to estimate the object masks and the room layout. Since we rely on semantic segmentation, we perform connected component analysis to resolve potentially different instances and split each segmentation map into multiple per-class and object masks. To improve robustness, we use the convex hull for each mask in an attempt to decouple the diminished region

shape from the result (the inpainting model is trained similarly). Likewise, the junction estimates are post-processed to generate a dense layout map by first connecting the top and bottom boundaries, and then identifying the corresponding structural labels across each column. Finally, for all object masks, the inpainting service is called, with its result fed into the super-resolution service and then composited on the original panorama. The outputs are then ready to be queried by the AR component that positions the 3D object, whose renders interact with the masks on the image domain to retrieve the appropriate result.

3 EXPERIMENTAL SETUP METHODOLOGY

The evaluation undertaken for the presented AI-based DR-enhanced AR system follows two routes. On the one hand, we seek to assess the DR sub-system's behaviour (Section 3.1), while on the other hand, we aim at validating the complete system's efficacy and goals (Section 3.2). For the former, we opt for an objective evaluation using photo-consistency metrics using complete-diminished pairs, while for the latter we employ subjective scoring using pre-authored scenes.

Given that our system's components have been trained on the Structured3D dataset [23], we use samples from the corresponding test set for both the objective and subjective experiments. Structured3D provides photo-realistic panoramic images of residential rooms, room layout annotation, object segmentation masks, as well as an empty room configuration of each scene that has all foreground (i.e. furniture) removed. The latter data address the most challenging part of objectively evaluating DR systems, which is the lack of paired data where the objects of interest are removed. To simulate indoor (re-)planning/design settings, we focus our evaluation on the {chair, bed, sofa, table, cabinet} class set.

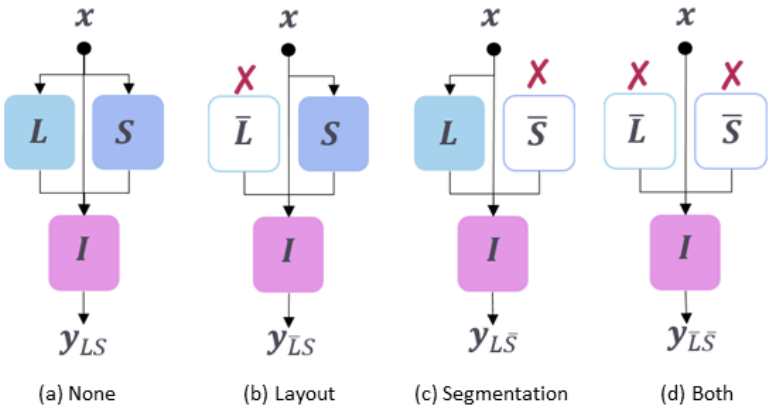

Figure 3: Component ablation experiments setup visualized with a vertical macro view of the DR sub-system of Figure 2. (a) refers to the experiments where both room layout and object masks are estimated by the system's corresponding data-driven components, (b) the layout path is ablated, by replacing the estimations with the annotated ground truth while preserving the segmentation mask estimates, (c) the dual configuration to (b), with the segmentation path ablated and the layout estimations preserved, and (d) where both components are replaced by the annotated ground truth layout and object masks.

3.1 Component Ablation

The DR sub-system comprises three different AI components. When considering it as a sum of its parts, we only need to evaluate the result of the diminished output against an empty scene. Yet two layout and segmentation components that operate in parallel, and cascade their outputs to the inpainting component, can also propagate their errors. Those errors for each separate AI component can be accumulated, affecting the overall performance of the system. Given that each part of this sub-system is an AI model performing a distinct task, its performance can be evaluated in isolation from the complete sub-system. Reasonably, as performance improves, it is expected that the final result will also be improved. Nonetheless, from a system analysis perspective, it is important to identify the weakest link, the component whose system performance relies mostly upon, and thus affects the most outcomes of the system. As a result, we ablate the system's components using differential analysis, where the component is bypassed and instead, a perfect prediction is used (the annotated metadata). Consequently, Figure 3 presents the component ablation setup for the DR sub-system, with the layout estimation and segmentation components ablated in isolation and jointly. The latter experiment allows us to assess the performance of the inpainting component both absolutely, using the metrics, as well as relatively, with respect to the other ablated components' performance degradation. We use the Mean Absolute Error (MAE), the Peak Signal-to-Noise Ratio (PSNR), the Structured Similarity Index Metric (SSIM), and the Perceptual Image Patch Similarity (LPIPS) [21] metrics on the results and compare over the objects' masked regions. LPIPS measures the perceptual similarity between two images based on a VGG pre-trained network [18]. It has been shown that it accounts for several parts of human perception, in contrast with PSNR. Due to the nature of DR to hallucinate realistic content, we consider it as our primary evaluation metric. For the pixel-wise metrics (PSNR & MAE) the union of the ground truth and predicted masks was used to more strictly penalize erroneous segmentations under these photo-consistency metrics. Still, for the local (window-based SSIM) and global (CNN-based LPIPS) metrics, the entire images were used.

Figure 4: Example survey scene types. The first column depicts the original panorama, the second column the panorama with the object removed (i.e. pure DR), the third column, the one with the virtual furniture added in the diminished scene (i.e. DR-enhanced AR), and the final column, the one with the virtual object added without previously removing the existing object (i.e. pure AR).

3.2 User Study

While objective analysis can help in identifying critical components and assessing the system's overall performance, the end result cannot be quantitatively assessed. This is either because ground truth is not necessarily available, or due to the subjectivity of the results. Still, end-user appreciation is the ultimate goal and as a result, we additionally performed a user survey for the entire system's outputs. We used remote questionnaires that were distributed to 38 users split into two sub-groups, one having no knowledge regarding its inner workings (i.e. Group A), and the other knowledgeable regarding AI (i.e. Group B). The questionnaires required the participants to rate the appearance of a masked area in each one of 5 different scenes. An interactive panorama viewer was used, with the initial viewpoint bearing looking at the object to be removed. For each scene, users first were allowed to freely navigate the entire scene in 3 degrees of freedom, then an annotated panorama with the object to be removed or replaced was presented to them. This process ensures that users will not get lost within the 360° field-of-view, and will understand the task at hand. Afterward, users were asked to score the appearance of the previously marked area, once presented with the object removed (i.e. pure DR), and then once with a virtual object replacing the previous one (i.e. DR-enhanced AR). After all, scenes were evaluated, users were asked to rate the scenes again, this time without DR, scoring the result of the pure virtual object superimposition on the existing real object (i.e. pure AR). This last step was isolated from the previous ones to remove any bias when scoring
DR results. Scoring was based on a 5-point Liker scale, resulting in aggregated mean opinion scores (MOS). Figure 4 depicts samples used in the survey.

4 RESULTS DISCUSSION

Before presenting the results of our experiments, it is worth noticing the potential sources of errors. Since the inpainting component is dependent on the results of the layout and segmentation models, it is expected that any errors in these components will be accumulated in the final diminished result. Under-segmenting an object may result in erroneous diminishing of scenes, since artifacts of the old object will be present around the inpainted region. Similarly, over-segmenting may potentially remove important relevant objects like chairs next to a table, resulting in uncanny visuals. Another potential source of error is the layout junction localization. The inpainting model heavily depends on the layout of the input, as described in Section2.3. Given that the boundaries reconstructed from the junctions are used to generate the dense layout segmentation map used to drive the SEAN decoding blocks, such errors will propagate into both style code generation, as well as the diminished area boundary separating the different structural areas. As a consequence, even slight errors in the junctions' coordinates will translate to large miss-classified regions, manifesting in severe diminishing distortions.

Table 1: Quantitative results assessing the DR sub-system output by ablating its components. Arrows denote direction of better performance

Experiment	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓
$\mathbf{y}_{\overline{LS}}$	**29.61**	**0.9393**	**0.0131**	**0.1127**
$\mathbf{y}_{\overline{L}S}$	29.13	0.9353	0.0134	0.1149
$\mathbf{y}_{L\overline{S}}$	27.37	0.9126	0.0166	0.1259
$\mathbf{y}_{LS}$	27.86	0.9189	0.0158	0.1225

4.1 Objective Evaluation

Table 1 shows the quantitative results for the experiments described in Section 3.1. The first row which showcases the best performance is the case (d) of Figure 3, where both models are replaced with perfect estimates. This is in contrast to the last row, corresponding to case (a) of Figure 3, which relies on all models' predictions. Interestingly, cases (b) and (c) are the most interesting ones as they present us with the weakest link of the DR sub-system, which is the layout estimation model, given that when replaced with the annotated layouts, performance consistently increases. As the segmentation model produces reasonable results, the sparser junction localization errors propagate deeper into the diminished result, which is reasonable as the structural segmentation is responsible for both style code extraction and boundary preservation.

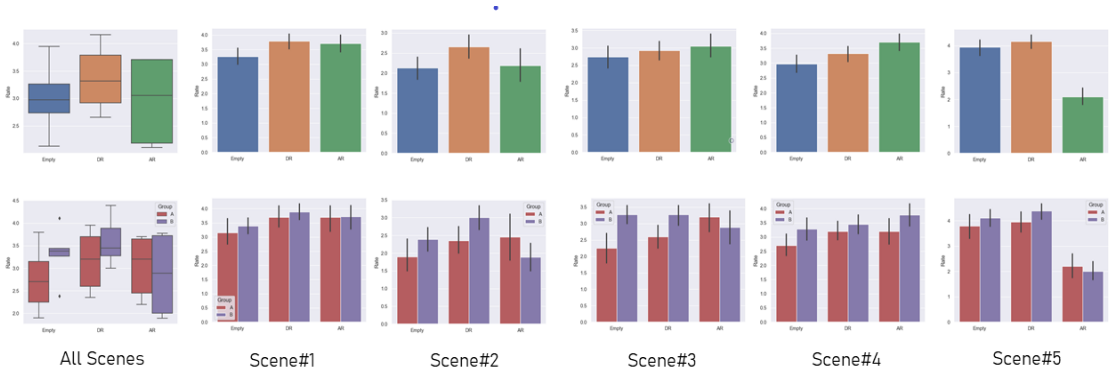

Figure 5: Results of the user survey. The first row depicts the total average rating for all the three cases, i.e. pure DR (empty), DR-enhanced AR (DR), pure AR (AR), across all scenes (first column) as well as for each scene separately in the following columns. Similarly, the second row depicts the same average rating split per each sub-group participating in the study (A & B, not familiar with AI and those knowledgeable about it respectively).

4.2 Subjective Evaluation

Figure 5 presents the results of the user survey. The left columns aggregate MOS scores across all scenes, while the remaining columns present the results for each scene in sequence. The top row presents the results for all subjects, while the bottom row splits them into two different groups, those not familiar with AI (i.e. Group A) and those experienced with it (i.e. Group B). From these results, it is evident that purely diminished scenes were rated lower than diminished scenes with augmentations overlaid. This is expected as superimposing content on the DR result may potentially hide defects. Further, the final scenes without DR where the virtual object was simply overlaid on the actual ones, without removing them, scored lower than the scenes where the real objects had been diminished/removed. Nevertheless, the statistical confidence is lower and this is partly expected as not all scenes may require DR. Indeed there are cases when the objects are of similar size and shape that render DR as not that important. The availability of the functionality, however, is very important for the remaining cases and may even outweigh the need to deliver high-quality DR results.

Regarding the two user groups, those familiar with AI presented with larger discrepancies between the different scene types, albeit the ranking across both groups remained the same.

5 CONCLUSION

In this work, we present a system that can drive user-facing applications for interior design. The focus of our system is on usability as it relies on 360° image acquisition of scenes, compared to scanning processes that tax users and are more error-prone. Further, we lift the requirement for manually marking the diminished region and seek to preserve the room structure during diminishing which is highly relevant for the targeted application domain. Our system is purely AI-based, a fact that introduces the need for assessing error propagation between its different components. To that end, we present a system ablation analysis, accompanied by a user survey that showcases the need for DR in indoor AR planning. Nonetheless, our work operates directly on the image domain (i.e. 2D), and besides the benefits, this introduces, it inevitably only offers perspective views and neglects occlusion effects. Another limitation is that the current system has been only verified with synthetic data. The Structure3D dataset offers annotations for all sub-tasks apart from the super-resolution one, a trait that real-world datasets will not easily provide. Apart from that, the application to in-the-wild real-world data is expected to reduce performance, which will require revisiting our analysis. Future work will focus on overcoming these challenges by integrated geometric inference (e.g. depth) to support more advanced features like occlusions and lighting, and transitioning to real-world domain training data and validation.

ACKNOWLEDGMENTS

We thank Werner Bailer (JOANNEUM RESEARCH), Georg Thallinger (JOANNEUM RESEARCH), Vladimiros Sterzentsenko (Information Technologies Institute / Centre for Research and Technology) and Suzana Farokhian (Usability Partners) for insightful discussion and feedback. This work was supported by the EC funded H2020 project ATLANTIS [GA 951900].

REFERENCES

[1] Torchserve. https://github.com/pytorch/serve, 2021.

[2] Anonymous. Anonymized for peer reviewing.

[3] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.

[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, pp. 770–778, 2016.

[5] J. Herling and W. Broll. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In 2010 IEEE International Symposium on Mixed and Augmented Reality, pp. 207–212. IEEE, 2010.

[6] J. Jam, C. Kendrick, K. Walker, V. Drouard, J. G.-S. Hsu, and M. H. Yap. A comprehensive review of past and present image inpainting methods. Computer Vision and Image Understanding, p. 103147, 2020.

[7] S. Jiddi, B. Pugh, Q. Dai, L. Puig, N. Lianos, P. Gauthier, B. Totty, A. Dorbie, J. Yin, and K. Wong. An end-to-end mixed reality product for interior home furnishing. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2020.

[8] N. Kawai, T. Sato, and N. Yokoya. Diminished reality based on image inpainting considering background geometry. IEEE transactions on visualization and computer graphics, 22(3):1236–1247, 2015.

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[10] S. Mori, S. Ikeda, and H. Saito. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. IPSJ Transactions on Computer Vision and Applications, 9(1):1–14, 2017.

[11] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703, 2019.

[12] V. T. Phan and S. Y. Choo. Interior design in augmented reality environment. International Journal of Computer Applications, 5(5):16–21, 2010.

[13] G. Queguiner, M. Fradet, and M. Rouhani. Towards mobile diminished reality. In 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 226–231. IEEE, 2018.

[14] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park, and I. S. Kweon. A realtime augmented reality system to see-through cars. IEEE transactions on visualization and computer graphics, 22(11):2395–2404, 2016.

[15] P. Reuksupasompon, M. Aruncharathorn, and S. Vittayakorn. Ar development for room design. In 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–6. IEEE, 2018.

[16] D. Schmalstieg, S. Mori, O. Erat, D. Kalkofen, W. Broll, and H. Saito. Inpaintfusion: Incremental rgb-d inpainting for 3d scenes. In IEEE Transactions on Visualization and Computer Graphics, 2020.

[17] S. Siltanen. Diminished reality for augmented reality interior design. The Visual Computer, 33(2):193–208, 2017.

[18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[19] C. Sun, C. Hsiao, M. Sun, and H. Chen. Horizonnet: Learning room layout with 1d representation and pano stretch data augmentation. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pp. 1047–1056, 2019.

[20] K. Wong, S. Jiddi, Y. Alami, P. Guindi, B. Totty, Q. Guo, M. Otrada, and P. Gauthier. Exploiting arkit depth maps for mixed reality home design. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2020.

[21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586–595, 2018.

[22] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong. Efficient image superresolution using pixel attention. arXiv preprint arXiv:2010.01073, 2020.

[23] J. Zheng, J. Zhang, J. Li, R. Tang, S. Gao, and Z. Zhou. Structured3d: A large photo-realistic dataset for structured 3d modeling. In Proceedings of The European Conference on Computer Vision (ECCV), 2020.

[24] P. Zhu, R. Abdal, Y. Qin, and P. Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5104–5113, 2020.