This full-text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Deep Learning Based Light-Field Refocusing of Burner Flames

Bamidele Ogunjumelo, Md. Moinul Hossain*, *Senior Member, IEEE*, Qi Qi, Gang Lu, *Senior Member, IEEE*
School of Engineering, University of Kent, Canterbury, Kent, CT2 7NT, UK
bao21@kent.ac.uk, *m.hossain@kent.ac.uk, q.qi@kent.ac.uk, g.lu@kent.ac.uk

***Abstract*— A light-field (LF) camera with refocusing algorithms enables the reconstruction of images at different depths. Using images of various depths, the intensity distribution of each image section can then be calculated, which is particularly useful for analyzing complex scenes such as flames. However, LF images cannot be refocused in real time using conventional algorithms due to high computational demands. In this paper, we present a deep learning (DL)-based method for real-time refocusing of LF images of a burner flame, leveraging depth information obtained from the captured LF images. A Convolutional Neural Network (CNN) model is developed using a transfer learning approach, where a pre-trained ResNet-50 model is extended with custom convolutional layers to perform the refocusing task. Synthetic datasets are generated using a ray tracing simulation based on the ray transfer matrix (RTM) method to train the model. The trained model produces four refocused outputs corresponding to distinct depth planes. The proposed method eliminates the need for the computationally intensive shift-and-sum algorithm traditionally applied to LF images. Simulation results show that the method accurately refocuses both geometric and flame structures, preserving depth-aware detail across planes. This approach has strong potential for enabling real-time, non-intrusive diagnostics in combustion systems and other dynamic, depth-varying environments.**

Keywords- light-field imaging, refocusing algorithm, deep learning, flame

I. INTRODUCTION

Understanding the characteristics of flames is crucial for enhancing energy efficiency and minimizing harmful emissions in combustion systems. Optical imaging techniques have become indispensable approaches for non-intrusive, 2-dimensional (2-D) and 3-dimensional (3-D) monitoring of the flames [1]. As a result, valuable insight into flame, such as structure, turbulence, temperature distribution, and chemical reactions, can be gained [2]. Among optical approaches, sectioning techniques such as Planar Laser-Induced Fluorescence (PLIF) [3], Laser Sheet Imaging [4], and Tomographic Schlieren Imaging [5] are commonly used to capture 2-D slices of a flame at different depths, which can then be used to reconstruct the temperature distribution, etc. Optical Sectioning Tomography (OST) [6] enables the capture of images at multiple depth layers. This 3-D distribution of luminosity of the flame can be reconstructed as a series of 2-D images, each focused on a different depth of the flame. The 3-D temperature distribution of the flame can then be reconstructed based on the relationship between luminosity and temperature. However, these optical sectioning methods often require complex and costly setups [7].

To address the above issues, researchers have employed single LF cameras to reconstruct 3-D flames [8]. By capturing both the spatial and angular distributions of light rays in a single exposure [9], LF imaging allows for computationally refocusing the images after capture [10]. This capability is especially valuable in flame studies, as it allows for acquiring 3-D characteristics of the flame with minimal physical intervention. Traditionally, LF refocusing algorithms such as the shift-and-sum technique [11] and the Fourier Slice method [10] are used to reconstruct sharp images at different depths from LF data. Despite their effectiveness, these traditional methods are constrained by high computational demands, including intensive memory usage and complex data handling, rendering them impractical for real-time LF processing tasks [12].

Recent advances in DL, including the development of more efficient network architectures, improved training strategies, and access to high-performance computing hardware, have the potential to improve both the efficiency and quality of LF processing tasks such as refocusing through DL-based methods [13]. DL models can be trained to learn optimal refocusing from large image datasets, making them an alternative for faster, more accurate, and real-time LF refocusing. Hedayati et al. [12] introduced RefNet, a DL-based technique for real-time LF refocusing, capable of generating 16 refocused images from a single LF capture. RefNet processes a 7×7 angular resolution LF input and predicts refocused images across a range of refocusing slopes (α = 0.125 to 2.0), where α is the refocus parameter used to adjust the synthetic focal plane distance during digital LF refocusing. Trained with ground truth data from shift-and-sum and Fourier slice methods, RefNet achieves mean Structural Similarity Index Measure (SSIM) values over 92.5% compared to the Fourier slice ground truth, proving high image quality. Dayan et al. [14] proposed RefocusNet, a DL model designed for real-time LF refocusing using sparse sub-aperture views. A RefocusNet operates with four sparse views arranged in a 2x2 grid, achieving high-quality refocusing without the need for dense angular data. Furthermore, Samarakoon et al. [15] demonstrated a method for arbitrary volumetric refocusing of both dense and sparse LF. They use pixel-dependent shifts within the shift-and-sum framework, complemented by a U-Net-based DL model, enabling simultaneous refocusing of multiple planar or volumetric regions within a scene. Collectively, these studies demonstrate that DL can be effectively used to generate refocused images from LF data.

In this paper, a DL method for real-time LF field refocusing of burner flames is presented. A transfer learning strategy using a pretrained ResNet-50[16] followed by custom convolutional layers is proposed. To train the network, synthetic datasets were generated using ray tracing simulations based on the RTM method. The trained model generates four refocused flame images at distinct depth planes. The simulation results show that the proposed method not only accelerates the refocusing process but also demonstrates the potential for real-time, non-intrusive diagnostics in flames in combustion systems.

II. METHODOLOGY

The block diagram of the proposed approach for real-time LF refocusing is illustrated in Fig.1. The approach involves generating a LF image dataset through a ray tracing simulation and extracting refocused images at various depths by the traditional shift-and-sum algorithm. These images serve as a ground truth for training the DL model. During the training process, each input LF image is passed through a pretrained ResNet-50, fine-tuned with custom convolutional and fully connected layers to predict four images corresponding to distinct focal planes. This enables the model to learn depth-aware refocusing from a single LF input. Supervision is provided using the ground truth images. To evaluate the model's generalization ability, the model is trained on synthetically generated LF data and then tested on a separate test dataset of unseen but similar samples.

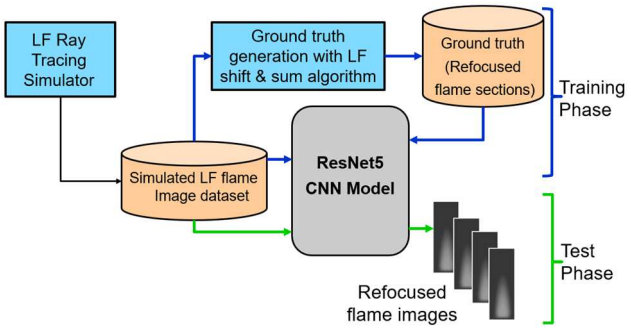

Fig. 1. Block diagram of the proposed DL model for real-time LF refocusing.

A. Light field imaging

LF imaging theory has been well documented [10]. For readers' convenience, only a brief introduction is given here. To capture an LF image, the LF camera incorporates an MLA between the main lens and the image sensor, as illustrated in Fig. 2(a). Each microlens captures a range of angular light rays arriving from different directions, enabling the camera to record both spatial and angular components of incoming light in a single exposure.

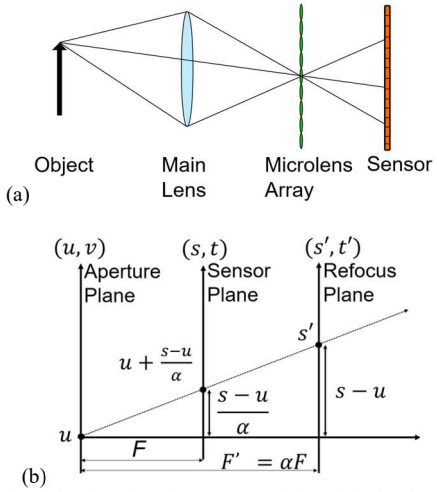

Fig. 2. (a) Principle of LF imaging setup and (b) Digital refocusing of LF.

This captured information is represented by a 4-D LF function, *L(u,v,s,t)*, where *(u,v)* are coordinates on the aperture plane (main lens), and *(s,t)* are coordinates on the sensor plane. In contrast to conventional cameras—which capture only 2-D spatial intensity images—this richer 4-D representation preserves directional information. As shown in Fig.2(b), a light ray is parameterized by its intersections with two planes: the aperture plane (*u, v*) and the sensor plane (*s, t*), separated by a reference distance *F*. This two-plane parameterization enables synthetic refocusing of the captured LF by virtually shifting the sensor plane to a new location (*s′, t′*) along the optical axis. The refocused plane is located at depth $F' = \alpha F$, where α controls the position of the virtual focal plane. Refocusing is implemented using the shift-and-sum algorithm, in which each sub-aperture image is shifted by an amount proportional to its angular coordinates (*u, v*) and the selected depth parameter α. These shifted views are then summed to generate a new image that is synthetically focused at the desired depth. A focused image at the sensor plane is given by [10]:

$$E(s,t) = \frac{1}{F^2}\iint L(u,v,s,t)\, A(u,v) cos^4\,\theta du\, dv \tag{1}$$

where *A (u, v)* is the aperture function and θ is the angle between a ray and the optical axis. For refocusing to a plane at αF, the image is computed as [10]:

$$E(s',t') = \frac{1}{F^2}\iint L\left(u,v,u+\frac{s'-u}{\alpha}, v+\frac{t'-v}{\alpha}\right) du\, dv \tag{2}$$

B. Light field ray tracing

To model light propagation within the LF imaging system, the RTM method, commonly known as the ABCD matrix formalism, was employed [17]. This approach models the behaviour of paraxial rays as they travel through the optical elements of the system.

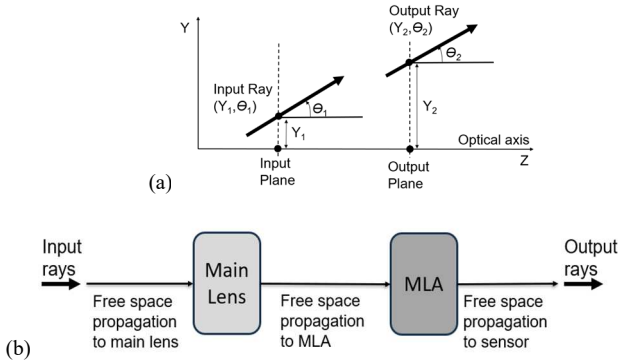

Fig. 3. RTM-based ray propagation in an LF imaging system (a) Geometric representation of a paraxial ray entering the system at the input plane and exiting at the output plane (b) Schematic of ray propagation through the optical system, including the main lens and MLA, with free-space intervals between components.

As illustrated in Fig.3(a), the diagram shows a single ray entering the optical setup at the input plane with an initial height Y_1 and angle θ_1 relative to the optical axis. Assuming a single lens lies between the input and output planes, the ray exits with an updated height Y_2 and angle θ_2. In the RTM framework, free space is treated as an optical element, as it alters the ray's position over distance without changing its angle. Each stage of the ray's propagation is governed by a 2×2 matrix, which updates the ray's position and angle:

$$\begin{bmatrix} Y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} Y_1 \\ \theta_1 \end{bmatrix}, \tag{3}$$

Here, the matrix elements A, B, C, and D represent the optical properties of each element in the system. Free-space propagation over a distance d is represented by the RTM:

$$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}, \tag{4}$$

and the effect of a thin lens with focal length f is modeled by:

$$\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \tag{5}$$

Each component in the system—free space, the main lens, and the MLA—is modeled using such matrices. By chaining them, the full ray path of the LFC is simulated from the object scene, through the lens and MLA, to the sensor as shown in Fig.3(b).

C. Deep learning model

In this study, refocusing is treated as a supervised image-to-image regression task where the goal is to map a single LF image $X \in \mathbb{R}^{H\times W\times C}$ to a set of four refocused outputs:

$$\hat{Y} = \{\hat{I}^{(n)}\}_{n=1}^{4}, \quad \hat{I}^{(n)} \in \mathbb{R}^{H\times W\times C}, \tag{6}$$

where, $\hat{Y}$ denotes the set of predicted refocused images, $\hat{I}^{(n)}$ represents the nth image corresponding to a specific refocus depth, H and W are the spatial dimensions, and C=1 indicates a single greyscale channel. Each output image $\hat{I}^{(n)}$ corresponds to a different focal depth slice, approximating traditional shift-and-sum refocusing. The network learns a parametric function $f_\varphi : X \rightarrow \hat{Y}$, where φ denotes the set of learnable parameters. This follows the standard function approximation paradigm in DL [18].

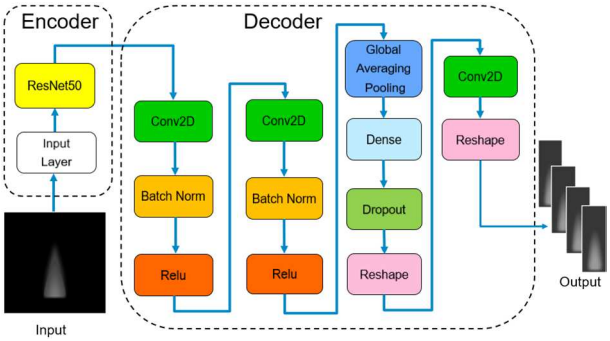

Fig. 4. Block diagram of the proposed DL model combining a ResNet-50 encoder with a custom decoder for refocused image generation

The proposed DL model uses a ResNet-50 backbone, pretrained on ImageNet [19], as an encoder to extract meaningful features from the input data. A lightweight decoder is subsequently applied to transform the extracted features into a set of depth-wise refocused outputs. The block diagram of the DL model, including this decoder, is illustrated in Fig.4. The decoder consists of a sequence of layers that transform high-level features into four refocused images, each corresponding to a distinct depth. It begins with a 2-D convolutional layer with 512 filters and a 3×3 kernel, followed by batch normalization to improve training stability and a Leaky ReLU activation to avoid dead neurons and preserve gradient flow [20, 21]. A second 2-D convolutional layer with 256 filters and a 3×3 kernel further refines the feature maps and captures localized spatial patterns. This is again followed by batch normalization and Leaky ReLU activation.

Next, global average pooling (GAP) is applied to compress the spatial dimensions of the feature maps into a compact global feature vector [22]. This vector is passed through a fully connected layer and reshaped into a tensor with a size of (90×180×4), where (90×180) represents the spatial resolution of each image, and the third dimension corresponds to four grayscale refocused images at different depths. To reduce overfitting, dropout regularization with a rate of 0.3 is applied [23]. The final convolutional layer applies 4 filters with a sigmoid activation to normalize the output to the range [0,1]. This produces four grayscale feature maps, each corresponding to a different refocused depth. The output is then reshaped to the format (4, 90, 100, 1), where 4 represents the number of focal planes, 90 and 100 denote the image height and width in pixels, respectively, and 1 indicates a single grayscale channel.

III. Model Construction

A. Multi-Shape Dataset

To train the DL model for LF refocusing, a synthetic dataset was generated through a ray tracing simulation. A 2-D scene template composed of basic geometric shapes (i.e., square, circle, rectangle, and triangle) was arranged in discrete depth layers along the optical axis, with a fixed 10 mm gap between adjacent layers as shown in Fig. 5. In the simulation, an MLA was used, where each microlens had a focal length of 0.4 mm and a pitch of 0.112 mm—that is, the center-to-center spacing between adjacent microlenses. The main lens had a focal length of 150 mm. The simulated LF setup achieves a horizontal and vertical field of view (FOV) of approximately 19.5°. The circular shape of the template is positioned at the focal plane, while the other shapes are placed at fixed intervals of 10 mm along the optical axis, both in front of and behind the focal plane.

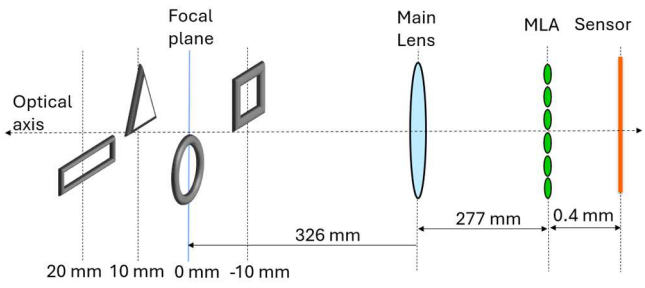

Fig. 5. Setup for LF camera simulation using multi-shape 2-D object scene templates.

Using the shift-and-sum algorithm, a range of slope values was applied to refocus the images, and sharpness metrics were used to identify the sharpest image for each shape. The square produced the sharpest image at a refocusing slope of −1.25, while the circle—positioned at the focal plane—was sharpest at a slope of 0.00. The triangle and rectangle, located behind the focal plane, were sharpest at slopes of 1.15 and 2.10, respectively.

B. Simulated Flame Dataset

Simulated LF flame images were also generated using the same approach as the multi-shape template. As illustrated in Fig.6, four flame images with different heights were

positioned at four different depths along the optical axis, creating distinct depth layers useful for evaluating the model's refocusing capability. In each dataset, the average flame height ranged from approximately 14.47 mm to 29.48 mm across different depth positions. Each flame exhibited minimal variation, with standard deviations between 1.41 mm and 1.48 mm. These values were computed from known 3-D object-space coordinates defined during the simulation.

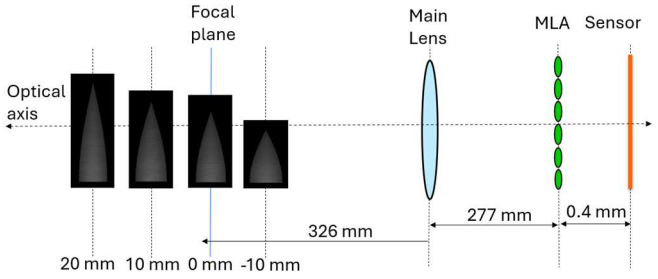

Fig. 6. Setup for LF camera simulation using simulated flame 2-D object scene templates.

C. Ground Truth

For the multi-shape dataset, a total of 1,600 simulated LF images were used to generate ground truth data. Each LF image produced four refocused views using the shift-and-sum refocusing algorithm, resulting in 6,400 ground truth samples in total. In contrast, the flame dataset consisted of 200 LF images, which similarly produced four refocused views each, yielding 800 ground truth samples. Each sample contains four grayscale refocused images, one for each depth, with a spatial resolution of 180×90. This corresponds to a tensor shape of (4,180,90), where 4 denotes the number of refocused depths.

D. Performance Metrics

Two image quality assessment metrics were used to evaluate the perceptual fidelity of the refocused images: the SSIM and the Peak Signal-to-Noise Ratio (PSNR) [24, 25]. The SSIM between two images x and y is calculated as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}, \quad (6)$$

where μ_x and μ_y are the mean intensities of images x and y, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance between x and y, and c_1 and c_2 are constants to stabilize the division.

The PSNR is given by:

$$PSNR = 10log_{10}\left(\frac{MAX_I^2}{MSE}\right), \ (dB) \quad (7)$$

where MAX_I is the maximum possible pixel value (e.g., 255 for an 8-bit image) and MSE is the mean squared error between the predicted and reference images.

The SSIM assesses the structural similarity between the predicted and reference images by considering luminance, contrast, and structural information. The SSIM index usually has a value that ranges from 0 to 1, where 1 implies that the reconstructed image perfectly matches the original one. The PSNR measures the ratio between the maximum pixel intensity and the power of the distortion. The higher the PSNR value the better the image quality.

E. Model Construction

The DL model was implemented in Keras with TensorFlow as the backend [26] using Python. A range of hyperparameters, including the number of convolutional layers, learning rates, and batch sizes, was evaluated to determine the optimal configuration. The finalized model employed a batch size of 2, a learning rate of 0.0001, and was trained using the Adam optimizer [27]. Backpropagation was used to update the trainable layers while keeping the ResNet-50 parameters fixed. The dataset was partitioned into training (80%), validation (10%), and testing (10%) subsets. The model was trained in a supervised manner using mean squared error as the loss function. All training was conducted on a Lenovo ThinkStation P620 equipped with an AMD Ryzen Threadripper PRO 3955WX CPU, 128 GB of RAM, and an NVIDIA RTX A6000 GPU. The model was trained separately on the multi-shape and simulated flame datasets. For the multi-shape dataset, training was performed on a GPU for 5,000 epochs, taking approximately 10.7 hours. The model trained on the simulated flame dataset used a CPU for 1,000 epochs and required approximately 6.35 hours. It is important to note that these training times are not directly comparable, as the models were trained using different epoch counts and hardware configurations.

IV. Results and Discussion

A. Refocused Multi-Shape Images

Refocused images were generated at depths of −10 mm, 0 mm, 10 mm, and 20 mm as shown in Fig.7. The model produced consistently high image quality across all planes, with SSIM values ≥ 0.96 and PSNR exceeding 43 dB, as shown in Table I. In the focal plane at 0 mm, which contains a circular object, the highest SSIM of 0.97 was observed. The PSNR at this depth was 45.31 dB, slightly lower than that of the square shape at 10 mm, which reached 47.87 dB, this may reflect the influence of the object's geometry. Circular edges are more sensitive to localization variations, which can impact pixel-level error metrics such as PSNR, as described in [28].

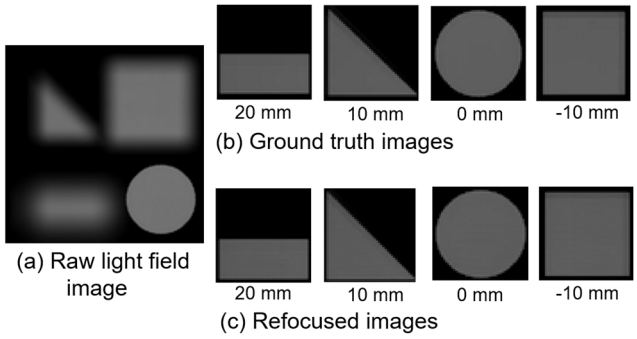

Fig. 7. (a) Raw LF image of a multi-shape template. (b) Ground truth slices at different depths. (c) Refocused images generated by the DL model.

TABLE I. PERFORMANCE EVALUATION OF REFOCUSING MULTI-SHAPE IMAGES

	Rectangle at depth 20 mm	**Triangle at depth 10 mm**	**Circle at depth 0 mm (Focal plane)**	**Square at depth -10 mm**
SSIM	0.96	0.96	0.97	0.96
PSNR	45.05dB	43.73dB	45.31dB	47.87dB

B. Refocused Flame Images

Refocused images of a simulated flame were evaluated at −10 mm, 0 mm, 10 mm, and 20 mm as shown in Fig.8. As shown in Table I, SSIM values ranged from 0.91 to 0.96, while PSNR values exceeded 39 dB across all depths for the DL model. The highest SSIM, 0.96, was achieved at the focal plane at 0 mm, indicating excellent structural similarity. At 10 mm, the SSIM remained high at 0.94, showing that the flame structure was well preserved despite the lower PSNR of 39.41 dB. This confirms the model's robustness in maintaining image quality even when the flame is outside the focal plane.

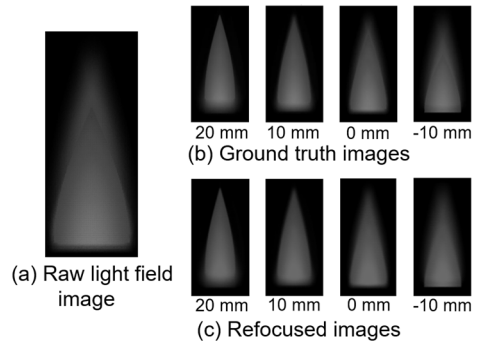

Fig. 8. (a) Raw LF image of a simulated flame. (b) Ground truth slices at different depths. (c) Refocused images generated by the DL model.

TABLE II. PERFORMANCE EVALUATION OF REFOCUSING FLAME IMAGES

	Flame at depth 20 mm	**Flame at depth 10 mm**	**Flame at depth 0 mm (Focal plane)**	**Flame at depth -20 mm**
SSIM	0.91	0.94	0.96	0.95
PSNR	41.64dB	39.416dB	40.06dB	41.87dB

V. Conclusion

In this paper, we presented a DL-based method for LF refocusing of burner flame images, with the potential for real-time application. The proposed approach eliminates the need for the computationally intensive shift-and-sum algorithm traditionally applied after capture. The key contributions include the implementation of a ray-tracing simulator based on the RTM to generate synthetic LF data and the development of a transfer learning–based DL model using a ResNet-50 backbone with custom convolutional layers. This model replaces the traditional shift-and-sum algorithm with an optimized ResNet-50-based architecture capable of real-time depth-aware refocusing. The model was evaluated on both geometric test scenes and simulated flame data. Geometric samples validated accurate refocusing at known depths, while flame evaluations demonstrated the model's ability to preserve flame structure across multiple focal planes. The proposed method has the potential to enable real-time LF refocusing, making it well suited for practical deployment in combustion diagnostics. Future work will focus on expanding the training dataset to include a wider range of flame conditions, incorporating real LF camera data, and integrating post-processing pipelines for the automated extraction of radiation-intensity profiles from the refocused outputs.

VI. Acknowledgement

The authors wish to express their gratitude to the Engineering & Physical Sciences Research Council (EPSRC) Project Reference: EP/X020789/1.

References

[1] L. Zheng, Combustion Visualisation Monitoring Using High Speed Imaging, Ph.D. dissertation, Univ. of Sheffield, Sheffield, U.K., 2018.

[2] Y. Ying and G. Situ, "A survey for 3D flame chemiluminescence tomography: Theory, algorithms, and applications (Invited)," Frontiers in Photonics, 2022.

[3] M. G. Allen et al.,"Planar laser-induced-fluorescence imaging measurements of OH and hydrocarbon fuel fragments in high-pressure spray-flame combustion," Appl. Opt., vol. 34, no. 27, pp. 6287–6300, 1995.

[4] M. Horio et al., "Instrumentation and measurements," in Handbook of Fluidization and Fluid-Particle Systems, W.-C. Yang, Ed. Boca Raton, FL: CRC Press, 2003, pp. 643–704.

[5] S. J. Grauer, A. Unterberger, A. Rittler, K. J. Daun, A. M. Kempf, and K. Mohri, "Instantaneous 3D flame imaging by background-oriented schlieren tomography," Combust. Flame, vol. 196, pp. 284–299, 2018.

[6] B. Zhou, S. Wang, C. Xu, and J. Zhang, "3-D flame temperature reconstruction in optical sectioning tomography," in Proc. IEEE Int. Workshop Imag. Syst. Techn., May 2009, pp. 313–318.

[7] M. M. Hossain, Tomographic Characterization of Burner Flames Through Digital Imaging and Image Processing, Ph.D. dissertation, Sch. Eng. Digit. Arts, Univ. Kent, Canterbury, U.K., 2014.

[8] J. Sun et al., "Investigation of flame radiation sampling and temperature measurement through light field camera," Int. J. Heat Mass Transf., vol. 121, pp. 1281–1296, 2018.

[9] M. Levoy and P. Hanrahan, "Light field rendering," in Proc. SIGGRAPH, 1996.

[10] R. Ng, Digital Light Field Photography. Ph.D. dissertation, Stanford University, Stanford, CA, USA, 2006.

[11] W. Fu et al., "Implementing light field image refocusing algorithm," in Proc. 2015 2nd Int. Conf. Opto-Electron. Appl. Opt. (IEM OPTRONIX), Vancouver, BC, Canada, 2015, pp. 1–8.

[12] E. Hedayati et al., "Machine learning method for light field refocusing," arXiv preprint arXiv:2103.00056, 2021.

[13] J. Y. Chang, Y. Wang, and J. Zhang, "Light field image processing: An overview," IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 95–106, Nov. 2017.

[14] S. Dayan et al., "Deep sparse light field refocusing," in Proc. 31st British Mach. Vis. Conf. (BMVC), 2020.

[15] T. Samarakoon et al., "Arbitrary volumetric refocusing of dense and sparse light fields," arXiv preprint arXiv:2502.19238, 2025.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Microsoft Research, pp. 770–778, Dec. 2015.

[17] B. E. A. Saleh and M. C. Teich, Fundamentals of Photonics, 2nd ed. Hoboken, NJ, USA: Wiley, 2007.

[18] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[19] J. Deng et al, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, June 2009, pp. 248–255.

[20] A. Maas et al., "Rectifier Nonlinearities Improve Neural Network Acoustic Models," ICML, 2013.

[21] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," ICML, 2010.

[22] M. Lin, Q. Chen, and S. Yan, "Network in network," ICLR, 2014

[23] N. Srivastava et al., "Dropout: A simple way to prevent neural networks from overfitting," JMLR, 2014

[24] Z. Wang et al., "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[25] A. Hore and D. Ziou,"Image quality metrics: PSNR vs. SSIM,"2010 20th International Conference on Pattern Recognition (ICPR), pp. 2366–2369, IEEE, 2010.

[26] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," arXiv preprint arXiv:1603.04467, 2016.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations (ICLR), 2015.

[28] P. R. S. Mendonça et al., "Bias in the localization of curved edges," in Computer Vision – ECCV 2004, vol. 3024, T. Pajdla and J. Matas, Eds. Berlin, Heidelberg: Springer, 2004, pp. 162–174.