IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 30, 2022, accepted February 10, 2022, date of publication February 14, 2022, date of current version February 22, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3151375

FHI-Unet: Faster Heterogeneous Images Semantic Segmentation Design and Edge AI Implementation for Visible and Thermal Images Processing

MING-HWA SHEU[1], (Member, IEEE), S. M. SALAHUDDIN MORSALIN[1], (Graduate Student Member, IEEE), SZU-HONG WANG[1], (Member, IEEE), LIN-KENG WEI[1], SHIH-CHANG HSIA[1], (Member, IEEE), AND CHUAN-YU CHANG[2], (Senior Member, IEEE)

[1]Department of Electronic Engineering, National Yunlin University of Science and Technology, Douliu, Yunlin 64002, Taiwan
[2]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu, Yunlin 64002, Taiwan

Corresponding author: S. M. Salahuddin Morsalin (s.morsalin10@gmail.com)

This work was supported in part by the Ministry of Science & Technology, Taiwan, under Project 110-2221-E224-052; and in part by the National Yunlin University of Science and Technology, Taiwan.

ABSTRACT The same class of objects clustering process in a frame is known as semantic segmentation. The deep convolutional neural network-based semantic segmentation needs large-scale computations and annotations for data training to reach real-time inference speeds. The heterogeneous image segmentation is a more challenging task to categorize each pixel of an image. However, the heterogeneous image semantic segmentation method extracts the features of visible and thermal images separately. We designed an efficient architecture with the multi-hybrid-autoencoder and decoder for Faster Heterogeneous Image (FHI) Semantic Segmentation. The proposed corresponding architecture has fewer layers resulting in lower parameters, higher inference speed, and Intersection over Union (IoU). The specialty of this architecture is the discrete autonomous feature extraction framework for RGB image and Thermal (T) image inputs with individual convolutional layers. Later, we combined the 4-channels (RGBT) convolution features to reduce computational complexity and robust the model performances. The proposed FHI-Unet semantic segmentation model experimented on NVIDIA Xavier NX edge AI platforms with standard accuracy under the real-time inference requirement. The proposed FHI-Unet model has the highest mIoU of 43.67 and the fastest real-time inference of 83.39 frames per second on edge AI implementation. The proposed approach improves 31.36% inference speed, 7.16% mAcc, and 5.1% mIoU on the Multi-spectral Semantic Segmentation Dataset compared with the existing works.

INDEX TERMS Heterogeneous image, semantic segmentation, edge AI platform, deep convolution, multi-hybrid-autoencoder, autonomous feature extraction, feature fusion.

I. INTRODUCTION

Semantic segmentation is the fundamental technique for autonomous application. The human-computer interaction, virtual reality, and medical representation analysis rely on semantic segmentation. The semantic segmentation does fine-grained reason by employing compact categorization and labeling for each pixel. The use of the convolutional neural network has been increased considerably in recent years with the rapid development of deep learning applications. The computer vision-based object detection model identifies the location and detects object, classifies each object, determines their class number, predicts the direction, and many other things [1]. However, the object detection model marks a bounding box corresponding to each class in the frame, but nothing explains the object shape. The image segmentation technique creates the pixel-wise mask for each object in an image, which is a grainier understanding of

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

the picture. Image segmentation techniques make a massive impact on autonomous security systems [2], military surveillance [3], self-driving cars [4], traffic congestion systems [5], the granular manner in the medical image [6], and so on. The research community has made encouraging progress on Convolutional Neural Network (CNN) model architectures for semantic segmentation process, for example, coarse-to-fine semantic segmentation [7], DeepLab [8], PASCAL VOC [9] to learn image illustration methods. The Fully Convolutional Network (FCN) [10] is a foundational work that accepts input images for semantic segmentation and modifies the fully connected layers into convolution layers for classification networks such as the AlexNet [11], the VGG16 [12], and the GoogLeNet [13]. Whereas FCN network reduces the sizes of the final predictions because of several convolutional strides and spatial pooling functions resulting in the loss of granular picture information and erroneous predictions.

The ability of deep neural networks, amount of training data, quality of input images, and the lighting source of image and video inputs all aspect have a significant role in robust performance. Some neural networks [14] and [15] have been built on 3-channel RGB input images from near-infrared visible cameras. Unfavorable lighting conditions, such as darkness, cloudy or foggy weather, and glare from automobile headlights, make significant obstacles for RGB picture deterioration [16]. The thermal imaging camera creates heat radiation pictures, which can see in various light conditions [17]. In this work, we acquire images fusion of thermal and visible images to obtain more accurate semantic segmentation. Commercial appliances such as remote sensing, autonomous surveillance, automotive driving assistance systems, military surveillance, embedded module systems require faster inference speed and appropriate object segmentation.

In this study, the various image data are divided into two categories based on camera functions: thermal images and visible-light images. A thermal image employs an object with fluctuating degrees of thermal radiation energy to create the temperature distribution map. Furthermore, its perception range makes it suitable for usage at night view, cloudy and foggy weather, and in the presence of glare from opposing headlights at absolute and abnormal temperatures [18]. The RGB visible image contains rich information such as object color, texture, and clear boundary, which is comparatively easy to extract features in a lighter environment. Although, the image discrimination ability decreases in a dim environment. Therefore, the combination of both image characteristics of these two inputs complements each other to alleviate the environmental interference and obtain better semantic segmentation results, which is called heterogeneous image semantic segmentation. If the RGB and thermal images convolute directly without processing, it's hard to improve the precision. Therefore, we adopted the concept of Unet architecture and extended it through the proposed FHI-Unet model. We have developed the independent convolutional network with the multi-hybrid-autoencoder for RGB and Thermal (T) image inputs feature extraction separately. The visual and thermal pictures are semantically segmented using the multi-hybrid-autoencoder and decoder through the proposed heterogeneous image segmentation architecture. We utilized 4-channels (RGBT) inputs autonomous encoder and feature fusion encoder to match the heterogeneous dual image features and extracts the thermal and visible image features particularly. The following is a list of significant contributions of this work.

1. We designed the independent convolutional network with the multi-hybrid-autoencoder for RGB and Thermal (T) image inputs feature extraction separately, which reduces the computational complexity of the proposed architecture.
2. A feature fusion encoder combines and fuses the 4-channel (RGBT) convolution features that enhance inference speeds.
3. The proposed FHI-Unet model has fewer layers, lower parameters, lower-rung read and write memory which increases the FPS and accuracy.
4. The proposed design has been implemented on NVIDIA Xavier NX edge AI platforms for investigating the faster heterogeneous image semantic segmentation.

II. RELATAD WORK

The semantic segmentation algorithms require large-scale and high-quality data to robots the performance while dealing with numerous instabilities. The semantic segmentation methods are categorized into traditional approaches and deep learning algorithms. The sparse representation approaches [19], k-means clustering [20], Markov random fields [21], and the random forest [22], clustering [23] are counted as traditional approaches. The traditional techniques are replaced by convolutional neural networks (CNNs) progressively. In recent years, researchers are investigating the CNNs based algorithms for semantic segmentation with the rapid growth of deep learning algorithms. The PSPNet [24] techniques represented the dilation convolution method. The state-of-the-art network DeepLabv3+ [25], two-part of neural networks aggregated multi-scale contexts to enlarge the receptive field and lead to a higher-resolution and compact FCN based pixel prediction. The Visible and Thermal image fusion [26] for few-short semantic segmentation based on bimodal images. The Edge conditioned convolutional neural network [27] for thermal image semantic segmentation built on a feature-wise transform layer. The GMNet [28] categorized feature extraction to the multilevel for feature fusion.

In recent years, the encoder-decoder base models have been investigated actively in semantic segmentation with the popularity of deep learning algorithms. The dependability and flexibility of encoder-decoder-based models are more suitable in real-world applications such as robots and autonomous applications. ABMDRNet [29] multimodality feature fusion network employs a bi-directional image-to-image translation through two-stage networks. The SegNet [30] is a deep convolutional neural network that

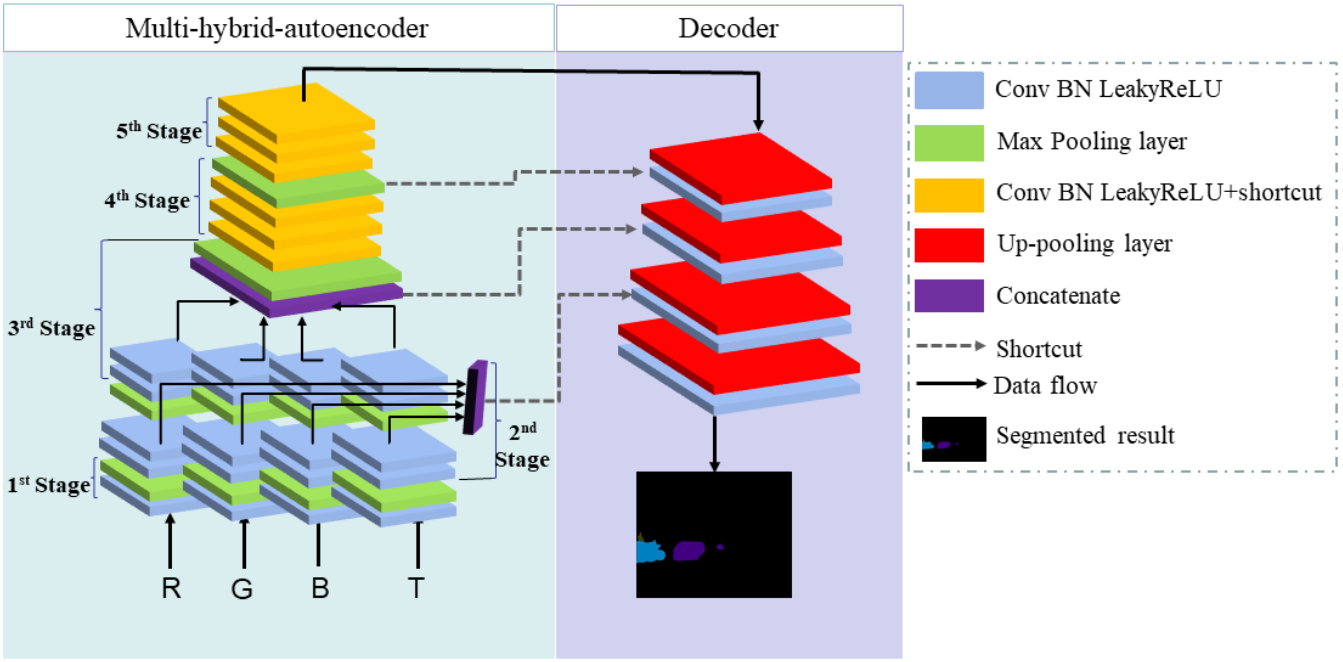

FIGURE 1. The proposed FHI-Unet architecture is presented with the multi-hybrid-autoencoder on left side and a decoder respectively.

employs the encoder and decoder to conduct semantic pixel-wise segmentation. The model encoder consists of 13 convolutional layers of VGG16, which serves for down-sampling and max-pooling. In addition, the pooling coordinates solve the pixel location information loss causes by multiple pooling layers. Besides, the decoder employs the associated max-pooling index value for up-sampling. Finally, the Softmax classifier predicts each pixel's class output feature map. The Unet [31] model predicts tiny medical picture segmentation by linking the symmetric relationship between encoder-decoder. Two convolutions use at the encoder to perform the four down-samplings by max pooling. The decoder uses up-convolution to perform up-sampling and concatenation of the corresponding size of the encoder feature map. The MFNet [32] was designed based on the dual-encoder architecture, with RGB and Thermal images two parallel branches being simulations by the encoders. The encoder fuses the RGB feature, and Thermal feature maps using element summations and is sent to the decoder for convolutional operation through the nearest-neighbor interpolation for up-sampling. The RTFNet [16] consists of an RGB encoder, a thermal encoder, and a decoder to extract features using RGB and thermal data fusion whereas, the ResNet [33] is the backbone network. The thermal feature maps fuse the RGB encoder through the element-wise summation. The MMNet [34] consists of two-stage networks to feature extraction and refine details. The GMFNet [35] is composed of three parallel Unet for modality and multimodal fusion. The FuseSeg [36] architecture used a dense native representation for laser range scanner data introduction. The effectiveness of the method is LiDAR and RGB data fusion for segmenting the LiDAR point clouds. The dual attention network for image segmentation [37] method extracts the feature map spatial dependency through the location channel attention mechanism.

Some of the semantic segmentation models mentioned above have good performance. However, the complexity of the architecture and frame structure leads to computational problems which require costly graphics cards for the implementation and the inability to use embedded systems for real-time preceding to trade-off the accuracy and higher speed.

III. THE PROPOSED NETWORK

Figure 1 displays the proposed Faster Heterogeneous Image (FHI) Semantic Segmentation architecture. The proposed FHI-Unet architecture consists of 2 modules: the multi-hybrid-autoencoder for feature extraction and decoder for feature map sampling. The first autoencoder extracts independent 4-channels RGB and Thermal (T) input image features separately at the initial stage, whereas needs to do several convolutional operations, batch normalization with Leaky Relu activation function, and max-pooling at the next steps. Later, another feature fusion autoencoder combines the 4-channel (RGBT) convolution features for further process. The convolutional feature fusion speeds up the model operation and computation. The individual input feature extraction autoencoder saves operation time and speeds up the performance. We employed the customize convolution to implement this experiment. The customize convolutional computation speed is faster than the typical deep convolutional architecture [38]–[41].

The typical convolution complexity is measured by adding all input and output channels together. If the input image

TABLE 1. Multi-Hybrid-Autoencoder operation analysis.

Multi-hybrid-autoencoder operation					
Stage	R	G	B	T	Feature size ($H\times W\times C$)
1st Stage	CBL	CBL	CBL	CBL	480×640×8 (Individual)
	Max pooling	Max pooling	Max pooling	Max pooling	480×640×8 (Individual)
2nd Stage	2× CBL	2× CBL	2× CBL	2× CBL	240×320×16 (Individual)
	Hybrid Concatenate output				240×320×64
	Max pooling	Max pooling	Max pooling	Max pooling	120×160×16 (Individual)
3rd Stage	2× CBL	2× CBL	2× CBL	2× CBL	120×160×32 (Individual)
	Hybrid Concatenate output				120x160x128
	Max pooling				60×80×128
4th Stage	3×Conv + Shortcut				60×80×128
	Max pooling				30×40×128
5th Stage	3×Conv + Shortcut				30×40×128

TABLE 2. Decoder operation analysis.

Decoder operation		
Stage	R G B T	Feature size (H×W×C)
4th Stage	Up-pooling	60×80×128
	Shortcut block	60×80×128
	CBL	60×80×128
3rd Stage	Up-pooling	120×160×128
	Shortcut block	120×160×64
	CBL	120×160×64
2nd Stage	Up-pooling	240×320×64
	Shortcut block	240×320×32
	CBL	240×320×32
1st Stage	Up-pooling	480×640×32
	CBL	480×640×9

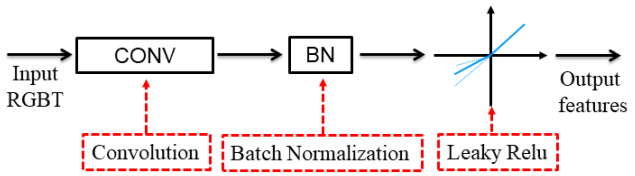

FIGURE 2. Convolution, BN and Leaky Relu operation.

size is 240 × 320 with 4-channel RGBT data, and the output channel is 32 with the kernel size (K) is 3 × 3, the total flops are 88.473 for the initial layer. The proposed FHI-Unet uses a multi-hybrid-autoencoder to extract RGB and Thermal (T) picture features individually. The proposed architecture generates 8 output feature channels for each input channel based on the same kernel size. The proposed FHI-Unet has 22.118M flops for the first stage. Following the process, the second and third stages also reduce the overall flops.

Table 1 illustrates the multi-hybrid-autoencoder operation details for the proposed FHI-Unet semantic segmentation architecture. Each stage has different convolutional layers, different input & output channels, and different feature sizes. Whereas, H, W & C stands for the height, weight, and channel number of each image separately. The 1st stage, 2nd stage, and 3rd Stage has individual convolutional operation. The 1st stage extracts RGB and Thermal image features separately. The 2nd stage and 3rd stage create hybrid concatenate layer for output features. The 4th stages and 5th stage do the combine convolution and pass information to the decoder for further process.

The decoder performs for feature map up-sampling and restores the target to 480 × 640 resolution. The design of the decoder is mainly for recovering the input image feature up-sampling size. Since the encoder used four down-sampling operations, the decoder also performed four up-sampling operations to recovery the same size feature map. There is different way to calculate the recovery up-samples and reduce the computational effort. We used the nearest-neighbor interpolation method to reduce the computational complexity. In addition, the performance of up-sampling used same size feature map for the decoder and encoder to reduce the loss by adding them together. Table 2 demonstrates the detailed operational computational function of the decoder. Whereas up-pooling and Conv BN Leaky Relu blocks utilized for the operation.

IV. MULTI-HYBRID-AUTOENCODER AND DECODER COMPUTING INSTRUCTIONS

The encoder and decoder computing instructions belong to down-sampling and up-sampling. The convolution kernel settings, layers computation, shortcut blocks, and all other aspects of the multi-hybrid-autoencoder, decoder operation are described in depth at these sections.

A. CBL (CONV BN LEAKY RELU) LAYER

An autoencoder extracts the R, G, B, and T 4-channels input features independently. The autoencoder operation for stage 1 and stage 2, extracts features using individual convolution, batch normalized, and the Leaky Relu activation function for the output feature. The Batch Normalization (BN) calculates the mean and variance values. The Leaky Relu properties backpropagation [42], one of the possible newer activation functions. The Leaky Relu mitigates the dying

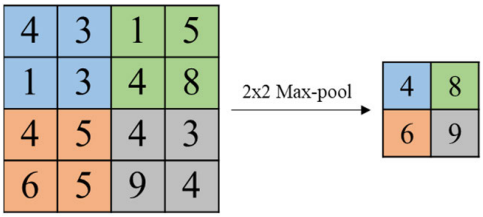

FIGURE 3. The max-Pooling operation using 2 × 2 kernel size.

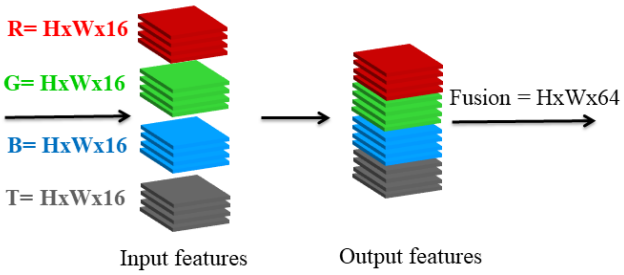

FIGURE 4. The concatenate operation and features fusions.

Relu problem that prevents backpropagation from being terminated if the value is less than 0, which calculating by the equation (1). Figure 2 shows the convolution, batch normalized, and activation function details. Wearers, the Leaky Relu negative slope = 0.2.

$$LeakyRELU\ (x) = \begin{cases} x, & x \geq 0 \\ x \times negative\ slope, & x < 0 \end{cases} \quad (1)$$

B. MAX LAYER (MAX POOLING 2 × 2)

The max-Pooling function picks the maximum value from each kernel, the highest value creates a significant impact in the image [43]. When the kernel size is 2 × 2, half of the values denote the actual value, which increases the receptive field. In this study, the max-pooling operation reduced the feature maps by the convolution as down sampling. Besides, the input feature map H x W is scaled down by a 2 × 2 pooling layer in the encoder. The output feature map becomes half of H and half of W with the maximum of 2 × 2 kernel filters. Figure 3 shows how to shrink the feature map from stage 1 to stage 5 in the autoencoder operation while saving computing time.

C. CONCATENATE LAYER

A concatenation layer accepts many inputs and concatenates them along with specific dimension. However, the entries must have the same size in all aspects, which increases the precision of learning [44]. The autoencoder concatenates thermal (T) image characteristics with the visible RGB image features as the same size, and the channel features are integrated and simplified, as shown in Figure 4. In stage 3, the autoencoder combined RGBT features at same size and joined along the axis and dimension make it easy to perform decoder convolution operations.

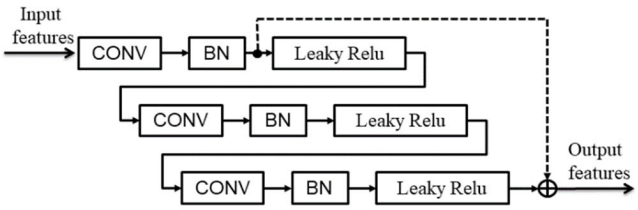

FIGURE 5. Conv + short segmentation layer operation.

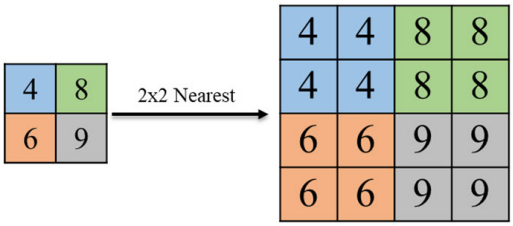

FIGURE 6. The nearest-neighbor interpolation for Up-sampling.

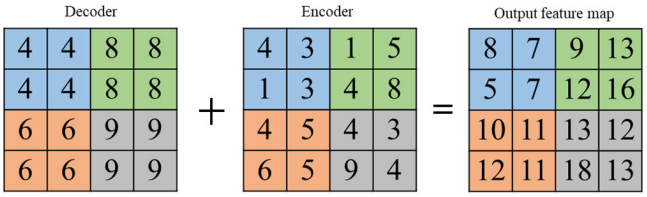

FIGURE 7. The encoder and decoder addition for Shortcut block operation.

D. CONV + SHORTCUT LAYER

The shortcut connections skip the imperfect low-level features training layers by transferring immediately to high-level features [33] and solves the gradient drifting problems. We incorporated the shortcut connections in the designed architecture. The purpose of the Conv and Shortcut layer is to perform one convolution of the input feature map followed by residual joining. Besides, avoids the disappearance of the backpropagation gradient during training. The Conv and Shortcut layer speeds up the convergence of the architecture shown in Figure 5. The residual function overcome the vanishing gradients problem and mitigates the deterioration problem during stages four and five of multi-hybrid-autoencoder operations.

E. UP-SAMPLING

The up-sampling operation transforms a small image input into a large image output. The feature maps up sampling works by repeating the rows and columns features at the decoder for restoring the target resolution. The up-sampling rate can be considerably high while guaranteeing the higher quality of the up-sampled results [45]. In this work, the nearest interpolation method employed for features up-sampling, resulting in a doubling of each row and column for input data (see Figure 6). From stage 5 to stage 1, the decoder used 2 × 2 nearest-neighbor interpolation to reduce the complexity and speed up the computational calculation.

F. SHORTCUT BLOCK

The shortcut block in the decoder acquires context information, creates semantic characteristics, and enables features fusion between multiple output resolutions [46]. The shortcut block maintained the detail features of the encoders and added with the decoder feature map shown in Figure 7. From stage 5 to stage 1, added all feature resolution instead of concatenating. This shortcut block technique took less time to compute the output feature maps and significantly reduced the memory requirements in the system.

V. DATA ANALISYS AND EXPERIMENT

A. THE DATASET

The dataset is one of the most important parts of machine learning performance while dealing with deep neural networks. It's critical to collect and construct a comprehensive turbulence-degraded image dataset before designing a semantic segmentation model in degraded conditions. The RGB-Thermal image dataset with pixel-level annotation and multi-spectral semantic segmentation dataset [47] was used for this experiment, which execute pixel-by-pixel labeling for visible and thermal images. The image dataset consists of three channels of viewable image with a horizontal field viewing angle of 100 degrees and a one-channel thermal image with a viewing angle of 32 degrees. The dataset is stored in 4-channel PNG format whereas, 1568 training data (820 for daytime and 749 at nighttime), 392 validation data, and 393 test data. In general, most of the road picture segmentation is available in the dataset. The dataset has nine category objects: background, car, person, bike, curve, car stop, guardrail, color cone, and bump, with each component having its different color.

B. TRAINING DETAILS

We used PyTorch frameworks for the proposed faster heterogeneous image semantic segmentation architecture to conduct the experiments. The AMD 5600, Intel Core i7 CPU, NVIDIA 3090 with 24GB graphics card, CUDA 11.1, and cuDNN v8.0.4 are all employed in the training procedure. For the FHI-Unet experimental evaluation, we used the Frames Per Second (FPS), Mean Accuracy (mAcc), and Mean Intersection over Union (mIoU) as evaluation metrics, as well as an Adam optimizer (Adaptive moment estimation) for weight update, Cross-Entropy Loss (CEL) function for training loss calculation, and Batch size parameter is 4.

C. EVALUATION METRICS

Many real-world applications demand for firster inference speed into the production environment; hence network latency time correctly is one of the most significant aspects of installing a deep network. To calculate the multispectral image semantic segmentation inference speed for the proposed FHI-Unet model the equation 2 is the following.

$$Inference\ speed = \frac{\text{Running test time}}{\text{Test numbers}} \tag{2}$$

Two validation measurement models are used to assess the heterogenous image semantic segmentation performance. The first one is Accuracy (Acc) per class of pixels (equation 3), and the second one is Intersection over Union (IoU) per class of pixels, as calculated in equation 4. The TN (True Negative) refers to negative samples that are wrongly stated to as positive samples by the FP (False Positive). Likewise, the TP (True Positive) is the positive samples and erroneously sorted into FN (False Negative) to calculate the Accuracy and Intersection over Union.

TP = True Positive, TN = True Negative,
FP = False Positive, FN = False Negative,

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

The "mAcc" indicates the mean value of the accuracy function, and "mIoU" represents the mean value of Intersection over Union. The values of mAcc and mIoU can be calculated by the following equations 5 and 6, where the total number of object categories denoted by $\mu = 9$.

$$mAcc = \frac{1}{\mu}\sum_{0}^{\mu}\frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$mIoU = \frac{1}{\mu}\sum_{0}^{\mu}\frac{TP}{TP + FP + FN} \tag{6}$$

The fewer parameter refers less computational complexity to the convolutional neural network's whereas, the number of parameters influences the memory size. Besides, more computations make the network's more complex and corresponded to the model execution time. Assuming that the input channel is C_{in}, the convolution kernel size is $K \times K$, and the output channel is C_{out}, the input feature map size is $H_{in} \times W_{in}$, the output feature map is $H_{out} \times W_{out}$, then the number of parameters and the computational quantity as in Equations 7 and 8. However, the value of G refers to 1 without making any groups.

$$\text{Parameters: } C_{in} \times K \times K \times C_{out} \times \frac{1}{G} \tag{7}$$

Computational quantity:

$$H_{out} \times W_{out} \times C_{in} \times K \times K \times C_{out} \times \frac{1}{G} \tag{8}$$

VI. EXPERIMENTAL RESULTS

In this experiment, we have considered the real time inference speed, Intersection over Union, and accuracy on edge AI platform. The six models, namely the RTFNet, the FuseSeg, the FuseNet, the MFNet, the SegNet, the U-net, and proposed FHI-Unet performance compared on the GPU as well as Nvidia Xavier NX Edge AI platform.

A. PERFORMANCE COMPARISON

Table 3 displays the performance of RTFNet, FuseSeg, FuseNet, MFNet, SegNet, Unet, and proposed FHI-Unet

TABLE 3. Performance compares on GPU and edge AI platform.

Architecture	Parameter (M)	Memory (MB)	Madd (GMAdd)	Complexity (GFlops)	Memory read write (GB)	FPS (GPU)	FPS (Edge)
RTFNet	185.23	2336.72	398.04	183.71	4.93	44.63	13.16
*FuseSeg	100	N/A	N/A	N/A	N/A	N/A	3.13
FuseNet	44.17	2565.82	566.27	283.47	5.0	136.42	34.85
MFNet	0.73	238.44	16.72	8.39	0.48	121.6	34.96
SegNet (4C)	1.98	466.41	48.55	24.34	0.93	185.58	51.59
Unet (4C)	17.33	2088.28	377.67	189.11	3.96	258.13	63.48
Proposed FHIU-Net	1.22	286.52	22.97	11.52	0.59	**293.85**	**83.39**

* The FuseSeg doesn't have open access code and above data in their paper.

TABLE 4. Comparison of the accuracy rate of semantic segmentation architecture.

Models	Background		Car		Person		Bike		Curve		Car Stop		Guard rail		Color Cone		Bump		Average	
	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	m-Acc	m-IoU
FuseNet	98.82	98.82	80.96	68.09	75.21	59.03	64.51	37.47	51.25	31.75	17.4	6.31	0.0	0.0	31.1	11.38	51.9	45.48	**52.35**	**39.81**
MFNet	99.31	99.30	77.02	67.53	67.12	57.62	53.91	47.7	36.25	32.17	12.5	13.77	0.0	0.0	30.3	28.88	30.11	32.25	**45.16**	**42.13**
SegNet (4C)	99.33	99.33	64.96	58.19	45.94	42.26	27.46	24.32	33.71	27.85	14.86	10.85	0.0	0.0	3.97	2.09	2.00	2.00	**32.47**	**29.65**
Unet (4C)	99.36	99.36	71.11	62.89	67.21	56.24	51.43	38.33	47.85	39.85	13.87	11	0.0	0.0	29.67	25.45	46.48	40.14	**47.44**	**41.47**
Proposed FHI-Unet	99.38	99.47	78.6	67.4	72.36	61.36	61.99	48.99	48.31	33.31	17.5	16.11	0.0	0.0	30.58	24.48	48.88	41.35	**50.84**	**43.60**

models. Among these models the FuseSeg and RTFNet had slowest inference speed on GPU and Nvidia Xavier NX Edge AI platforms, which is marked as red color, both models aren't suitable for real-time proceeding applications. The MFNet architecture has good performance on Parameter, Memory, Madd, Flops, and read-write memory but the FPS performance is still similar as the FuseNet and lower than SegNet, as well as Unet. The proposed FHI-Unet semantic segmentation model achieved height and fastest inference speed on GPU and Edge AI platforms. Besides, the Parameter, Memory, Madd, Flops, and read-write memory performances are better than the RTFNet, FuseSeg, FuseNet, SegNet, and Unet. Considering the real-time applications, the inference speed accelerates the performance of devises. The proposed FHI-Unet semantic segmentation model has achieved maximum FPS on both platform (marked as green color) among those models. Therefore, the proposed FHI-Unet model could be a good solution for real-time applications for Edge AI devises.

B. EDGE AI IMPLEMENTATION

For the Nvidia Xavier NX Edge AI Implementation, Table 4 demonstrations the performance of FuseNet, MFNet, SegNet (4C), Unet (4C), and proposed FHI-Unet segmentation model. The FuseNet achieved best result for object detection accuracy (mAcc) of 52.35 which is noticeable in green color. Besides, the proposed FHI-Unet model achieved second-highest accuracy on 50.84 that marked as purple color. The MFNet and The Unet (4C) has decent presentation for object detection accuracy. However, the SegNet (4C) performed lowest result (red color) for object detection in terms of accuracy. On the other hand, the Intersection over Union (IoU) for object detection performance, the proposed FHI-Unet model has achieved the best mIoU of 43.60 (marked as green color) and MFNet has the second-highest accuracy of mIoU value whish shown in purple color. The FuesNet and Unet (4C) shows an average performance for intersection over union of object detection. Whereas, the SegNet (4C) has lowest mIoU only 29 that marked as red color. All models show different values for each class of object segmentation, but the 'Guardrail' color pixels are only 0.0095 percentage among other classes, which is difficult to classify for all models.

Figure 8 shows the FHI-Unet model implementation system on Nvidia Xavier NX edge AI platform for heterogenous image semantic segmentation. We have connected the Nvidia Xavier NX devise with our desktop computer for system implementation. The RGB-Thermal image dataset

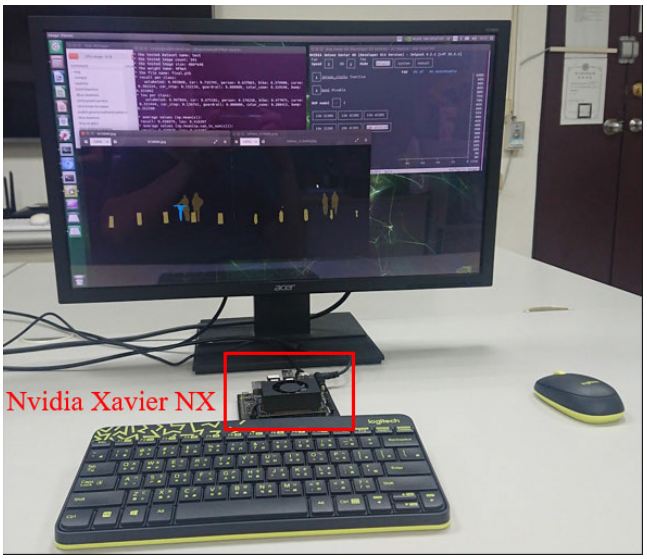

FIGURE 8. The Edge AI system implementation on Nvidia Xavier NX devise.

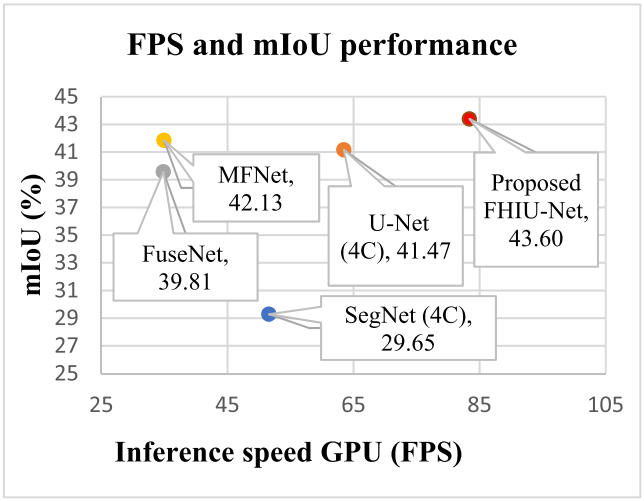

FIGURE 9. The FPS and mIoU performance for semantic segmentation on Nvidia Xavier NX Edge AI platform.

"multi-spectral semantic segmentation dataset" introduced by MFNet model was used for this experiment. Table 4 illustrated the details results of proposed FHI-Unet performance, which achieved best FPS of 83.39 and mIoU of 43.59.

Figure 9 illustrates the performance of the FuseNet, MFNet, the SegNet (4C), the Unet (4C), and the proposed FHI-Unet model on the Nvidia Xavier NX Edge AI platform in terms of mIoU and inference speed comparison. The FuseNet has the lowest mIoU and inference speeds. Alternatively, the proposed FHI-Unet image semantic segmentation model has state-of-the-art performance than the rest of models.

Table 5 illustrates the performance (flops, mAcc, mIoU, and FPS) comparison of Unet (4C) and proposed FHI-Unet on Nvidia Xaviar XN edge AI platform. The proposed FHI-Unet model has less computation and higher mAcc and mIoU value than Unet (4C) model. Furthermore, the proposed

TABLE 5. Model size and inference speed performance compares on GPU and edge AI platform.

Model	Flops (G)	Average		FPS
		Acc	IoU	
Unet (4C)	189.11	47.44	41.47	63.48
Proposed FHI-Unet	11.52	50.84	43.60	83.39

FHI-Unet achieved better inference speed on edge AI platform, which is state-of -the art performance.

To evaluate the segmentation results of different models, we considered four RGB images and four Thermal images as inputs with night views and daytimes perspectives which shown in first and second rows at Figure 10. In addition, the third row demonstrates the ground truth of RGBT images fusion results. The FuseNet, the MFNet, the SegNet (4C), the Unet, and the proposed FHI-Unet models' performance are evaluated based on segmentation results in the columns (a), (b), (c), and (d). The wrong prediction and failure segmentation are marked by red circle of those columns.

The columns (a) and (b) represent the segmentation results of night view images, besides columns (c) and (d) signify the segmentation results of daytime images. The FuseNet did some wrong prediction and segmentation in the column (a) and (d). Besides that, the model is unable to predict and segmentation of the bicycles at column (c). Similarly, the MFNet also did wrong prediction and segmentation for example, the model is unable to predict person and full car segmentation at the column (b). In addition, the model did some wrong prediction and segmentation in the column (a) and (d) for car and person segmentation. The SegNet (4C) model has missed some objects in the columns (a), (b), and (d) while doing the segmentation. The color temperature of the bicycle and the background appear to be similar in the RGB and Thermal image at the column (c); as a result, the Unet (4C) model is incapable for bicycle segmentation that shown in the column (c). In addition, the U-net (4C) model did some wrong segmentation in the column (a) and (d). However, the proposed FHI-Unet model has excellent performance for the heterogenous image semantic segmentation similar as ground truth objects, whereas the other models couldn't dose the proper way.

The FuseNet has highest accuracy result, but still did some wrong prediction and segmentation. Whereas the U-Net (4C) model has better segmentation performance than the FuseNet, MFNet and SegNet (4C). Considering to the object prediction and segmentation, the proposed FHI-Unet model achieved better performance than the Unet (4C). Whereases, the Unet (4C) model has adopted to design the proposed FHI-Unet model and an extended form of that model. However, the segmentation result shows that the proposed FHI-Unet model has better performance than other models. Furthermore, the proposed FHI-Unet model achieved second-higher accuracy, best inference speed, intersection over union as well as object segmentation on Nvidia Xavier NX platform compares to other models.

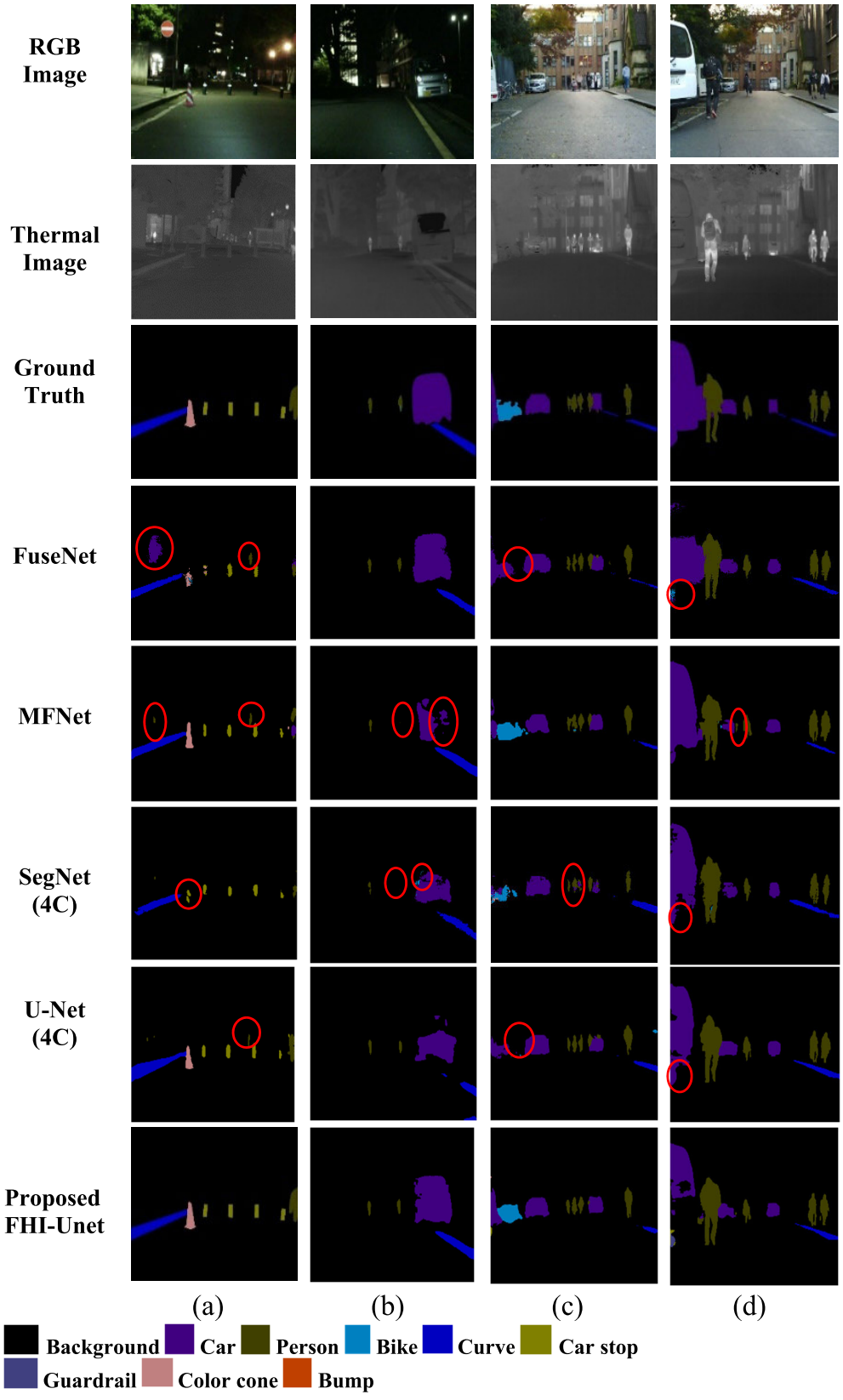

FIGURE 10. The Visual segmentation results comparisons of the proposed FHI-Unet with different models. The first and second rows represent the RGB and Thermal images. Third row denotes ground truth of the RGBT images. The column (a) and (b) shows the performance of night views. The column (c) and (d) demonstrations the performance of daytime images respectively. The wrong detection and missing information are marked by red circle of each model.

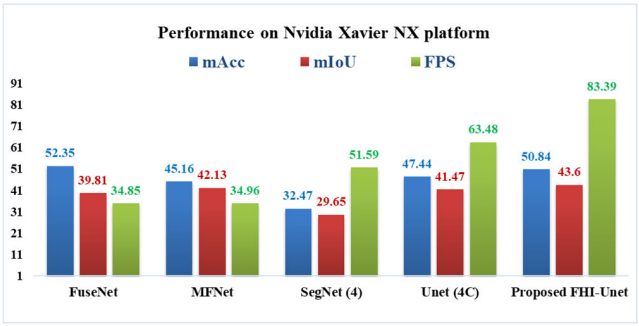

FIGURE 11. Proposed FHI-Unet model feasibility comparison.

Figure 11 illustrates the performance of proposed FHI-Unet and other models on the Nvidia Xavier NX platform. The FuseNet has better accuracy performance than the MFNet, and Unet while the SegNet has the lowest accuracy, whereas the proposed FHI-Unet achieved the second-highest accuracy. Furthermore, the proposed FHI-Unet displays best inference speed and highest mIoU among those approaches. On the other hand, the FuseNet and MFNet have the slowest inference speeds. The SegNet has the smallest mIoU compared to other models. Finally, in terms of FPS, mIoU, and mAcc, the proposed FHI-Unet model beats the state-of-the-art performance on the Nvidia Xavier NX platform.

VII. DISCUSSION

The Pytorch 1.6 framework has been employed for proposed FHI-Unet which is implemented on Nvidia Xavier NX edge AI platform. For considering the higher speed on real-time applications, the proposed FHI-Unet model has less computation and higher inference speed. The proposed model accomplishes edge computing for heterogeneous image segmentation and reduced computational complexity. We intend RGB and Thermal images for daytime and nighttime as training datasets which improve the FPS performance. However, the background data makes up most of the total pixels in the dataset, and the number of object category was imbalanced. The frequency of each item category was not modified individually for each image as a result the accuracy was bit low. For improving the accuracy rate, need to increase the amount of training data and balances the number of training categories in practice. In future, we may increase the convolutional layer to higher accuracy performance of the proposed FHI-Unet semantic segmentation model.

VIII. CONCLUSION

The proposed FHI-Unet semantic segmentation model for visual and thermal image feature fusion minimizes the computational complexity and speeds up the real-time operation. A multi-hybrid-autoencoder is included with architecture for individual RGB and Thermal image input feature extraction and down sampling operations. Later, another feature fusion encoder combines the 4-channel feature maps for further process. An efficient decoder is utilized to recover the feature map to compensate of feature loss during up sampling, which reduces the number of parameters and computational complexity. The convolutional layers were generated using the Leaky Relu activation function to avoid back-propagation errors. The experimental result shows the proposed FHI-Unet model has the highest mean Intersection over Union value (43.39) and inference speed of 83.39 FPS for the multi-spectral semantic segmentation dataset. The proposed FHI-Unet model could be a suitable approach for real-time application on edge AI platforms.

REFERENCES

[1] M.-H. Sheu, S. M. S. Morsalin, J.-X. Zheng, S.-C. Hsia, C.-J. Lin, and C.-Y. Chang, "FGSC: Fuzzy guided scale choice SSD model for edge AI design on real-time vehicle detection and class counting," *Sensors*, vol. 21, no. 21, p. 7399, Nov. 2021, doi: 10.3390/s21217399.

[2] T. Omar and M. L. Nehdi, "Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography," *Autom. Construct.*, vol. 83, pp. 360–371, Nov. 2017.

[3] A. Carrio, Y. Lin, S. Saripalli, and P. Campoy, "Obstacle detection system for small UAVs using ADS-B and thermal imaging," *J. Intell. Robotic Syst.*, vol. 88, nos. 2–4, pp. 583–595, Dec. 2017.

[4] J. S. Yoon, K. Park, S. Hwang, N. Kim, Y. Choi, F. Rameau, and I. S. Kweon, "Thermal-infrared based drivable region detection," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 978–985.

[5] Y. Jin, W. Hao, P. Wang, and J. Wang, "Fast detection of traffic congestion from ultra-low frame rate image based on semantic segmentation," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 528–532, doi: 10.1109/ICIEA.2019.8834159.

[6] H. Asma-Ull, I. D. Yun, and D. Han, "Data efficient segmentation of various 3D medical images using guided generative adversarial networks," *IEEE Access*, vol. 8, pp. 102022–102031, 2020, doi: 10.1109/ACCESS.2020.2998735.

[7] L. Jing, Y. Chen, and Y. Tian, "Coarse-to-fine semantic segmentation from image-level labels," *IEEE Trans. Image Process.*, vol. 29, pp. 225–236, 2020.

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.

[10] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–12.

[14] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, Jan. 2018.

[15] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, Apr. 2019.

[16] Y. Sun, W. Zuo, and M. Liu, "RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2576–2583, Jul. 2019.

[17] R. Gade and T. B. Moeslund, "Thermal cameras and applications: A survey," *Mach. Vis. Appl.*, vol. 25, no. 1, pp. 245–262, 2014.

[18] M. Vollmer, *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[19] S. Minaee and Y. Wang, "An ADMM approach to masked signal decomposition using subspace representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3192–3204, Jul. 2019.
[20] L. He and H. Zhang, "Kernel K-means sampling for Nyström approximation," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2108–2120, May 2018.
[21] H. Bi, L. Xu, X. Cao, Y. Xue, and Z. Xu, "Polarimetric SAR image semantic segmentation with 3D discrete wavelet transform and Markov random field," *IEEE Trans. Image Process.*, vol. 29, pp. 6601–6614, 2020.
[22] B. Kang and T. Q. Nguyen, "Random forest with learned representations for semantic segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3542–3555, Jul. 2019.
[23] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, "Image segmentation based on adaptive K-means algorithm," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, p. 68, Dec. 2018.
[24] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 2881–2890.
[25] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 801–818.
[26] Y. Bao, K. Song, J. Wang, L. Huang, H. Dong, and Y. Yan, "Visible and thermal images fusion architecture for few-shot semantic segmentation," *J. Vis. Commun. Image Represent.*, vol. 80, Oct. 2021, Art. no. 103306, doi: 10.1016/j.jvcir.2021.103306.
[27] C. Li, W. Xia, Y. Yan, B. Luo, and J. Tang, "Segmenting objects in day and night: Edge-conditioned CNN for thermal image semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3069–3082, Jul. 2021, doi: 10.1109/TNNLS.2020.3009373.
[28] W. Zhou, J. Liu, J. Lei, L. Yu, and J.-N. Hwang, "GMNet: Graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 7790–7802, 2021, doi: 10.1109/TIP.2021.3109518.
[29] Q. Zhang, S. Zhao, Y. Luo, D. Zhang, N. Huang, and J. Han, "ABMDRNet: Adaptive-weighted bi-directional modality difference reduction network for RGB-T semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2633–2642, doi: 10.1109/CVPR46437.2021.00266.
[30] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (Lecture Notes in Computer Science), vol. 9351. 2015, pp. 234–241.
[32] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku, and T. Harada, "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 5108–5115.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[34] X. Lan, X. Gu, and X. Gu, "MMNet: Multi-modal multi-stage network for RGB-T image semantic segmentation," *Appl. Intell.*, 2021, doi: 10.1007/s10489-021-02687-7.
[35] E. Balit and A. Chadli, "GMFNet: Gated multimodal fusion network for visible-thermal semantic segmentation," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 1–4. [Online]. Available: https://drive.google.com/file/d/1pvXwxx4OD9ABD2gKYgcGQFvnns8Nh0e1/view
[36] G. Krispel, M. Opitz, G. Waltner, H. Possegger, and H. Bischof, "FuseSeg: LiDAR point cloud segmentation fusing multi-modal data," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1874–1883.
[37] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.
[38] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
[39] M. Tan and Q. V. Le, "EfficientnetV2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.
[40] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10428–10436.
[41] S. S. Shivakumar, N. Rodrigues, A. Zhou, I. D. Miller, V. Kumar, and C. J. Taylor, "PST900: RGB-thermal calibration, dataset and segmentation network," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 9441–9447.
[42] R. Parhi and R. D. Nowak, "The role of neural network activation functions," *IEEE Signal Process. Lett.*, vol. 27, pp. 1779–1783, 2020, doi: 10.1109/LSP.2020.3027517.
[43] C. Shi, Y. Zhou, B. Qiu, D. Guo, and M. Li, "CloudU-Net: A deep convolutional neural network architecture for daytime and nighttime cloud images' segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1688–1692, Oct. 2021, doi: 10.1109/LGRS.2020.3009227.
[44] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55135–55144, 2020, doi: 10.1109/ACCESS.2020.2978629.
[45] Y. Qiao, L. Jiao, S. Yang, and B. Hou, "A novel segmentation based depth map up-sampling," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 1–14, Jan. 2019, doi: 10.1109/TMM.2018.2845699.
[46] P. Bilinski and V. Prisacariu, "Dense decoder shortcut connections for single-pass semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6596–6605, doi: 10.1109/CVPR.2018.00690.
[47] Q. Ha, K. Watanabe, T. Karasawa, U. Yoshitaka, and T. Harada. *Multi-Spectral Semantic Segmentation Dataset*. [Online]. Available: https://drive.google.com/drive/folders/18BQFWRfhXzSuMloUmtiBRFrr6NSrf8Fw

MING-HWA SHEU (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the National Cheng Kung University, Taiwan, in 1989 and 1993, respectively. He is currently a Full Professor with the Department of Electronic Engineering, National Yunlin University of Science & Technology, Taiwan. From 2015 to 2018, he has worked as a Supervisor of the Taiwan IC Design Association. From 2008 to 2011, he worked as the Chairman of the Department of Electronic Engineering. His research interests include CAD/VLSI, digital signal process, algorithm analysis, edge AI, and embedded systems. He has served as the Committee Chair of the E.E. Course Planning for Technical High School, Ministry of Education, Taiwan. He has served as a Review Committee of the Engineering Department, Ministry of Science & Technology (MOST).

S. M. SALAHUDDIN MORSALIN (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from Daffodil International University, Bangladesh, in 2015, and the M.Sc. degree in green technology for sustainability (major in electronics) from Nanhua University, Taiwan, in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, National Yunlin University of Science and Technology, Taiwan. He is also working as a Research Assistant with the Department of Electronic Engineering, National Yunlin University of Science and Technology. He is also an Adjunct Lecturer with the Department of Computer Science and Information Engineering, Nanhua University, Taiwan. His research interests include image and video processing, data analytics, deep learning, artificial intelligence, and edge AI systems design.

SZU-HONG WANG (Member, IEEE) received the M.S. degree from the Department of Computer and Communication Engineering, National Kaohsiung University of Science and Technology, Taiwan, in 2005, and the Ph.D. degree from the Institute of Engineering Science and Technology, National Kaohsiung First University of Science and Technology, in 2010. He is currently an Associate Professor with Bachelor Program in interdisciplinary studies with the National Yunlin University of Science and Technology. His research interests include image processing, DSP/VLSI architecture design, and embedded systems.

LIN-KENG WEI received the B.Sc. degree from the Department of Aeronautical Engineering, National Formosa University, Taiwan, in 2013. He is currently pursuing the M.Sc. degree with the Department of Electronic Engineering, National Yunlin University of Science and Technology, Taiwan. His research interests include digital signal process, image processing, object detection, semantic segmentation, deep learning, embedded systems, and their applications.

SHIH-CHANG HSIA (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, National Cheng Kung University, Taiwan, in 1996. From 1986 to 1989, he was an Engineer with the R&D Department, Microtek International Inc. He was an Instructor and an Associate Professor with the Department of Electronic Engineering, Chung Chou Institute of Technology, from 1991 to 1998. He worked as a Professor with the Department of Computer and Communication Engineering and the Department of Electronics Engineering, National Kaohsiung First University of Science and Technology Kaohsiung, from 1998 to 2010. He was elected as the Chairman with the Department of Electronics Engineering, in 2007. He is currently a Professor with the Department of Electronics Engineering, National Yunlin University of Science and Technology. His research interests include VLSI/SOC designs, video/image processing, HDTV/Stereo TV systems, LED lighting systems, and electrical sensors.

CHUAN-YU CHANG (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the National Cheng Kung University, Taiwan, in 2000. He is currently the Deputy General Director of the Service Systems Technology Center, Industrial Technology Research Institute, Taiwan. He is also a Distinguished Professor with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Taiwan. He was the Chair of the Department of Computer Science and Information Engineering, from 2009 to 2011. From 2011 to 2019, he worked as the Dean of Research and Development and the Director of the Incubation Center for Academia-Industry Collaboration and Intellectual Property. His current research interests include computational intelligence and their applications to medical image processing, automated optical inspection, emotion recognition, and pattern recognition. In the above areas, he has more than 200 publications in journals and conference proceedings. He served as the Program Co-Chair of TAAI 2007, CVGIP 2009, 2010–2019 International Workshop on Intelligent Sensors and Smart Environments, and the Third International Conference on Robot, Vision and Signal Processing. He served as the General Co-Chair of 2012 International Conference on Information Security and Intelligent Control, 2011–2013 Workshop on Digital Life Technologies, CVGIP2017, WIC2018, ICS2018, and WIC2019. He is an IET Fellow, a Life Member of IPPR, and TAAI. From 2015 to 2017, he was the Chair of IEEE Signal Processing Society Tainan Chapter and the Representative for Region 10 of IEEE SPS Chapters Committee. He is currently the President of Taiwan Association for Web Intelligence Consortium.