Research Context

Background

Extended Reality (XR), which includes Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), creates a digital platform where users can interact with computer-generated virtual environments. As the perceptual eyes of XR systems, 3D vision technology enables XR systems to perceive and interpret the physical world, and reconstruct it within digital spaces. In general, 3D vision technology plays a crucial role in converting our physical world into spatially-aware high-fidelity digital world, enhancing the utility of XR systems.

In the context of 3D vision technology, perception and reconstruction techniques are two fundamental components. Basically, 3D perception is about capturing spatial and temporal information from the surroundings, as well as understanding the environment given captured data. In addition, 3D reconstruction aims to digitally present the perceived environment, creating a virtual mirror of the physical world for downstream applications. Considering the fact that 3D reconstruction utilizes the data obtained from 3D perception to build accurate 3D models of the environment, the synergy between them can allow the XR system to comprehend the user's physical space and interactions with virtual contents, impacting the immersion and interactivity of XR experiences.

Motivation

While research and development in 3D perception and reconstruction have been active in the deep learning era [17, 21, 29, 31, 50, 52], their successful utilization in XR systems has been limited. The existing challenges, such as acquiring specialized, labeled large-scale data for diverse XR scenarios, ensuring robustness of visual algorithms against environmental variables, managing the computational cost of deep neural networks on XR devices, and handling incomplete or noisy data from 3D sensors, have posed significant barriers to current XR systems. Moreover, previous XR related research such as [3, 19] explored individual algorithms for 3D perception and reconstruction without a specific focus on their synergy and adaptability for general XR applications and platforms.

To unlock the full potential of XR across various domains like healthcare, education, or entertainment, we aim to develop a modular framework that integrates 3D perception and reconstruction for XR applications. The general design of our framework is illustrated and annotated in Fig. 1 and Fig. 2, respectively. As far as we are concerned, integrating 3D perception and reconstruction is crucial for XR to secure its coherency: the perception process makes the XR system understand the user's physical world, and the reconstruction process utilizes this understanding to create immersive and interactive virtual environments that are consistent with the physical world. By achieving and utilizing this integration, XR systems will not only present high-fidelity virtual content to the user but also comprehensively understand and precisely adapt to the user's real-world interactions, providing better quality of XR immersion and interactivity.

The PI's Related Work

The PI's prior academic and industrial experience has provided a solid foundation for the proposed project on integrating 3D perception and reconstruction for XR. Academically, the PI's research has focused on 3D data processing/object recognition/scene understanding, leading to top-tier journal and conference publications such as CVPR/ICCV/TPAMI/TMM/etc. Specifically, the PI has proposed a few methods to the fundamental problems of 3D perception research, including a geometric back-projection network for 3D object classification [35], a bilateral augmentation and adaptive fusion model for 3D scene understanding [36] (shown in Fig. 4), and a context plug-and-play for 3D data's feature representation enhancement [38] (shown in Fig. 5), etc. The PI has also contributed to the main topics in 3D reconstruction research: for example, a self-supervised learning method for completing single 3D scans [9] (shown in Fig. 6), an energy-based generative model for point cloud inpainting [10], and a transformer-based network for 3D upsampling [37], etc. These explorations have benefited the development of our 3D research community, directly informing the methodologies proposed in the current project.

In addition to academic research, the PI's previous involvement in XR industries has provided a practical perspective on the application of 3D perception and reconstruction in real-world scenarios. Particularly, the PI's development of XR platform (shown in Fig. 3) and implementation of 3D perception and reconstruction pipeline prototype (shown in Fig. 7) provide critical insights into the practical challenges and solutions of this research project.

Potential Contribution and Impact

This research project aims to develop an integrated framework of 3D perception and reconstruction, which is specifically optimized for XR applications. As we intend to modularize the framework with high efficiency, robustness, and adaptability in practice, our research will focus on the development, implementation, and evaluation of the proposed framework in XR. By exploring data-efficient learning scheme, investigating robust and adaptable 3D vision algorithms, and enhancing the framework's applicability and performance across various XR applications and platforms, the anticipated impacts of this project can be found in academia, society, economy, environment, healthcare, and the broader XR community, both locally and globally. More details are provided in the document of "Pathways to Impact Statement".

Ethical Considerations

The use of 3D vision and XR technologies in different applications raises important ethical concerns. As such, we plan to implement the following ethical considerations in this project:

- Informed Consent. We will obtain informed consent from all participants in the research project. This includes informing participants about the nature of the study, the potential benefits and risks when interacting with XR systems, and their right to withdraw at any time.

- Data Privacy and Security. We will ensure that all participants data collected during the study is kept secure and confidential. All data will be anonymous and only accessible to authorized personnel.

- Responsibility and Accountability. We will ensure that all aspects of the research project are conducted in a transparent and ethical manner, where the potential risks and benefits are carefully considered and communicated to all participants.

Research Questions and Research Methods

Current methods for 3D perception and reconstruction in Extended Reality (XR) often operate separately, addressing individual aspects of these complex processes without a coherent, integrated approach tailored to diverse XR applications and platforms. The challenges are multifold: from acquiring specialized, large-scale labeled data for varied XR scenarios to ensuring algorithmic robustness against dynamic environmental variables in XR development. The aim of this project is to develop an integrated framework that symbiotically unites 3D perception and reconstruction, leveraging AI-driven 3D vision technology to enhance the interactivity and immersion of XR experiences across various applications. This involves not only addressing the aforementioned challenges but also ensuring the modular, adaptable, and efficient functionality of the proposed framework in real-world XR scenarios. In the following content, we outline the main research questions and corresponding methods that will be investigated in this project.

Research Question #1: How can large-scale, labeled data for diverse XR scenarios be efficiently acquired and utilized?

Conventional XR systems usually use single modality of data, e.g., multi-view RGB images, to perceive the spatial information of target 3D objects given camera poses. However, multi-view images might not provide accurate depth information, especially in scenes with repetitive textures or occlusions. Also, calculating disparities between multiple views can be computationally intensive. In this project, we will explore data acquisition techniques that are efficient and scalable for XR applications. In general, a multi-modal data acquisition system that utilizes various sensors, e.g., RGB-D cameras [49], ToF cameras [23], or LiDAR [22], will be introduced to comprehensively capture spatial and temporal data of different modalities. Accordingly, we focus on investigating: (i) learning-based intelligent fusion algorithms to integrate data from different modalities, enhancing robustness and reliability of perception; and (ii) context-aware fusion strategies that dynamically adjust based on the application scenario and available modalities. The proposed multi-modal data acquisition system will be comprehensively evaluated within real-world XR scenarios, given the quality, efficiency, and applicability of acquired data in 3D perception and reconstruction processes.

As a complementary to acquiring real-world data, we can also consider generating synthetic data that is representative and applicable for training visual models of 3D perception and reconstruction. Unlike existing methods that basically follow the Computer-Generated Imagery (CGI) technology to render high-quality virtual 3D scenes and record useful data via virtual sensors [5], we prefer using Generative Adversarial Networks (GANs) [33] and recent Diffusion Models [8] to facilitate the creation of synthetic data representing various XR scenarios. Compared to the CGI-driven methods that rely on high rendering computation and strong empirical settings, the generative models can be deployed in an easier and more intelligent manner without heavy manual manipulations.

Although generating synthetic data can enrich the pool of training data, a consequent problem is about mitigating the domain gap between synthetic data and real-world datas. To tackle this problem, we will basically leverage and optimize Domain Adaptation (DA) techniques [43, 45, 46] to realize an effective generalization of models trained on synthetic data to real-world XR scenarios. Particularly, Domain Adaptation via Prompt Learning (DAPL) [15] is a novel prompt learning paradigm for Unsupervised Domain Adaptation (UDA), which leverages embedded domain information to prompt pre-trained large vision-language models [26, 24] for the generation in target domain. This DAPL scheme will be further studied for XR applications, since it is efficient to train and easy to implement where only a few parameters need to be optimized. Moreover, to alleviate the dependency on large-scale labeled data, self-supervised and weakly-supervised models will be developed to utilize fewer or simply annotated labels. For example, we plan to explore self-supervised 3D learning given the promising results of Masked Auto-Encoder (MAE) [18] in 2D domain. The effectiveness of proposed self-supervised and weakly-supervised 3D learning methods will be quantivatively and qualitatively evaluated in various XR scenarios.

Research Question #2: How to enhance the robustness of data-driven visual algorithms in XR development?

Considering the vulnerability of data-driven visual algorithms to complex environmental variables, first of all, we will research how to enhance the robustness of 3D perception algorithms against varying and dynamic lighting conditions in real-world XR applications. Basically, we aim to develop adaptive algorithms that can easily handle diverse lighting scenarios. Specifically, physical lighting models [16] can inspire the main structure of intended algorithms, where both synthetic/simulated corner cases and data augmentation techniques [40, 6] can be utilized to further strengthen the robustness of algorithms. Furthermore, the strategy of pre-training on large-scale common data and then fine-tuning on specific application can be exploited to maintain both the robustness and specificity of the algorithms.

As previously discussed, different data modalities are needed to obtain a comprehensive perception of the real-world surroundings. Thus, we expect to research a robust 3D reconstruction approach that can achieve accurate and consistent performances given different data representations such as RGB-D, point cloud, or LiDAR data, etc. This research will firstly lead to a deep analysis of each data type's unique characteristics and the corresponding reconstruction methods. Then, a unified data processing pipeline will be developed to modulate the different data representations. Particularly, the key of this pipeline is to seek an intermediate data representation that minimizes the variances of original data types, and then construct a robust and accurate 3D reconstruction method in this intermediate representation domain. In general, there are two ways to achieve this intention: (i) we can implicitly explore a learnable space; or (ii) explicitly leverage geometric estimations such as projecting the data into hyperbolic space [12, 28]. In addition, continuous refinement will be facilitated through iteratively evaluating our proposed 3D reconstruction approach, in order to satisfy the dynamic needs of indoor and outdoor XR scenes.

Research Question #3: How to optimize computational cost of deep neural networks for XR devices?

Due to expensive computational cost of neural networks, it is practically difficult to deploy large and deep models running on standalone XR devices that only utilize commercial-level system on a chip (SoC), such as Qualcomm Snapdragon XR2 for the latest Meta Quest 3 and HTC VIVE XR Elite. On this front, our research will particularly focus on developing lightweight yet effective 3D perception and reconstruction algorithms affordable for mobile processors. The research will start from identifying the computational and memory constraints of target XR devices including VIVE, Quest, etc. Then, we will incorporate algorithmic efficiency and model compression techniques such as quantization [34] and pruning [51] with tailored 3D perception and reconstruction algorithms adhering to hardware constraints. On the other hand, we will utilize efficient convolutional operations to devise the architectures of 3D perception and reconstruction models. In particular, Minkowski Engine [7] is a CUDA-based library that implements sparse convolutions [25] for efficient large-scale 3D data learning. Based on this library, we will develop novel 3D learning networks that are highly compatible and efficient for mobile XR platforms. For the XR systems that deploys dedicated GPUs or AI accelerators, we will not only consider commonly used strategies such as network pruning [27] and knowledge distillation [34], but also focus on leveraging hardware accelerated computing techniques, e.g., multi-GPU parallelization [13, 32], to achieve real-time 3D perception and reconstruction inferences. More recently, cloud computing [42] and edge computing [11] have been applied to facilitate the deployment of deep learning models on resource-constrained XR devices. In this project, we aim to develop a hybrid computational approach: for computationally intensive tasks such as model training and complex inference, the processes will be performed through cloud or edge nodes [42, 44]; while light processings can be conducted on the XR device. The approach will utilize efficient and adaptive strategies to optimize the balance between local and remote computation, realizing minimal latency and optimal utilization of available resources.

Research Question #4: How to amend perceived 3D data for better reconstruction results?

Considering the fact that 3D sensors of XR systems inevitably capture incomplete, irregular, and noisy data due to occlusion and hardware restrictions, we need to further improve the quality of perceived 3D data for the reconstruction purpose. In general, we will focus on researching cutting-edge 3D low-level vision techniques including but not limited to denoising [20], upsampling [48], completion [14], to effectively amend 3D data for the down-stream tasks related to 3D reconstruction. In addition to developing lightweight and efficient algorithms that can mitigate common 3D data issues of noise and irregularity, we expect to introduce a unified model capable of handling various low-level 3D data processing tasks, given the success of Pre-Trained Image Processing Transformer (IPT) [4] in 2D domain. Basically, we intend to pre-train the model in a self-supervised contrastive learning approach [1], which can enhance the model's capability to understand and process 3D data effectively; then, a multi-head and multi-tail architecture will be enabled to simultaneously manage different low-level 3D processing tasks, showcasing its generalization ability and adaptability. Beyond serving as a functional module that amends perceived data for high-quality 3D reconstruction in XR, this unified model will fill the gaps in large vision foundation models for 3D low-level data processing.

In the context of 3D perception and reconstruction for XR, accurate 3D registration is crucial to ensure that the reconstructed or perceived 3D data is spatially coherent and aligned, providing a stable and consistent XR experience. Conventional ICP-based methods [2], which primarily consider basic point geometries and lack the incorporation of more representative features like semantic/structural context for global registration, prove to be insufficient as they allow misalignment to accumulate over time, particularly when performing online scan-by-scan registration. In this project, we aim to develop a robust 3D data registration solution for XR systems, where effective 3D networks will be integrated to enhance feature matching and sophisticated 3D geometry techniques will be exploited to refine transformation parameters, realizing accurate and robust alignment of 3D data in real-time XR environments.

Research Question #5: How can 3D perception and reconstruction be effectively integrated for XR systems?

Since 3D perception and reconstruction techniques are often separately explored, the synergy between them is yet to be fully utilized for XR applications, consequently leading to a high complexity of the whole system. To figure out the solutions to this issue, firstly our research will focus on understanding the computational and algorithmic requirements of each individual process, as well as identifying commonalities and divergences between different processes. As for a modular framework, 3D perception and reconstruction modules are expected to achieve their synergy given smooth data flow. On this front, we will explore unified data management and communication protocols tailored to deep learning models [39], where the techniques of asynchronous processing, data compression, and caching, can be used to optimize data transmission and resolve data bottlenecks. Particularly, based on the findings in [30], we will mainly research the data management issues of deep learning models raised in the deployment and post-deployment phases, including overfitting, data drifts, etc. The integration will be further facilitated through an optimization unit and a utility unit, allowing the modules to not only share data but also operate collaboratively.

The modularity and adaptability of the whole framework will be tested by using it across various XR applications and platforms. Basically, we will explore a comprehensive evaluation strategy. Quantitative metrics, such as interaction latency, system response time, and accuracies of 3D perception and reconstruction, will be measured under various scenarios to objectively assess the framework's performance. In addition, participants will engage with XR environments empowered by the proposed framework, and their immersion and interactivity experiences can be qualitatively evaluated via review and feedback [41, 47]. By collecting quantitative and qualitative data, we will comprehensively analyze the effects of our design and accordingly refine the framework modules.

References

[1] Mohamed Afham, Isuru Dissanayake, Dinithi Dissanayake, Amaya Dharmasiri, Kanchana Thilakarathna, and Ranga Rodrigo. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9902–9912, 2022.

[2] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie, 1992.

[3] Fabio Bruno, Stefano Bruno, Giovanna De Sensi, Maria-Laura Luchi, Stefania Mancuso, and Maurizio Muzzupappa. From 3d reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage*, 11(1):42–49, 2010.

[4] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12299–12310, 2021.

[5] Ziang Cheng, Jiayu Yang, and Hongdong Li. Stereo matching in time: 100+ fps video stereo matching for extended reality. *arXiv preprint arXiv:2309.04183*, 2023.

[6] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.

[7] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3075–3084, 2019.

[8] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[9] Ruikai Cui, Shi Qiu, Saeed Anwar, Jiawei Liu, Chaoyue Xing, Jing Zhang, and Nick Barnes. P2c: Self-supervised point cloud completion from single partial clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14351–14360, 2023.

[10] Ruikai Cui, Shi Qiu, Saeed Anwar, Jing Zhang, and Nick Barnes. Energy-based residual latent transport for unsupervised point cloud completion. *arXiv preprint arXiv:2211.06820*, 2022.

[11] Sahraoui Dhelim, Tahar Kechadi, Liming Chen, Nyothiri Aung, Huansheng Ning, and Luigi Atzori. Edge-enabled metaverse: The convergence of metaverse and mobile edge computing. *arXiv preprint arXiv:2205.02764*, 2022.

[12] Aleksandr Ermolov, Leyla Mirvakhabova, Valentin Khrulkov, Nicu Sebe, and Ivan Oseledets. Hyperbolic vision transformers: Combining improvements in metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7409–7419, 2022.

[13] Mokhtar Essaid, Lhassane Idoumghar, Julien Lepagnot, and Mathieu Brévilliers. Gpu parallelization strategies for metaheuristics: a survey. *International Journal of Parallel, Emergent and Distributed Systems*, 34(5):497–522, 2019.

[14] Ben Fei, Weidong Yang, Wen-Ming Chen, Zhijun Li, Yikang Li, Tao Ma, Xing Hu, and Lipeng Ma. Comprehensive review of deep learning-based 3d point cloud completion processing and analysis. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[15] Chunjiang Ge, Rui Huang, Mixue Xie, Zihang Lai, Shiji Song, Shuang Li, and Gao Huang. Domain adaptation via prompt learning. *arXiv preprint arXiv:2202.06687*, 2022.

[16] Amy Gooch, Bruce Gooch, Peter Shirley, and Elaine Cohen. A non-photorealistic lighting model for automatic technical illustration. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 447–452, 1998.

[17] Hanry Ham, Julian Wesley, and Hendra Hendra. Computer vision based 3d reconstruction: A review. *International Journal of Electrical and Computer Engineering*, 9(4):2394, 2019.

[18] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[19] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568, 2011.

[20] Alireza Javaheri, Catarina Brites, Fernando Pereira, and João Ascenso. Subjective and objective quality evaluation of 3d point cloud denoising algorithms. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2017.

[21] Yiwei Jin, Diqiong Jiang, and Ming Cai. 3d reconstruction using deep learning: a survey. *Communications in Information and Systems*, 20(4):389–413, 2020.

[22] A Karji, S Rokooei, and E Mainzer. Integration of lidar and augmented reality for construction quality control, a conceptual framework. In *Proceedings of the CSCE 2021 Annual Conference, Las Vegas, NV, USA*, pages 26–29, 2021.

[23] Reinhard Koch, Ingo Schiller, Bogumil Bartczak, Falko Kellner, and Kevin Koeser. Mixin3d: 3d mixed reality with tof-camera. In *Dynamic 3D Imaging: DAGM 2009 Workshop, Dyn3D 2009, Jena, Germany, September 9, 2009. Proceedings*, pages 126–141. Springer, 2009.

[24] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023.

[25] Baoyuan Liu, Min Wang, Hassan Foroosh, Marshall Tappen, and Marianna Pensky. Sparse convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 806–814, 2015.

[26] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023.

[27] Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, and Trevor Darrell. Rethinking the value of network pruning. *arXiv preprint arXiv:1810.05270*, 2018.

[28] Pascal Mettes, Mina Ghadimi Atigh, Martin Keller-Ressel, Jeffrey Gu, and Serena Yeung. Hyperbolic deep learning in computer vision: A survey. *arXiv preprint arXiv:2305.06611*, 2023.

[29] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438:14–33, 2021.

[30] Aiswarya Raj Munappy, Jan Bosch, Helena Holmström Olsson, Anders Arpteg, and Björn Brinne. Data management for production quality deep learning models: Challenges and solutions. *Journal of Systems and Software*, 191:111359, 2022.

[31] Niall O'Mahony, Sean Campbell, Lenka Krpalkova, Daniel Riordan, Joseph Walsh, Aidan Murphy, and Conor Ryan. Computer vision for 3d perception: A review. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 2*, pages 788–804. Springer, 2019.

[32] Saptadeep Pal, Eiman Ebrahimi, Arslan Zulfiqar, Yaosheng Fu, Victor Zhang, Szymon Migacz, David Nellans, and Puneet Gupta. Optimizing multi-gpu parallelization strategies for deep learning training. *Ieee Micro*, 39(5):91–101, 2019.

[33] Zhaoqing Pan, Weijie Yu, Xiaokai Yi, Asifullah Khan, Feng Yuan, and Yuhui Zheng. Recent progress on generative adversarial networks (gans): A survey. *IEEE access*, 7:36322–36333, 2019.

[34] Antonio Polino, Razvan Pascanu, and Dan Alistarh. Model compression via distillation and quantization. *arXiv preprint arXiv:1802.05668*, 2018.

[35] Shi Qiu, Saeed Anwar, and Nick Barnes. Geometric back-projection network for point cloud classification. *IEEE Transactions on Multimedia*, 24:1943–1955, 2021.

[36] Shi Qiu, Saeed Anwar, and Nick Barnes. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1757–1767, 2021.

[37] Shi Qiu, Saeed Anwar, and Nick Barnes. Pu-transformer: Point cloud upsampling transformer. In *Proceedings of the Asian Conference on Computer Vision*, pages 2475–2493, 2022.

[38] Shi Qiu, Saeed Anwar, and Nick Barnes. Pnp-3d: A plug-and-play for 3d point clouds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):1312–1319, 2023.

[39] P Rohini, Sachin Tripathi, CM Preeti, A Renuka, José Luis Arias Gonzales, and Durgaprasad Gangodkar. A study on the adoption of wireless communication in big data analytics using neural networks and deep learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pages 1071–1076. IEEE, 2022.

[40] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[41] Mel Slater et al. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: teleoperators and virtual environments*, 8(5):560–565, 1999.

[42] Theodoros Theodoropoulos, Antonios Makris, Abderrahmane Boudi, Tarik Taleb, Uwe Herzog, Luis Rosa, Luis Cordeiro, Konstantinos Tserpes, Elena Spatafora, Alessandro Romussi, et al. Cloud-based xr services: A survey on relevant challenges and enabling technologies. *Journal of Networking and Network Applications*, 2(1):1–22, 2022.

[43] Larissa T Triess, Mariella Dreissig, Christoph B Rist, and J Marius Zöllner. A survey on deep domain adaptation for lidar perception. In *2021 IEEE intelligent vehicles symposium workshops (IV workshops)*, pages 350–357. IEEE, 2021.

[44] Hongjia Wu, Hongxin Zhang, Jiang Cheng, Jianwei Guo, and Wei Chen. Perspectives on point cloud-based 3d scene modeling and xr presentation within the cloud-edge-client architecture. *Visual Informatics*, 2023.

[45] Qiangeng Xu, Yin Zhou, Weiyue Wang, Charles R Qi, and Dragomir Anguelov. Spg: Unsupervised domain adaptation for 3d object detection via semantic point generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15446–15456, 2021.

[46] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10368–10378, 2021.

[47] Mark Yi-Cheon Yim, Shu-Chuan Chu, and Paul L Sauer. Is augmented reality technology an effective tool for e-commerce? an interactivity and vividness perspective. *Journal of interactive marketing*, 39(1):89–103, 2017.

[48] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2790–2799, 2018.

[49] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.

[50] Chaoqiang Zhao, Qiyu Sun, Chongzhen Zhang, Yang Tang, and Feng Qian. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences*, 63(9):1612–1627, 2020.

[51] Michael Zhu and Suyog Gupta. To prune, or not to prune: exploring the efficacy of pruning for model compression. *arXiv preprint arXiv:1710.01878*, 2017.

[52] Michael Zollhöfer, Patrick Stotko, Andreas Görlitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. State of the art on 3d reconstruction with rgb-d cameras. In *Computer graphics forum*, volume 37, pages 625–652. Wiley Online Library, 2018.