Journal of Robotics and Control (JRC)
Volume 3, Issue 2, March 2022
ISSN: 2715-5072, DOI: 10.18196/jrc.vxix.xxxxx

Optimizing Indoor Camera Localization: An Efficient Approach Integrating Recurrent Neural Networks and Motion Blur Elimination

Muhammad S. Alam[1*], Farhan B. Mohamed [2], Ali Selamat [3], AKM B. Hossain[4]
[1,2,4] Department of Emergent Computing, School of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[2] Media and Game Innovation Centre of Excellence (MaGICX), Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[3] Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[1,4] Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia
Email: [1]shamsul20@graduate.utm.my, [2]farhan@utm.my, [3]aselamat@utm.my, [4]k.m.a@graduate.utm.my
*Corresponding Author

***Abstract*—Rapid growth and technological improvements in computer vision have enabled indoor camera localization in the environment. Indoor camera localization is helpful in various applications, including robotic systems, game playing, and vehicle navigation. The accurate localization of a camera in a large indoor environment is challenging. Solving this challenge is necessary for autonomous robots such as indoor mobile robots. Existing approaches for solving camera-localization problems have limitations in terms of accuracy and speed. Indoor environments have many complex problems, motion blur is one of them. Motion blur introduces significant errors and affects feature matching, making it difficult to accurately determine camera pose. Object and camera movements create motion blur. Motion blur degrades the image quality, complicates the calculation of the indoor camera pose, and reduces the performance. Therefore, it is imperative to develop indoor robot navigation systems that require high accuracy. It is still necessary to improve the camera localization performance, even though CNN provides improved performance in positioning applications. In this research study, we propose a deep architecture to solve the indoor camera localization problem using a deep architecture trained by minimizing motion blur. A recurrent deep architecture improves the camera localization accuracy and speed by** 20-30%

***Keywords*—Camera pose estimation, Indoor camera localization, Indoor robot navigation, Motion blur, SLAM**

I. INTRODUCTION

The position and orientation of the camera with respect to the world coordinate system are called camera localization. The initial step in interpreting the surrounding scene to determine the camera pose. In structure-from-motion (SfM), which a camera moves across an environment, it is necessary to determine the following camera poses at different instants to reconstruct the surrounding scene. It is essential to define the relative camera pose or connect one camera to another in a multicamera system [1]. In robotics and computer vision, indoor camera localization is a significant problem [2]. Various methods have been developed for indoor camera localization, which are crucial for applications [3], such as indoor robot navigation, SfM, and SLAM. One approach involves point-based techniques that use image descriptors and 3D scene point clouds obtained from SfM to establish a camera pose based on 2D-3D matches. However, this method may not be accurate in some situations, such as when motion blur [4] is present. To address this problem, researchers have applied machine learning algorithms [5], [6] such as SCoRF, which generates camera pose hypotheses based on the predicted 3D locations of four input image pixels and then refines them using RANSAC. However, SCoRF requires depth maps to be matched with input images during training, which must be performed within a limited time. Researchers used convolutional neural networks (CNNs) to calculate the camera pose. Most indoor camera localization algorithms are regression problems, with the CNN pretrained architecture directly estimating the camera pose. Despite addressing several issues with point-based systems, there are still certain limitations to learning-based techniques. As a result, the training architecture uses a mapping from pixel to pose [7] that is dependent on the system of coordinates used by the training data for a certain scene. Multi-scenario localization complicates the process and prevents the transfer of geometric knowledge between locations. As pointed out, the second issue is the apparent limitation of scaling to large indoor environments, because a limited neural network limits the physical region it can learn. Developing a comprehensive indoor positioning system has been a challenging research area for many years, with precise pose data being essential for various applications, such as autonomous robot navigation [8].

A significant amount of research has been conducted on indoor positioning [9]–[11]. For example, when capturing images

CC BY SA

of moving objects or in poor light, motion blur is a typical problem in indoor camera localization. The accuracy of localization algorithms can be significantly decreased by producing distortions and ambiguities in the collected images. For camera localization systems to operate more effectively, motion blur must be addressed. To achieve precise and trustworthy indoor camera localization, it is essential to select the appropriate deep learning architecture. Traditional methods frequently use single-image algorithms, which may not fully exploit the temporal information present in the image sequences. It is possible to improve localization outcomes using temporal data collected over a number of subsequent frames as helpful cues for motion estimates and blur removal. We used sophisticated image deblurring techniques that successfully restore the sharpness and clarity of blurred images to remove motion blur. To estimate the blur kernel and subsequently eliminate blur-induced distortions, these algorithms use motion estimation techniques. We performed extensive evaluations on various indoor datasets to assess the performance of the proposed strategy, in contrast to cutting-edge camera localization techniques. Our findings demonstrate how the approach can increase localization stability and accuracy, especially in difficult motion blur circumstances. In this research study, we eliminate motion blur from a sequence of images and introduce an innovative recurrent deep architecture for indoor camera localization. The recurrent deep architecture employed a deblurring approach to minimize the impact of motion blur on the input images as shown in Figure 1.

A recurrent neural network evaluates the input images and estimates the camera pose. To train the proposed approach, we used a significant number of indoor images along with ground-truth camera poses, combining supervised and unsupervised learning strategies. It uses an LSTM (long short-term memory) network to identify temporal correlations in the input sequence. Our contributions summarize as below:

1) We propose a innovative technique for eliminating motion blur in indoor camera localization. We can significantly minimize the effect of motion blur on the precision of camera localization by evaluating the blur patterns and utilizing cutting-edge image processing techniques.
2) We presented a recurrent deep architecture specifically designed for evaluating image sequences in the context of camera localization. To increase the precision and stability of the camera pose estimation, our model takes advantage of the temporal data recorded in the succeeding frames.
3) We integrated a unified indoor camera localization system with a motion blur removal technique and a recurrent deep architecture. The experimental results reveal the efficiency of the proposed method, displaying considerable gains in robustness and localization accuracy.

The research study has applications in robotics, augmented reality, and indoor navigation systems, among other fields that rely on indoor camera localization. Although the recurrent deep architecture improves the overall performance of camera localization systems, enabling more accurate and reliable localization in real-world applications, the motion blur removal technique can assist in overcoming obstacles associated with dynamic environments and moving cameras. The proposed approach represents a substantial development in the field of indoor camera localization by addressing the problems of motion blur and utilizing a recurrent deep architecture. Combining these contributions strengthens the dependability and accuracy of camera localization systems, creating opportunities for further developments in associated research and real-world applications.

II. Related Works

This section describes the literature review of motion blur, indoor camera localization, and recurrent neural network.

A. Motion Blur

In the past decade, blurred images have received considerable attention for camera localization in indoor environments. Researchers have addressed blur caused by camera movements and plane translations. They successfully removed blurs when the motion assumptions were followed. A novel approach [12] addressed the challenge of evaluating and improving a blurry image through removing motion blur. To calculate the probability distribution of motion blur, researchers have adopted a CNN-based deep-learning approach. Image deblurring aims to restore sharpness to a blurry image caused by camera shake or object motion [6]. The researchers focused on measuring and minimizing non-uniform motion blur [13]. To provide a unique deep learning-based strategy for predicting motion blur accompanied by a patch survey data deblurring approach tailored to motion blur. Using CNNs' powerful feature learning capabilities of CNNs, they can accurately predict complex motion blur [14]. Motion blur is pervasive in indoor camera localization, particularly when using small, transportable devices, such as cellphones and hidden cameras. Considering a few constraints on a particular type of blur, many strategies [15] have recently been developed to minimize the blur caused by camera object movements. Motion blur [16] in actual images can be caused by multiple factors involving the camera [17] and object movement, resulting in complicated blur patterns. Uniform deblurring approaches cannot eliminate non-uniform blur [18]. Some end-to-end approaches have been proposed for reconstructing blur-free images [19]. However, they can tolerate only minor Gaussian blur. Recently published a feed-forward neural network-based deblurring approach [17] that employs information from image sequences with minor errors. A patch-based system [20] proposed to expect frequency information uniform motion blur reduction. The most significant research [21] focused on employing an update-level blurred-type classification approach based on a CNN to predict movement flow

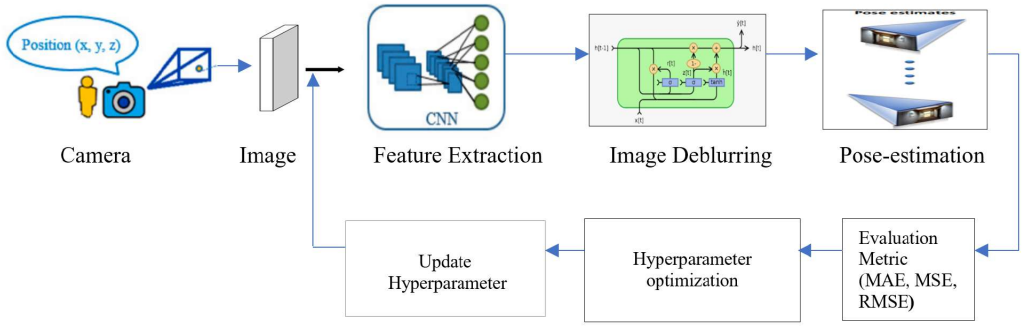

Fig. 1. Overview of an improved indoor camera localization through motion blur elimination and recurrent neural network.

from a single blurry image. An efficient and flexible deep-learning-based technique [22] have been proposed to predict and reduce heterogeneous motion blur. Motion blur is one of the most noticeable flaws in images taken with handheld cameras [23]. Camera shaking caused by extended exposure times in low-light circumstances and quick object movements in a dynamic image causes the blur artefact. An internetwork was created by converting the classic coarse-to-fine technique to a CNN [24], using a recurrent neural network (RNN) with a nonlinear architecture, dramatically improving its performance. DeblurGAN [25], which is a deep deblurring network, is influenced by research on Generative Adversarial Nets (GAN). Researchers added a dark channel to the loss function to reduce pattern artifacts, and lightweight U-nets [26] were used to replace the residual net DeblurGAN. A framework proposed for recovering from a combination of noisy and unclear images, a sharp and clear image [27]. A novel recurrent network [28] designs that operate on arbitrary duration films. In [27], proposed an adaptive temporal blending component on a fast RNN, whereas Wieschollek et al. [28] used information from the previous frame by simply copying features. Researchers have used an iterative hidden layer update approach inside a single inter-frame time step to ensure that the transmitted hidden state fits the target frame. As a result, effective motion deblurring [29] techniques enhance the dependability of associated industries, such as aerospace, traffic monitoring, army search, satellite, and space imagery. In some of the most recent algorithms, deep learning [30] predicts the probability dispersion of motion blur and restores the damaged images [31]. Divided these deep-learning methods into two groups. The first method uses multi-frame images containing a complex network foundation, and the second uses only a single image [32] to deblur the degraded image. Deep-learning algorithms are not ideal for single-image deblurring [33] because they require a long time to compute complicated building structures or have particular criteria for blur conditions. Non-uniform single-image deblurring or predicting unknown non-uniform blur kernels remains a problematic ill-posed inverse issue for recovering a clear image using a blurred image [34]. The optical flow describes the displacement of nearby frames [35] to assist in learning future neural network models.

B. Indoor Camera Localization

The indoor camera localization measures the camera poses of the query image in a random scene. In indoor camera localization, a single image or image sequence is the input and the predicted camera poses are the outputs [36]. The localization of indoor cameras comprises an approach based on scene representation. The camera localization [37] problem was first implemented as a localization detection problem [38]. The image was located using an image retrieval system. When matches between a queried image and an image in the database, the image closest to the query is found. Convolutional neural networks have been utilized in deep learning-driven camera pose estimation to determine how to transform RGB images into related poses. PoseNet [39] was the first to employ CNNs to predict straight 6-DOF camera pose estimation [71]. Motion blur is the leading cause of performance degradation. Although some images contain texture surfaces that are free from motion blur, many missing ground-truth scene coordinate labels might cause issues. In Bayesian-PoseNet [40] researchers introduced PoseNet to account for the uncertainty in pose estimation. Localization fails because of constraints, such as motion blur and illumination changes. The localization performance is improved by the LSTM-PoseNet architecture, which reduces the structured dimensionality and addresses challenges such as motion blur and illumination changes. In [41], proposed an hourglass architecture as the basis for pose regression. This study demonstrates that the method works on data with motion blur and lighting change problems. Camera localization is restricted [42] employing conditional generative adversarial networks and the regression model to achieve pose estimation. Some other research has focused on frameworks to increase the performance of camera localization [43]. In [44], proposed a lightweight CNN for real-time camera localization. Several approaches have been proposed to predict poses using computer vision. In [45], suggest a method that focuses on important geometric features [69] through multitasking, which uses in-

formation from related activities. In [46], present a technique based on ConvNet that dynamically predict the reat-time camera pose estimation. In [47], use CNN-RNN networks with image sequences to enforce temporal smoothness in camera motion. Another approach developed [48] a viewgraph using graph neural networks, which enables the sharing of information from non-consecutive frames [49]. Various studies have employed this approach, such as Peng and Saenko [50], who utilized deep correlation alignment networks to recognize 3D CAD modelsIn [51], proposed a combined dataset of both natural and artificial images for object recognition; and in [52], mapped synthetic to real images in benchmark representations to bridge the gap between synthetic and real images in pattern representations. Additionally, some researchers, such as [53], [70], have synthesized images using 3D models to evaluate search images against a database of synthetic images using in-depth features, specifically for the geolocation image dataset. Inspired by this research, in [54], proposed using syntactic and actual image sequences as a training dataset and the PoseNet [39] network to directly predict the camera pose. In [55], used a recurrent neural network, domain adaptation and addressed the enormous visual and domain-specific differences between artificial and real images, which significantly degrade localization accuracy [56]. Indoor camera location based on deep learning remains challenging because indoor scenes have motion blur. Deep learning-based approaches overcome the limitation of local feature-based methods but are still far from an actual value. Some indoor positioning applications such as mobile guide robots [72] and indoor robot navigation require more accurate camera positioning. Therefore, there is a need to improve the location of RNN-based indoor cameras.

C. Recurrent Neural Network

The ability of RNNs (Recurrent Neural Networks) to handle sequential data has led to their widespread application in camera localization tasks. For camera localization, three common RNN architecture options include LSTM (Long Short-Term Memory, Bi-directional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU). LSTM models have been used successfully to determine the exact location inside cameras by taking advantage of their capacity to collect long-range dependencies and handle sequential inputs [57]. These models are able to anticipate the camera's position with accuracy because they make use of the temporal information included in the camera data. Bidirectional Long Short-Term Memory (Bi-LSTM) networks have been used to great success in a number of computer vision tasks, including indoor camera localization. Bi-LSTM models can successfully learn and predict the position of a camera inside an indoor environment by capturing the temporal interdependence and spatial context in sequential data [58]. GRUs have demonstrated potential for modeling sequential data and identifying temporal dependencies, which makes them suitable for evaluating video feeds from indoor cameras. Researchers have used GRUs to achieve precise and reliable localization results, thereby enabling accurate tracking and surveillance in indoor situations.

III. Material and Methods

A novel recurrent deep architecture that minimizes motion blur for indoor camera localization is completely experimental research. We collected data from secondary sources to identify the image features, recognition techniques, recurrent neural network techniques, and optimization techniques. Primary data were gathered from publicly available databases. In addition, we analyzed the image database using a deep-learning system. We developed a recurrent neural network model to estimate and evaluate the camera pose.

A. Recurrent Deep Model

First, we started by using blurry images as input for a camera pose model. Then, employed a Convolutional Neural Network (CNN) to extract features from these blurry images. After that, we used Recurrent Neural Networks (RNNs) to reduce the blurriness and generate clear and accurate images without any errors. To identify motion blur in an image analyzed its frequency spectrum. A presence of low-frequency components indicates motion blur, and we can determine the extent and direction of the blur by examining the orientation of these low-frequency contents. Once we identify the motion blur, we need to determine the blur kernel using a blind deconvolution. Then, we can apply a deblurring technique, such as total variation deconvolution, to remove the blur and obtain a clear copy of the original image. Additionally, we can enhance the image quality further by using denoising techniques and improving overall image quality. Even after deblurring, there might still be some noise present in the image. We can evaluate the performance of the motion blur reduction algorithm by comparing the original and deblurred images or by using the peak signal-to-noise ratio (PSNR). Through a comprehensive analysis, we can determine whether the algorithm successfully reduces motion blur. We calculated the camera pose using a recurrent neural network (RNN), and the optimization process helps improve the results.

Figure 2 illustrates the entire procedure of camera localization, while Figure 3 demonstrates the complete deblurring process. Recurrent neural networks retain information from the previous state in a single hidden layer, allowing them to improve the removal of motion blur. By discarding less important information and incorporating new data from the input, the hidden state at each time step provides relevant information with long-term connections. We proposed a modified RNN with an added recurrent unit for these operations. Equation 1 performs the blur removal operation.

$$\tilde{h_t} = RU(\tilde{h}_{t-1}, \tilde{h_t}) \tag{1}$$

As a series of higher-order memory states, the output of recurring unit (RU), $\tilde{h_t}$ is created by examining the variations of

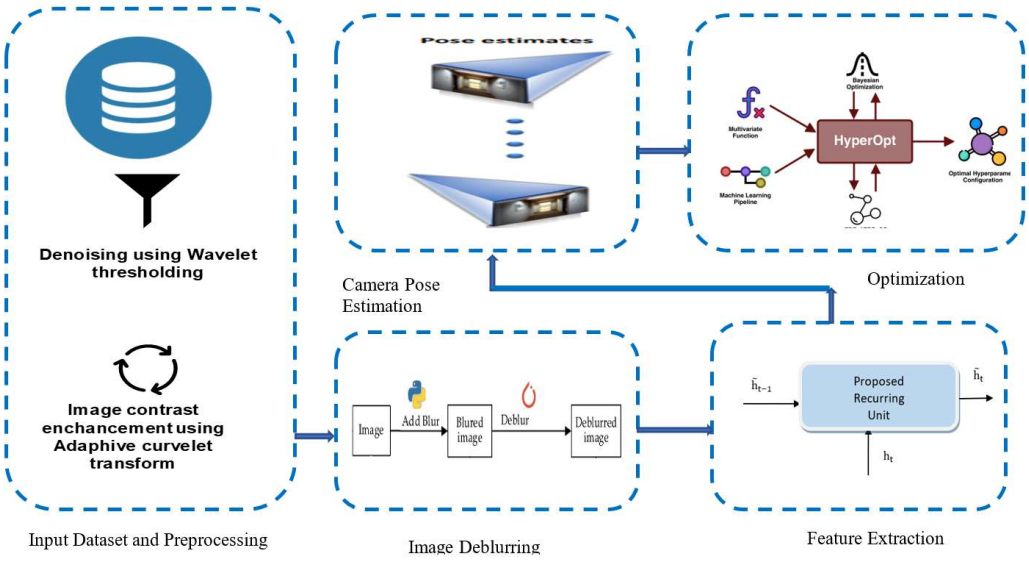

Fig. 2. Camera pose estimation from deblurred images

h_t. The recurring unit (RU) does not require additional memory to hold the information in a long connection. In this recurrent architecture $\tilde{h_t}$ is perform the blur removal performance.

B. Experimental Setup

The network is trained uniformly on different datasets, such as 7-Scenes [39], by resizing the images to 256 pixels. Subsequently, the input images were adjusted to have intensity values between -1 and 1. The ResNet34 component of the network was pretrained on the ImageNet dataset, whereas the other elements were randomly initialized. Throughout the training and testing phases, we shrank 256×256 pixel images for the network by applying an arbitrary and centralized cropping mechanism. The augmentation phase was necessary to increase the generalization capabilities of the architecture under various meteorological scenarios and the duration of the scenarios. Our methods are in Python 3.10 using PyTorch [64]and Adam solvers with a learning rate of 5×10^{-5}. On a GPU, we trained the network with a few hyperparameters, such as epoch is 20, batch size is 64, train dropout is 0.5, test dropout is 0.0, validation frequency is 5, weight decay is 0.0005, learning rate is $5e^{-05}$, weight initialization β is 0.9, and γ is 0.3.

C. Datasets

The availability of public databases for localization activities has recently increased, with a focus on utilizing deep learning approaches for image processing. Currently, only a few public databases exist, such as Microsoft Researchers 7-Scenes [59], TUM LSI [57], and InLoc [60]. To assess advanced indoor camera localization algorithms, it is crucial to have large-scale multidimensional datasets that incorporate diverse collection platforms, environments, and images. The Microsoft 7-Scenes dataset developed by Microsoft Research was utilized in this study. It contains seven indoor environments and is a widely used RGB-D dataset. The images were captured using a hand-held Kinect camera with a resolution of 640×480 pixels, and the ground-truth camera positions were obtained through Kinect fusion. Each scene was accompanied by a detailed 3D model, and there were multiple sequences of tracked RGB-D camera frames split into training and testing data. However, this dataset presents significant challenges, owing to factors such as motion blur, textureless surfaces, and repetitive structures [61]. Figure 4 shows the input, ground truth and predicted image.

D. Model Evolution Metrics

Error calculation techniques were applied to evaluate the effectiveness of the recurrent neural network model Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) are most of them.

E. Experimental Results

1) Accuracy: The recurrent deep architecture was validated by comparing its performance with those of other existing approaches. Evaluate the 7-Scenes dataset [59], we used PoseNet [39], LSTM-PoseNet [57], MapNet [62], AtLoc [63], Epi-Locb66, and CGAPoseNetb67. The expected result shows better performance than the existing approaches. Compared with the previous baseline, the precision of the position increased by 20%, and the rotation increased by 25%. Our study decreases the positional error from 0.18 m to 0.16 m and the rotational error from 6.68° to 5.31°, improving previous research as shown in table 1. Our deep architecture obtained higher accuracy using recurrent neural networks and motion blur eliination than the edge-cutting architectures.

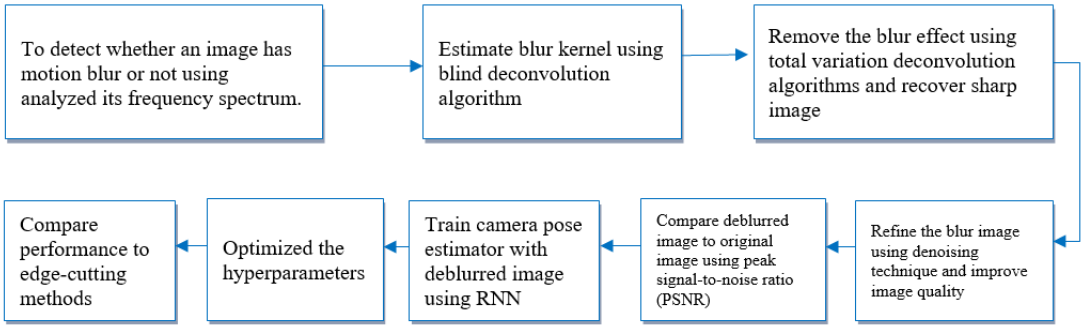

Fig. 3. Image deblurring process

TABLE I
AVERAGE POSE ERRORS OF EXISTING ALGORITHMS AND OUR PROPOSED ARCHITECTURE

Network and Ref.	Chess	Fire	Head	Office	Pumpkin	RedKitchen	Stairs	Avg. Pose Error
PoseNet [39]	0.32m,4.06°	0.47m,7.33°	0.29m, 12.00°	0.48m, 6.00°	0.47m, 4.21°	0.59m, 4.32°	0.47m, 06.93°	0.45m, 9.84°
LSTM-PoseNet [57]	0.24m, 5.77°	0.34m, 11.90°	0.21m, 13.70°	0.31m, 8.07°	0.33m, 7.00°	0.37m, 8.82°	0.41m, 13.70°	0.31m, 9.85°
MapNet [62]	0.08m, 3.25°	0.27m, 1169°	0.18m, 13.25°	0.17m, 5.15°	0.22m, 4.02°	0.23m, 4.93°	0.30m, 12.08°	0.21m, 7.78°
AtLoc [63]	0.10m, 4.07°	0.25m, 11.40°	0.16m, 11.80°	0.17m, 5.34°	0.21m, 4.37°	0.23m, 5.42°	0.26m, 10.50°	0.20m, 7.56°
EpiLoc [66]	0.07m, 2.71°	0.24m, 9.18°	0.14m, 12.6°	0.18m, 4.45°	0.18m, 3.32°	0.23m, 4.60°	0.24m, 11.00°	0.18m, 6.82°
CGAPoseNet [67]	0.26m, 6.34°	0.28m, 10.03°	0.17m, 7.98°	0.26m, 7.23°	0.22m, 5.18°	0.55m, 16.7°	0.17m, 12.00°	0.27m, 9.39°
Proposed	0.14m, 3.76°	0.18m,10.15°	0.15m,9.75°	0.14m, 5.33°	0.15m,3.01°	0.17m, 3.04°	0.19m,4.13°	0.16m, 5.31°

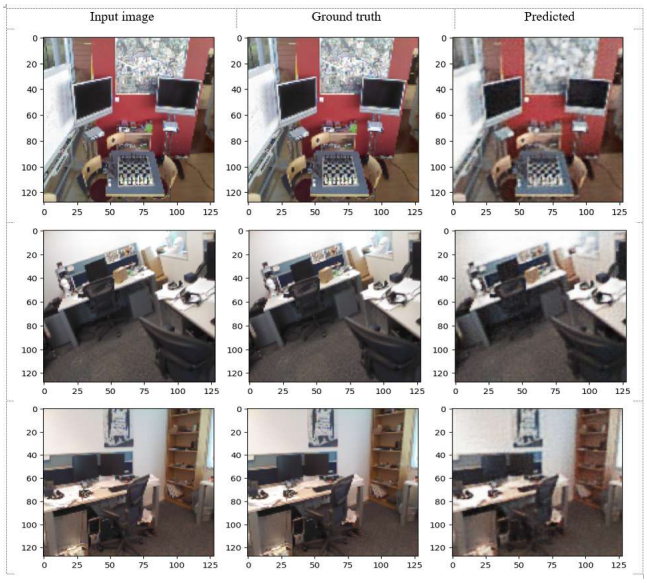

Fig. 4. Input images, ground truth images, and predicted images of 7-scenes dataset.

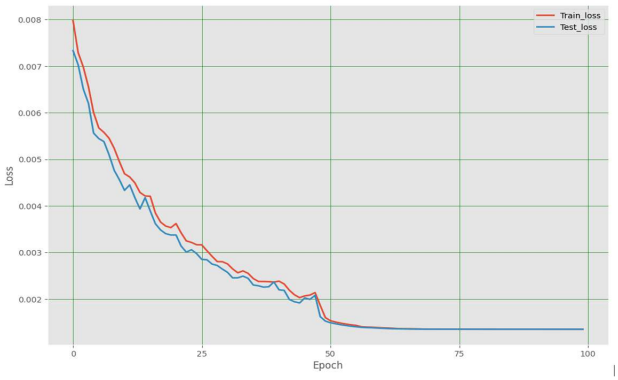

Fig. 5. Training and testing loss for proposed model

Figure 5 illustrates a plot depicting the train loss and test loss for the motion blur elimination model. The x-axis represents the number of epochs, while the y-axis displays the train and test loss. The plot reveals a pattern where the error is initially high but gradually decreases as the number of epochs increases. Eventually, the error reaches a minimum level, indicating regularization. Initially, the accuracy is low, but over time, it progressively improves and eventually stabilizes at a consistent level. Our proposed architecture achieves the best localization performance compared to existing methods, as illustrated in Table 2.

2) Speed: We examined the average running times of PoseNet [39], MapNet [57], LSTM-PoseNet [62], AtLoc [63], and our proposed recurrent deep model. MapNet takes the slowest to operate, at 9.4 ms for each frame, because it must interpret additional data from a good deal of sensory information and a sequence of images to apply geometric restrictions. LSTM-

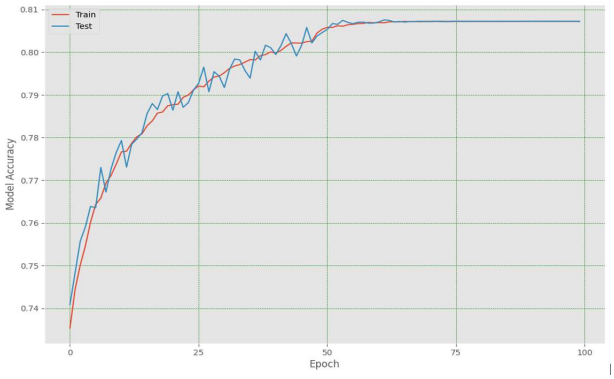

Fig. 6. Training and testing accuracy for proposed model

PoseNet has a running time of 9.2 ms per frame, which is 3.7 ms quicker than PoseNet, because of the time-consuming recursive procedures in LSTMs. AtLoc uses 6.3 milliseconds of each frame to achieve computational efficiency and localization accuracy. Our recurrent deep architecture running time is 6.0 ms.

TABLE II
RUNNING TIME COMPARISON OF EXISTING RESEARCHES AND OUR PROPOSED ARCHITECTURE.

Network	References	Running Time/Frame
PoseNet	[39]	12.9 ms
LSTM-PoseNet	[57]	9.4 ms
MapNet	[62]	9.2 ms
AtLoc	[63]	6.3 ms
Proposed	-	6.0 ms

3) Metric Evaluation: We evaluated the recurrent deep architecture using 7-Scene datasets [39], [65] using the MAE, MSE, and RMSE. The evaluation results for the three distinct metrics are presented in Table 3. MSE generated the lowest error value compared to the other two methods. Figure 6 demonstrate the training loss ands the validation loss of the GRU and Figure 6 shows the training and testing accuracy for the proposed architecture. We applied the test set after each epoch of the architecture learned during training, and evaluated the test data after each epoch. The goodness-of-fit analysis of the three architectures shows that their overall performance is fairly similar. However, compared with LSTM and Bi-LSTM, the GRU architecture is more accurate. Therefore, we deployed the GRU architecture to train the deep architecture.

IV. DISCUSSION

In recent advancements, image-based indoor-camera localization still faces several challenges. Achieving high localization accuracy requires accounting for training costs and offline training, making online implementation difficult and impractical. While 3D structure-based localization approaches

TABLE III
EVALUATION METRIC FOR DIFFERENT RECURRENT NEURAL NETWORKS.

Error	Error for different RNN architecture		
	LSTM	***Bi-LSTM***	***GRU***
MAE	0.2135	0.1964	0.1617
MSE	0.0841	0.0730	0.0699
RMSE	0.2876	0.2766	0.2644

can achieve excellent indoor performance, they struggle to adapt to changing environments, repeated elements, textures with low surface area, illumination variations, motion blur, and limited viewpoint adjustments, which significantly affects localization results. The need for motion deburring to reduce motion blur in the input images is one of the limitations of this architecture. Although the architecture works well with regulated motion blur in synthetic data, it may not adapt well to real-world situations with unexpected motion blur. The proposed architecture additionally requires an extensive collection of indoor scenes with real camera poses for training. Such a dataset can be challenging to gather, which may restrict the application of the architecture in certain circumstances. Finally, changes in the environment, including changes in lighting or scene complexity, may affect architecture performance. The architecture may find it challenging to predict camera poses effectively because of these modifications, which could restrict its applicability in certain situations.

V. CONCLUSION

Image-based indoor camera localization is a complex and vital issue. This research provides an indoor camera localization deep architecture using a recurrent neural network. Our approach allows the architecture to learn geometrically robust features and reduce motion blur issues. Using the selected datasets, we showed that our method outperforms state-of-the-art techniques. Because of the large variety of scene appearances and movements, our research provides a new approach to camera localization by minimizing motion blur. The proposed recurrent deep architecture can help learn robust geometrical features and reduce the effects of shifting motion blurs. We have shown state-of-the-art results in demand settings. In the future, we will refine the recurrent deep architecture further and test it to see whether it can enhance multi-sensor camera pose regression.

VI. FUTURE WORK

The utilization of various sensors, such as LiDAR, Wi-Fi, IMU, and Bluetooth, can provide detailed indoor localization information, but the challenge lies in handling the diverse features of each sensor to achieve precise positioning. To address this challenge, a more powerful and accurate localization system can be developed by effectively combining the data from multiple sensors. The semantics of features in real-world scenes plays a crucial role in achieving accurate

positioning outcomes by simplifying the exclusion of dynamic object attributes. Thus, semantic architecture can be useful for indoor positioning systems in the future. In addition, a multi-camera setup can enhance robotic applications by providing a 360-degree panoramic field of view. To expedite the feature calculation time, a potential direction for future research is to develop a recurrent neural network system to assess features from numerous cameras end-to-end. The proposed architecture uses a series of images and requires considerable computing time. Future research could investigate making the architecture more efficient for real-time use on low-power gadgets, such as cell phones or embedded systems. This would allow for a more extensive range of architecture applications.

ACKNOWLEDGMENT

The authors wish to thank to Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

REFERENCES

[1] M. Sewtz, X. Luo, J. Landgraf, T. Bodenmüller, and R. Triebel, "Robust approaches for localization on multi-camera systems in dynamic environments," in 2021 7th International Conference on Automation, Robotics and Applications (ICARA), IEEE, 2021, pp. 211–215.

[2] M. S. Alam, F. B. Mohamed, A. Selamat, and A. B. Hossain, "A Review of Recurrent Neural Network Based Camera Localization for Indoor Environments," IEEE Access, vol. 11, 2023.

[3] R. Brylka, U. Schwanecke, B. Bierwirth, "Camera based barcode localization and decoding in real-world applications," in 2020 International Conference on Omnilayer Intelligent Systems (COINS), 2020, pp. 1–8, IEEE.

[4] R. Brylka, U. Schwanecke, and B. Bierwirth, "Camera based barcode localization and decoding in real-world applications," in 2020 International Conference on Omni-layer Intelligent Systems (COINS), IEEE, 2020, pp. 1–8. [

[5] E. Brachmann et al., "Dsac-differentiable ransac for camera localization," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 6684–6692.

[6] H. Yao, R. W. Stidham, Z. Gao, J. Gryak, and K. Najarian, "Motion-based camera localization system in colonoscopy videos," Med Image Anal, vol. 73, p. 102180, 2021.

[7] P.-E. Sarlin, A. Unagar, M. Larsson, H. Germain, C. Toft, V. Larsson, M. Pollefeys, V. Lepetit, L. Hammarstrand, F. Kahl, et al., "Back to the feature: Learning robust camera localization from pixels to pose," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 3247–3257.

[8] I. Arrouch, N. S. Ahmad, P. Goh, J. MohamadSaleh, "Close proximity time-to-collision prediction for autonomous robot navigation: An exponential gpr approach," Alexandria Engineering Journal, vol. 61, no. 12, pp. 11171–11183, 2022.

[9] A. Raza, L. Lolic, S. Akhter, and M. Liut, "Comparing and evaluating indoor positioning techniques," in 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2021, pp. 1–8.

[10] J. Zhang, H. Mao, "Wknn indoor positioning method based on spatial feature partition and basketball motion capture," Alexandria engineering journal, vol. 61, no. 1, pp. 125–134, 2022.

[11] C. E. A. Bundak, M. A. Abd Rahman, M. K. A. Karim, N. H. Osman, "Fuzzy rank cluster top k euclidean distance and triangle based algorithm for magnetic field indoor positioning system," Alexandria Engineering Journal, vol. 61, no. 5, pp. 3645–3655, 2022.

[12] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang, "Deep video deblurring for hand-held cameras," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1279–1288.

[13] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 769–777.

[14] D. Rozumnyi, M. R. Oswald, V. Ferrari, M. Pollefeys, "Motion-from-blur: 3d shape and motion estimation of motion-blurred objects in videos," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 15990–15999.

[15] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1628–1636.

[16] J. F. Schmid, S. F. Simon, R. Radhakrishnan, S. Frintrop, and R. Mester, "HD Ground-A Database for Ground Texture Based Localization," in 2022 International Conference on Robotics and Automation (ICRA), IEEE, 2022, pp. 7628–7634.

[17] P. Wieschollek, B. Schölkopf, H. Lensch, and M. Hirsch, "End-to-end learning for image burst deblurring," in asian conference on computer vision, Springer, 2016, pp. 35–51.

[18] W. Yang, X. Zhang, H. Ma, G. Zhang, "Laser beamsbased localization methods for boom-type roadheader using underground camera non-uniform blur model," IEEE Access, vol. 8, pp. 190327–190341, 2020.

[19] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," Adv Neural Inf Process Syst, vol. 29, 2016.

[20] A. Chakrabarti, "A neural approach to blind motion deblurring," in European conference on computer vision, Springer, 2016, pp. 221–235.

[21] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in IEEE international conference on computational photography (ICCP), IEEE, 2013, pp. 1–8.

[22] D. Gong et al., "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2319–2328.

[23] F. Pirahansiah, S. Sahran, S. N. H. S. Abdullah, "Camera calibration and video stabilization framework for robot localization," Control Engineering in Robotics and Industrial Automation: Malaysian Society for Automatic Control Engineers (MACE) Technical Series 2018, pp. 267–287, 2022.

[24] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8174–8182.

[25] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8183–8192.

[26] S. Zhang, A. Zhen, R. L. Stevenson, "Deep motion blur removal using noisy/blurry image pairs," Journal of Electronic Imaging, vol. 30, no. 3, pp. 033022–033022, 2021.

[27] T. H. Kim, S. Nah, K. M. Lee, "Dynamic video deblurring using a locally adaptive blur model," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 10, pp. 2374–2387, 2017.

[28] P. Wieschollek, M. Hirsch, B. Scholkopf, H. Lensch, "Learning blind motion deblurring," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 231–240.

[29] M. D. Kim, J. Ueda, "Real-time image de-blurring and image processing for a robotic vision system," in 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 1899–1904, IEEE.

[30] M. Tian, Q. Nie, H. Shen, "3d scene geometry-aware constraint for camera localization with deep learning," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 4211–4217, IEEE.

[31] J. Gast, A. Sellent, S. Roth, "Parametric object motion from blur," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1846– 1854.

[32] K. Purohit, S. Vasu, M. P. Rao, A. Rajagopalan, "Multiplanar geometry and latent image recovery from a single motion-blurred image," Machine Vision and Applications, vol. 33, no. 1, p. 10, 2022.

[33] S. Klenk, L. Koestler, D. Scaramuzza, D. Cremers, "E-nerf: Neural radiance fields from a moving event camera," IEEE Robotics and Automation Letters, 2023.

[34] D. Park, D. U. Kang, J. Kim, S. Y. Chun, "Multitemporal recurrent neural networks for progressive non-uniform single image deblurring with

incremental temporal training," in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23– 28, 2020, Proceedings, Part VI 16, 2020, pp. 327–343, Springer.
[35] J. Pan, H. Bai, J. Tang, "Cascaded deep video deblurring using temporal sharpness prior," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3043–3051.
[36] T. Xie, K. Dai, K. Wang, R. Li, J. Wang, X. Tang, L. Zhao, "A deep feature aggregation network for accurate indoor camera localization," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 3687–3694, 2022.
[37] J. Yu, F. Gao, J. Cao, C. Yu, Z. Zhang, Z. Huang, Y. Wang, H. Yang, "Cnn-based monocular decentralized slam on embedded fpga," in 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2020, pp. 66–73, IEEE.
[38] W. Zhang, J. Kosecka, "Image based localization in urban environments," in Third international symposium on 3D data processing, visualization, and transmission (3DPVT'06), 2006, pp. 33–40, IEEE.
[39] A. Kendall, M. Grimes, R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2938–2946.
[40] A. Kendall, R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization," in 2016 IEEE international conference on Robotics and Automation (ICRA), 2016, pp. 4762–4769, IEEE.
[41] I. Melekhov, J. Ylioinas, J. Kannala, E. Rahtu, "Imagebased localization using hourglass networks," in Proceedings of the IEEE international conference on computer vision workshops, 2017, pp. 879–886.
[42] K. Liu, Q. Li, G. Qiu, "Posegan: A pose-to-image translation framework for camera localization," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 166, pp. 308–315, 2020.
[43] F. Ott, T. Feigl, C. Loffler, C. Mutschler, "Vipr: visualodometry-aided pose regression for 6dof camera localization," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 42–43.
[44] M. Müller, S. Urban, B. Jutzi, "Squeezeposenet: Image based pose regression with small convolutional neural networks for real time uas navigation," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 4, p. 49, 2017.
[45] C. Wang, B. Luo, Y. Zhang, Q. Zhao, L. Yin, W. Wang, X. Su, Y. Wang, C. Li, "Dymslam: 4d dynamic scene reconstruction based on geometrical motion segmentation," IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 550–557, 2020. 10
[46] M. Li, J. Qin, D. Li, R. Chen, X. Liao, B. Guo, "Vnlstmposenet: A novel deep convnet for real-time 6-dof camera relocalization in urban streets," Geo-Spatial Information Science, vol. 24, no. 3, pp. 422–437, 2021.
[47] R. Clark, S. Wang, A. Markham, N. Trigoni, H. Wen, "Vidloc: A deep spatio-temporal model for 6-dof video-clip relocalization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6856–6864.
[48] F. Xue, X. Wu, S. Cai, J. Wang, "Learning multi-view camera relocalization with graph neural networks," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11372–11381, IEEE.
[49] X. Huang, M. Halwani, R. Muthusamy, A. Ayyad, D. Swart, L. Seneviratne, D. Gan, Y. Zweiri, "Realtime grasping strategies using event camera," Journal of Intelligent Manufacturing, vol. 33, no. 2, pp. 593–615, 2022.
[50] X. Peng, K. Saenko, "Synthetic to real adaptation with generative correlation alignment networks," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 1982–1991, IEEE.
[51] E. Bayraktar, C. B. Yigit, P. Boyraz, "A hybrid image dataset toward bridging the gap between real and simulation environments for robotics: Annotated desktop objects real and synthetic images dataset: Adoreset," Machine Vision and Applications, vol. 30, no. 1, pp. 23–40, 2019.
[52] M. Rad, M. Oberweger, V. Lepetit, "Feature mapping for learning fast and accurate 3d pose inference from synthetic images," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4663–4672.
[53] I. Ha, H. Kim, S. Park, H. Kim, "Image retrieval using bim and features from pretrained vgg network for indoor localization," Building and Environment, vol. 140, pp. 23–31, 2018.
[54] D. Acharya, K. Khoshelham, S. Winter, "Bim-posenet: Indoor camera localisation using a 3d indoor model and deep learning from synthetic images," ISPRS journal of photogrammetry and remote sensing, vol. 150, pp. 245–258, 2019.
[55] D. Acharya, S. Singha Roy, K. Khoshelham, S. Winter, "A recurrent deep network for estimating the pose of real indoor images from synthetic image sequences," Sensors, vol. 20, no. 19, p. 5492, 2020.
[56] N. Li, H. Ai, "Efiloc: large-scale visual indoor localization with efficient correlation between sparse features and 3d points," The Visual Computer, pp. 1–16, 2022.
[57] F. Walch, C. Hazirbas, L. Leal-Taixe, T. Sattler, S. Hilsenbeck, D. Cremers, "Image-based localization using lstms for structured feature correlation," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 627–637.
[58] M. S. Alam, A. K. M. B. Hossain, and F. B. Mohamed, "Performance Evaluation of Recurrent Neural Networks Applied to Indoor Camera Localization," International Journal of Emerging Technology and Advanced Engineering, vol. 12, no. 8, 2022
[59] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, A. Fitzgibbon, "Scene coordinate regression forests for camera relocalization in rgb-d images," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2930–2937.
[60]] H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, A. Torii, "Inloc: Indoor visual localization with dense matching and view synthesis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7199–7209.
[61] N. Radwan, A. Valada, W. Burgard, "Vlocnet++: Deep multitask learning for semantic visual localization and odometry," IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 4407–4414, 2018.
[62] S. Brahmbhatt, J. Gu, K. Kim, J. Hays, J. Kautz, "Geometry-aware learning of maps for camera localization," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2616–2625.
[63] B. Wang, C. Chen, C. X. Lu, P. Zhao, N. Trigoni, A. Markham, "Atloc: Attention guided camera localization," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 10393–10401.
[64] S. Imambi, K. B. Prakash, and G. R. Kanagachidambaresan, "PyTorch," Programming with TensorFlow: Solution for Edge Computing Applications, pp. 87–104, 2021.
[65] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi, "Real-time RGB-D camera relocalization," in 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, 2013, pp. 173–179.
[66] L. Xu, T. Guan, Y. Luo, Y. Wang, Z. Chen, and W. Liu, "EpiLoc: Deep Camera Localization Under Epipolar Constraint.," KSII Transactions on Internet & Information Systems, vol. 16, no. 6, 2022.
[67] A. Pepe and J. Lasenby, "Cga-posenet: Camera pose regression via a 1d-up approach to conformal geometric algebra," arXiv preprint arXiv:2302.05211, 2023.
[68] A. Abozeid, A. I. Taloba, R. M. Abd El-Aziz, A. F. Alwaghid, M. Salem, and A. Elhadad, "An Efficient Indoor Localization Based on Deep Attention Learning Model.," Comput. Syst. Sci. Eng., vol. 46, no. 2, pp. 2637–2650, 2023.
[69] Y. Cho, S. Eum, J. Im, Z. Ali, H. Choo, and U. Park, "Deep Photo-Geometric Loss for Relative Camera Pose Estimation," IEEE Access, 2023.
[70] M. Lyu, X. Guo, K. Zhang, and L. Zhang, "A Visual Indoor Localization Method Based on Efficient Image Retrieval," Journal of Computer and Communications, vol. 12, no. 2, pp. 47–66, 2024.
[71] Z. Xiao, C. Chen, S. Yang, and W. Wei, "EffLoc: Lightweight Vision Transformer for Efficient 6-DOF Camera Relocalization," arXiv preprint arXiv:2402.13537, 2024.
[72] M. S. Alam, F. B. Mohamed, and A. K. M. B. Hossain, "Self-Localization of Guide Robots Through Image Classification," Baghdad Science Journal, vol. 21, no. 2 (SI), p. 832, 2024.