IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 February 2024, accepted 1 March 2024, date of publication 4 March 2024, date of current version 20 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373459

APPLIED RESEARCH

Semantic Liquid Spray Understanding With Computer-Generated Images

WEI LUN LIM[1], (Graduate Student Member, IEEE), MATTHEW Y. W. TEOW[2], (Senior Member, IEEE), RICHARD T. K. WONG[3, 4], REFAT KHAN PATHAN[1], (Student Member, IEEE), CHIUNG CHING HO[1], (Senior Member, IEEE), RAHUL BABU KONERU[5], PRASHANT KHARE[6], LUIS BRAVO[7], AND SIAN LUN LAU[3,4], (Senior Member, IEEE)

[1]Department of Computing and Information Systems, Sunway University, Bandar Sunway, Petaling Jaya 47500, Malaysia
[2]University Partnership (Coventry University), PSB Academy, Singapore 039594
[3]Department of Engineering, Sunway University, Bandar Sunway, Petaling Jaya 47500, Malaysia
[4]Sunway DataXSight Research Cluster, Sunway University, Bandar Sunway, Petaling Jaya 47500, Malaysia
[5]Department of Aerospace Engineering, University of Maryland, College Park, MD 20742, USA
[6]Department of Aerospace Engineering, University of Cincinnati, Cincinnati, OH 45221, USA
[7]DEVCOM Army Research Laboratory, Adelphi, MD 21005, USA

Corresponding author: Sian Lun Lau (sianlunl@sunway.edu.my)

This work was supported by the U.S. Army International Technology Center Pacific funded project—Deep-Spray+ under Contract FA520920P0188. The work of Luis Bravo was supported by the U.S. Army Research Laboratory 6.1 Basic Research Program in Propulsion Sciences.

ABSTRACT Understanding liquid spray is essential for spray applications, including but not limited to designing fuel-efficient engines. Due to the challenges involved in collecting real-world liquid spray images, a synthetic liquid spray was generated using fluid simulation based on the atomization of the liquid jet. Semantic segmentation was chosen to analyze the liquid spray, as it reflects the precise location of the objects in the image. This paper presents a workflow to train a U-Net with a small sample (only 24 training images) dataset under the constraint that no ground truth is provided. An image is selected from the generated images of liquid spray and edited by randomly masking some objects. After the chosen image is annotated, a data augmentation technique that includes rotation and Gaussian smoothing is applied, resulting in 24 images available as the training set. The RGB original-sized images are fed to U-Net for training. Due to how the liquid spray images are obtained in the real world, Gaussian smoothing is explored as the inductive bias. Gaussian smoothing is incorporated between the convolutional layers of U-Net to enhance its feature extraction ability. The experiment results showed that the segmentation output improved when smoothing was incorporated into the U-Net. By visualizing the convolutional feature map of trained U-Net, we discover that smoothing makes convolution less biased to texture information. Going through this workflow, the trained U-Net is found to generalize well to the test images despite learning from few samples. Code is available at https://github.com/lynerlwl/spray-unet

INDEX TERMS Computer vision, image processing, image statistics, semantic segmentation, scene understanding, convolutional network.

I. INTRODUCTION

Efficient combustion is a fundamental aspect for optimizing engine performance. For efficient combustion to occur in a propulsion engine, the fuel must mix well with air, creating a homogeneous mixture that burns and releases energy to power the vehicle. In liquid-fuelled systems, atomization is crucial for creating this fuel-air mixtures. Atomization breaks liquid jets into a spray of smaller ligaments called droplets, that will be dispersed into gas [1]. If the fuel remains in large droplets, it will not mix properly with the air, leading to inefficient combustion. Therefore, understanding the spray formation process is essential for the development of air-breathing propulsion systems [2]. Thus, analysis of the size,

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

distribution, and spatial arrangement of the ligaments in the spray is necessary.

The spray was photographed using a high-speed camera [3]. Thus, image segmentation comes naturally as the suitable method to be applied to the photographed liquid spray image to analyze the ligaments. This research attempts to showcase the process of preparing liquid spray images for semantic segmentation under the condition that no ground truth is available. Since obtaining real-world liquid spray images is challenging, computer-generated liquid spray images are used for this research. The trained model can later be adapted to real images using domain adaptation technique without requiring the annotations [4]. Alternatively, augmenting the real images with synthetic images can also improve the segmentation accuracy of the trained model [5]. Four object classes exist in the computer-generated liquid spray images, differentiated by the ligament shape and aspect ratio.

Most of the works on liquid spray analysis performed segmentation on small droplet areas, and a simple U-Net trained on a limited number of computer-generated liquid spray images proved that convolutional networks is capable of recognizing the different ligament classes in the test images with dense object segmentation [6]. However, the convolutional network is found to be biased towards texture. Thus, increasing shape bias will increase the robustness of the network [7]. It is best if the network have a balance bias towards texture and shape, not just shape-biased [8]. As mentioned, the main feature of liquid spray images is shape. If the learning of the model relies less on texture and more on shape, it would produce a better segmentation model. One of the ways to achieve that is by applying smoothing to the image to emphasize the edge structure. Due to the nature of liquid spray images taken with high-speed cameras likely exposed to motion or defocus blur, Gaussian smoothing is applied to the test images to check the robustness of the trained model [9]. Through an empirical discovery, inducing an appropriate blurring artifact into the test image significantly maximize the contour boundary visibility. This inductive bias allows U-Net to recognize the initially undetected ligaments in the image. However, an adaptive blurring is necessary to provide an optimal recognition rate [10].

Gaussian smoothing applied to the image could enhance the contour visibility. What if Gaussian smoothing is incorporated into the network architecture? As the subsampling layers in the convolutional network, such as max-pooling or strided convolution, do not follow the Nyquist sampling theorem, the high-frequency components might not be adequately sampled, causing the convolutional feature map to have aliasing [11]. Inclusion of a pooling layer is essential to semantic segmentation as it enlarges the receptive field of the network. An information-preserving downsampling module is needed to minimize the information lost such that boundary, scale, and texture information can be preserved [12]. Though applying anti-aliasing before subsampling improves shift-equivariance in the convolutional network, anti-aliasing and data augmentation cannot achieve fully translation-invariant in the convolutional network due to non-linear activation functions [13]. A non-linear sampling layer that selects the sampling grid adaptively, namely adaptive polyphase sampling (APS), is proposed to replace the conventional pooling layer in the convolutional network, allowing the convolutional network to be truly shift invariant [14]. Since APS is a handcrafted downsampling method, a generalization of APS called learnable polyphase sampling (LPS), which is end-to-end trainable, is proposed [15].

Inspired by the fact that smoothing before downsampling is better for retaining spatial information, the pooling layer in U-Net is replaced by the LPS layer [15]. The new U-Net will be referred to as the improved U-Net. Fig. 1 shows the architecture of the improved U-Net used in this research. An image from the computer-generated liquid spray dataset is chosen randomly to create the training set. A few random parts of the chosen image are masked to be different from the original image. Then, the image is partially annotated. With the ground truth, rotation and Gaussian smoothing are applied to the image for extra image generation. Fig. 2 depicts the data preparation workflow. Now, all 24 new images are fed to improved U-Net for training. The visual comparison of the segmented outputs from basic and improved U-Net showed that the improved U-Net produces better results. Since the ground truth for testing images is unavailable, it is inaccurate to say a particular model is doing better based on the training loss or dice score. Thus, a metric that counts the number of contours in the segmented outputs is proposed to justify any improvement made to the model. Through evaluating the visual behavior of the convolutional feature map [16], the improved U-Net is found to have the feature map with a less jagged effect, indicating the Gaussian smoothing successfully reduce the texture bias.

In summary, we demonstrated the procedure to train a semantic segmentation model from a small sample dataset without any ground truth. With that, we addressed the research limitations. Firstly, semantic segmentation is known to be needing many training images. We showed that we can train a segmentation model using a small sample of 24 with correct data preprocessing. Secondly, the irregular-shaped liquid spray is hard to detect. We improved the detection rate with Gaussian smoothing. The rest of this paper is organized as follows. Section II explains the properties of the liquid spray images in the dataset. Section III shows the experiments using U-Net to validate different experimental settings. The results are presented along with a detailed discussion in section IV before concluding the paper in section V. The main contributions of this paper are as follows:

1) The image size is important in training a semantic segmentation model. The experiments show that training the model with the original image size yields a more accurate segmentation outcome than using the resized

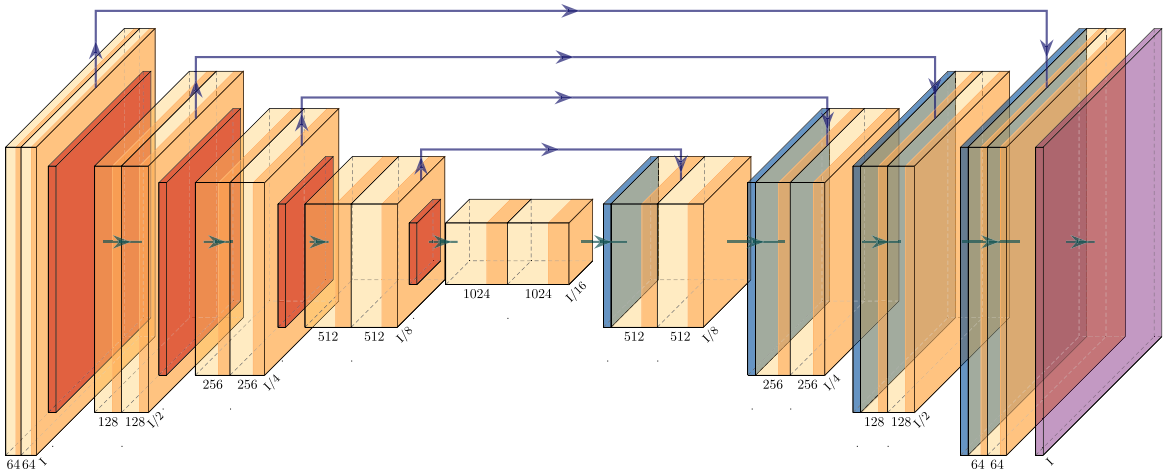


FIGURE 1. This is a description of the architecture of an improved U-Net. The light orange block represents the convolution and ReLU layer. The dark orange downsampling layer is referred to as LPS and is followed by Gaussian smoothing. The blue layer denotes the transpose convolution layer. The final layer in magenta is the output layer that consists of the Softmax activation function.

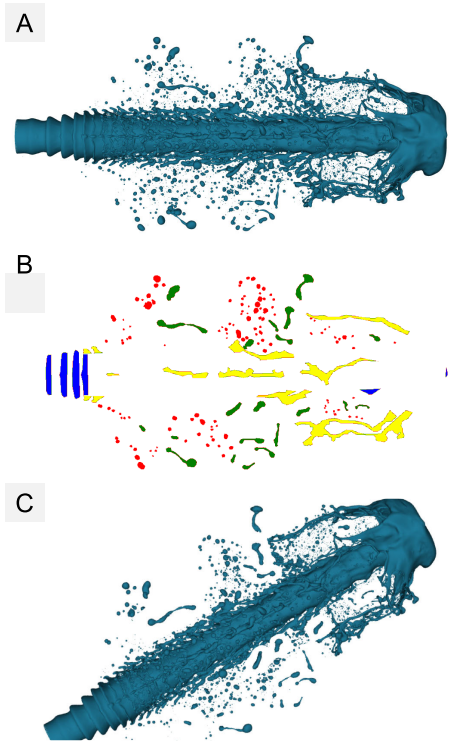


FIGURE 2. A pipeline for preparing training data. The unlabeled liquid spray in image A is annotated, producing the ground truth in image B. Then, data augmentation techniques involving rotation and Gaussian smoothing are applied to both images, creating the training set, as seen in image C.

image. The color information matters to the feature learning. Keeping the training images in RGB instead of grayscale to leverage the color information is better to differentiate the contour.

2) Incorporating blurring between the convolution layer of U-Net improves feature learning, thus leading to better segmentation performance.

II. DATASET – LIQUID SPRAY IMAGES

The dataset used in this research study contains computer-generated liquid spray images unique to the aerospace engineering domain, which is uncommon to natural scene images that can be found and largely used in most computer vision applications. For example, the ImageNet dataset [17], which is commonly used by both computer vision researchers and practitioners in their visual experiments, in contrast, this computer-generated liquid spray images dataset is primarily for aerospace engineers designing jet engines. They need to understand the distribution of ligaments in the spray. In addition to these images are domain-oriented, applying segmentation to these computer-generated liquid spray images is also beyond the usual foreground and background separation. Indeed, part segmentation is required to analyze different classes of ligaments specifically. Following on, the main challenge of segmenting images in this dataset is that no ground truth has been provided. Therefore, labeling objects in these images further posts another labor difficulty, especially when some objects are tiny in size fewer than ten pixels.

The computer-generated liquid spray dataset contains 1573 images of the liquid spray transition from a bulk liquid to liquid spray. The liquid spray images are named as f_n, where n is the frame number. For example, the first image is named f_00000, and the last image is named f_01572. All images in the dataset are horizontal-orientated with irregular-shaped liquid sprays and captured in different aspect ratios. Fig. 3 shows some images sampled from the dataset. These images are generated using Basilisk [18], an open-source computational fluid dynamics software. All red images in the dataset are 8-bit RGB images with a scale of 256 intensity levels and having a spatial size of 1200×600 pixels.

An image is selected randomly to be edited as the training image. This selected image is f_01213. The image histogram is used to study the color distribution. The primary colors in the image observed visually are blue (for objects) and

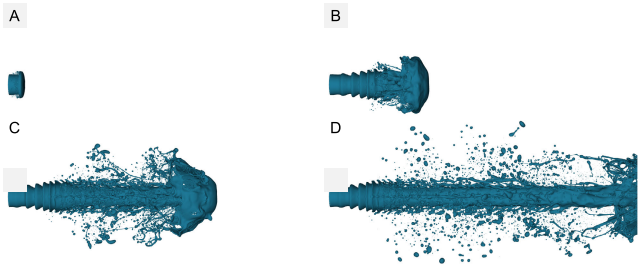


FIGURE 3. Liquid spray image generation. This figure shows the progressive transition of the liquid spray test image from the initial graphic generation stage in image A until the completion of image formation in image D.

white (for background). The intensity distribution of the histogram is highly imbalanced as the histogram without scaling shows the white color as the majority composition of the image, with the pixel value 255 around 70% of the image. Thus, the y-axis of the following histogram is limited to 0.025 to observe other pixel values, as shown in Fig. 4. The red channel in the image histogram is significantly higher than the green and blue channels. This color imbalance suggests that color information is essential to this image. Thus, continuing to work on this image in grayscale, which removes the color information, is not recommended. The range between the darkest and brightest intensities is clear, which makes the object easily distinguishable from the background. However, this also means that the objects are less distinct, making it harder to separate the objects from each other.

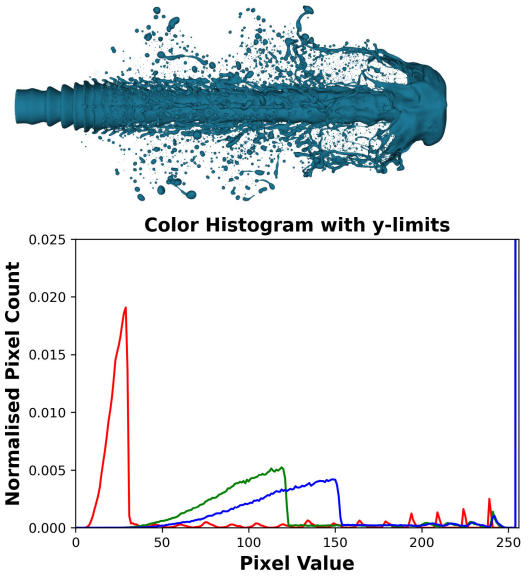


FIGURE 4. The image histogram for image f_01213 of the liquid spray image dataset.

After exploring the color information, we look at the image texture in Fig. 5 using the entropy. From different entropy values, we observe that the object has uneven texture. Take the droplets in the image D as an example. The region near the edge has a rougher texture than the hollow inner region due to low values filtered. By comparing image A to image D, more objects with low texture can be seen in images A and B.

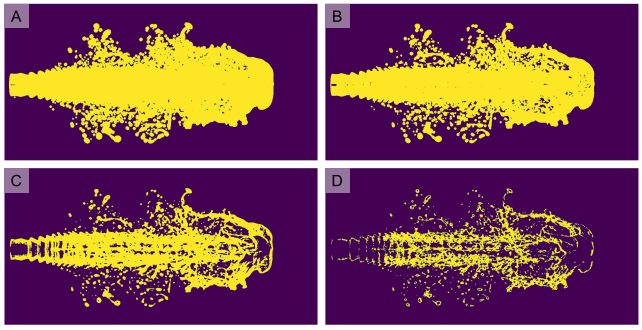


FIGURE 5. Entropy image for image f_01213 with 0.01, 1.00, 3.00, and 4.00 threshold settings, respectively.

A. SELECTION OF TRAINING IMAGE

This dataset is prepared for training and validating a semantic segmentation model. The objects of interest or object classes in this dataset are droplet, detached ligament, attached ligament, and lobe. All these classes will be explained in the ground truth annotation paragraph. An image with all four object classes clearly visible is chosen from the dataset and edited as the training image. The edited image has some objects removed. See Fig. 6 for the comparison. The selected image is edited to avoid using the same image in the dataset to train the semantic segmentation model.

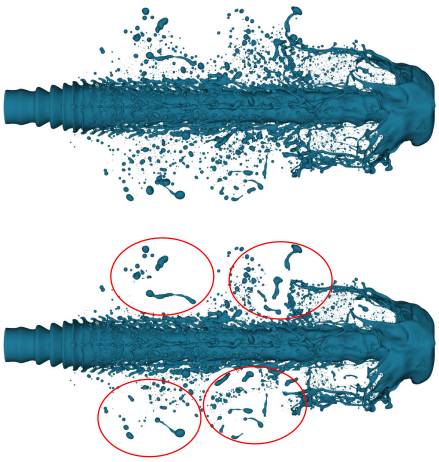


FIGURE 6. The droplet masking process used in the dataset preparation. The example is the image f_01213. The top image is the original, unedited image, and the bottom image is the edited image with masking applied. The red circles in the edited image indicate the masked areas.

B. ANNOTATING LABEL

A pixel-wise label is needed for semantic segmentation. To obtain the pixel-wise label, we annotate the label on the edited version of that image using labelme [19]. The partially

labeled image is shown in Fig. 7, as indicated by different colors in the label. The descriptions of the four object classes are summarised in Table 1.

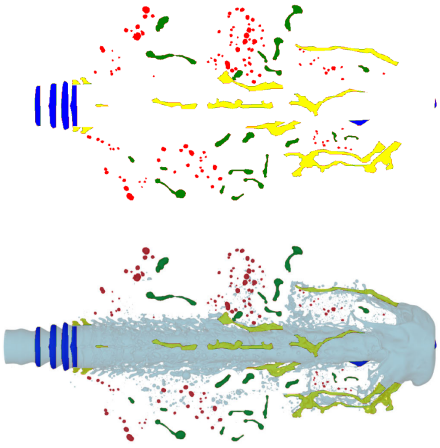


FIGURE 7. Ground truth of the training image. The top image is the label for training, and the bottom image overlays the training image f_01213 with the label.

C. DATA AUGMENTATION

The only preprocessing done to the training image is rotation and Gaussian smoothing. The training image is rotated in 30-degree intervals eleven times to complete a full 360-degree rotation. The rotation allows the model to learn the orientation of the spray in the image when the camera position is off-center. Smoothing highlights the coarse contour structure in the image, indicating that the edge or shape is an important feature. Gaussian blurring with 0.5 standard division is applied to the images after rotation, resulting in a training dataset of 24 images.

D. SEGMENTATION CHALLENGE

Based on the ground truth of the image, the ligaments are differentiated by their shape. Using conventional image segmentation algorithms to segment the object is difficult due to the complexity of the scene. Fig. 8 shows a few segmentation results in which the image is hardly partitioned into the correct segments. The segmentation challenges come from no identifiable main object appearing significantly on the image surface for easy segmentation. The objects are sparse and scattered around the surface of the test images to seek successful segmentation. In addition, the success of object segmentation also depends on the formation of the object's shape and its spatial distance across the neighborhood objects. U-Net as a variant of CNN, trained with a small amount of data, had successfully segmented the shadowgraph liquid spray [20]. Thus, by using a convolutional network and the ground truth in supervised learning, we can develop a computation model to classify the ligaments in the spray image. However, CNN is known to be biased towards texture, a less important feature in our dataset.

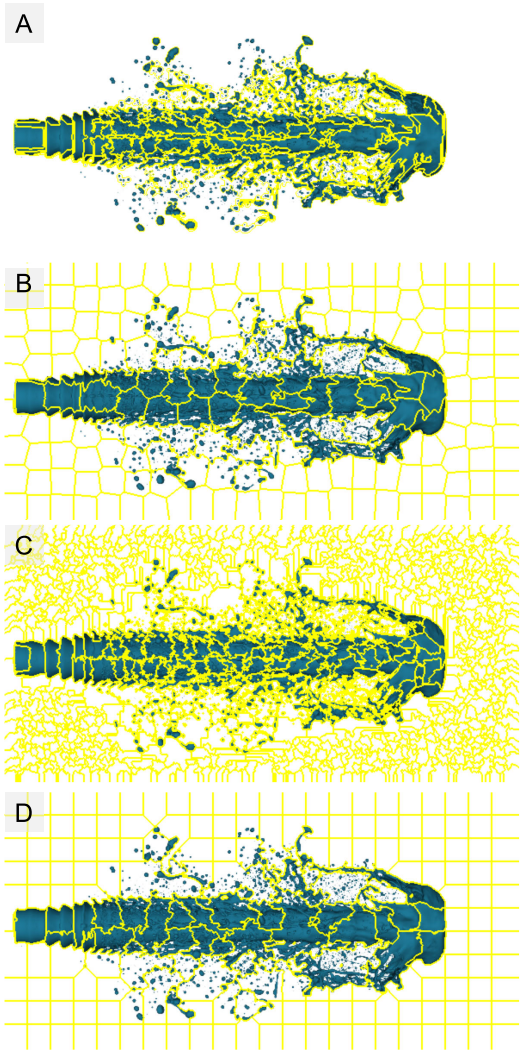


FIGURE 8. Four conventional segmentation algorithms were applied to the training image to showcase the difficulties of the segmentation. Image A is Felzenszwalbs's Algorithm, image B is SLIC, image C is Quickshift, and image D is Compact Watershed.

III. EXPERIMENT

A. U-NET

U-Net [21] is an encoder-decoder network in which skip connections link shallow layers with deeper ones. The encoder is responsible for feature learning, containing convolutional and pooling layers. On the other hand, the decoder reconstructs the feature learned in the encoder, and it contains transpose convolution as the upsampling layer.

Two types of U-Net are developed: a **basic U-Net** and an **improved U-Net** with the shift-invariant mechanism. For the shift-invariant U-Net, the pooling layers in the encoder part are replaced with the LPS layer [15]. The reason for having two models is to observe how the blurring layer in the feature extraction layer affects the segmentation result.

The U-Net is trained from scratch, meaning no pre-trained backbone is used in the feature extraction layer. The reason is that the visual objects of the image used in this research

TABLE 1. Object classes and their descriptions in training image.

Object name	Color on image	Class pixel value	Object description
Background	Black	0	Uninterested object that including the core.
Drop	Red	1	Ligaments with a small aspect ratio detached from the liquid accumulation.
Detached ligament	Green	2	Ligaments with a high aspect ratio that is detached from the liquid accumulation.
Attached ligament	Yellow	3	Ligaments with a high aspect ratio that is still attached to the liquid accumulation.
Lobe	Blue	4	The wavy shape of the liquid accumulation, it can be seen as the precursor of a ligament before it elongates axially.

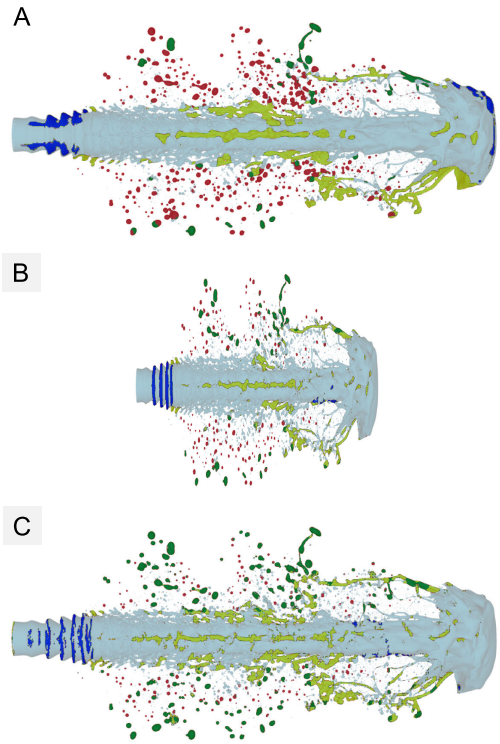


FIGURE 9. Image A is the predicted image from basic U-Net trained with the original size image. Image B and C are the predicted image from basic U-Net trained with resized image. Both are the same image with different resolutions.

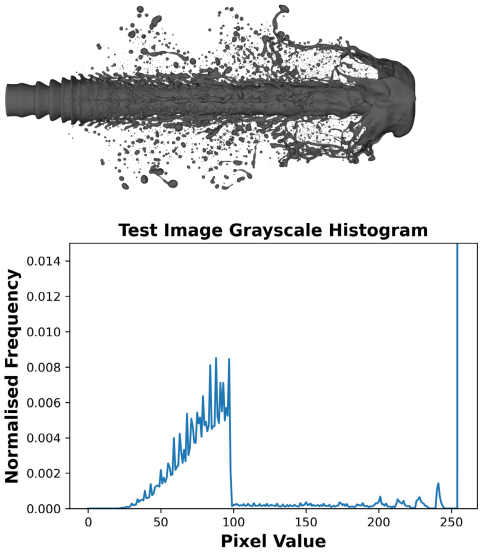


FIGURE 10. Test image in grayscale and its image histogram.

for segmentation are visually different from the natural scene images widely used in most AI/ML applications regarding visual pattern formation and spatial geometrical appearances. We will go through the elements that affect model training next.

B. IMAGE PROPERTIES

The original resolution of the training images is 1200 × 600. We tried resizing it to 600 × 600 to reduce the size to fit more images per batch size. The images are fed to basic U-Net for training. As the image is not resized following the aspect ratio, the U-Net wrongly predicted the objects in the testing image, as shown in image C in Fig. 9.

This subsection shows that the semantic segmentation of this dataset requires the color information for better segmentation. The histogram of the grayscale image, as shown in Fig. 10, contains less information than the RGB version, as shown in Fig. 4. We trained two improved U-Net, with RGB images and grayscale images, respectively. Fig. 11 is the predicted output showing that the model with color information predicts better than the model without color information. The model trained with grayscale images loses information about the lobe class, as the prediction only contains four classes out of the five, of which the blue-colored pixel is missing.

C. FINAL SETTINGS ON MODEL TRAINING

Both U-Net are trained with original-sized RGB images. With a 10% train-test split, the train set is 22, and the validation set is two. The deep learning framework used in the experiment

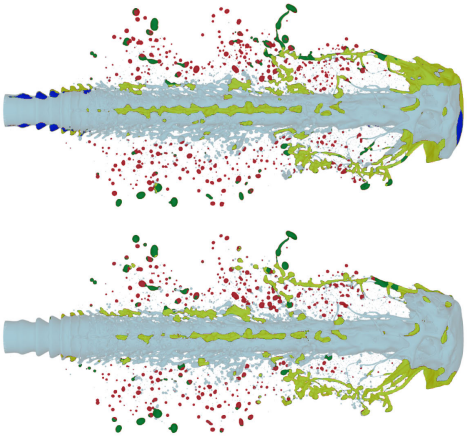


FIGURE 11. Prediction outcome of improved U-Net. The top image is from a model trained with RBG images, and the bottom image is the prediction outcome of the model trained with grayscale images.

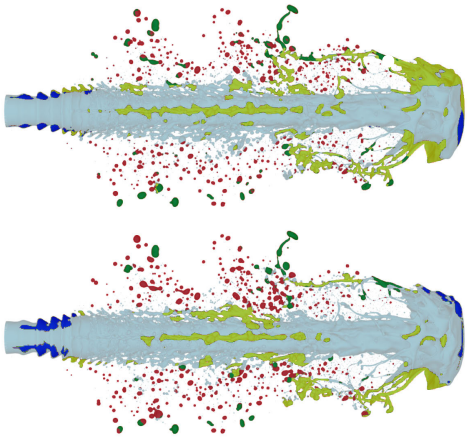


FIGURE 12. Prediction result of horizontal image f_01340. The top image is the result of improved U-Net that detected 493 droplets while the bottom is the result of basic U-Net that detected 382 droplets.

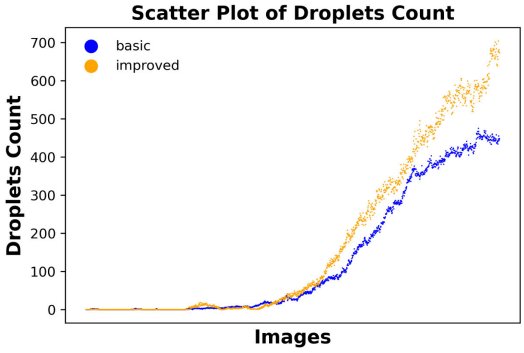


FIGURE 13. Scatter plot of droplets count across the dataset.

is PyTorch. The model is trained with ten epochs per run and one batch size, initialized by a 10-e4 learning rate. The loss function is the sum of cross entropy loss and dice loss. The backpropagation algorithm is RMSprop. Both models are trained on Google Colab with T4 GPU. The basic version of Google Colab comes with the Intel(R) Xeon(R) CPU @ 2.2GHz and 13 GB RAM. The training time per run for basic U-Net is around 5 mins, and 15 mins for improved U-Net.

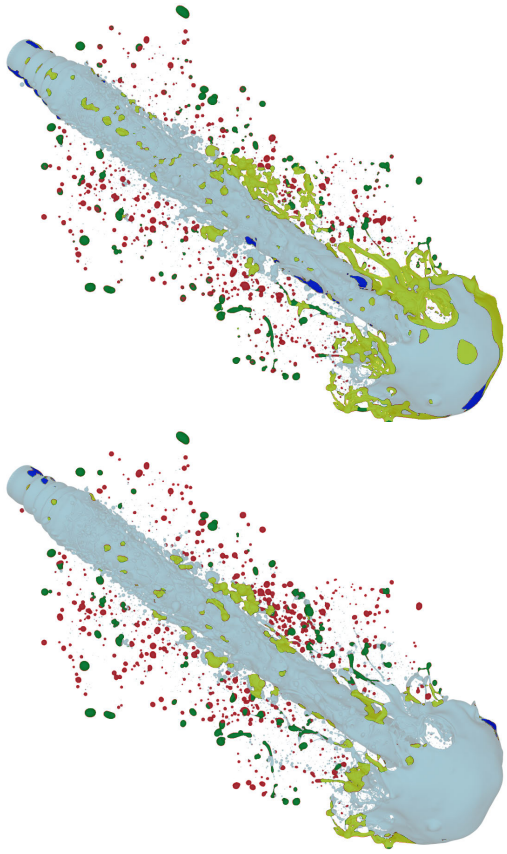


FIGURE 14. Prediction result of tilt image f_01522. The top image is the result of the improved U-Net that detected 701 droplets, while the bottom is the result of the basic U-Net that detected 443 droplets.

D. COUNTING OF OBJECT AS EVALUATION

The trained U-Net performed inference on the test images to get the predicted image. The predicted image is also called the mask. Since no ground truth is available for the test images, it is not easy to quantify the result. To evaluate the quality of the segmentation, we counted the object's occurrences, specifically the droplets. With a plain assumption that with more droplets detected, the model learned better.

How to identify an object from the mask? An assumption that one contour represents one object is made. The contour of the droplets class is detected from the mask using a border following algorithm proposed by Suzuki and Abe [22]. Note that the droplet is represented by value one in the pixel intensity. Inspired by inverted binary thresholding, pixels with an intensity equal to one in the mask will become black, and pixels with an intensity not equal to one in the mask will become white. After this process, a binary image is obtained and used in a border following algorithm. All of the contours in the image will be retrieved with the approximation that compresses horizontal, vertical, and diagonal segments and leaves only their endpoints. Lastly, the number of elements in

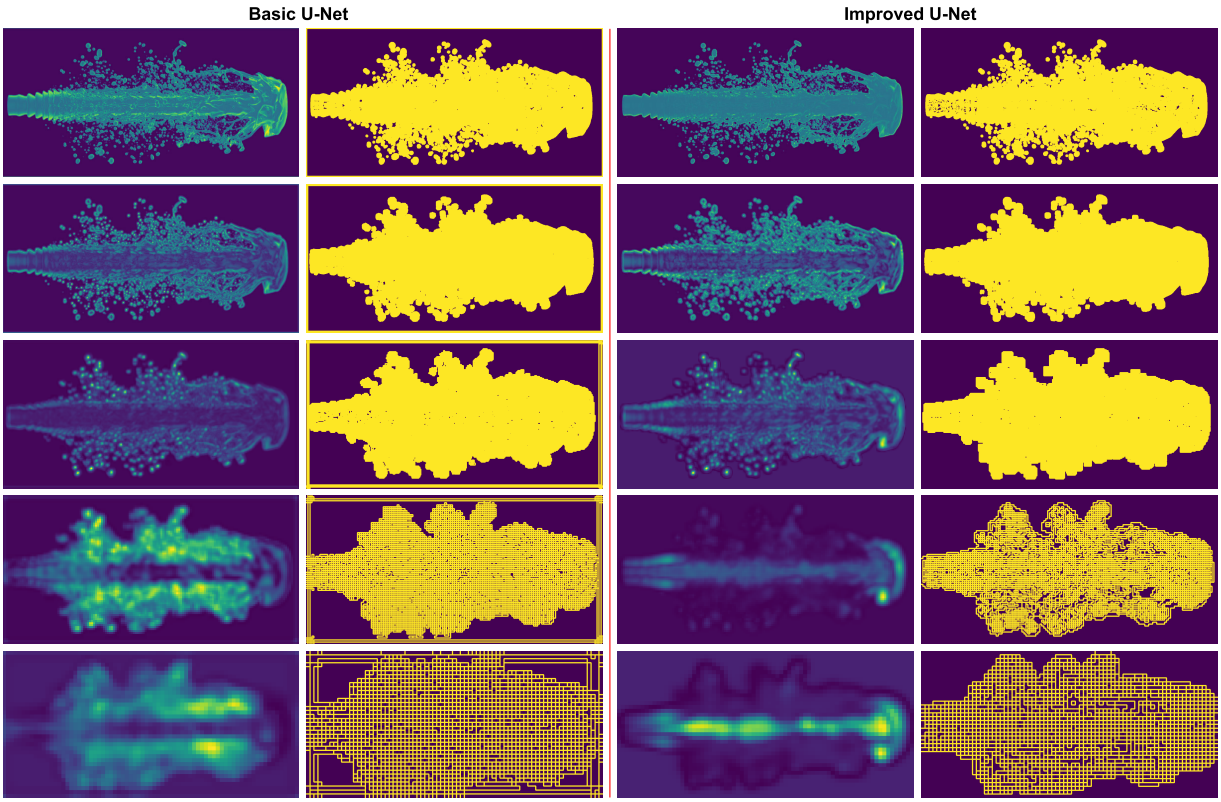


FIGURE 15. The figure displays the convolutional feature maps and their corresponding entropy images of a horizontal test image. The two left columns present the results from a basic U-Net, while the two right columns present the results from the improved U-Net.

the contour list is counted to get the final number of detected droplets.

IV. RESULTS AND DISCUSSION

The images in the testing set have horizontal and tilt orientations. The results and discussion will be based on the predicted outcome with the number of objects detected as the metric. Fig. 12 is the result of a horizontal test image where both the models perform well on the test image, with all four classes detected. The priority here is the number of droplet class (the objects in red) detected. The improved U-Net provides a more promising visual outcome with more droplets detected, which indicates it performs better than the basic U-Net in liquid spray segmentation.

Both models are inferred on the complete dataset, and the plot of droplet detection count is shown in Fig. 13. The result indicates that the improved U-Net detected many more droplets than the basic U-Net. The shortcoming of the current counting evaluation is the inaccurate counts affected by false contouring in the objects. Supposedly, there is only one class per object. In the predicted outcome, some objects have two classes predicted, mainly seen in droplet prediction. To further evaluate the generalization of the model, the model is used to predict the tilt orientation of liquid spray with a resolution of 1600×1200. Fig. 14 is the result of the tilt test image. The improved U-Net is better than the basic U-Net in this visual comparison.

The only difference between improved U-Net and basic U-Net is the pooling layer. The improved U-Net used the LPS layer, a non-linear trainable downsampling layer with anti-aliasing applied. As aforementioned, CNN is biased to the texture of the image [7]. We had an assumption that reducing the texture bias guides convolution to rely more on the edge/shape information. Since both models are trained with blurred images, the LPS layer with anti-aliasing applied plays an important role in proving better segmentation output. Fig. 15 provides a clear visual comparison of the convolutional feature map between the two models. The third layer of the basic U-Net starts to show a jagged effect. The corresponding image entropy also generates irrelevant texture around the image border. The irrelevant texture becomes occupied in the last layer of the basic U-Net. We suspect this is the reason the convolutional network makes decisions based on the wrong features. The improved U-Net performs better because the blurring layer between the convolutional feature map reduces texture information and increases contour information, which is the success factor of segmentation. The convolutional feature map of the improved U-Net contains texture information without noise, so the classifier is less prone to wrong predictions. The result also shows that texture is still the prominent feature that convolution learned, which also justifies the occurrence of false contouring among the objects. For example, in some droplets, the outer texture differs from the inner texture. Since

the convolutional network classifies the objects based on the texture, two classes are predicted in one droplet.

V. CONCLUSION

In general experimental practices, obtaining acceptable segmentation results for computer-generated liquid spray images using a semantic segmentation deep learning model such as U-Net requires intensive pixel-wise annotation and a large number of training samples. Through our research, we have demonstrated that with a methodical approach of pre-processing training images, it is possible to train a U-Net semantic segmentation model with only partial ground truth annotations. Our findings suggest that even with limited sample sizes, such as the 24 training images used in this research, it is possible to achieve satisfactory segmentation outcomes for computer-generated liquid spray images, as measured by visual evaluation. This workflow is helpful to be applied to a niche dataset, with some prospective applications such as detecting cultural patterns, differentiating fruit types, and segmenting medical images. According to our research findings, we collected some practical directions to train a U-Net semantic segmentation deep model. We suggest keeping the original image resolution during training to retain all spatial information unaffected by resizing and keeping the image in RGB mode to leverage color information for the convolutional feature extractor to learn a rich image representation. Another step to improve the model is to apply blurring after the convolution layer to highlight the contour information of the object to enhance the segmentation rate. Through the visualization of the convolutional feature map, we found that Gaussian smoothing reduces the jagged effect caused by downsampling in the convolutional network, thus improving the segmentation outcome. However, the price of small sample learning is the occurrence of false contours in some objects, which falsify the number of predicted objects.

ACKNOWLEDGMENT

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies or positions, either expressed or implied, of the DEVCOM Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

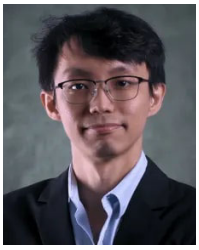
REFERENCES

[1] A. H. Lefebvre and V. G. McDonell, *Atomization and Sprays*. Boca Raton, FL, USA: CRC Press, 2017.
[2] R. K. Pathan, W. L. Lim, S. L. Lau, C. C. Ho, L. Bravo, R. B. Koneru, and P. Khare, "Application of computer vision techniques to identify spray primary breakup structures in high-speed flow," in *Proc. IEEE Int. Conf. Comput. (ICOCO)*, Oct. 2023, pp. 468–474.
[3] L. Bayvel and Z. Orzechowski, *Liquid Atomization*. Evanston, IL, USA: Routledge, 1993.
[4] L. Hoyer, D. Dai, and L. Van Gool, "DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9914–9925.
[5] M. Ivanovs, K. Ozols, A. Dobrajs, and R. Kadikis, "Improving semantic segmentation of urban scenes for self-driving cars with synthetic images," *Sensors*, vol. 22, no. 6, p. 2252, Mar. 2022.
[6] R. K. Pathan, W. L. Lim, S. L. Lau, C. C. Ho, P. Khare, and R. B. Koneru, "Experimental analysis of U-Net and mask R-CNN for segmentation of synthetic liquid spray," in *Proc. IEEE Int. Conf. Comput. (ICOCO)*, Nov. 2022, pp. 237–242.
[7] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 9234–9255.
[8] H. Chung and K. H. Park, "Shape prior is not all you need: Discovering balance between texture and shape bias in CNN," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2022, pp. 4160–4175.
[9] W. L. Lim, M. Y. Teow, R. T. Wong, R. K. Pathan, S. L. Lau, C. C. Ho, L. Bravo, R. B. Koneru, and P. Khare, "Performance assessment of U-Net for semantic segmentation of liquid spray images with Gaussian blurring," in *Proc. IEEE Int. Conf. Comput. (ICOCO)*, Oct. 2023, pp. 462–467.
[10] X. Zou, F. Xiao, Z. Yu, Y. Li, and Y. J. Lee, "Delving deeper into anti-aliasing in ConvNets," *Int. J. Comput. Vis.*, vol. 131, no. 1, pp. 67–81, Jan. 2023.
[11] R. Zhang, "Making convolutional networks shift-invariant again," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 97, 2019, pp. 7324–7334.
[12] G. Xu, W. Liao, X. Zhang, C. Li, X. He, and X. Wu, "Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109819.
[13] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?" *J. Mach. Learn. Res.*, vol. 20, no. 184, pp. 1–25, 2019.
[14] A. Chaman and I. Dokmanic, "Truly shift-invariant convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3772–3782.
[15] R. A. Rojas-Gomez, T.-Y. Lim, A. Schwing, M. Do, and R. A. Yeh, "Learnable polyphase sampling for shift invariant and equivariant convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022, p. 35.
[16] M. Y. W. Teow, "Convolutional autoencoder for image denoising: A compositional subspace representation perspective," in *Proc. IEEE Int. Conf. Artif. Intell. Eng. Technol. (IICAIET)*, Sep. 2021, pp. 1–6.
[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[18] S. Popinet. (2013). *Basilisk*. [Online]. Available: http://basilisk.fr
[19] K. Wada. (2016). *Labelme: Image Polygonal Annotation With Python*. [Online]. Available: https://github.com/wkentaro/labelme
[20] G. Chaussonnet, C. Lieber, Y. Yikang, W. Gu, A. Bartschat, M. Reischl, R. Koch, R. Mikut, and H.-J. Bauer, "Towards deepspray: Using convolutional neural network to post-process shadowgraphy images of liquid atomization," Karlsruhe Inst. Technol., Karlsruhe, Germany, Tech. Rep. 1000097897, 2019.
[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.
[22] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 3, p. 396, Mar. 1985.

WEI LUN LIM (Graduate Student Member, IEEE) received the B.I.T. degree (Hons.) in artificial intelligence from Multimedia University, Melaka, Malaysia, in 2018, and the M.Sc. degree in IT from Multimedia University, Cyberjaya, Malaysia, in 2021. He is currently pursuing the Ph.D. degree in computing with Sunway University, Selangor, Malaysia. His research interests include semantic segmentation and representation learning.

MATTHEW Y. W. TEOW (Senior Member, IEEE) received the B.Sc. degree in electronic and electrical engineering from Robert Gordon University, U.K., the M.Eng. degree in electrical engineering from Universiti Teknologi Malaysia, Malaysia, and the Ph.D. degree in engineering from Multimedia University, Malaysia.

He is currently a Lecturer with University Partnership (Coventry University), PSB Academy, Singapore. He has published more than 30 papers in conferences and journals. His current research contributes to the scientific foundations of artificial intelligence in the areas of representation theory, generative learning, and visual inference.



RICHARD T. K. WONG received the B.Eng. degree (Hons.) in mechatronic engineering from Universiti Sains Malaysia, Malaysia, in 2012, and the D.Phil. degree in engineering science from the University of Oxford, U.K., in 2017

He started his academic position with Sunway University, in 2018. He is currently a Senior Lecturer with the Department of Engineering, School of Engineering and Technology, Sunway University. His research interests include the application of artificial intelligence approaches, more specifically optimization algorithms, in the context of smart transportation systems and swarm robotic systems.

REFAT KHAN PATHAN (Student Member, IEEE) received the B.Sc. degree (Hons.) in computer science and engineering from BGC Trust University Bangladesh, in 2020. He is currently pursuing the M.Sc. degree in computer science with Sunway University, Malaysia.

His research interests include image processing, instance segmentation, and data mining.

CHIUNG CHING HO (Senior Member, IEEE) is currently a Senior Adjunct Research Fellow with the Department of Computing and Information Systems, Sunway University. His research interests include location intelligence, time series forecasting, natural language processing, and multimodal biometrics.

RAHUL BABU KONERU is currently a Computational Multiphase Modeling Scientist with the Leidos Research Support Team, working as an On-Site Contractor with the National Energy Technology Laboratory (NETL), Morgantown, WV, USA. His research interests include dispersed multiphase flows, mesoscopic methods, and high-performance computing.

PRASHANT KHARE is currently an Associate Professor and the Associate Department Head and the Program Chair of aerospace engineering with the University of Cincinnati. His research interests include multiphase fluid dynamics, hypersonic flows, combustion, high-performance computing, and machine learning.

LUIS BRAVO is currently a Senior Aerospace Engineer with the DEVCOM Army Research Laboratory. He leads several research efforts in partnership with academia, industry, and cross-service agencies that span basic and early applied research in propulsion sciences. He is also an Associate Fellow of the American Institute of Aeronautics and Astronautics (AIAA), holds academic appointments as an Adjunct Professor of aerospace engineering with the University of Cincinnati and the University of Maryland, and the Chair of modeling and simulation with the Propulsion and Power System Alliance. He was a recipient of multiple prestigious awards, including the DoD HPCMP Frontier, OSD SERDP, OUSD ARAP, and the Army Modeling and Simulation Award.

SIAN LUN LAU (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electronics and telecommunications from Universiti Malaysia Sarawak, Malaysia, in 2000, and the M.Sc. and Dr.Ing. degrees in electrical communication engineering from the University of Kassel, Germany, in 2004 and 2011, respectively.

Since 2004, he has been a full-time Researcher with the University of Kassel. He has contributed and managed various German national and EU-funded projects, such as the EU FP6 MobiLife, the ITEA S4ALL, the BMBF MATRIX, and the EU FP7 SEAM4US. Currently, he is a Professor with the Department of Engineering, School of Engineering and Technology, Sunway University, Malaysia. He continues to be involved in active research and has published over 90 publications in conferences, workshops, book chapters, and journals. His research interests include ubiquitous computing, sustainable smart cities, context-awareness, and applied machine learning.