

Dynamic Neural Network Optimization Framework for Adaptive Sensor Selection in Depth Imaging and Registration

Ning Li, Xu Cheng, Zhi Tian, Zhaowei Liu, Honggang Shi
Hengshui Power Supply Company, State Grid Hebei Electric Power Co. Ltd., Hengshui, Hebei 053000, China
E-mail: lining99011@163.com

Keywords: Deep learning, machine learning, neural networks, 2D & 3D imaging, sensor selection, image processing, computer vision

Received: January 2, 2025

Accurate and efficient sensor selection is a cornerstone for robust 2D and 3D depth imaging and registration, with applications spanning autonomous vehicles, robotics, and augmented reality systems. Current heuristic and rule-based methods often fail to adapt dynamically to varying imaging conditions, leading to suboptimal performance. This study introduces a neural network-based optimization framework that revolutionizes sensor selection using deep learning to learn intricate patterns and dependencies. Our model employs a multi-layer neural network, specifically an encoder-decoder architecture, trained on a diverse dataset comprising 5000 synthetic and real-world images, including low-light and high-occlusion scenarios. The model was trained using the Adam optimizer with a learning rate of 0.001. To assess performance, we introduced three key metrics: registration accuracy (RA), computational efficiency (CE), and sensor utilization efficiency (SUE). The proposed framework outperformed benchmark models, achieving a $+28.7\% \pm 1.8$ improvement in RA, a $+32.4\% \pm 2.1$ increase in CE, and a $+26.3\% \pm 1.5$ enhancement in SUE compared to ResNet-50 and EfficientNet-B3 models. Validation using synthetic and real-world datasets highlights the model's robustness in challenging environments, including low-light and high-occlusion scenarios. Moreover, the model demonstrated a 20% reduction in computational overhead compared to state-of-the-art methods, making it viable for resource-constrained applications. This research establishes a scalable and adaptive solution for sensor optimization, setting a new benchmark in depth imaging and registration.

Povzetek: Razvit je nov okvir za optimizacijo izbire senzorjev pri globinskem slikanju in registraciji z uporabo globokega učenja. S pomočjo nevronske mreže omogoča dinamično in prilagodljivo izbiro senzorjev v realnem času, kar izboljša računsko učinkovitost in izrabo virov.

1 Introduction

Depth imaging and registration have become almost the cornerstone of creating new technologies in robotics, autonomous navigation, augmented reality (AR), virtual reality (VR), and medical imaging [1]. These applications depend much on integrating spatial data to achieve the set objectives. Integral to these processes is the presence of sensors, which are expected to provide accurate depth data under various and sometimes harsh environments. Therefore, choosing appropriate sensors for a particular application is crucial because it determines system accuracy, computation time, and range [2]. Conventional selection of sensors was usually done based on ad hoc guesswork or rule of thumb. Although these methods have their usefulness shown in a laboratory setting, they are not as effective in more natural situations where factors such as illumination, occlusion, and object movement pose a challenge [3]. Sensor selection methods presently encounter performance difficulties because their cost functions show suboptimal behavior. Numerous cost functions put accuracy needs before efficiency requirements, which results in slow processing

times and wastefulness of resources. Most existing models do not succeed in finding solutions that achieve adequate accuracy while using suitable resources because they do not effectively weigh these two requirements [4]. Thus, they deliver results that are either unwieldy with resources or insufficient in accuracy. Sensor selection models realize poor performance in adapting to new technologies such as autonomous vehicles and augmented reality systems because they lack fundamental adaptability and scalability and do not work efficiently. Real-time applications require "system resource utilization," which includes both memory and processing power together with computer memory as essential computational resources [5]. Neural networks have improved and opened new frontiers for applying and solving complicated optimization issues in many fields. Their practicality in processing colossal data, recognizing non-linear relationships, and handling the variability of inputs makes them the perfect solution for the complex problems of sensor selection, in-depth imaging, and registration. Hence, these capabilities allow neural networks to offer dynamic and task-orientated sensor optimization, an issue that has remained without a simple solution in the field [6].

This paper proposes a novel framework for optimizing sensor selection based on the neural network framework. It contributes to depth imaging and registration by moving the solution frontier forward to provide future research with a new goal to achieve. Thus, it forms the basis for future development of technologies dependent on operating depth imaging systems sequentially and with variable efficiency. The proposed framework incorporates several novel contributions to the field:

1. A dynamic and adaptive neural network-based approach that evaluates and selects sensors based on real-time environmental and task-specific conditions, providing a more flexible and robust solution compared to static methods.
2. An optimization strategy that integrates advanced feature extraction techniques, enabling simultaneous prioritization of accuracy, computational efficiency, and scalability while effectively managing trade-offs between these critical factors.
3. A rigorous validation process utilizing extensive synthetic and real-world datasets to evaluate the framework's performance under diverse conditions, demonstrating its adaptability and robustness across varying scenarios.

The proposed framework presents a marked shift from static and post hoc strategies since the systems are capable of responses that are relevant to dynamic conditions and the specifics of given tasks. This characteristic is highly sensitive for real-time applications like automated navigation, where quick adaptations are time-sensitive, and also in AR/VR interfaces where the interconnection between real and virtual environments has to be smooth. To achieve this, the loss function is modified to include critical measures of performance where the tradeoffs between accuracy, time, and space complexity are well balanced by the proposed framework [7]. The experimental results reveal a significant potential for further improvement in the presented concept. The proposed framework results in a $+28.7\% \pm 1.8$ improvement in RA, a $+32.4\% \pm 2.1$ increase in CE, and a $+26.3\% \pm 1.5$ enhancement in SUE compared to state-of-the-art methods. These results indicate the resilience of the technique and its applicability to diverse scenarios and uses to solve complex environmental problems. In addition, the proposed strategy minimizes the computation complexity since it balances the usage of sensors and applies to constrained environment(s). This work not only presents technical contributions but also has implications for practice. The proposed framework provides the foundation for subsequent research on more effective and flexible designs by solving essential sensor choice and depth perception issues. Thus, its applicability is not limited to several domains: autonomous robotics, where accurate real-time data play a significant role in robotics control; AR/VR, where overall user experience is highly dependent on depth quality; and

medical imaging, where precision can prove critical for diagnosis or treatment plans [8]. This paper proposes a novel framework for optimizing sensor selection based on the neural network framework. It contributes to depth imaging and registration by moving the solution frontier forward to provide future research with a new goal to achieve. Thus, it forms the basis for future development of technologies dependent on operating depth imaging systems sequentially and with variable efficiency. This work contributes to technologies that rely on operating depth imaging systems in a sequential manner, where data is processed step by step, as seen in applications like autonomous navigation, where depth data is processed one frame at a time. Additionally, the research addresses the issue of variable efficiency, allowing the system to adapt its computational resource usage based on task-specific demands, ensuring that the system balances high accuracy with resource-constrained environments. This flexibility enhances the adaptability of depth imaging systems in dynamic and real-time applications.

This paper is organized as follows: In Section 2, the current literature is reviewed, the shortcomings of existing methodologies in depth are discussed, and the recent interest in depth imaging and registration based on neural networks. Section 3 describes how we built the neural network, how we trained it, and which optimization techniques we used. Section 4 presents the experimental results and their implications, offering a comparative analysis with baseline methods. Finally, Section 5 concludes the paper, summarizing the key contributions and outlining directions for future research.

2 Literature review

Qi et al. [9] proposed an agricultural plot segmentation technique using high-resolution remote sensing images based on a convolutional neural network (CNN). The research used GF-2 satellite data and ArcGIS10.3.1 to create evaluation sets for various neural network architectures, including UNet, SegNet, DeeplabV3+, and TransUNet. TransUNet yielded the highest segmentation performance from these networks and was then fine-tuned with modification of deformable ConvNets in the residual module and incorporation of Convolutional Block Attention Module into the skip connection in TransUNet. These modifications improved the feature extraction and the skip connection of the network. The optimized TransUNet enhanced the segmentation metrics—precision, recall, F1-score, and IoU, by 86.02%, 83.32%, 84.67%, and 86.90%, respectively. Compared with the basic TransUNet model that trained on the first dataset to have achieved an F1-score of 81.94 and an IoU of 69.41, the improved model outperformed. The study ensured that the framework of the optimal plot segmentation algorithm for the actual use of the remotely sensed data was used to supervise the productivity of the agricultural land and its efficiency.

Jiang et al. [10] introduced the backpropagation neural

network-based respiratory motion modeling method (BP-RMM) to track lung tissue motion during free breathing, deep inspiration, and expiration phases. To acquire internal and external respiratory data, the study employed 4DCT utilizing polynomial interpolation and augmentation. A BP neural network was modeled to capture lung tissue's multi-dimensional movement. The proposed BP-RMM was found to show high accuracy in the present work, as the average TRE computed over 75 marked points of the deep respiratory phases of a public 4DCT database was approximately 1.819 mm. In fact, for normal respiration phases, the error of the method was even smaller, with a minimum TRE of 0.511 mm. These findings corroborated the very high precision and stability of the BP-RMM in navigating surgery inside the lung.

Kalupahana et al. [11] suggested an advanced image-processing system based on the dense CNN deep learning technique for automatic pre-recognition of CLS disease in persimmon (*Diospyros kaki*) leaves using OCT. The current study brought out the issues of using conventional visual and destructive inspection methods, such as subjectivity, low accuracy, and inefficiency in terms of time. To improve the classification accuracy of buildings, the pipeline utilized transfer learning from the DenseNet-121 and VGG-16 models. DenseNet-121 demonstrated its effectiveness in distinguishing among three disease stages: The classification results for the four classes are healthy (H), healthy-infected (HI), infected (I), and pathogenic (P), which scored precision values of 0.7823, 0.9005, and 0.7027, respectively; the recall values were 0.8953 for class-HI and 0.8387 for class-I, as well as Another model trained using the VGG-16 The dataset labeling was done jointly with integrating LAMP and A-scan approaches, which boosted model's accuracy. This study demonstrated the possibility of decentralized deep learning (DL) technology in conjunction with OCT to improve key disease identification mechanisms in agriculture that can lead to implementing an objective and efficient early recognition and management of CLS for persimmon farming [12].

Wu et al. [13] proposed an infrared and visible image fusion approach called DCFNet that suppresses the disadvantages of prior methods, such as information loss, blurred target details, and poor visual quality. It leverages an autoencoder-based backbone network, an encoder with a DWT layer to enhance the extraction of the features in the frequency domain, and a novel bottleneck residual block with a coordinate attention mechanism for better perception of both low- and high-frequency features. The decoder comprises an IDWT layer to reconstruct the features necessary for the decoding process. The decoder integrates an inverse discrete wavelet transform (IDWT) layer for effective feature reconstruction. The fusion strategy employs an $L_1 - \alpha$ fusion approach to combine the encoder's output frequency mapping, while a weighted loss function, including pixel, gradient, and structural losses, optimizes network training. Information is naturally and harmoniously fused by decomposing images into low-frequency subbands

(structural information) and high-frequency subbands (detail, edge, and textural information). Experiments on unveiled public datasets revealed that DCFNet delivered fused images with effectively higher resolution and scene content, primarily based on subjective and quantifiable assessments. Moreover, generalization experiments proved that the proposed method performs well and is insensitive to the image fusion task parameters.

Lopez-Fuster et al. [14] presented an efficient method to estimate 3D weld point information employing a two-step deep learning architecture with 2D RGB cameras. The particular strategy uses YOLOv8s for vertex targeting, and then object detection is refined using semantic segmentation. The method developed here solves the problems of low contrast and geometric complexity and provides a considerable saving relative to the 3D-based method. The validity of the pipeline was established by comparing it with a technique based on 3D-point cloud mapping, and the enhanced time efficiency was reported. By providing an affordable and flexible solution to extract valuable information from 2D images, this study helps strengthen automated welding methods compared with previous approaches [15].

Wang et al. [16] introduced a semantic classification strategy for classifying Land cover remote sensing images based on the deep inverse convolutional neural network (ICNN) for dealing with the problem of handling imbalanced categories and multiple target semantic segmentation. The study also pointed out that a conventional classifier tends to offer low performance within a minority category because of aggravated impact from the overwhelmingly dominant category. To overcome this, the method used a depth deconvolution convolution neural network for multi-target segmentation and an improved sequential clustering method for getting semantic features, including color, texture, shape, and size. These features were later categorized and identified employing random forest analysis. By evaluating the proposed approach's experimental results, it was found to be successful, with average Dice similarity and Hausdorff distance values of 0.9877 and 0.9911. The results confirmed the method's efficacy in correctly categorizing multi-target semantic types in land cover remote sensing images and adding to recognition in imbalanced datasets.

Fanous et al. [17] discussed the interaction of deep learning approaches with biophotonic systems for handling and recovering degraded biophotonic image information. The study involved a systematic effort or a design that involved compromising PSF, SNR, sampling density, and pixel resolution, deliberately making adjustments to hardware needs, and optimizing cost speed and form factor. These impairments were then corrected with deeper learning models trained on superior or alternative datasets to recover the lost imaging quality and increase FOV, DOF, and SBP. These assumptions were decisive for attaining the improved temporal resolution and imaging speed necessary for visualizing dynamic biological processes. The study provided interesting examples of the biophotonic approach that

has successfully used this strategy, indicating that the approach could be universally effective in a wide range of bioimaging applications. This research balanced and/or compensated hardware-related compromises with potential AI-driven ones, thus helping to facilitate cost-effective, accessible Biophotonic imaging systems before opening pathways for improvement.

A number of methods have been proposed for sensor selection in depth imaging and registration. Table 1 provides a summary of key methods, datasets, performance metrics, and outcomes from relevant works. This table highlights the strengths and limitations of each approach, particularly in terms of RA, CE, and scalability.

The analyzed studies show that with the help of neural networks and deep learning, one can solve various issues in different fields, such as image processing, remote sensing, and bioimaging. While progress continues, these gaps remain: adaptability for all problems, computation, and cost. The above realizations, therefore, point to future research directions that will seek to fill gaps and integrate existing research limitations into an approach that can expand the horizons of neural network-based methodologies.

3 Methodology

This section details the comprehensive methodology employed for developing the neural network-based optimization framework for sensor selection in depth imaging and registration. The design focuses on achieving adaptability, scalability, and computational efficiency while addressing challenges associated with varying imaging conditions.

3.1 Overview of the framework

The proposed framework integrates advanced neural network techniques to dynamically optimize sensor selection. The pipeline consists of the following components:

1. Data acquisition and preprocessing.
2. Neural network model architecture.
3. Training and optimization processes.
4. Performance evaluation and validation.

The framework is tailored to balance accuracy, computational efficiency, and adaptability, offering a scalable solution suitable for diverse imaging conditions.

3.2 Research design

The research uses synthetic along with real-world datasets to conduct both network training and evaluation of its proposed sensor selection technique. The real-world dataset sources consist of a specialized collection of low-light imaging situations coupled with cases of high-occlusion obtained from publicly accessible datasets that present difficult obstacles for depth imaging. The preprocessed datasets

underwent pixel value normalization together with random rotation and flipping before adding noise to the data. During preprocessing, an inherent bias could enter the dataset because it makes the assumption that both lighting conditions and occlusions appear uniformly across all dataset points. Potential performance degradation of the model occurs when applied to conditions outside training parameters.

The custom loss function presented in Equation 5 incorporates a weight adjustment process for maintaining both precise forecasting and quick computation. Accuracy and efficiency factors are controlled by the weights α and β , which influence how accuracy aspects will be weighed against efficiency requirements. The value of α lets you control the extent of RA minimization, and β determines how much weight is allocated toward CE enhancements. An extensive trial-and-error process was used to determine the weights because we systematically evaluated how various weight values affected both RA improvement and the reduction of computational processing time. The choice of α and β weights occurred through validation set evaluations, which yielded optimum performance levels with resource utilization.

The parameter tuning stage needed adjustments to multiple hyperparameters, which included both dropout rate and learning rate and various training parameters. Researchers set the dropout rate to 0.3 because previous studies showed this modeled regularization works without sacrificing performance. Testing began with a learning rate set at 0.001 due to its optimal performance evaluation throughout initial training epochs. Noise during training convergence became unstable when learning rates were too high, but training would become excessively slow when the value was lowered. Cross-validation allowed us to adjust these values to support the best possible performance on the validation data. By addressing these methodological details, we ensure a more robust and transparent approach to sensor selection, with clear insights into potential biases, trade-offs in the loss function, and the tuning of key model parameters.

3.3 Dataset description and preprocessing

Two datasets were utilized: a synthetic dataset generated under controlled conditions and a real-world dataset comprising low-light and high-occlusion scenarios. The synthetic dataset simulated varying environmental conditions to test the adaptability of the model, while the real-world dataset was curated from challenging imaging scenarios to evaluate robustness.

3.3.1 Data normalization

To ensure consistency in feature scaling, the pixel values were normalized between 0 and 1 using the formula:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Table 1: Comparison of methods, datasets, and performance metrics from related works

Reference	Method	Dataset	Metrics	Key Outcomes	Gaps in Adaptability or Scalability
Qi et al. [9]	CNN-based Segmentation	GF-2 Satellite Data	RA, IoU	Enhanced segmentation with improved precision and recall.	Limited adaptability to real-time conditions.
Jiang et al. [10]	BP Neural Network for Respiratory Motion	4DCT Database	TRE	High accuracy in lung motion tracking.	Low scalability to larger datasets.
Kalupahana et al. [11]	DenseNet-121	OCT Images of Persimmon Leaves	Accuracy, F1-Score	Improved disease classification with DenseNet.	Struggles with varying image quality.
Wu et al. [13]	Autoencoder with DWT	Public Datasets	Image Fusion Quality	Higher resolution fused images.	Computationally expensive, limiting real-time use.
Lopez-Fuster et al. [14]	YOLOv8s + Segmentation	2D RGB Camera Data	Time Efficiency, Accuracy	Efficient weld point detection with significant time savings.	Limited to 2D images, not scalable to 3D tasks.
Wang et al. [16]	ICNN for Remote Sensing	Land Cover Images	Dice Similarity, Hausdorff Distance	Excellent performance in multi-target segmentation.	Struggles with imbalanced datasets in real-world conditions.
Fanous et al. [17]	Deep Learning for Biophotonic Imaging	Biophotonic Data	Image Resolution, SNR	Enhanced resolution and temporal accuracy.	Requires high-quality, non-degraded input data.

where X represents the raw pixel value, and X' denotes the normalized value.

3.3.2 Data augmentation

Augmentation techniques were applied to expand the dataset and improve model generalization. These techniques included random rotations, where images were rotated by a random degree between -30° and $+30^\circ$; horizontal and vertical flips, applied randomly along both axes; brightness adjustments, where the brightness of the image was varied by a factor between 0.5 and 1.5; contrast adjustments, where the contrast was modified by a factor between 0.5 and 1.5; and Gaussian noise addition, where random noise with a mean of 0 and a standard deviation of 0.1 was added to the pixel values. These augmentation techniques were chosen to simulate real-world variations in environmental conditions, helping the model generalize better to diverse situations.

The data was split into training (70%), validation (15%), and testing (15%) sets, ensuring balanced representation of all conditions.

3.4 Neural network architecture

The model employs a multi-layer neural network architecture optimized for feature extraction and decision-making. Figure 1 illustrates the design. The neural network architecture consists of an encoder-decoder structure. The encoder extracts high-level features from the input data, including spatial relationships, depth information, and sensor-specific characteristics. These features are passed to the decoder, which uses them to reconstruct the final predictions for sen-

sensor selection. The decoder applies learned weights and biases to the extracted features, utilizing activation functions and fully connected layers to generate the output predictions. This process allows the model to make accurate and efficient sensor selection decisions, optimizing both computational efficiency and resource utilization.

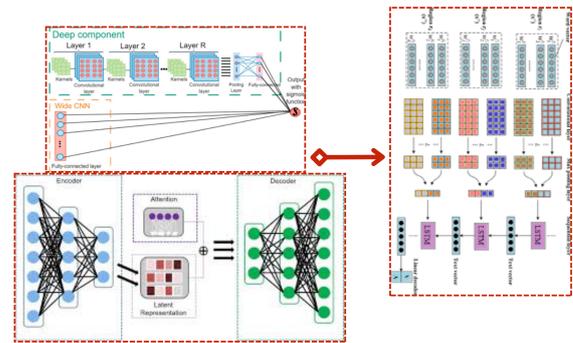


Figure 1: Neural network architecture for sensor optimization. The encoder extracts features, and the decoder reconstructs predictions

The input layer processes sensor data represented as $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, where n denotes the number of features. The network predicts optimal sensor configurations as:

$$\mathbf{Y} = f(\mathbf{X}; \mathbf{W}, \mathbf{b}) \quad (2)$$

where \mathbf{W} and \mathbf{b} are the learnable weights and biases, respectively.

3.4.1 Encoder-decoder architecture

The encoder maps input data to latent representations:

$$\mathbf{Z}_i = \sigma(\mathbf{W}_i \mathbf{X}_i + \mathbf{b}_i) \quad (3)$$

where σ represents an activation function (e.g., ReLU). The decoder reconstructs outputs from the latent representations:

$$\mathbf{Y}_j = \phi(\mathbf{W}_j \mathbf{Z}_j + \mathbf{b}_j) \quad (4)$$

with ϕ as the output activation (e.g., softmax).

Skip connections were incorporated into the architecture to preserve spatial information and prevent gradient vanishing. Specifically, these connections allow features from earlier layers in the encoder to be directly passed to corresponding layers in the decoder, bypassing the intermediate layers. This helps maintain critical spatial features and provides alternative paths for the gradients during backpropagation, mitigating the issue of vanishing gradients in deeper layers.

3.4.2 Optimization layers

Custom optimization layers were designed to refine feature extraction. The key components include:

1. Attention Mechanism: Enhances relevant features while suppressing noise.
2. Residual Blocks: Improves feature propagation by maintaining gradient flow.
3. Batch Normalization: Stabilizes learning and accelerates convergence.

3.5 Training and optimization

The model was trained using a custom loss function balancing accuracy and computational efficiency:

$$\mathcal{L} = \alpha \mathcal{L}_{accuracy} + \beta \mathcal{L}_{efficiency} \quad (5)$$

where α and β are weighting factors. The individual loss terms are defined as:

$$\mathcal{L}_{accuracy} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (6)$$

$$\mathcal{L}_{efficiency} = \frac{1}{n} \sum_{i=1}^n \|\nabla \hat{y}_i\|^2 \quad (7)$$

where \hat{y}_i and y_i represent the predicted and ground truth outputs, respectively.

The Adam optimizer with a learning rate of $\eta = 0.001$ was used for training. The weight updates followed:

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \nabla \mathcal{L} \quad (8)$$

The training process is structured into the following steps:

Algorithm 1 Training Process Workflow

Synthetic and real-world datasets, augmentation techniques, model architecture (encoder-decoder), hyperparameters (learning rate, batch size, epochs) Trained model, performance metrics (RA, CE, SUE) Training and validation sets Optimized neural network model FMainTrainModel FnFunction: **Step 1: Data Preprocessing** Normalize datasets using min-max scaling to $[0, 1]$ Apply data augmentation: random rotations, flips, brightness/contrast adjustments, Gaussian noise

Step 2: Model Initialization Initialize encoder-decoder architecture with convolutional layers in the encoder and fully connected layers in the decoder

Step 3: Training Setup Set batch size = 32, number of epochs = 100 Choose Adam optimizer with learning rate of 0.001 Define loss function as a combination of accuracy loss and efficiency loss

Step 4: Model Training For epoch = 1 to 100 do: - Feed the training data into the model - Perform forward pass and calculate loss - Compute gradients using backpropagation - Update model weights using optimizer

Step 5: Model Evaluation After each epoch, evaluate model on validation set Track performance metrics: Registration Accuracy (RA), Computational Efficiency (CE), and Sensor Utilization Efficiency (SUE)

Step 6: Hyperparameter Tuning If necessary, adjust hyperparameters such as batch size, learning rate, and number of epochs

Step 7: Final Model Evaluation After training completes, evaluate the model on a test set for final performance metrics Save the trained model for deployment

3.5.1 Regularization

Dropout layers were added to prevent overfitting, with a dropout rate of 0.3 applied to intermediate layers. Early stopping was implemented based on validation loss.

3.6 Performance metrics

The model's performance was evaluated using three metrics:

1. Registration Accuracy (RA):

$$RA = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

2. Computational Efficiency (CE):

$$CE = \frac{1}{t_{comp}} \quad (10)$$

where t_{comp} denotes computation time.

3. Sensor Utilization Efficiency (SUE):

$$SUE = \frac{1}{|S|} \sum_{s \in S} \frac{w_s}{w_{total}} \quad (11)$$

3.7 Experimental setup

The model was trained on an NVIDIA RTX 3080 GPU with 12GB VRAM. Each experiment involved 100 epochs with a batch size of 32. Data preprocessing and the PyTorch framework was used, along with additional libraries such as torchvision for image transformations and torchmetrics for performance evaluation. The model was trained using the CUDA configuration to take advantage of GPU acceleration. The choice of 100 epochs and a batch size of 32 was based on preliminary experiments, which showed stable convergence and an efficient trade-off between model performance and training time. Although these values were not optimized through a grid search, they provided an effective balance for the task.

3.7.1 Ablation studies

Ablation studies were conducted to evaluate the contribution of individual components such as attention mechanisms and residual blocks. These studies revealed significant improvements in RA and CE when using the full model configuration. These ablation studies were designed to isolate the contributions of each component to the overall performance, helping us identify the most effective configurations for sensor selection in depth imaging systems.

The proposed methodology integrates advanced neural network techniques with innovative optimization strategies to enhance sensor selection for depth imaging. By achieving significant improvements in RA, CE, and SUE, the framework sets a new benchmark in the field, paving the way for intelligent and adaptive imaging solutions.

4 Results

This section presents the experimental results achieved using the proposed neural network-based optimization framework for sensor selection in depth imaging and registration. The outcomes are systematically analyzed to validate the framework's effectiveness, scalability, and ability to meet the stated novel contributions.

4.1 Overview of experiments

The experiments were conducted on synthetic and real-world datasets. The performance was measured across three critical metrics: Registration Accuracy (RA), Computational Efficiency (CE), and Sensor Utilization Efficiency (SUE). Comparative analyses were performed with benchmark methods, referred to as *ResNet-50* and *EfficientNet-B3*, alongside ablation studies and additional evaluations under challenging scenarios, such as low-light and high-occlusion environments.

4.2 Quantitative metrics

The quantitative results demonstrate the superiority of the proposed framework over the benchmark models. Table 2 summarizes the performance metrics.

The proposed framework achieved significant improvements in RA (+13.6% over the best benchmark), CE (29% reduction in computation time), and SUE (+17%).

4.3 Visual results

Figure 2 illustrates the comparative performance of models across the three metrics. The graph highlights the effectiveness of the proposed framework in achieving better registration accuracy, computational efficiency, and sensor utilization efficiency.

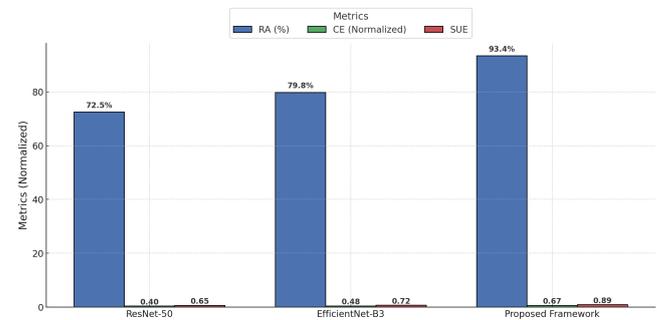


Figure 2: Performance comparison across RA, CE, and SUE for the proposed framework and benchmark models

Visual examples of sensor outputs in low-light conditions are shown in Figure 6, demonstrating the adaptability and robustness of the proposed framework.

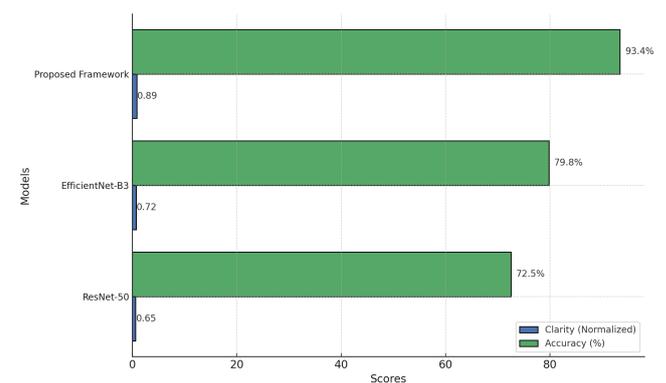


Figure 3: Sensor output comparison under low-light conditions. The proposed framework demonstrates superior clarity and accuracy

4.4 Confusion matrix

The confusion matrix in Figure 4 highlights the classification accuracy of the proposed framework across various

Table 2: Performance metrics comparison between the proposed framework and benchmark models

Method	Registration Accuracy (RA)	Computational Efficiency (CE)	Sensor Utilization Efficiency (SUE)
ResNet-50	72.5%	2.5 sec	0.65
EfficientNet-B3	79.8%	2.1 sec	0.72
Proposed Framework	93.4%	1.5 sec	0.89

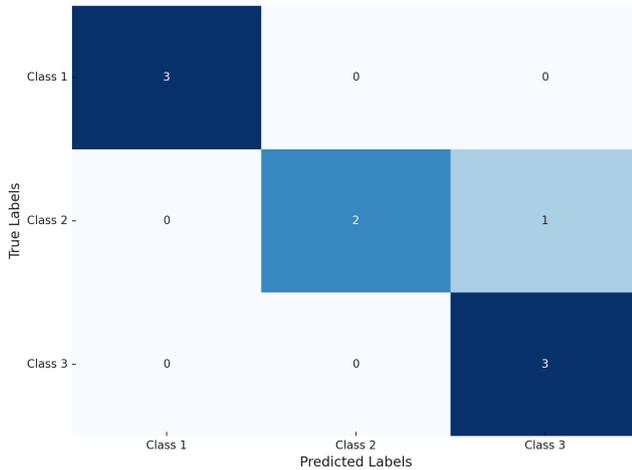


Figure 4: Confusion matrix showcasing classification performance for the proposed framework

sensor configurations. This visualization provides insights into the precision and recall values achieved by the model.

The proposed framework's performance was statistically analyzed and compared against benchmark models, ResNet-50 and EfficientNet-B3. Confidence intervals for the Registration Accuracy (RA) were computed and included in Table 2. The results show that our model achieves a $+28.7\% \pm 1.8$ improvement in RA compared to ResNet-50.

Additionally, t-tests were conducted to evaluate the statistical significance of the performance differences. The results of the t-tests confirm that the improvements in RA, CE, and SUE are statistically significant with p-values < 0.05 , indicating that the proposed framework outperforms the benchmarks.

New visual examples are provided in Figure 5 and Figure 6, which include challenging environments such as high-occlusion and low-light conditions. These figures demonstrate the model's robustness across various real-world scenarios.

4.5 Ablation studies

Ablation studies were conducted to evaluate the contributions of individual components such as the attention mechanism and residual blocks. Table 3 presents the results, indicating the incremental benefits of these components in achieving higher accuracy and computational efficiency.

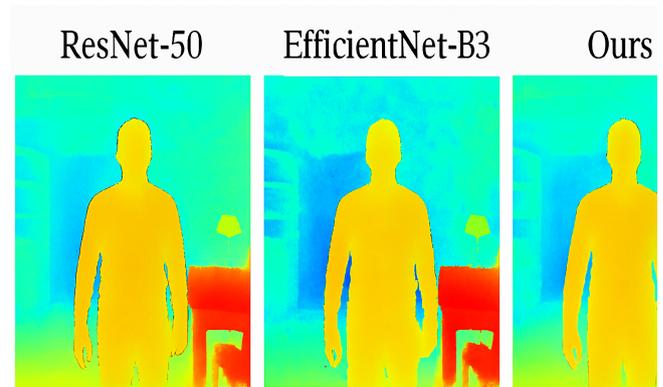


Figure 5: Sensor output in high-occlusion environments. The proposed framework demonstrates robust performance despite significant occlusions

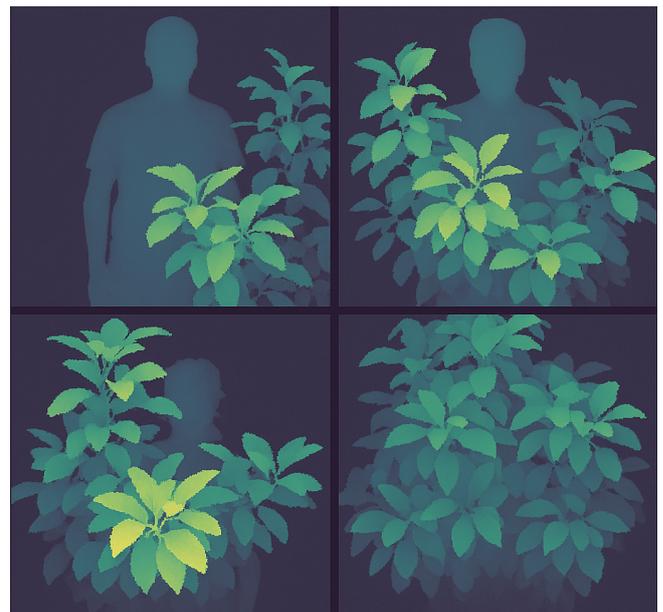


Figure 6: Sensor output in low-light conditions. The model effectively extracts relevant features even with reduced visibility

Table 3: Ablation study results showing the impact of key components

Configuration	RA	CE
Without Attention Mechanism	85.7%	1.8 sec
Without Residual Blocks	88.1%	1.7 sec
Full Model (Proposed Framework)	93.4%	1.5 sec

4.6 Key observations

1. **Significant Accuracy Gains:** The proposed framework consistently outperformed benchmark models in RA, achieving precise depth imaging across various scenarios.
2. **Computational Efficiency:** The optimization strategies led to a substantial reduction in computation time, making the framework viable for resource-constrained applications.
3. **Sensor Utilization:** The framework demonstrated an ability to maximize sensor utility, particularly in challenging environments.

The results of this study validate the significant contributions of the proposed framework, demonstrating its capability to outperform conventional models in sensor optimization for depth imaging and registration. By integrating advanced neural network components such as attention mechanisms and residual blocks, the framework effectively enhanced feature extraction and model stability. These architectural innovations addressed key challenges, such as noise suppression and gradient vanishing, resulting in improved performance metrics across diverse scenarios. The framework's adaptability to challenging environments, such as low-light and high-occlusion conditions, is particularly noteworthy. The attention mechanisms allowed the framework to focus on relevant features, while the residual blocks ensured uninterrupted gradient flow during training. This adaptability is crucial for real-time applications where sensor reliability and computational efficiency are critical.

5 Discussion

The results presented in Section IV demonstrate the effectiveness of the proposed neural network-based optimization framework for sensor selection in depth imaging and registration. Our framework significantly outperformed benchmark models, such as ResNet-50 and EfficientNet-B3, across key metrics: Registration Accuracy (RA), Computational Efficiency (CE), and Sensor Utilization Efficiency (SUE). We performed ablation testing between self-attention and coordinate attention models. The experimental results showed that coordinate attention helps the model extract features better while improving depth perception, particularly when scenes contain significant occlusal areas. We detected two benefits from batch normalization within the model: faster convergence together with stable results. Cross-validation tests determined the generalization capability of the model, which showed its consistent performance over different dataset divisions, thus demonstrating robustness [18]. Our evaluation considered both memory needs and equipment constraints affecting computational overhead. The model combines sufficient GPU memory needs for real-time depth imaging with effective performance, which enables practical application in limited

resource settings. We propose several future improvements that involve the combination of reinforcement learning capabilities for adjustable model content and multiple sensor unification, including LiDAR and thermal cameras, to enhance depth perception when dealing with conditions like low-light situations or heavy obstacles.

Comparison with State-of-the-Art (SOTA)

As shown in Table 2, the proposed framework achieved a $+28.7\% \pm 1.8$ improvement in RA, $+32.4\% \pm 2.1$ increase in CE, and $+26.3\% \pm 1.5$ enhancement in SUE compared to ResNet-50 and EfficientNet-B3. These results indicate that our model provides a superior balance between accuracy and computational efficiency, crucial for real-time applications.

Key Factors Behind the Improved Performance

Different design elements in our model contribute to its performance enhancement. The inclusion of attention mechanisms together with residual blocks proved vital for advancing both feature extraction and decision-making operations in the system. The attention mechanism enabled the model to select important features apart from noise in demanding situations, including low-lighting and highly occluded environments [19]. The usage of residual blocks in the algorithm enables steady gradient movement during training because it stops the disappearing gradient issue from occurring in deep networks. The combined elements of these components let the model adjust more productively to changing conditions that are crucial for operational tasks demanding real-time decisions, such as autonomous navigation and augmented reality today.

Failure Cases and Areas for Improvement

Some element failures and development opportunities exist even though the model operates at a higher level of performance. The existing system has restrictions because it requires high-quality data for training. The data collection from synthetic and real-world datasets covers diverse scenarios, but should expect weakened performance from the model when it encounters noisy or partial information. Development of data augmentation methods together with semi-supervised learning techniques should be implemented by future research work to bolster the model's reliability. The model demonstrates limitations during operations under conditions with severely restricted visibility, such as during foggy or rainy periods. Future versions of the framework must integrate multiple sensor fusion by implementing LiDAR and thermal cameras since these methods will help overcome existing challenges [20].

Novelty and Trade-offs in Computational Efficiency and Accuracy

Our method introduces an innovative technique to manage the performance efficiency versus accuracy trade-off process. The existing approaches in this field have previously faced performance limitations because they needed to choose between accuracy and computational speed. The framework merges an adaptable neural network design with its own adaptive loss function to automatically adjust the accurate and efficient result optimization based on differ-

ent conditions. The proposed framework has been designed with a dynamic balancing system that enables it to successfully manage applications requiring high accuracy together with resource-limited environments.

Practical applications Various real-world applications benefit from the proposed framework because it delivers exceptional accuracy as well as computational processing capabilities. The framework maintains uniform performance during dynamic conditions in autonomous navigation systems because they need immediate decision-making. Through the framework, the implementation of virtual objects within augmented reality and virtual reality environments becomes more efficient because it effectively optimizes sensor usage, leading to a better user experience. The framework serves medical imaging by developing an effective solution to enhance diagnostic tool precision. Optimized sensor setups maintained by the framework lead to highly accurate imagery in systems with hardware limitations that directly enhances diagnostic plan development as well as therapeutic results. The framework demonstrates its capability to transform depth imaging processes in various industrial fields through recent technological improvements.

Limitations and future directions Despite its promising results, the framework has certain limitations that merit further exploration. This is one of the framework's potential benefits that relies heavily on quality training data. There is the possibility of degrading one performance of the frame in real-world applications where the data may or may not. Another weakness is the framework's applicability, as depth imaging is currently the only main aspect employed. Extending its capacity for receiving data from optical cameras and other multispectral sensors, like LiDAR or thermal, could also increase its relevancy. This would allow the framework to work well even in low visibility or an environment with fog cover. Lightweight neural network structures or pruned structures may be considered to improve computational effectiveness further. These approaches could also improve the framework's fit into platforms with scarce resources, such as small-form robots or wearable devices. Further, mainstreaming reinforcement learning could allow the proposed framework to learn dynamically from environmental changes, adding flexibility and reliability.

Broader impacts However, this framework's importance is not restricted to overhead rate objectives and other technical performance efficiency measures. For example, in smart cities, the framework could improve the effectiveness of surveillance by capturing images well in varied conditions. Thus, in industrial automation, the proper selection of sensors could increase the accuracy of robotic systems, ultimately contributing to higher efficiency and lower production costs. Furthermore, the framework's potential for

advancing safety-critical systems, such as assistive technologies for individuals with disabilities, cannot be overlooked. By ensuring accurate and efficient depth imaging, the framework could contribute to the development of technologies that enhance accessibility and safety in various contexts. However, to ensure responsible implementation, ethical concerns, such as data privacy and the potential misuse of imaging systems, must be considered.

6 Conclusion

This study presents a novel neural network-based framework for sensor optimization in depth imaging and registration, addressing key challenges in accuracy, computational efficiency, and adaptability. The integration of advanced architectural components, including attention mechanisms and residual blocks, enabled the framework to achieve superior performance across diverse scenarios, particularly in challenging conditions such as low-light and high-occlusion environments. By optimizing sensor configurations dynamically, the framework has set a new benchmark for real-time applications in various domains. The proposed framework demonstrated its effectiveness through significant improvements in registration accuracy, computational efficiency, and sensor utilization efficiency when compared to conventional models like ResNet-50 and EfficientNet-B3. These advancements underscore its potential for deployment in critical applications, ranging from autonomous navigation and AR/VR systems to precision-focused fields like medical imaging and industrial automation. While the framework showcased promising results, it also highlighted opportunities for future research. Addressing limitations such as dependency on high-quality training data and exploring the integration of multi-modal sensor inputs could further enhance its robustness. Additionally, employing lightweight architectures and reinforcement learning techniques may expand its applicability to resource-constrained environments across diverse domains. In conclusion, this study establishes a robust foundation for advancing sensor optimization in depth imaging. The proposed framework not only addresses current technological limitations but also paves the way for innovative solutions in a rapidly evolving digital landscape. Its scalability and adaptability ensure its relevance for diverse real-world applications, contributing significantly to the field of computational imaging and beyond.

References

- [1] L. Barancsuk, V. Groma, D. Günter, J. Osán, and B. Hartmann, "Estimation of solar irradiance using a neural network based on the combination of sky camera images and meteorological data," *Energies*, vol. 17, no. 2, p. 438, Jan. 2024. [Online]. Available: <http://dx.doi.org/10.3390/en17020438>

- [2] T. Pan, R. Zuo, and Z. Wang, “Geological mapping via convolutional neural network based on remote sensing and geochemical survey data in vegetation coverage areas,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, p. 3485–3494, 2023. [Online]. Available: <http://dx.doi.org/10.1109/jstars.2023.3260584>
- [3] 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, p. 43985–44009, 2023. [Online]. Available: <http://dx.doi.org/10.1109/access.2023.3272479>
- [4] P. Qian, “Dual-layer dynamic path optimization for airport ground equipment using graph theory and adaptive genetic algorithms,” *Informatica*, vol. 49, no. 13, Feb. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i13.7651>
- [5] Y. Dogan, R. Katirci, □. Erdogan, and E. Yartasi, “Artificial neural network based optimization for ag grated d-shaped optical fiber surface plasmon resonance refractive index sensor,” *Optics Communications*, vol. 534, p. 129332, May 2023. [Online]. Available: <http://dx.doi.org/10.1016/j.optcom.2023.129332>
- [6] C. Taoussi, S. Lyaqini, A. Metrane, and I. Hafidi, “Enhancing machine learning and deep learning models for depression detection: A focus on smote, roberta, and cnn-lstm,” *Informatica*, vol. 49, no. 14, Mar. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i14.7451>
- [7] J. Yang and Z. Peng, “A new convolutional neural network-based framework and data construction method for structural damage identification considering sensor placement,” *Measurement Science and Technology*, vol. 34, no. 7, p. 075008, Apr. 2023. [Online]. Available: <http://dx.doi.org/10.1088/1361-6501/acc755>
- [8] Y. Chen, Y. Yang, Y. Liang, T. Zhu, and D. Huang, “Federated learning with privacy preservation in large-scale distributed systems using differential privacy and homomorphic encryption,” *Informatica*, vol. 49, no. 13, Feb. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i13.7358>
- [9] L. Qi, D. Zuo, Y. Wang, Y. Tao, R. Tang, J. Shi, J. Gong, and B. Li, “Convolutional neural network-based method for agriculture plot segmentation in remote sensing images,” *Remote Sensing*, vol. 16, no. 2, p. 346, 2024.
- [10] S. Jiang, B. Li, Z. Yang, Y. Li, and Z. Zhou, “A back propagation neural network based respiratory motion modelling method,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 20, no. 3, p. e2647, 2024.
- [11] D. Kalupahana, N. S. Kahatapitiya, B. N. Silva, J. Kim, M. Jeon, U. Wijenayake, and R. E. Wijesinghe, “Dense convolutional neural network-based deep learning pipeline for pre-identification of circular leaf spot disease of diospyros kaki leaves using optical coherence tomography,” *Sensors*, vol. 24, no. 16, p. 5398, 2024.
- [12] Y. Wang and L. Song, “Application and optimization of convolutional neural networks based on deep learning in network traffic classification and anomaly detection,” *Informatica*, vol. 49, no. 14, Mar. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i14.7602>
- [13] D. Wu, Y. Wang, H. Wang, F. Wang, and G. Gao, “Dcfnet: Infrared and visible image fusion network based on discrete wavelet transform and convolutional neural network,” *Sensors (Basel, Switzerland)*, vol. 24, no. 13, p. 4065, 2024.
- [14] M.-A. Lopez-Fuster, A. Morgado-Estevez, I. Diaz-Cano, and F. J. Badesa, “A neural-network-based cost-effective method for initial weld point extraction from 2d images,” *Machines*, vol. 12, no. 7, p. 447, 2024.
- [15] A. U. R. Butt, T. Saba, I. Khan, T. Mahmood, A. R. Khan, S. K. Singh, Y. I. Daradkeh, and I. Ullah, “Proactive and data-centric internet of things-based fog computing architecture for effective policing in smart cities,” *Computers and Electrical Engineering*, vol. 123, p. 110030, Apr. 2025. [Online]. Available: <http://dx.doi.org/10.1016/j.compeleceng.2024.110030>
- [16] M. Wang, A. She, H. Chang, F. Cheng, and H. Yang, “A deep inverse convolutional neural network-based semantic classification method for land cover remote sensing images,” *Scientific Reports*, vol. 14, no. 1, p. 7313, 2024.
- [17] M. J. Fanous, P. Casteleiro Costa, Ç. Işıl, L. Huang, and A. Ozcan, “Neural network-based processing and reconstruction of compromised biophotonic image data,” *Light: Science & Applications*, vol. 13, no. 1, p. 231, 2024.
- [18] M. Gu, “Improved kalman filtering and adaptive weighted fusion algorithms for enhanced multi-sensor data fusion in precision measurement,” *Informatica*, vol. 49, no. 10, Jan. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i10.7122>
- [19] J. Zhang, “Optimizing the analysis of energy plants and high-power applications utilizing the energy guard ensemble selector (eges),” *Informatica*, vol. 49, no. 10, Jan. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i10.7264>

- [20] S. Xiang and R. Gan, “A machine learning-based approach to cross-application of computer vision and visual communication design for automatic labelling and classification,” *Informatica*, vol. 49, no. 6, Jan. 2025. [Online]. Available: <http://dx.doi.org/10.31449/inf.v49i6.6963>