

RESE-CNN: Residual Squeeze-and-Excitation Network for High-Contrast Optical Tomography Reconstruction

Yingkuang Zhu, Zhenhua Pan, Huajun Li, Jianyang Chen and Yihao Sheng

College of Automation, Hangzhou Dianzi University, Hangzhou, 315000, China

Abstract: Optical tomography, a critical component of process tomography, is an important tool for determining the absorption coefficient of cross-sectional media, with significant applications in multi-phase flow analysis, chemical processing, and combustion studies. However, the precise reconstruction of these media distributions is a major challenge. In this work, a sophisticated optical tomography (OT) system coupled with an innovative reconstruction algorithm is presented. The system architecture includes 25 light sources and 25 strategically placed fan-beam receivers. We present a convolutional neural network (CNN) with an encode-decode configuration, augmented by residual connections and a squeeze-and-excitation (SE) attention mechanism. Initial evaluations performed using MATLAB simulations showed the algorithm's superior performance compared to existing methods, with notable improvements in relative error (RE) and correlation coefficient metrics. Subsequent practical experiments validated these findings and emphasized the efficiency of the residual and SE components in improving reconstruction accuracy. While this study focuses on high-contrast binary scenarios, the proposed RESE-CNN framework provides a basic architecture for future extensions to weakly absorbing and scattering media where nonlinear reconstruction problems dominate.

Keywords: optical tomography, deep learning, reconstruction algorithm

1. INTRODUCTION

The acquisition of internal data on dynamic multi-phase flows in pipelines and the accurate reconstruction of cross-sectional images are persistent and critical challenges in numerous industrial fields, including chemical engineering, mineral oil processing, and nuclear engineering [1]. Optical tomography (OT) emerges as a formidable tomography technique that offers unparalleled advantages such as high-frame rates, exceptional spatial resolution, and robustness to electromagnetic interference. These characteristics make OT a highly promising solution for the complex problems associated with dynamic multiphase flow measurements. In addition, its capabilities can significantly improve industrial visualization and process control. Given its potential, OT is capable of revolutionizing the way internal flow characteristics are monitored and analyzed, paving the way for advances in efficiency and safety in these sectors [1], [2].

Image reconstruction is a crucial process in OT that is essential for converting raw absorption data into detailed cross-sectional images. Conventional image reconstruction algorithms address the problem by simplifying the complex non-linear mapping from absorption coefficient distribution vectors to projection vectors into a simple linear mapping, neglecting the non-linear relationship between the two, which ultimately affects the accuracy of the imaging results [3]. Therefore, there is a need to explore alternative methods for effective image reconstruction. In recent years, extensive research has been conducted on image reconstruction

algorithms [4], [5]. Traditional methods such as linear back projection (LBP), regularization methods, Landweber iteration, and algebraic reconstruction technique (ART) [6], are widely used in practice. However, these traditional algorithms often simplify the non-linear inverse problem to a linear one, failing to reflect the non-linear relationship between the projection vector and the true distribution. Moreover, the line-of-sight (LOS) nature of OT and the sensitive field distribution — resulting from the specific emitter and receiver placements — pose additional challenges that further degrade image quality [2], [7]. Innovative approaches using deep learning have emerged as promising solutions to overcome these limitations. For example, Ben et al. [8] proposed a multitask deep learning method for simultaneous image reconstruction and lesion localization in limited-angle diffuse optical tomography (DOT), which shows significant improvements over conventional iterative methods. This approach shows that it is suitable for specific medical applications by integrating tasks that require precise localization in addition to reconstruction. However, our proposed RESE-CNN focuses on improving the quality of reconstructed images through an innovative combination of residual connections and SE attention mechanisms. In contrast to the multitasking approach that includes lesion localization, RESE-CNN is designed to provide superior performance in absorption coefficient distribution reconstruction, making it suitable for a wider range of industrial process monitoring scenarios.

To overcome these limitations, deep learning techniques encompassing a wide range of neural network architectures such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have shown promise. CNNs are particularly well suited for OT because of their ability to model spatial hierarchies in imaging data [9]. In this work, we focus on the integration of residual connections and Squeeze-and-Excitation (SE) attention mechanisms within a U-Net architecture to improve the performance of OT image reconstruction. We compare our proposed method with traditional reconstruction algorithms, including LBP and Landweber iteration, as well as with other CNN-based approaches. In addition, for specific medical imaging problems, such as OCT image analysis in glaucoma assessment, there are studies using deep learning models to improve diagnostic accuracy. However, there are still relatively few CNN approaches in the field of OT, especially involving attention mechanisms. Although some studies have proposed CNN methods based on multi-scale and multi-branch weakly supervised attention mechanisms for atherosclerosis grading of fundus images, these methods are usually not directly applied to OT image reconstruction. To overcome these limitations, there is an urgent need for new methods that can effectively address these non-linearities and adapt to the complexities of the system's geometry.

Neural networks as data-driven machine learning algorithms are characterized by the fact that they capture complicated non-linear relationships between input and output through training. In a recent study, deep belief networks (DBNs) and RNNs were used to enhance non-linear tomographic absorption spectroscopy [10] and their performance was evaluated in comparison to different algorithms. While these methods provide viable solutions to obtain more accurate reconstructed images, they primarily focus on solving the problem in a one-dimensional context and may overlook important spatial information within the images. This limitation may limit their effectiveness in fully solving the complexities of the reconstruction process.

In recent years, driven by the rise of big data and advances in computer hardware, deep learning has experienced rapid growth and achieved significant milestones in various domains. One of the most successful deep learning algorithms is the convolutional neural network (CNN), which is often used for image processing tasks such as classification and segmentation [11], [12], [13]. In addition, CNNs have shown strong performance in the field of image reconstruction. For example, in a recent study, an optimized Visual Geometry Group Network (VGG Net) variant, known as CNN-RBF, was used for the reconstructions of images in electrical impedance tomography (EIT). This approach generates directly reconstructed images from voltage measurements and demonstrates the effectiveness of CNNs in improving imaging techniques. [14]. In addition, in the field of nonlinear tomographic absorption spectroscopy (NTAS), an investigation was performed using an optimized CNN to learn the non-linear mapping between sinograms and temperature distributions [15]. These methods typically use CNNs with relatively few hidden layers, which limits their ability to accurately approximate complex non-linear mappings. To address this

limitation, there is a recognized need for deeper CNN models. Intuitively, deeper CNN architectures — those with more stacked layers — often show better performance. However, previous studies have identified a phenomenon known as the degradation problem: as network depth increases, there is an initial improvement in accuracy, followed by a rapid degradation in performance. This challenge emphasizes the importance of carefully balancing the depth of CNN architectures to achieve optimal performance in tasks such as non-linear mapping in tomographic absorption spectroscopy [16]. Residual connections were specifically designed to mitigate the degradation problem in deep CNNs. By reformulating the mapping relationship between layers, residual connections allow convolutional layers to focus on learning residual features rather than learning entirely new features. This approach simplifies the learning task as the network can directly adapt its outputs through residual learning, which improves the overall training process and model performance [16].

Many neural networks that incorporate attention as an effective mechanism for filtering information and focusing on what is important have shown excellent performance in image classification [17], [18], natural language processing [19], and neuroscience [20], [21]. It helps to process information efficiently and make informed decisions. SE attention is a type of spatial attention mechanism used in deep learning models for computer vision tasks [17]. This technique effectively directs the attention of the model to important areas within the feature maps, improving overall performance. Typically, convolutional layers do not fully account for inter-channel dependencies in their outputs. SE attention allows the network to dynamically recalibrate the channel-wise feature responses by focusing more on informative features and suppressing less relevant ones. This adaptive mechanism greatly improves the CNN's ability to capture intricate patterns and correlations in the data, resulting in improved accuracy and efficiency across a variety of tasks.

In this work, we develop an OT system with fan-beam mode and propose a model that integrates CNN, Residual connection, and SE attention for an OT image reconstruction task, called RESE-CNN. A deep CNN has sufficient capabilities to perform complex non-linear mappings. The integration of residual connections and SE attention mechanisms addresses two critical challenges in deep CNN-based OT. First, residual connections mitigate gradient vanishing and degradation issues in deep networks by enabling the direct propagation of feature information through skip connections [22]. Second, the SE attention mechanism dynamically recalibrates the channel-wise feature responses and prioritizes informative regions in the feature maps [23]. In contrast to previous studies that limited the image resolution to 60×60 pixels due to computational constraints [10], [24], [25], our RESE-CNN reconstructs images with a resolution of 128×128 pixels. This improvement enables precise identification and enhances the detection capability for micro-scale process anomalies. Section II introduces the development of our OT system and the basic concept of OT image reconstruction. Section III introduces the structure of residual connection, SE attention, and RESE-CNN and explains the principle

of residual connection and SE attention. The details of neural work training can be found in Section IV. The results of the test data and the analysis can be found in Section V.

2. DEVELOPMENT OF AN OPTICAL TOMOGRAPHY SYSTEM

A. Hardware design

The OT system with fan-beam mode is shown in Fig. 1. It consists of a microcomputer, a controlling unit, an optical sensor, a photodiode array, and a sampling unit. The images of our tomography system and the optical sensor are shown in Fig. 2.

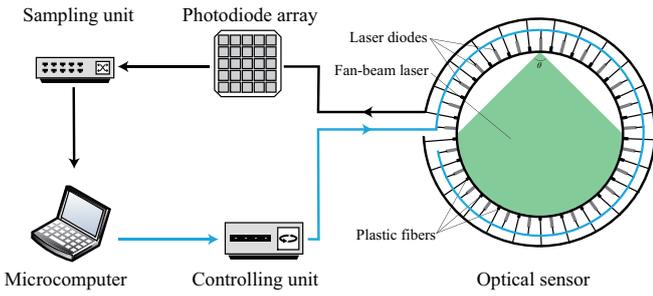


Fig. 1. OT system.

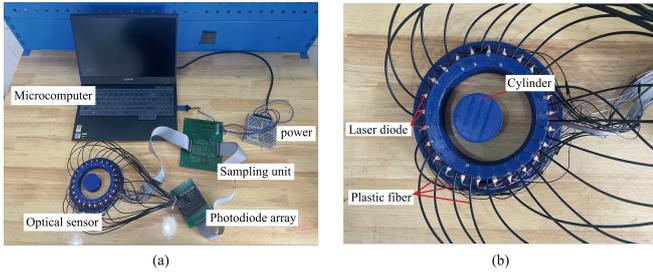


Fig. 2. Tomography system and optical sensor. (a) Tomography system. (b) Optical sensor.

The optical sensor consists of 25 emitters and 25 receivers arranged in a circular pattern around an inner section with a diameter of 100.0 mm. This arrangement ensures uniform measurement coverage of the cross-sectional area [1]. The emitters are Anford T80L514 laser diodes with a wavelength of 632 nm and a divergence angle of 1.60 radians, and the outer diameter is only 5 mm. These laser diodes have a compact outer diameter of 5 mm and operate with a power of 500 mW. On the receiver side, plastic optical fibers with a 1 mm inner diameter are used. These fibers are connected to the photodiode array (Vishay BPW34) at the other end to detect the light intensity.

The controlling unit and the sampling unit are homemade circuits based on STM32. The control circuit sequentially activates the laser diodes, while the sampling circuit collects the optical signals from the photodiode array. The cylinder located in the test section is used to simulate different distributions. Both the basic structure of the optical sensor and

the cylinders are manufactured using a commercial 3D printer (RAISE 3D N2) with a precision of 0.2 mm. The details of the development of the OT can be found in our previous work [26].

B. OT image reconstruction

According to the Lambert-Beer law [24], the intensity of light passing through the medium in which the absorption coefficient has the form of the function $\mu(p, q)$ follows the equation:

$$\Psi = \Psi_0 e^{-\int_r \mu(p, q) dl}, \quad (1)$$

where:

$r = \int dl$ represents the total path length (m) of the light beam through the medium;

$\mu(p, q)$ denotes the absorption coefficient distribution (m^{-1}) at the spatial coordinates (p, q) ;

Ψ_0 is the incident light intensity (W/m^2);

Ψ is the transmitted light intensity (W/m^2) after passing through the medium.

The integral $\int \mu(p, q) dl$ calculated the path-averaged absorption along the beam trajectory.

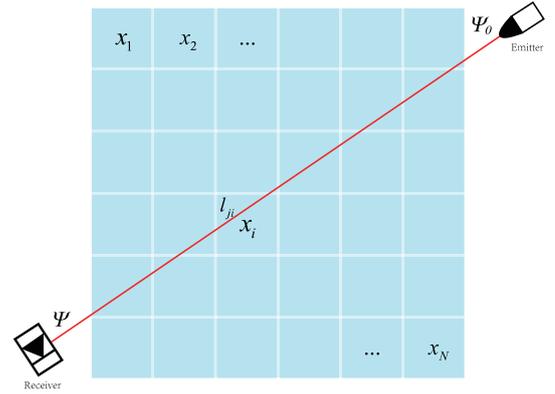


Fig. 3. Geometry of one laser beam through the tomography field.

As shown in Fig. 3, we can reconstruct the cross-sectional image of the measured area by dividing the region into N square pixels, where each pixel's value is the absorption coefficient of the area it represents. Consequently, we can derive a discretized form of the Lambert-Beer law:

$$\Psi_j = \Psi_0 e^{-\sum x_i l_{ji}}, \quad (2)$$

where Ψ_j is the intensity of the j^{th} beam, x_i is the absorption coefficient of the i^{th} pixel, and l_{ji} is the corresponding path of the laser in the i^{th} pixel. If we take the natural logarithm of both sides of the equation, it can be written as:

$$\ln \left(\frac{\Psi_0}{\Psi_j} \right) = \sum x_i l_{ji}. \quad (3)$$

The entire forward projection problem can be expressed as a matrix:

$$b = Ax, \quad (4)$$

where projection $\ln\left(\frac{\Psi_0}{\Psi_j}\right)$ is the element in vector b and x is the absorption coefficient distribution. A is the sensitivity matrix:

$$b = \left[\ln\left(\frac{\Psi_0}{\Psi_1}\right) \ln\left(\frac{\Psi_0}{\Psi_2}\right) \cdots \ln\left(\frac{\Psi_0}{\Psi_m}\right) \right]^T \quad (5)$$

$$x = [x_1 \ x_2 \ \cdots \ x_n]^T \quad (6)$$

$$A = \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \ddots & \vdots \\ l_{m1} & \cdots & l_{mn} \end{bmatrix}, \quad (7)$$

where m is the number of projections and n is the number of pixels in the measurement area.

The task of image reconstruction is to find the distribution vector x of the absorption coefficient from the projection vector b and the constant sensitivity matrix A . If the inverse of A exists, the forward problem can be solved directly by:

$$x = A^{-1}b. \quad (8)$$

However, this equation assumes ideal conditions with perfect sensors and no noise. In practical scenarios, errors due to sensor non-ideality and noise are unavoidable. Therefore, the actual solution contains an additional error term, which is represented as:

$$\hat{x} = A^{-1}b + \beta, \quad (9)$$

where β takes into account the deviations caused by these imperfections. Accounting for and mitigating such errors is crucial for accurate implementation and analysis. In this work, we have revised our model to include the noise term β . We also propose methods to handle and minimize the effects of β to ensure the robustness and accuracy of our algorithm.

But the value of n is in most cases much larger than m , which means that A is singular. If A is considered as a linear mapping from the absorption-coefficient distribution vector space to the projection vector space, A^T can be considered as a related mapping from the projection vector space to the absorption-coefficient distribution vector space, which gives an approximate solution:

$$\hat{x} = A^T b. \quad (10)$$

This approach is called linear back projection (LBP) and is the simplest way to reconstruct image in this situation.

In our OT system, the pixel grid of the reconstruction the image is set to 128×128 and covers a circular region of interest (ROI) with a 100 mm diameter. Within this ROI, the 12,596 effective pixels ensure that features as small as 1.6 mm (with a span of about 2 pixels) can be resolved, as shown in Fig. 4(a). Also, the beam arrangement and the sensitive field based on this mesh grid are shown in Fig. 4(b) and Fig. 4(c). A ring pattern can be seen in which the pixels have higher

sensitivity, and the center gains the highest sensitivity. In contrast, the pixels close to the boundary have a lower sensitivity.

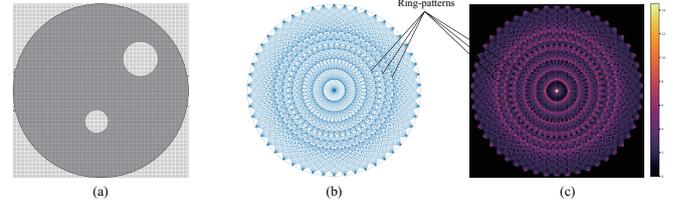


Fig. 4. Mesh grid, beam arrangement, and sensitive map. (a) Mesh grid. (b) Beam arrangement. (c) Sensitive field.

3. IMAGE RECONSTRUCTION METHOD

A. Structure

U-Net is a deep CNN with Encode-Decode-model that has proven successful in medical image segmentations [27]. In addition, many deep learning models in the U-Net series have also achieved great success [28], [29], [30]. Certain network architectures are inherently more robust. In particular, CNNs are known to be more resistant to local distortions due to their ability to learn spatial hierarchies of features. This robustness makes them particularly suitable for the task of OT image reconstruction, where the input data can be noisy and the distributions complex. Inspired by these studies, we propose RESE-CNN for OT image reconstruction.

The overall structure of RESE-CNN consists of two parts, as shown in Fig. 5. The first part aims to convert the input vector into a matrix so that it is compatible with the input format of the subsequent part. In this process, a projection vector S with 625 elements is used as input for the first part. Instead of using a fully connected layer or even a simple one-dimensional neural network, we decided to use the LBP to finish the first part. In contrast, the LBP algorithm transforms this raw data into a structured representation that captures local patterns and textures within the medium. This transformation not only preserves but also enhances the spatial structure, allowing for more accurate and efficient recognition. LBP is simple and fast to obtain an approximation of the medium distribution. Unlike fully connected layers that treat each input element independently without considering the spatial relationships, LBP produces a structured representation that captures local patterns and textures within the medium.

The second part consists of a deep CNN with nine convolutional blocks, which is used to further extract spatial features from the original image and transform it into the desired high-precision cross-sectional reconstruction image. Each convolutional block contains three hidden layers, as shown in Fig. 6. The first to fourth convolutional blocks perform down-sampling operations on the input, while the sixth to ninth convolutional blocks perform up-sampling operations. The fifth convolutional block serves as an intermediate step and connects the down-sampling section with the up-sampling section.

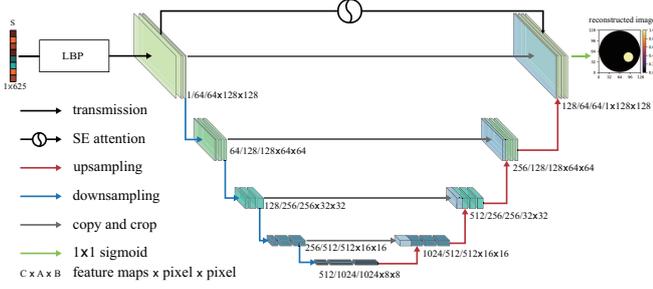


Fig. 5. Main structure of RESE-CNN.

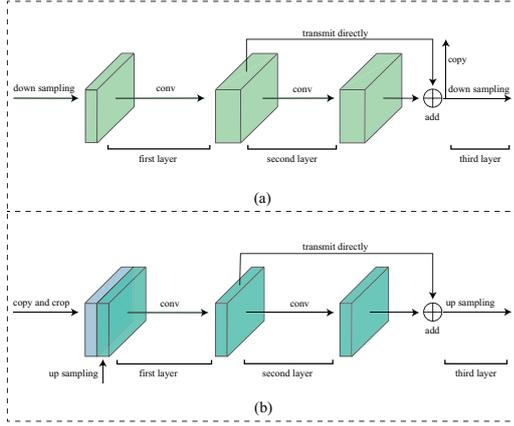


Fig. 6. Substructure of the blocks. (a) Block in down sampling module. (b) Block in up sampling module.

In the first to fourth convolutional blocks, the first and second layers are convolutional layers. In the second to fourth convolutional blocks, the first convolutional layer doubles the number of input channels, while the dimension of the height and width of the input and output remain the same. For the first convolutional block in particular the single-channel input is extended to 64 channels. The second convolutional layer in the first to fourth blocks retains the dimensions of the input-output tensors. The output of the second convolutional layer is added to its input and serves as the input for the third layer, which signifies the existence of a residual connection between the input and the output of the second layer. The third layer of the convolutional block is a down-sampling layer, which is achieved by max-pooling. This reduces the height and width dimensions of the input tensor to half their original size.

In the sixth to ninth convolutional blocks, the output of the previous block is concatenated with the output of the second convolutional layer from the corresponding fourth to first block. This retains the encoder's low-level features and combines them with decoder information. Each of these blocks has a convolutional layer that halves the channel dimensions, followed by another convolutional layer with residual connections that maintain the dimensions. The third layer in blocks six to eight is an up-sampling layer by transposed convolution, which doubles the tensor's height and width. The ninth block instead has a convolutional layer that reduces the channel dimension to match the input image. The fifth block, which serves as an intermediate step, has a similar structure, but with an up-sampling layer in the third layer.

B. Residual connection

The residual connection serves to overcome the degradation problem and enables the direct transmission of information to the deeper layers. We denote the basic mapping to be learned by the convolutional layer as $H(x)$. But in the residual block, we let the convolutional layer learn the residual function:

$$F(x) = H(x) - x. \quad (11)$$

And the original function can be expressed as:

$$H(x) = F(x) + x. \quad (12)$$

Although we could obtain the expected output by allowing the convolutional layer to learn $H(x)$ directly or learn the residual function and add the identity mapping, the ease of learning might be different. It has been proven that learning the residual function is easier than learning the original function [16].

C. Squeeze-and-excitation attention

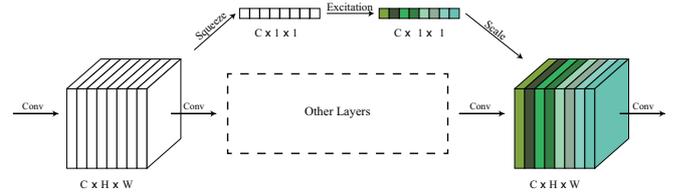


Fig. 7. Structure of the SE attention.

The SE attention mechanism use the channel information from the output of the second layer in the first block to compute the weight for each channel and scales the input of the third layer in the ninth block. Empirical studies have shown that this configuration leads to superior performance in our particular application context. By placing SE modules at these strategic points, we improve the network's ability to recalibrate channel-wise feature responses, which is particularly beneficial for capturing complicated spatial relationships within the data. Layers deeper in the network tend to capture higher-level abstractions, making them prime candidates for attention mechanisms such as the SE modules that can selectively emphasize informative features. This strategic placement allows the network to focus on relevant information while suppressing noise, improving overall accuracy and generalization. The detail of the SE attention mechanism is shown in Fig. 7. It consists of three steps: squeeze, excitation, and scale [17]. In the first step, the information of each channel is aggregated by global average pooling and channel-wise $Z \in R^C$ is generated from the input $X \in R^{C \times H \times W}$. The i^{th} element of Z is calculated by:

$$z_i = \frac{1}{H \times W} \sum_{j=1}^H \sum_{k=1}^W x_{ijk}, \quad (13)$$

where C , H , and W are the dimension of the channel, the height of the feature maps, and the width of the feature maps, respectively. The second step is to adaptively recalibrate the

obtained statistics, and fully capture the channel-wise dependencies. This is achieved by a simple gating mechanism:

$$\tilde{Z} = \text{Sigmoid}(W_2 \text{ReLU}(W_1 Z)), \quad (14)$$

where $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$ are the weight matrices to be learned and r is the reduction ratio. The i^{th} element of \tilde{Z} is the weight of the i^{th} channel. The last operation is to scale the input X :

$$x_i^c = \tilde{z}_i \cdot x_i^c. \quad (15)$$

4. TRAINING OF THE NETWORK

A reliable dataset is important for neural network-supervised learning. However, it is difficult to collect numerous samples from experiments. Therefore, we use MATLAB for finite element programming to create the dataset. The simulated OT system has the same structure as described in Section II. The ROI was discretized into a 128×128 pixel grid.

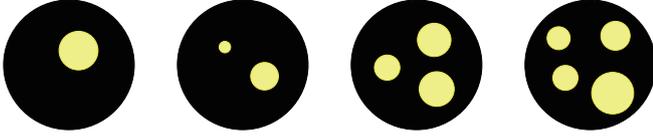


Fig. 8. Simulated absorption coefficient distributions and reconstruction results in a 128×128 pixel grid.

An individual sample includes the projection and the distribution of the medium. We activate one light emitter at a time and collect the projection value from all light receivers until each light emitter has been activated. With this measurement mode, we solve the forward problem and obtain the complete projection vector with 625 elements. As shown in Fig. 8, the ROI is discretized into a 128×128 pixel grid, which corresponds to a physical area of 100 mm in diameter. The reconstructed distributions show four characteristic bubble configurations: (a) single-bubble, (b) two-bubble, (c) three-bubble, and (d) four-bubble patterns. Each axis shows a dual-unit scaling (pixels and millimeters) to facilitate quantitative analysis of feature dimensions. In total, we randomly generated 15,000 samples with different numbers of bubbles and distributions and divided them into training sets and test sets at a ratio of 8:2. To help the network learn the mapping between the projection and medium distribution, we simulate high-contrast absorption scenarios by assigning binary absorption coefficients: 1 for bubble regions (full absorption) and 0 for the background (no absorption). This simplification allows us to validate the RESE-CNN architecture under controlled conditions while maintaining relevance for practical applications such as gas-solid flow monitoring.

The essence of processing the preliminary image with RESE-CNN is to predict the value of each pixel of the preliminary image. Therefore, the task of image reconstruction can be considered as a regression problem. Since the values of the pixels are between 0 and 1, we choose the binary cross entropy loss function (BCEloss) to calculate the loss between the actual absorption coefficient distribution and the predicted absorption coefficient distribution. Although this choice is

consistent with the current high-contrast experimental setup, we recognize that weakly absorbing media with continuous absorption distributions would require alternative loss functions (e.g., mean square error) to capture non-linear relationships between projection data and absorption coefficients. Future studies will investigate these extensions. To improve the network's generalization ability and prevent overfitting, we also include the L_2 regularization term in the loss function. The equation can be expressed as follows:

$$\mathcal{L}(x^{(pre)} | x^{(real)}, w) = \text{BCEloss}(x^{(pre)}, x^{(real)}) + \frac{\lambda}{2} L_2, \quad (16)$$

$$\begin{aligned} & \text{BCEloss}(x^{(pre)}, x^{(real)}) \\ &= -\frac{1}{N} \sum \left[x_i^{(real)} \log x_i^{(pre)} + (1 - x_i^{(real)}) \log (1 - x_i^{(pre)}) \right], \end{aligned} \quad (17)$$

$$L_2(w) = \|w\|_2, \quad (18)$$

where $x_i^{(pre)}$ is the i^{th} element of the predicted absorption coefficient distribution $x^{(pre)}$, $x_i^{(real)}$ is the i^{th} element of the real absorption-coefficient distribution $x^{(real)}$, w is the weight parameter of the network, and λ is the regularization parameter, which is 0.0001.

The optimization of the weight parameters is a crucial part of the training process. Instead of the stochastic gradient descent algorithm, we use the Adaptive Moment Estimation (Adam) optimizer to optimize the network's weight parameters, which leads to faster convergence. The learning rate η is set to 0.0001 and the decay rate of the learning rate is set to 0.99. RESE-CNN is trained with a batch size of 32 for 100 epochs on the train set.

5. RESULTS AND ANALYSIS

A. Evaluation indicators

The image relative error (RE) and the correlation coefficient (CC) are important metrics for the evaluation of the quality of the reconstructed images. RE measures the relative error between a reconstructed distribution and a real distribution. CC reflects the similarity between a reconstructed image and a true distribution. The formulas for RE and CC are as follows:

$$RE = \frac{\|x^{(pre)} - x^{(real)}\|_2}{\|x^{(real)}\|_2}, \quad (19)$$

$$CC = \frac{\sum_{i=1}^N (x_i^{(pre)} - \bar{x}^{(pre)})(x_i^{(real)} - \bar{x}^{(real)})}{\sqrt{\sum_{i=1}^N (x_i^{(pre)} - \bar{x}^{(pre)})^2 \sum_{i=1}^N (x_i^{(real)} - \bar{x}^{(real)})^2}}, \quad (20)$$

where $\bar{x}^{(pre)}$ is the mean of the predicted absorption coefficient distribution $x^{(pre)}$ and $\bar{x}^{(real)}$ is the mean of the real absorption coefficient distribution $x^{(real)}$.

B. Verification on simulation data

We use the test set with 3000 samples to test the network's performance. Also, different image reconstruction methods are used to compare with the RESE-CNN method. The average RE and CC is presented in Table 1, and some of the results of the different distributions types are shown in Fig. 9, Fig. 10, Fig. 11, and Fig. 12, respectively.

Table 1. Average RE and CC of different methods.

Method	Average RE	Average CC
LBP	0.9152	0.4656
Landweber	0.6727	0.7256
U-Net	0.3575	0.9341
RE-CNN	0.2164	0.9772
RESE-CNN	0.1873	0.9776

We use the LBP and Landweber algorithms in the experiments to compare them with the RESE-CNN method. To emphasize the contribution of the residual connection and the SE attention, we also compared RESE-CNN with the original U-Net and RESE-CNN without the SE attention (RE-CNN). As shown in column 2 of Fig. 9 to Fig. 12, the reconstructed LBP images exhibit significant artifacts. They cannot accurately reflect the boundary of the medium. LBP achieves the worst results with an average RE of 0.9152 and a CC of 0.4656. The result of the Landweber method are much better, but there are still a number of artifacts, and the reconstruction accuracy of the medium does not reach a high level. The average RE of Landweber is 0.6727 and CC is 0.7256. These metrics correlate with the system's ability to resolve features at a scale of 1.6 mm, as shown by the intensity profiles in Fig. 9. In addition, the pixels of the ROI, especially the pixels inside the bubbles through which little or no light beam passes, hardly obtain the true value by the traditional algorithms, which is also a potential factor related to the low image quality. This indicates that the distribution of the sensitive field significantly affects the quality of the reconstructed images produced by traditional algorithms. The comparison between the results of a traditional algorithm and the sensitive field is shown in Fig. 13.

Deep learning methods solve this problem well. It turns out that the difference between three deep learning methods and traditional algorithms is obvious. The reconstruction images of U-Net show that the pixels inside the bubble are almost unaffected by the distribution of the sensitive field, the average RE decreases to 0.3575 and the average CC increases to 0.9341. But there are still a few artifacts near the boundaries of the bubbles. Moreover, the boundaries of the bubbles are not very precise, especially the relatively small bubbles and the bubbles near the edge of the ROI, as shown in Fig. 11, column 4, row 4. This indicates that the influence on the bubbles' boundary caused by the distribution of the sensitive fields has not been removed. The RE-CNN shows a remarkable improvement compared to the U-Net. The average RE of the RE-CNN is 0.2164, which is a reduction of 39.46 % compared to the RE of U-Net. Its average CC is 0.9772, reach-

ing a relatively high level. The reconstruction images of this method have much sharper and more precise boundaries. And the relatively small bubbles and the bubbles near the edge of the ROI which U-Net could not reconstruct well, were clearly reconstructed. For example, in Fig. 9 (column 5, single bubble) and Fig. 11 (column 5, edge bubble), the sharp transitions at the bubble boundaries confirm the effective resolution of less than 2 mm. This is also confirmed by the accurate reconstruction of simulated bubbles with a 1.6 mm diameter (see Section IV), which occupy about two pixels in width. In contrast, traditional methods (LBP, Landweber) cannot resolve features below 3 mm due to sensitivity field artifacts (Fig. 13). These results confirm the efficiency of residual connection, and the inclusion of residual connection is a great advantage for the networks to learn a more accurate mapping between the absorption coefficient distribution and the projection vector.

Adding SE attention to the model could further improve the network's accuracy. The RE of RESE-CNN is 0.1873, which is a reduction of 13.45 % compared to the RE-CNN, and the average of CC is 0.9776. RESE-CNN has the lowest RE and the highest CC, and the artifacts at the bubbles' boundaries are lower. The inclusion of the SE attention improves the network's performance with minimal additional parameters.

We also investigate the effects of distribution complexity on the reconstruction results. The number of bubbles in the samples represents the complexity of the distribution. The variation of the average RE and CC of the reconstruction images with respect to the distribution complexity are shown in Fig. 14(a) and Fig. 14(b), respectively. It is obvious that the CC of the traditional algorithms decreases significantly with increasing distribution complexity. In contrast, the deep learning methods show more stable performance. Although the RE of RESE-CNN increases slightly compared to other methods when the number of bubbles increases from 1 to 2, it still remains at a relatively low level. Overall, RESE-CNN consistently achieves the best performance on different complexity distributions.

The reconstruction time affects the real-time performance of the imaging system. We tested the reconstruct-single-image time of four algorithms, including LBP, Landweber, U-Net, and RESE-CNN. The average imaging time of these algorithms is shown in Table 2. The LBP requires the shortest time due to its simplicity. The Landweber algorithm requires the most imaging time in experiments because it needs many iterations to converge. The imaging time of RESE-CNN is longer than that of U-Net, but the difference is very small. Considering both imaging time and image quality, this difference can be ignored.

Table 2. Reconstruction time of different methods.

Method	Average reconstruction time (s)
LBP	3.796×10^{-5}
Landweber	2.846
U-Net	7.627×10^{-3}
RESE-CNN	7.687×10^{-3}

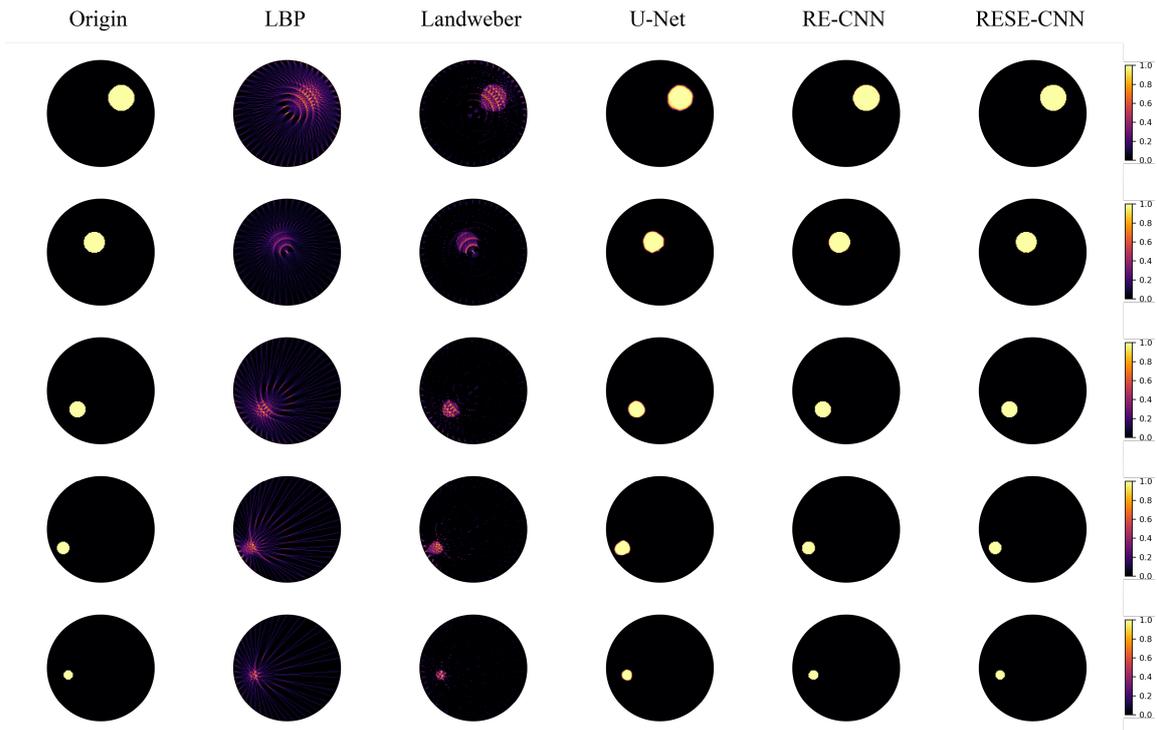


Fig. 9. Results of one-bubble distribution in a 128×128 pixel grid.

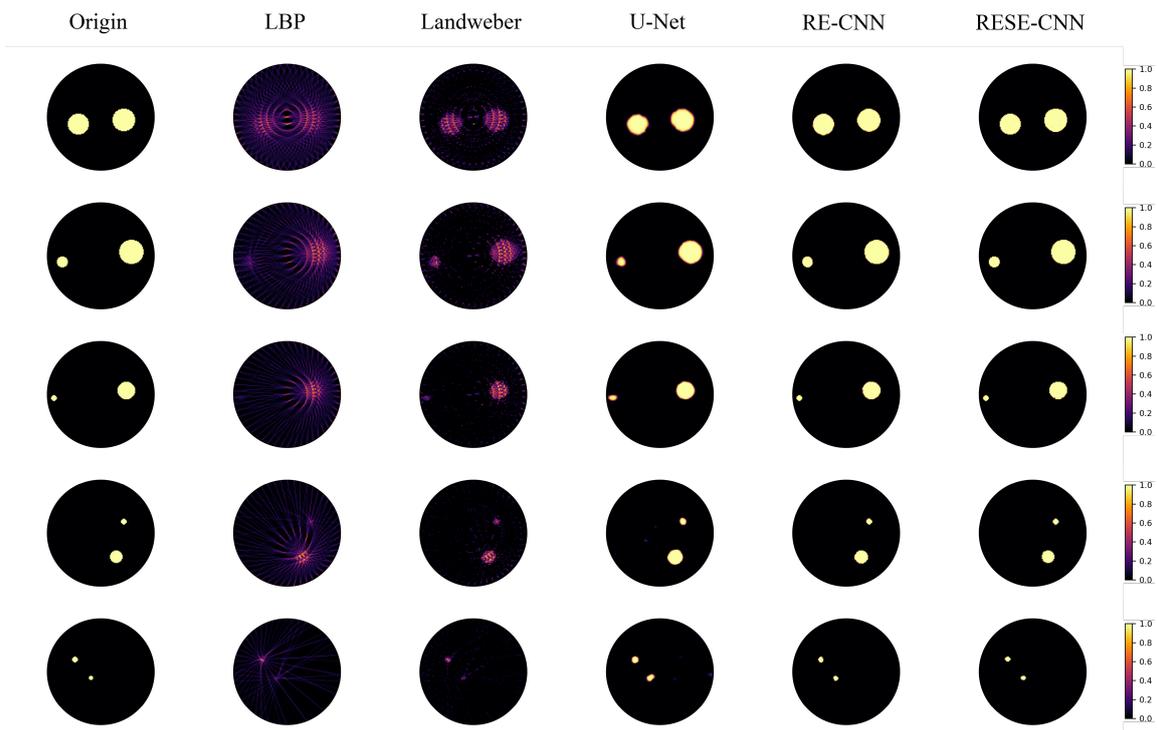


Fig. 10. Results of two-bubble distribution in a 128×128 pixel grid.

C. Verification on practical data

To validate the practicality of RESE-CNN, we have carried out feasibility tests with our developed OT system. The results of the different algorithms are shown in Fig. 15. Due to the inherent noise during actual measurements, the results

of all methods were affected differently. While RESE-CNN effectively reconstructed the images of single bubbles and detected only minor deviations for larger bubbles, the traditional algorithms exhibited poorer image quality. In particular, LBP produced excessively blurry reconstructions, while Landweber's results were inaccurate and exhibited significant arti-

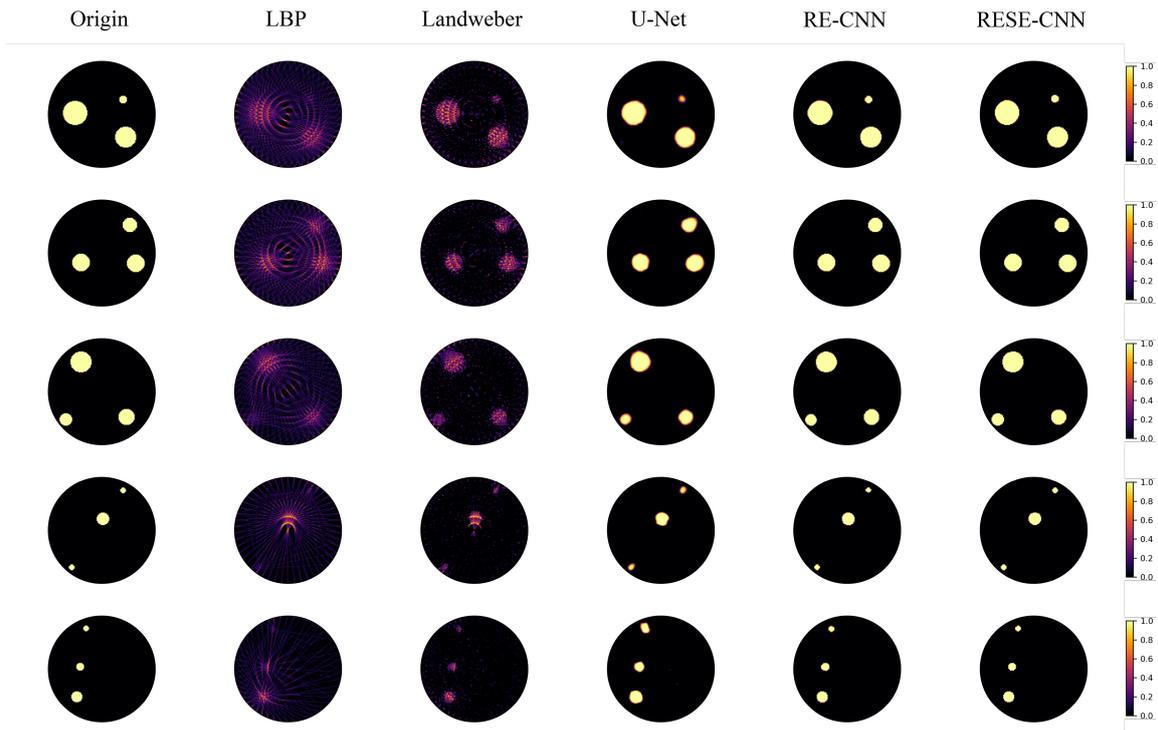


Fig. 11. Results of three-bubble distribution in a 128×128 pixel grid.

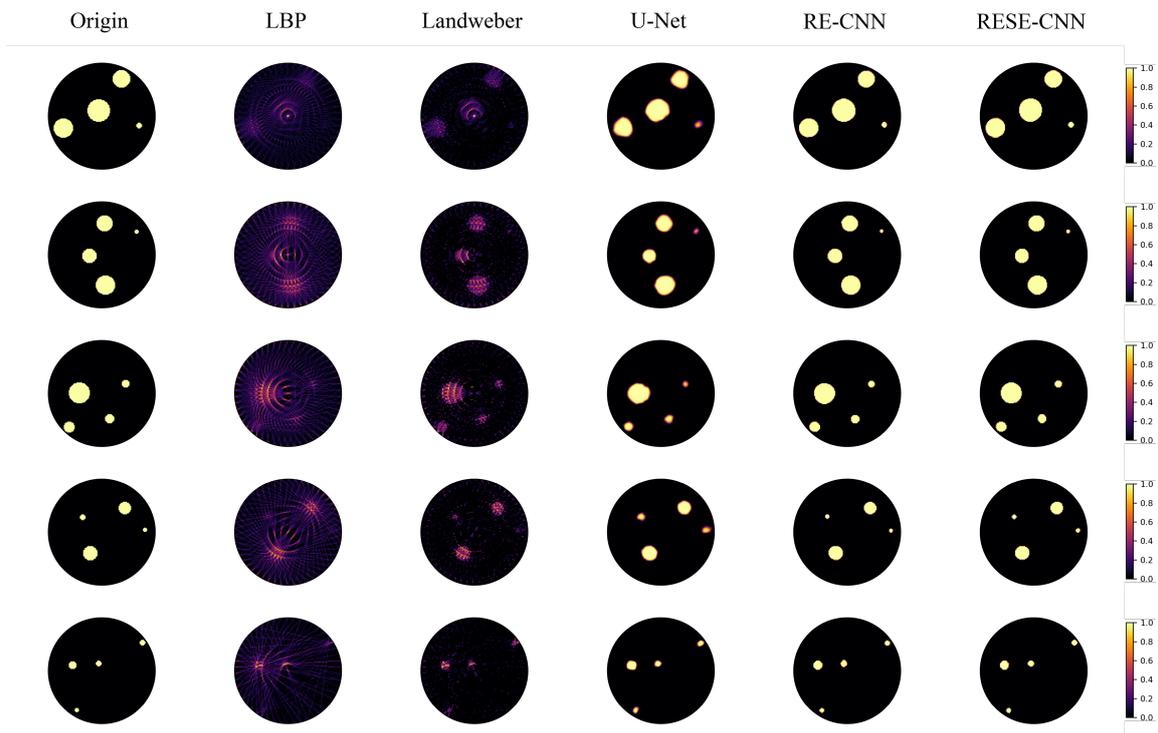


Fig. 12. Results of four-bubble distribution in a 128×128 pixel grid.

facts. Among the CNN methods, two alternatives were unable to adequately reconstruct images of larger bubbles. In scenarios with two bubbles, RESE-CNN, reconstructed the image but with slightly distorted boundaries and occasional unintended small bubble artifacts. In contrast, the compet-

ing methods consistently underperformed, with their results showing complete reconstruction of larger bubbles. In summary, RESE-CNN continues to deliver the best result of all the methods investigated, highlighting its superiority in OT image reconstruction tasks.



Fig. 13. Comparison between the results of traditional algorithms and sensitive field.

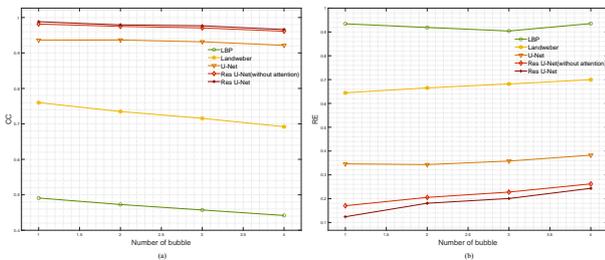


Fig. 14. RE and CC of different distributions.

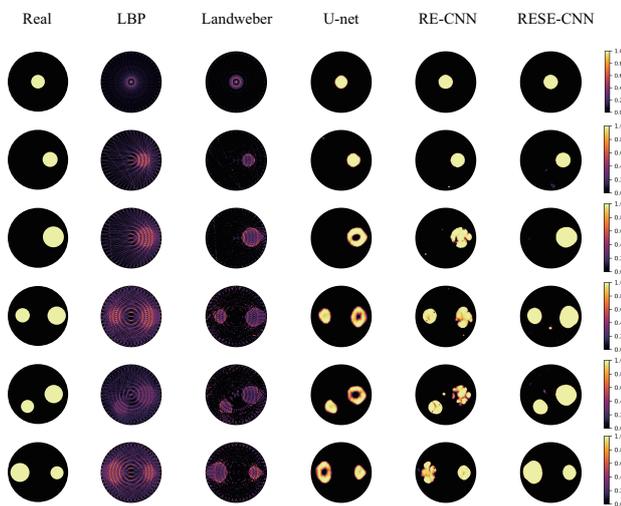


Fig. 15. Results of the experiments.

6. CONCLUSION

This paper presents an OT system using fan-beam lasers accompanied by the novel RESE-CNN method developed for high-resolution OT image reconstruction. RESE-CNN effectively extracts spatial information from initial, coarse images, significantly improving their accuracy. Experimental validations confirm that RESE-CNN is a viable approach for OT image reconstruction, characterized by high precision and fast imaging speeds. A notable challenge in OT imaging is the uneven distribution of sensitive fields, which leads to artifacts in reconstructed images that are difficult to mitigate using traditional methods such as LBP and Landweber. In contrast, deep learning techniques, especially RESE-CNN, are excellent at overcoming such challenges and almost eliminating their adverse effects. Practical verifications underline the method's reliability and show its superiority over alternatives in restoring object distributions from raw projection signals. Both in simulation and real-world experiments, the superior performance of RESE-CNN is repeatedly confirmed, affirming its efficiency in advancing OT image reconstruction capabilities.

RESE-CNN shows superior performance in reconstructing high-contrast binary absorption distributions, confirming its potential for applications such as gas-solid flow monitoring. However, the current system assumes idealized absorption conditions. Future work will focus on extending this method to weakly absorbing and scattering media where non-linear reconstruction problems dominate. Nevertheless, the overall results remain acceptable. Effectively combating noise in practical applications is of crucial importance for the future. Furthermore, improving the robustness of RESE-CNN by including more diverse and complex distributions — beyond the varying number of bubbles — in the training dataset would enable more precise mappings and extend the applicability of the method. These steps are essential to optimize the performance of RESE-CNN in a wider range of real-world scenarios.

REFERENCES

- [1] Schleicher, E., Da Silva, M., Thiele, S., Li, A., Wollrab, E., Hampel, U. (2008). Design of an optical tomograph for the investigation of single- and two-phase pipe flows. *Measurement Science and Technology*, 19(9), 094006. <https://doi.org/10.1088/0957-0233/19/9/094006>.
- [2] Li, H., Liu, G., Yu, S., Sun, C. (2022). Configuration optimization of optical tomography based on fan-beam lasers. *IEEE Transactions on Instrumentation and Measurement*, 71, 4505810. <https://doi.org/10.1109/TIM.2022.3201238>.
- [3] Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse problems*, 15(2), R41. <https://doi.org/10.1088/0266-5611/15/2/022>.
- [4] Cui, Z., Wang, Q., Xue, Q., Fan, W., Zhang, L., Cao, Z., Sun, B., Wang, H., Yang, W. (2016). A review on image reconstruction algorithms for electrical capacitance/resistance tomography. *Sensor Review*, 36(4), 429–445. <https://doi.org/10.1108/SR-01-2016-0027>.
- [5] Beck, M., Dyakowski, T., Williams, R. (1998). Process tomography—the state of the art. *Transactions of the Institute of Measurement and Control*, 20(4), 163–177. <https://doi.org/10.1177/014233129802000402>.
- [6] Yang, W., Peng, L. (2002). Image reconstruction algorithms for electrical capacitance tomography. *Measurement Science and Technology*, 14(1), R1. <https://doi.org/10.1088/0957-0233/14/1/201>.
- [7] Khairi, M. T. M., Ibrahim, S., Yunus, M. A. M., Faramarzia, M., Ayub, N. M. N. (2015). Application of optical tomography for monitoring gas bubbles flow based on independent component analysis algorithm. *Jurnal Teknologi*, 73(3), 145–152. <https://doi.org/10.11113/jt.v73.4260>.
- [8] Ben Yedder, H., Cardoen, B., Shokoufi, M., Golnaraghi, F., Hamarneh, G. (2022). Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 41(3), 515–530. <https://doi.org/10.1109/TMI.2021.3117276>.

- [9] Lecun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>.
- [10] Deng, A., Huang, J., Liu, H., Cai, W. (2020). Deep learning algorithms for temperature field reconstruction of nonlinear tomographic absorption spectroscopy. *Measurement: Sensors*, 10-12, 100024. <https://doi.org/10.1016/j.measen.2020.100024>.
- [11] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>.
- [12] Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [13] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4700–4708. <https://doi.org/10.1088/0957-0233/19/9/094006>.
- [14] Wu, Y., Chen, B., Liu, K., Zhu, C., Pan, H., Jia, J., Wu, H., Yao, J. (2021). Shape reconstruction with multi-phase conductivity for electrical impedance tomography using improved convolutional neural network method. *IEEE Sensors Journal*, 21(7), 9277–9287. <https://doi.org/10.1109/JSEN.2021.3050845>.
- [15] Huang, J., Liu, H., Dai, J., Cai, W. (2018). Reconstruction for limited-data nonlinear tomographic absorption spectroscopy via deep learning. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 218, 187–193. <https://doi.org/10.1016/j.jqsrt.2018.07.011>.
- [16] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [17] Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [18] Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H. (2019). Attention branch network: Learning of attention mechanism for visual explanation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 10705–10714. <https://doi.org/10.1109/CVPR.2019.01096>.
- [19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York, US: Curran Associates, 6000–6010. ISBN 978-1-5108-6096-4.
- [20] Itti, L., Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203. <https://doi.org/10.1038/35058500>.
- [21] Olshausen, B. A., Anderson, C. H., Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11), 4700–4719. <https://doi.org/10.1523/JNEUROSCI.13-11-04700.1993>.
- [22] Yun, J. (2024). Mitigating gradient overlap in deep residual networks with gradient normalization for improved non-convex optimization. In *2024 IEEE International Conference on Big Data (BigData)*. 3831–3837. <https://doi.ieeecomputersociety.org/10.1109/BigData62323.2024.10825094>.
- [23] Hayat, M. (2024). Squeeze & excitation joint with combined channel and spatial attention for pathology image super-resolution. *Franklin Open* 8, 100170. <https://doi.org/10.1016/j.fraope.2024.100170>.
- [24] Cai, W., Kaminski, C. F. (2017). Tomographic absorption spectroscopy for the study of gas dynamics and reactive flows. *Progress in Energy and Combustion Science*, 59, 1–31. <https://doi.org/10.1016/j.pecs.2016.11.002>.
- [25] Chen, Y., Li, K., Han, Y. (2020). Electrical resistance tomography with conditional generative adversarial networks. *Measurement Science and Technology*, 31(5), 055401. <https://doi.org/10.1088/1361-6501/ab62c4>.
- [26] Li, H., Yu, S., Liu, G., Zheng, X., Sun, C. (2022). Development of optical tomography system based on fan-beam lasers. *IEEE Sensors Journal*, 22(19), 18718–18725. <https://doi.org/10.1109/JSEN.2022.3198493>.
- [27] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- [28] Guan, S., Khan, A. A., Sikdar, S., Chitnis, P. V. (2020). Fully dense unet for 2-D sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health Informatics*, 24(2), 568–576. <https://doi.org/10.1109/JBHI.2019.2912935>.
- [29] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., Johansen, H. D. (2020). DoubleU-Net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 558–564. <https://doi.ieeecomputersociety.org/10.1109/CBMS49503.2020.00111>.
- [30] Valanarasu, J. M. J., Sindagi, V. A., Hacihaliloglu, I., Patel, V. M. (2021). KiU-Net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Transactions on Medical Imaging*, 41(4), 965–976. <https://doi.org/10.1109/TMI.2021.3130469>.

Received June 20, 2024
Accepted April 23, 2025