IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 November 2022, accepted 4 January 2023, date of publication 10 January 2023, date of current version 18 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3235965

RESEARCH ARTICLE

Space Object Recognition With Stacking of CoAtNets Using Fusion of RGB and Depth Images

NOUAR ALDAHOUL[1], HEZERUL ABDUL KARIM[2], (Senior Member, IEEE), MHD ADEL MOMO[3], FRANCESCA ISABELLE FLORES ESCOBARA[4], AND MYLES JOSHUA TOLEDO TAN[4,5], (Member, IEEE)

[1]Computer Science Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[2]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia
[3]Fleet Management Systems & Technologies, 34522 Esenyurt, Turkey
[4]Department of Natural Sciences, University of Saint La Salle, Bacolod 6100, Philippines
[5]Department of Chemical Engineering, University of Saint La Salle, Bacolod 6100, Philippines

Corresponding author: Nouar Aldahoul (naa9497@nyu.edu)

This work was supported in part by Multimedia University, Malaysian.

ABSTRACT Space situational awareness (SSA) system requires recognition of space objects that are varied in sizes, shapes, and types. The space images are challenging because of several factors such as illumination and noise and thus make the recognition task complex. Image fusion is an important area in image processing for various applications including RGB-D sensor fusion, remote sensing, medical diagnostics, and infrared and visible image fusion. Recently, various image fusion algorithms have been developed and they showed a superior performance to explore more information that are not available in single images. In this paper, we compared various methods of RGB and Depth image fusion for space object classification task. The experiments were carried out, and the performance was evaluated using 13 fusion performance metrics. It was found that the guided filter context enhancement (GFCE) outperformed other image fusion methods in terms of average gradient (8.2593), spatial frequency (28.4114), and entropy (6.9486). additionally, due to its ability to balance between good performance and inference speed (11.41 second), GFCE was selected for RGB and Depth image fusion stage before feature extraction and classification stage. The outcome of fusion method is fused images that were used to train a deep ensemble of CoAtNets to classify space objects into ten categories. The deep ensemble learning methods including bagging, boosting, and stacking were trained and evaluated for classification purposes. It was found that combination of fusion and stacking was able to improve classification accuracy largely compared to the baseline methods by producing an average accuracy of 89 % and average F1 score of 89 %.

INDEX TERMS CoAtNet, deep ensemble learning, RGB-D, image fusion, space object classification.

I. INTRODUCTION

The near-Earth space environment is known to be used for both commercial and scientific use. Satellites are launched on a regular basis in aid of space navigation, communication, and weather forecasting. As technology advances, space exploration and satellite launching become more feasible leading to increased activity in the near future. In the past decades there has been a rapid growth of space debris and objects measured with an estimated over 750,000 debris measuring beyond 1 cm found within the orbit, and more than 20,000 near-Earth objects [1], [2]. Space objects are vaguely defined to be hardware launched by man into outer space [3], [4] while debris include artificial objects that no longer serve purpose and fragments from collisions and anti-satellite tests [1]. Man-made space debris make up the majority of the objects found within the space environment, with explosions caused by residual energy in fuel and batteries being the largest contributor [5]. Near-Earth objects (NEO) encompass asteroids and comets whose orbit passes close, roughly 45 million

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

kilometers, to the Earth's orbit [6], [7]. These may range from a few meters to several kilometers in size, thus while chances are slim, regular tracking is essential. Due to the increasing prevalence of space environment use, possible collisions or approach of different objects should always be monitored in order to ensure the safety of the people and secure essential space satellites and shuttles. The field of space situational awareness (SSA) encompasses subjects concerning space environment, tackling the following subjects: space surveillance and tracking of man-made objects, space weather monitoring and forecast, and near-Earth objects monitoring [8]. In relation, efforts for space traffic management (STM) are also in effect for safe outer space projects. Detection, identification, and surveillance of different space objects is an important task for SSA. The use of computational methods for this mission can result in more efficient systems for observation as well as impact risk assessment and mitigation.

Documentation of the space environment and activity makes use of substantial data for analysis and investigation. Digital sky surveys are a prevalent source of data in astronomy. These consist of terabytes of image data with characterized attributes covering a full range of wavelengths [2]. With the progression of technology, the influx of such data opens a window for more effective digital analysis [9], [10] presenting the application of machine learning to the subject field. Machine learning is a branch of artificial intelligence built on computational learning theory and pattern recognition. Due to its eficiency, machine learning has been applied for data analysis across several aspects including computer vision, natural language processing, and data analysis [11], [12]. This provides aid in analyzing large-scale data and addresses the significance of pre-processing techniques as well as the problems pertaining to image artifacts and difficulties in distinguishing certain objects. Moreover, when compared to physical models, data-driven models have shown better performance in prediction tasks and are not limited by construction time and quality of structure [9]. Further advancements such as deep learning brings artificial intelligence to even greater heights exhibiting superior performance against older machine learning approaches. Currently, such models are implemented in many programs including space navigation, astronaut assistance, and Earth observatory data analysis with many projects under development [13]. With many existing applications of recognition tasks across healthcare and commercial industries, machine learning for space object recognition holds a promising opportunity that can bear positive impact in the management of the expanding use of satellites and space shuttles.

II. RELATED WORK

Traditional machine learning methods have been used over the years in application of classification, detection, and data selection. Artificial neural networks (ANNs), decision trees (DTs) and support vector machines (SVMs) have been used in space debris detection, galaxy morphology classification, and identification of asteroids [14], [15], [16], [17]. Principal component analysis (PCA) also aided in several classification problems [16], [18] for the reduction of dimensionality of data in which only information most essential is retained for increased interpretability [19]. Conventional models make use of signals extracted from sensors or features processed from image data. Detection and classification of satellites and space debris by Perez et al. explored traditional and deep learning techniques. When using SVM and multinomial SoftMax regression (MSR), the outcome was observed to have increased performance after applying PCA and cubic feature mapping. Overall, cubic feature vector and deep learning achieved 99.8% and 99.3% accuracy on distant objects respectively [20].

Deep learning is advantageous to traditional methods in many ways. Consisting of multiple layers, information is consecutively transformed into higher abstract representations [21]. This enables the learning of complex functions where raw data can be directly processed. However, by nature deep learning requires high amounts of data to execute well. This poses a challenge in space object recognition research as datasets are limited. To resolve this, synthetic datasets are formulated. Dung et al. [22] introduced an annotated spacecraft dataset made of real and synthetic images for detection, segmentation, and recognition tasks. Perez et al. [20] used images of experimental setups of space objects for machine learning. Other novel datasets include URSO [23] and SPEED [24], [25] for spacecraft pose estimation from laboratory-acquired or computer-generated images. SPARK [26] is a more recent dataset with 11 classes for spacecraft and debris recognition. This provides 150,000 images featuring RGB and depth data.

Currently, convolutional neural networks (CNNs) are most frequently used in the field of computer vision [2]. CNNs are composed of several layers made up of convolutional neurons with each carrying out convolution of the input to output a feature map [1]. Many studies have used CNNs for both space object classification and detection tasks. Transfer learning of pre trained CNNs is prevalent, using knowledge obtained from one task on a different but related problem. These were trained on extensive classification or detection datasets such as ImageNet or COCO dataset, where its learned weights can be used to train smaller datasets. Space object identification using ResNet18 [20] obtained an accuracy of 99.96%. Afshar and Lu [27] also implemented the same method for satellite classification and pose estimation achieving promising results. To further resolve limited data, they also used data augmentation to improve performance. For solar system object detection and classification [28], the model combined an Image Classification Network (ICN) and Tracklet Classification Network (TCN). For ICN, using pre-trained ResNet18 was discovered to be more advantageous than training a CNN from scratch. Other literature used well-known CNN architectures as part of their framework. Jia et al. [29] used ResNet50 as a backbone in feature extraction when using Faster R-CNN for astronomical target recognition; while it was able to surpass older frameworks, it still had lapses in

analyzing stimulated data. A CNN based on LeNet was used in debris classification as part of a space debris detection algorithm proposed by Xi et al. [30]. A convolutional kernel size of 5 × 5 was selected for testing across image classification of varying signal-to-noise ratios. Results ranged from 86-99% in accuracy. You only look once (YOLO) is another CNN that was used for satellite and satellite component recognition [31]. When tested under different distances and conditions, the model was able to have a recognition accuracy over 90%.

Efforts for refined computer vision, incorporation of separate image data presents a promising approach. RGB information can be restricted in two-dimensional space, limiting information for better analysis, thus the addition of depth information can significantly boost object recognition [32]. Multimodal models combine and analyze features obtained from separate RGB and depth images. In a survey by Gao et al. [32], object recognition was higher when using RGB-D fusion techniques across hand crafted-based methods, traditional feature learning-based methods, and multimodal CNN models in contrast to using depth or RGB images alone. In spacecraft recognition, two studies found improved performance using RGB-D-based techniques. RGB and depth images were classified [33] wherein a pre-trained ResNet50 with a SVM classifier was applied on RGB image classification while an end-to-end CNN was utilized for depth images; overall results measured 89% in accuracy. A separate paper [34] discusses the improvement in spacecraft and debris recognition using multi-modal methods combining transfer learning of pre-trained CNNs and vision transformers on RGB data with end-to-end CNN on depth images. The proposed method was superior to models that solely used vision transformers and CNN models.

Parallel to the improvements seen in technologies throughout the years that have resulted in large amounts of diverse data with greater quality and accessibility are equally significant improvements in the field of computer vision. Since as early as 1985, image information fusion has been on the rise, particularly for fusion of infrared and visible light images, image fusion in remote sensing, multi-focus imaging, and medical imaging.

Image fusion is another technique wherein images are combined in order to come up with an output containing information from multiple samples [35]. There are several methods for image fusion categorized into spatial and frequency domain. Spatial image fusion involves pixel manipulation while frequency domain covering decomposition and discrete transformation. A more advanced technique is using deep learning models for image fusion. Such processes have contributed to the improvement of medical, surveillance, and remote sensing. Li et al. [36] illustrates the efficacy of image fusion of RGB and infrared images using deep learning. The study used VGG19 for feature extraction, followed by fusion and image reconstruction. In comparison to previous methods, the proposed model produced the best outputs in image quality. Rather recent developments in deep learning have enabled the creation of methods for image fusion based on CNNs [37], [38], [39], recurrent neural networks [40], the U-Net architecture [41], [42], and generative adversarial networks (GANs) [43], [44], [45], [46]. Zhang et al. [46] introduced a visible and infrared image fusion benchmark (VIFB) containing a test set, 20 fusion algorithms with evaluation results featuring 13 metrics. Experimental results show that non-learning algorithms were superior over deep learning models with multi-scale image decomposition producing decent results in both flight and man light image pairs.

This paper highlights an interesting challenge for the research community. It contributes to the body of knowledge as follows:

1) A novel space related image fusion method to generate fused images from RGB and depth images using various image fusion algorithms. The fusion methods were evaluated and compared in terms of 13 different evaluation metrics.
2) The fused images generated have similar contents of original RGB and Depth images which leads to images that are rich of informative contents and comprehensive features which is significant to enhance the decision-making tasks such as classification of space objects.
3) To the best of our knowledge, this is the first paper that proposes an ensemble of CoAtNets (Convolution and Attention Networks) [50].
4) Utilization of deep ensemble learning such as bagging [47], bosting [48], and stacking [49] for classifying space objects into ten categories. Each weak learner in ensemble models is based on CoAtNet.

This paper is organized as follows: Section III describes the space objects dataset. Additionally, it discusses numerous image fusion methods. Additionally, it demonstrates deep ensemble learning models such as bagging, bosting, and stacking for classification tasks. In Section IV, the experiments conducted are described to analyze results in detail. We performed an evaluation and comparison between image fusion methods from one side and deep ensemble learning models in other side. Finally, Section V. summarizes the work presented in this paper by giving readers a glimpse into potential improvements in the future.

III. MATERIALS AND METHODS

In this section, we present an overview of the dataset utilized in this work to show the challenging contents of spacy images. Additionally, various image fusion methods are explored to highlight the difference between them in terms of various metrics. Moreover, numerous ensemble learning methods such as bagging, boosting, and stacking are demonstrated to study their efficiency to enhance the performance of space objects classification task.

A. DATASET OVERVIEW

A unique space object dataset was used in this research to address the space object recognition challenge presented in

ICIP 2021. This novel dataset is called SPARK (Spacecraft Recognition leveraging Knowledge of the Space Environment) [26], [51], [52]. A set of 300,000 images divided into 150,000 RGB and 150,000 depth was utilized to categorize 11 different objects classes, considering ten satellite objects (each with 12,500 images) and one piece of orbital debris [26], [51]. The debris items were classified into different five classes (5,000 photos for each) but were merged as one. AcrimSat, Aquarius, Aura, Calipso, CloudSat, CubeSat, Debris, Jason, Sentinel-6, Terra, and TRMM are the eleven classes, which were derived from 3D resources of NASA [26], [51]. The performance of space-based imagery was determined by several elements, including excellent contrast, variable lighting, and a minimum signal-to-noise ratio.

The 150,000 RGB images and another 150,000 depth images were fused using fusion method to generate new 150,000 fused images that were distributed as follows: 60% (90,000 photos) for training, 20% (30,000 photos) for validation, and 20% (30,000 photos) for testing [26], [51]. The RGB images have a resolution of 1024×1024, while depth images have a resolution of 256×256. This dataset presents the following challenges [26], [51]: (a) The objects are randomly located inside the range of vision of a chaser's equipment. Additionally, the chaser model comes in a variety of angles and distances and the earth and sun rotate randomly about their axes. (b) The collection contains photographs of scenes illuminated and images with increased contrast. Models of the sun's flares, beams, and reflective surfaces on earth from space were created. (c) The visual backdrops are varied, incorporating a variety of orbital settings. Spacecrafts are positioned and oriented in a variety of ways in the background. (d) There are a variety of ranges and separations among the chaser's image sensor and the target spacecraft. (e) Spaceborne photos have a substantial noise level due to the small sensor scale and the high-definition photography.

Figure 1 illustrates several pictures from the SPARK dataset [26], [51]. These pictures were selected to show various challenges presented in this dataset such as increased contrast and illuminated stars, random of objects, various object sizes, various locations, orientations, and positions of space objects in the background, a real noise level, the earth with clouds and oceans in the background, and a variety of orbital settings.

B. THE PROPOSED SOLUTION

The solution proposed in this work to recognize space objects in images includes a combination of image fusion method and ensemble learning model. Twenty image fusion methods have been evaluated using 13 metrics to select the best that can balance between performance and runtime speed [46]. The outcome of image fusion block is fused images that were applied to the inputs of ensemble learning models to learn patterns from them. Various ensemble methods including bagging [47], boosting [48], and stacking [49] were implemented and evaluated. Convolution and attention network [50] (CoAtNet) was used as a basic learner in each of ensemble models. The reason behind selection of CoAtNet is that it takes advantages of both convolutional layers and attention mechanism. The block diagram of the ensemble of CoAtNets connected to fusion of RGB and depth Images is shown in Figure 2.

C. IMAGE FUSION METHODS

Image fusion was found to fuse sources images and generate a fused image. Recently, several algorithms have been found for fusion of visible and infrared images. To do a performance comparison between these algorithms, VIFB benchmark was found with twenty most recent algorithms for visible and infrared image fusion [46]. The algorithms include gradient transfer (GTF) [53], convolutional neural network (CNN) [54], Anisotropic diffusion (ADF) [55], cross bilateral filter (CBF) [56], fourth order partial differential equations (FPDE) [57], guided filter context enhancement (GFCE) [58], hybrid multi-scale decomposition guided filter (HMSD GF) [58], Guided Filtering (GFF) [59], hybrid multi-scale decomposition (Hybrid-MSD) [60], infrared feature extraction and visual information preservation (IFEVIP) [61], latent low-rank representation (LatLRR) [62], Multi-scale guided image filter-based fusion (MGFF) [63], multi-scale transform and sparse representation (MST SR) [64], Laplacian pyramid and sparse representation (RP SR) [64], nonsubsampled contourlet transform and sparse representation (NSCT SR) [64], multi-resolution singular value decomposition (MSVD) [65], residual network (ResNet) [66], Two-scale image fusion (TIF) [67], visual saliency map and weighted least square (VSM WLS) [68], and deep learning framework (DLF) [36]. The algorithms were adjusted for color image fusion. Each RGB image channel was fused with corresponding infrared image. The algorithms were mainly divided into Multi-scale, Deep learning (DL) based, and Hybrid. ADF [55], CBF [56], GFCE [58], GFF [59], MGFF [63], HMSD GF [58], Hybrid MSD [60], and MSVD [65] are multi-scale methods. On the other hand, DLF [36], CNN [54], and ResNet [66] are DL based methods. Additionally, MST SR [64], NSCT SR [64], RP SR [64], and VSMWLS [68] are hybrid methods [46].

In this work, we implemented these twenty algorithms using RGB and depth space images instead of visible and infrared images. The aim is to extract informative details from each of these RGB and depth images and combine them in fused images to enhance the following stage of classification of space object types. Figure 3 shows samples of depth images, RGB images, and fused images for ten categories of space objects including AcrimSat, Aquarius, Aura, Calipso, CubeSat, Debris, Jason, Sentinel-6, Terra, and TRMM from left to right respectively.

D. DEEP ENSEMBLE LEARNING MODELS

This section discusses the three main ensemble learning models including bagging, boosting, and stacking. These models were utilized to recognize space objects and classify them

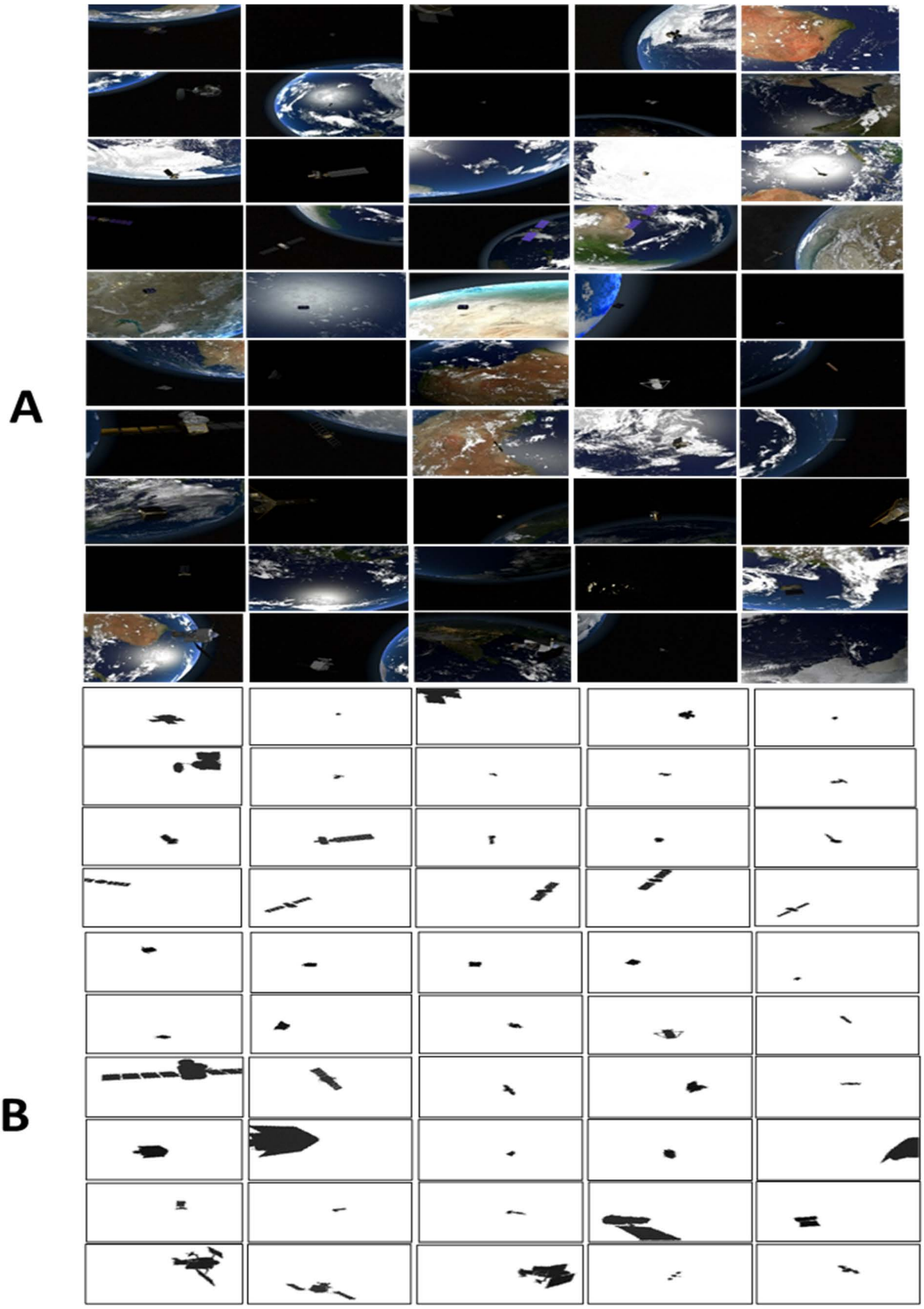


FIGURE 1. Few samples of RGB and depth space images with various object categories, sizes, and backgrounds from the SPARK dataset [26], [51], [52].

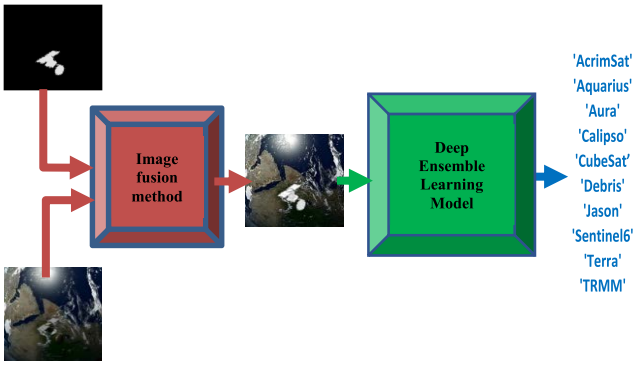


FIGURE 2. The block diagram of the proposed Solution. The images were taken from [26] and [52].

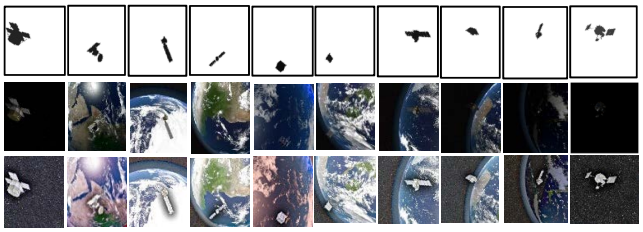


FIGURE 3. samples of source images (depth in the first row and RGB in the second row) and fused images (in the third row) The depth and RGB images were taken from [26] and [52]. The fused images were generated by the GFCE method.

TABLE 1. Learner architecture.

Input Layer (Image Size= (384,384,3))
CoAtNet0 (backbone)
GlobalAveragePooling2D layer
Dense layer (10 nodes)
SoftMax activation function

TABLE 2. Learner hyperparameters.

Batch size=16
Optimizer=Adam
Loss=Categorical Cross entropy

into ten categories. The input of these models were fused images that result from image fusion block. The architecture of each learner is shown in Table 1. All the layers in this architecture were trained including the layers of CoAtNet0. Additionally, the hyperparameters of each learner are shown in Table 2. Table 3 refers to training and testing distribution for each class.

1) CoAtNet

This network enjoys the strengths of both ConvNets and Transformers [50]. The light version of CoAtNet, namely CoAtNet0 was utilized as a pre-trained model (already trained on ImageNet dataset [83]) and all parameters of all layers

TABLE 3. Training and testing distribution for each class.

Class	Number of Samples training	Number of Samples testing
AcrimSat	7500	2500
Aquarius	7500	2500
Aura	7500	2500
Calipso	7500	2500
CubeSat	7500	2500
Debris	15000	5000
Jason	7500	2500
Sentinel-6	7500	2500
Terra	7500	2500
TRMM	7500	2500

TABLE 4. Training dataset distribution for each class in each learner in bagging technique.

Class	Number of Samples
AcrimSat	750
Aquarius	750
Aura	750
Calipso	750
CubeSat	750
Debris	1500
Jason	750
Sentinel-6	750
Terra	750
TRMM	750

were fine-tuned with SPARK dataset. We used CoAtNet0 as a weak learner in various ensemble learning models such as bagging, boosting, and stacking.

2) BAGGING

In this technique, ten weak learners were used. Each learner is CoAtNet version 0 [50]. The outputs of these ten learners were aggregated by maximum to find the final prediction as shown in figure 4. The experiment was run on GPU named RTX A6000 with GPU memory = 48 GB. The dataset was divided into ten subsets. Each learner was trained on each subset. Table 4 clarifies the distribution of training dataset for each class in each learner. Figure 4 represents both training and inference phases.

3) BOOSTING

In this technique, two weak learners were used. Each learner is CoAtNet version 0 [50]. In training phase, the distribution of training data used with learner 1 is shown in Table 5. The misclassified samples from the first learner were utilized to

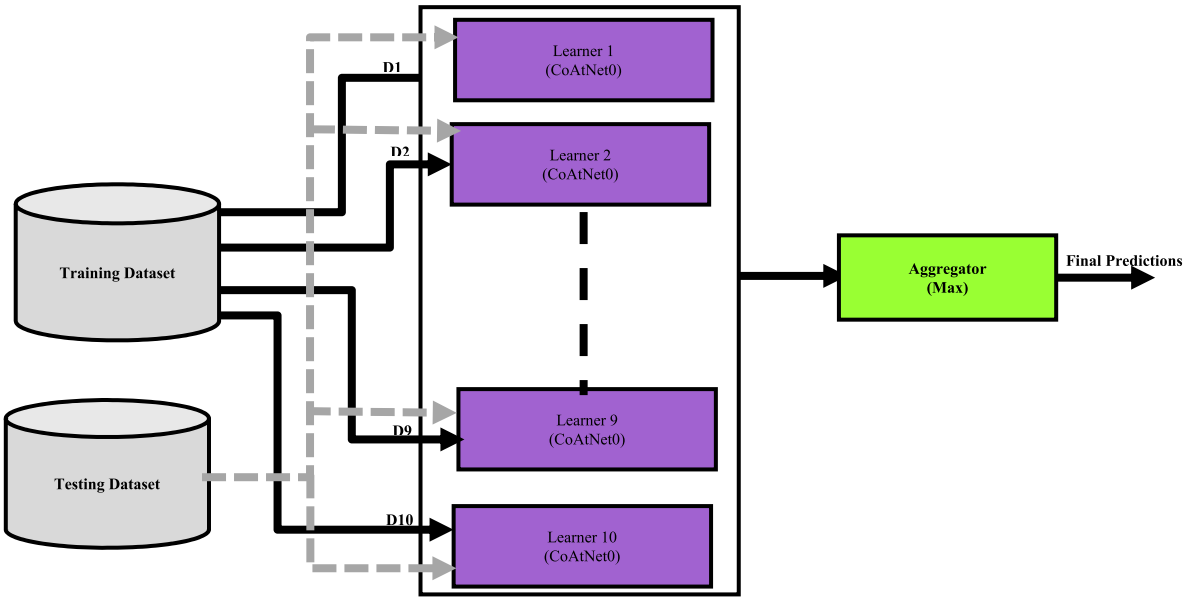


FIGURE 4. Ensemble learning using bagging technique.

TABLE 5. Training dataset distribution for each class for the first learner in boosting technique.

Class	Number of Samples
AcrimSat	2250
Aquarius	2250
Aura	2250
Calipso	2250
CubeSat	2250
Debris	4500
Jason	2250
Sentinel-6	2250
Terra	2250
TRMM	2250

TABLE 6. Training dataset distribution for each class in for the second learner in boosting technique.

Class	Number of Samples
AcrimSat	239
Aquarius	216
Aura	340
Calipso	270
CubeSat	110
Debris	331
Jason	356
Sentinel-6	320
Terra	507
TRMM	158

train the second learner as shown in Table 6. In inference phase, the outputs of two learners were aggregated by maximum to find the final prediction as shown in figure 5. The experiment was run on GPU named RTX A6000 with GPU memory = 48 GB. The architecture of each learner is shown in Table 1. Additionally, the hyperparameters of each learner are shown in Table 2. Figure 5 refers to each of training and inference phases.

4) STACKING

In this technique, three weak learners were used. Each learner is CoAtNet version 0 [50]. The outputs of these three learners were applied to meta learner to find the final prediction as shown in figure 6. The experiment was run on GPU named RTX A6000 with GPU memory = 48 GB. The architecture of each learner is shown in Table 1. Furthermore, the hyperparameters of each learner are shown in Table 2. The dataset was divided into three subsets. Each learner was trained on each subset. Table 7 clarifies the distribution of training dataset for each class in each learner. Additionally, Table 8 indicates the distribution of training dataset for each class in meta learner. Figure 6 represents both training and inference phases.

The architecture of meta learner is shown in Table 9. It includes support vector machine (SVM) [82] with regularization factor C = 1 and kernel function of radial basis function (RBF). The input of meta learner is 30 features (3 weak learners × 10 predictions of each learner).

IV. RESULTS AND DISCUSSION

This section discusses the experimental setup and results of image fusion methods from one side and results of ensemble learning models from the other side.

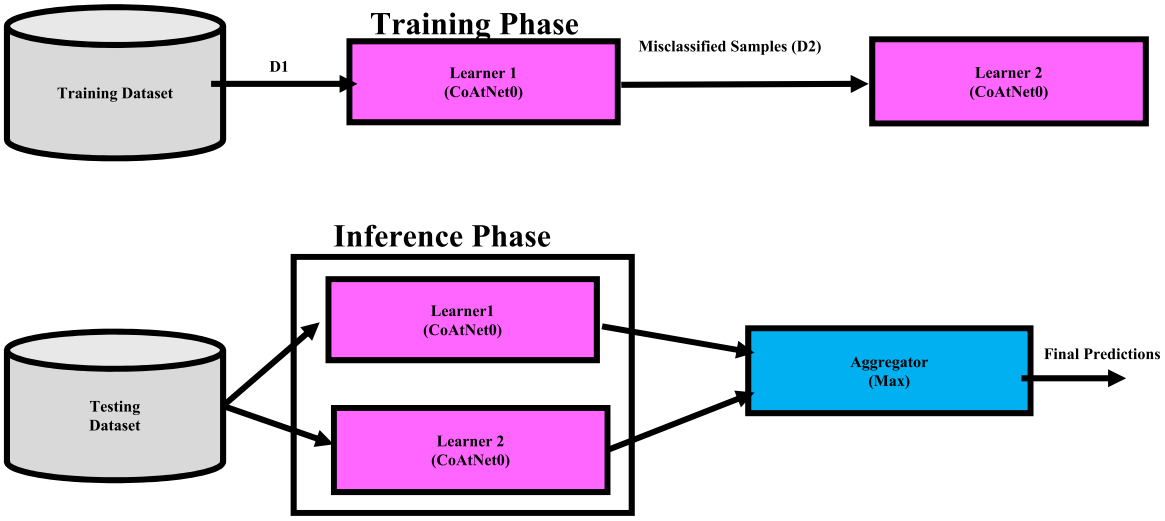


FIGURE 5. Ensemble learning using boosting technique.

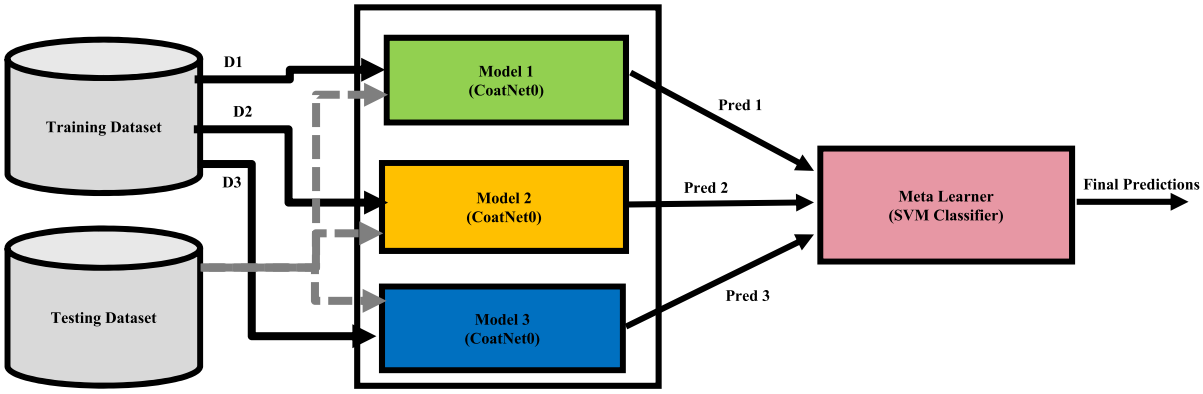


FIGURE 6. Ensemble learning using Stacking technique.

TABLE 7. Dataset distribution for each learner in stacking technique.

Class	Number of Samples
AcrimSat	1750
Aquarius	1750
Aura	1750
Calipso	1750
CubeSat	1750
Debris	3500
Jason	1750
Sentinel-6	1750
Terra	1750
TRMM	1750

TABLE 8. Meta learner dataset distribution in stacking technique.

Class	Number of Samples
AcrimSat	2250
Aquarius	2250
Aura	2250
Calipso	2250
CubeSat	2250
Debris	4500
Jason	2250
Sentinel-6	2250
Terra	2250
TRMM	2250

A. EXPERIMENTAL SETUP

For space image fusion using SPARK dataset, twenty image fusion methods were evaluated using thirteen various metrics. The results are shown in thirteen figures (from Figure 7 to Figure 19) and in Table 10. MATLAB program was used to run these 20 methods using 54 pairs of images of RGB and depth.

For space object recognition, the training set that were provided with labels were divided into two sets: 80 % (72,000 images) for training, and 20% (18,000 images) for validation. On the other hand, testing dataset has 30,000 RGB images

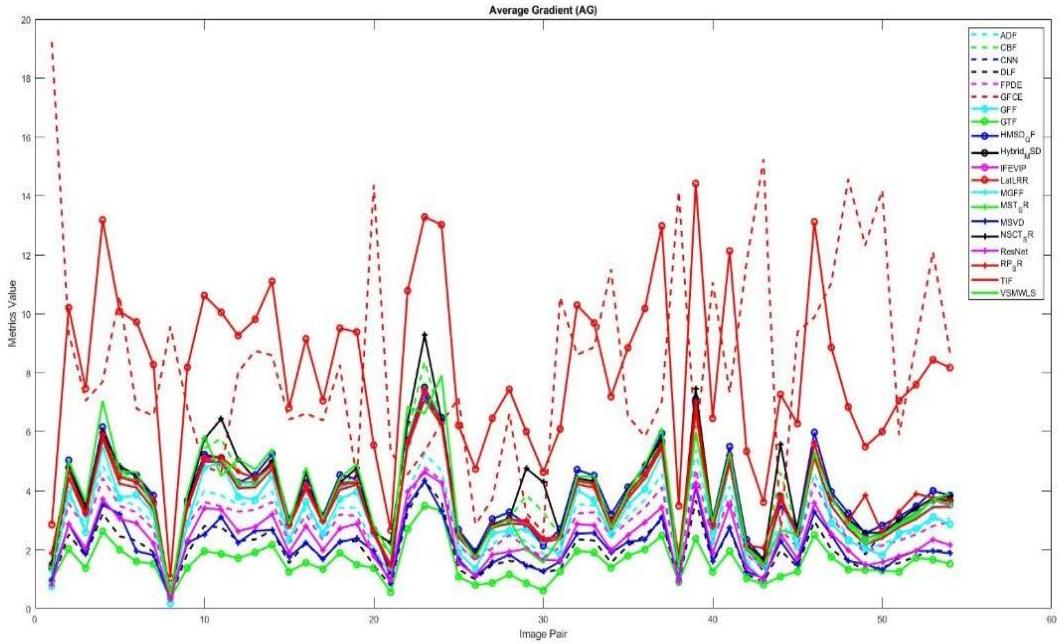


FIGURE 7. Quantitative comparisons of AG metric of 20 methods.

TABLE 9. Meta learner architecture.

Input (features= (30))
SVC (C=1.0, kernel=RBF)

and their correspondence of depth images. Tables 12, 13, 14 and Figure 20 refer to the results using testing dataset. In this work, we carried out several experiments for space object recognition using numerous ensemble learning methods such as bagging, boosting, and stacking. Python with TensorFlow and scikit learn frameworks were utilized for this task.

B. QUANTITATIVE RESULTS FOR IMAGE FUSION

In this work, various types of metrics [46] have been used for RGB and depth image fusion for comprehensive quantitative comparison. These metrics are classified into four categories including human perception-based such as Chen-Blum metric (QCB) [69] and chen-varshney metric (QCV) [70], image feature-based such as average gradient (AG) [71], edge intensity (EI) [72], standard deviation (SD) [73], spatial frequency (SF) [74], and gradient based fusion performance (QAB/F) [75], image structural similarity-based such as root mean squared error (RMSE) [76] and structural similarity index measure (SSIM) [77], and information theory-based such as Mutual information (MI) [78], peak signal-to-noise ratio (PSNR) [76], cross entropy (CE) [79], and entropy (EN) [80]. The description of 13 metrics are as follows:

1) SD measures image contrast. Higher value of SD indicates better fusion method.
2) EI preserves the edge detail information and presents a high image quality and more clearness. Higher value of EI indicates better fusion method.
3) AG represents the amount of texture variation in the image. Higher value of AG indicates more gradient information the image contains and better fusion method.
4) SF measures the amount of frequency content in the image to show the clarity or sharpness. Additionally, it preserves high frequency content. Higher value of SF indicates the richer edges and texture details the image contains and better fusion method.
5) EN measures the information content in an image. Higher value of EN indicates better fusion method.
6) CE evaluates similarity in information content between the fused image and the source images. Lower value of CE indicates better fusion method.
7) MI is an index that calculates the quantity of dependency between source and fused images. Higher value of MI indicates better fusion method.
8) PSNR measures the amount of noise available in the fused image. Higher value of PSNR indicates the less distortion the fusion process produces and better fusion method.
9) RMSE measures the error between two images. Lower value of RMSE indicates better fusion method.
10) SSIM is a measure of the similarity (perceptual difference) between two images. Higher value of SSIM indicates better fusion method.
11) QAB/F is used to measure the transferred edge information amount from source images to the fused one. Higher value of QAB/F indicates better fusion method.
12) QCV to measure the visual performance. Lower value of QCV indicates better fusion method.
13) QCB to measure the visual performance. Higher value of QCB indicates better fusion method.

TABLE 10. Average evaluation metric values for twenty fusion methods (human perception-based in light-green, image feature-based in light-blue, image structural similarity-based in orange, and information theory-based in lavender) on 54 image pairs. The highest three values are colored in red, green and blue, respectively. The best five methods are highlighted in grey color.

Method	ADF [55]	CBF [56]	CNN [54]	DLF [36]	FPDE [57]	GFCE [58]	GFF [59]	GTF [53]	HMSD_GF [58]	Hybrid_MSD [60]	IFEVIP [61]	LatLRR [62]	MGFF [63]	MST_SR [64]	MSVD [65]	NSCT_SR [64]	ResNet [66]	RP_SR [64]	TIF [67]	VSMWLS [68]
Average Gradient (AG)	2.9614	3.8243	3.6039	1.9901	2.7715	8.2593	3.2411	1.5778	3.8199	3.6735	3.6475	8.1515	3.5183	3.6620	2.1168	3.9726	2.3700	3.7438	3.5035	3.7526
Cross Entropy (CE)	3.1791	2.6099	2.5599	3.2646	3.1627	5.8847	4.2985	2.3556	3.1772	3.1329	2.3187	2.0937	2.7620	3.0278	3.5072	2.9421	2.7381	2.8154	2.5004	3.5515
Edge Intensity (EI)	27.5882	37.2670	35.0027	19.2129	25.6341	70.1874	33.9066	14.8345	36.9578	35.6445	35.3538	78.0973	34.1293	35.5414	19.5018	39.0023	23.0164	36.0981	34.0722	36.0215
Entropy (EN)	5.6914	6.3719	6.2785	5.5515	6.6759	6.9486	6.2726	2.2967	6.2551	6.3798	6.3466	6.2444	5.8034	6.4032	5.5929	6.4461	6.1053	6.3997	5.8089	5.9319
Mutual information (MI)	2.1627	4.2866	3.5677	3.3731	2.2684	1.8560	2.0640	0.3971	3.6249	3.7633	4.3275	1.6135	1.7406	3.8153	2.0892	4.2491	1.8950	3.8000	1.7196	2.3460
Peak Signal-to-Noise Ratio (PSNR)	61.8979	61.5026	61.2716	61.9097	61.9036	57.9895	61.6657	61.6943	61.2306	61.5729	61.3646	60.6443	61.7899	61.3605	61.9012	61.2367	61.8397	61.3250	61.8071	61.5855
Gradient based fusion performance ($Q^{AB/F}$)	0.7442	0.9402	0.9174	0.5542	0.6933	0.5178	0.6328	0.2857	0.9065	0.9392	0.9446	0.2899	0.8682	0.9411	0.3977	0.9450	0.7472	0.9027	0.8559	0.8112
Chen-Blum metric (Q_{cb})	0.3162	0.3473	0.3264	0.3248	0.3209	0.2107	0.2919	0.3901	0.3239	0.3375	0.3408	0.3224	0.3127	0.3385	0.3141	0.3486	0.3211	0.3389	0.3157	0.2854
Chen-Varshney metric (Q_{cv})	1000	381.5906	213.4802	922.8418	931.6496	572.9844	86.9583	5343	108.4371	81.3829	170.4730	258.6518	860.9435	190.1087	981.0992	1337	507.1603	227.5276	419.5640	392.3321
Root mean squared error (RMSE)	0.0477	0.0522	0.0542	0.0476	0.0476	0.1055	0.0499	0.0497	0.0549	0.0509	0.0529	0.0614	0.0487	0.0532	0.0477	0.0564	0.0484	0.0533	0.0485	0.0508
Spatial Frequency (SF)	8.8189	12.4233	11.3841	6.3269	8.0895	28.4114	10.4048	6.5214	12.1673	11.5383	11.4384	25.1193	10.9900	11.4713	7.9609	11.9192	7.5240	13.1268	11.0281	12.0382
Structural similarity index measure (SSIM)	0.9548	1.0790	1.0878	0.9377	0.9527	0.4969	1.0150	0.9264	1.0814	1.0928	1.0954	0.8688	0.9479	1.0887	0.9123	1.0743	1.0180	1.0797	0.9495	0.9950
Standard Deviation (SD)	32.2267	57.2712	61.3142	31.8411	31.9082	59.1283	58.5928	20.4218	62.2788	59.0618	60.7250	68.1338	37.8289	60.6173	31.6791	58.1875	40.1640	60.3683	37.0702	50.8166

TABLE 11. RGB and depth image fusion algorithms that have been compared in term of runtime.

Method	Average Time [s]
TIF[67]	0.121241
IFEVIP [61]	**0.150890**
MST_SR[64]	0.318490
GFF [59]	1.280850
RP_SR[64]	1.487353
MSVD[65]	3.569982
ADF [55]	3.654358
MGFF[63]	5.26262
ResNet [66]	9.601039
HMSD_GF [58]	9.735696
CBF [56]	114.405382
GFCE [58]	**11.414826**
FPDE [57]	14.860613
VSMWLS[68]	17.035625
Hybrid_MSD [60]	38.733202
GTF [53]	46.018902
DLF [36]	**47.407352**
CNN [54]	203.412232
NSCT_SR[64]	**297.913130**
LatLRR[62]	**1224.841291**

TABLE 12. Classification report of bagging technique.

class	precision	recall	F1-score	Number of samples
0	0.90	0.88	0.89	2500
1	0.97	0.75	0.85	2500
2	0.84	0.90	0.87	2500
3	0.81	0.84	0.83	2500
4	0.93	0.88	0.91	2500
5	0.87	0.90	0.89	5000
6	0.74	0.90	0.81	2500
7	0.92	0.91	0.91	2500
8	0.86	0.95	0.90	2500
9	0.83	0.66	0.74	2500

To have comprehensive and objective performance comparison, 13 metrics that cover all four categories were used in Table 10. The average values of 13 evaluation metrics for the twenty methods on 54 image pairs were demonstrated. It is obvious that the GFCE [58] gave the best values for three metrics including entropy, average gradient, and special frequency and the second-best values of edge intensity. Additionally, we can indicate from the Table 10 that IFEVIP method got the best values for mutual information and structural similarity index measure and the second-best values for cross entropy and gradient based fusion performance. On the other hand, deep learning-based method such as DLF was able to outperform others in terms of root mean squared error and peak signal-to-noise ratio. Finally, LatLRR produced the best value for cross entropy, edge intensity and standard deviation and the second-best values of average gradient and spatial frequency.

The results presented in this table indicate that there is no dominant fusion method that can produce the best values of

TABLE 13. Classification report of boosting technique.

class	precision	recall	F1-score	Number of samples
0	0.99	0.84	0.91	2500
1	0.93	0.78	0.85	2500
2	0.95	0.51	0.66	2500
3	0.79	0.81	0.80	2500
4	0.92	0.93	0.92	2500
5	0.80	0.94	0.86	5000
6	0.85	0.70	0.77	2500
7	0.88	0.85	0.87	2500
8	0.81	0.97	0.88	2500
9	0.61	0.85	0.71	2500

TABLE 14. Classification report of stacking technique.

class	precision	recall	F1-score	Number of samples
0	0.98	0.89	0.94	2500
1	0.93	0.86	0.90	2500
2	0.96	0.79	0.87	2500
3	0.86	0.83	0.84	2500
4	0.95	0.97	0.96	2500
5	0.78	0.98	0.87	5000
6	0.92	0.76	0.83	2500
7	0.95	0.86	0.90	2500
8	0.88	0.98	0.92	2500
9	0.84	0.85	0.84	2500

all metrics. The reason is that each method focused on specific type of information and metrics when it was designed.

To investigate more quantitative comparison between various image fusion methods, the values of 13 metrics for the twenty methods on 54 image pairs are shown in thirteen figures from Figure 7 to Figure 19. Figure 7 shows that LatLRR and GFCE produced the best (highest) AG for 54 image pairs. On the other hand, GTF gave the worst AG for 54 image pairs.

Figure 8 shows that LatLRR, IFEVIP, and GTF produced the best (smallest) CE for 54 image pairs. On the other hand, GFCE gave the worst CE for 54 image pairs. Additionally. Figure 9 shows that LatLRR and GFCE produced the best (highest) EI for 54 image pairs. On the other hand, GTF gave the worst EI for 54 image pairs.

Figure 10 shows that GFCE produced the best (highest) EN for 54 image pairs. On the other hand, GTF gave the worst EN for 54 image pairs. Additionally. Figure 11 shows that IFEVIP, CBF, and NSCT_SR produced the best (highest) MI for 54 image pairs. On the other hand, GTF gave the worst MI for 54 image pairs.

Figure 12 shows that almost all methods produced high PSNR for 54 image pairs. On the other hand, GFCE gave the worst PSNR for 54 image pairs. Additionally. Figure 13 shows that IFEVIP, NSCT_SR and MST_SR produced the best (highest) QAB/F for 54 image pairs. On the other hand, GTF and LatLRR gave the worst QAB/F for 54 image pairs.

Figure 14 shows that GTF produced the best (highest) Qcb for 54 image pairs. On the other hand, GFCE gave the worst Qcb for 54 image pairs. Additionally. Figure 15 shows that Hybrid MSD and GFF produced the best (smallest) Qcv for 54 image pairs. On the other hand, GTF gave the worst Qcv for 54 image pairs.

Figure 16 shows that DLF and FPDE produced the best (smallest) RMSE for 54 image pairs. On the other hand, GFCE gave the worst RMSE for 54 image pairs. Additionally. Figure 17 shows that Hybrid MSD and IFEVIP produced the best (highest) SSIM for 54 image pairs. On the other hand, GFCE gave the worst SSIM for 54 image pairs.

Figure 18 shows that GFCE produced the best (highest) SF for 54 image pairs. On the other hand, DLF gave the worst SF for 54 image pairs. Additionally. Figure 19 shows that LatLRR produced the best (highest) SD for 54 image pairs. On the other hand, GTF gave the worst SD for 54 image pairs.

C. RUNTIME COMPARISON

In this work, an experiment was carried out to compare the runtime of image fusion algorithms as shown in Table 11. Twenty algorithms were implemented in VIFB work [46]. We reran these methods with space object dataset to fuse RGB and depth images. As can be seen, most image fusion methods are computationally expensive. We can also infer large variations of runtime between various image fusion methods. Moreover, the fusion methods that belong to the same category have also varied runtimes. For example, even both CBF [56] and GFF [59] are multi-scale methods, but the runtime of GFF [59] is 1.28 second and the runtime of CBF [56] is 114.41 seconds. On the other hand, deep learning-based algorithms that used pretrained models have very long runtime. However, ResNet [66] is the fastest deep learning-based method with 9.60 seconds. Similarly, LatLRR [62] and NSCT SR [64] are very time-consuming with 1224.84, 297.91 seconds respectively.

A real-time image fusion is required in various applications. Therefore, we need to select fusion method that can balance between speed and fusion performance. GFCE [58] was able to get this balance. As seen in Table 11, GFCE produced low runtime of 11.41 second and at the same time as seen in Table 10, GFCE [58] gave the first best values for three metrics including entropy and average gradient and special frequency and the second-best values of edge intensity. Therefore, in this work, GFCE was selected to fuse image pairs to produce fused space images that were applied to the recognition stage. The methods in Table 11 which are shown in bold font are the best five methods in Table 10.

D. CLASSIFICATION RESULTS

The classification performance was evaluated using several metrics, namely accuracy, recall, precision, and F1 score. The performance indicators are as follows:

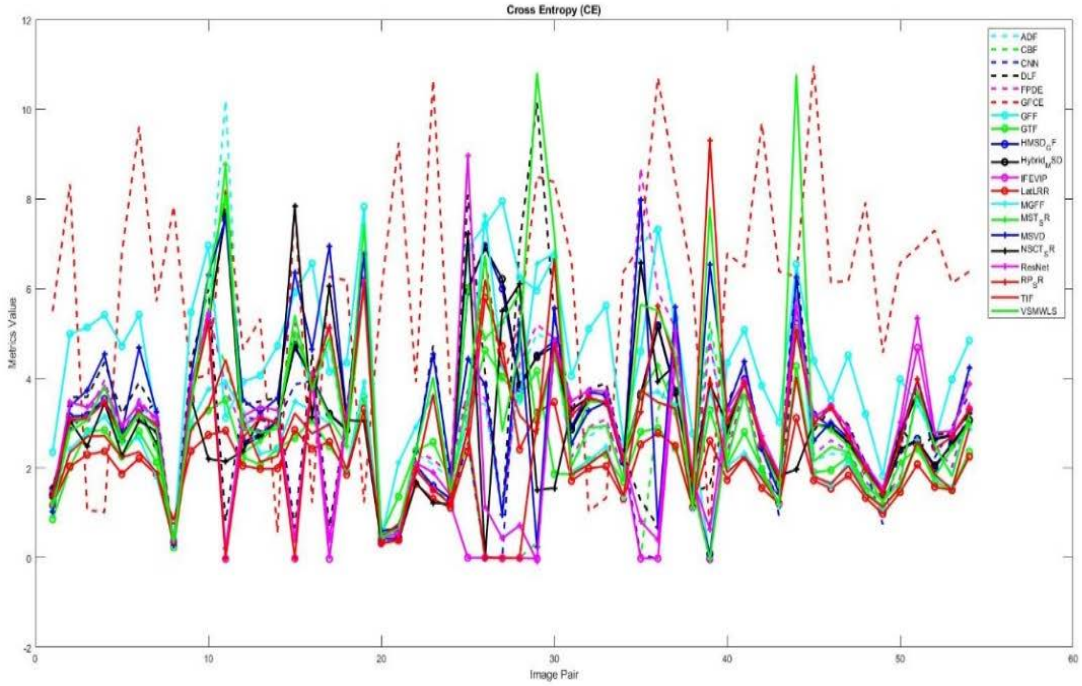


FIGURE 8. Quantitative comparisons of CE metric of 20 methods.

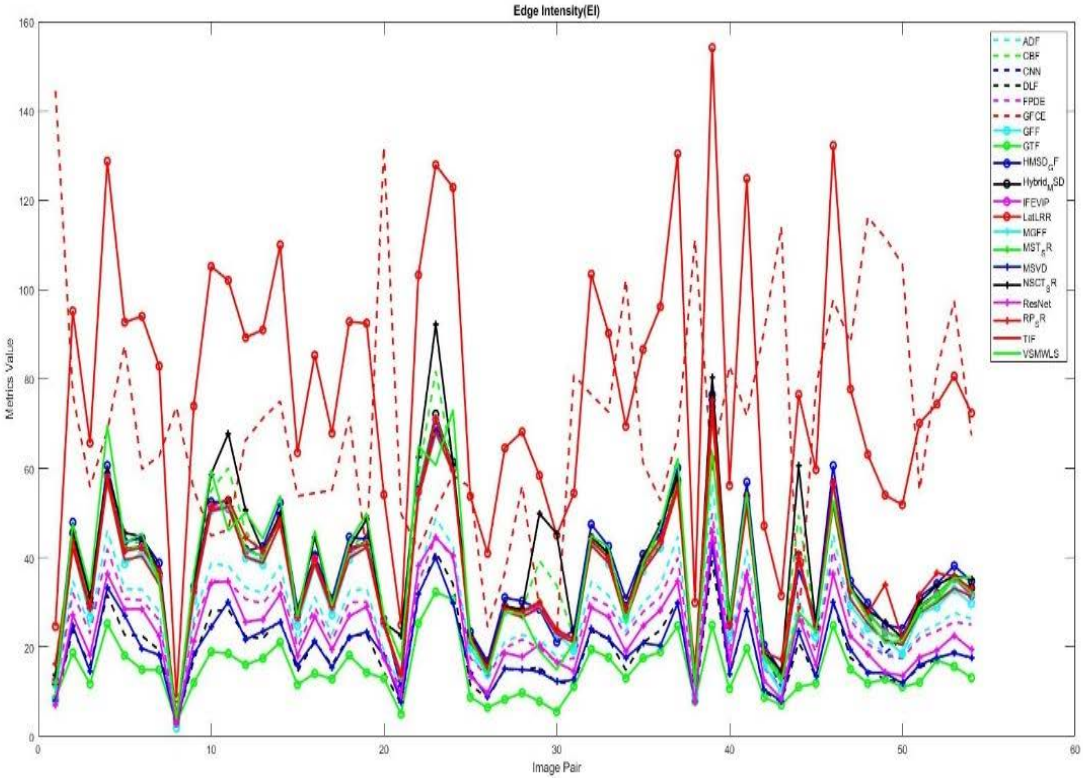


FIGURE 9. Quantitative comparisons of EI metric of 20 methods.

1) Accuracy is a metric to find proportion of samples that were predicted correctly.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2) Recall (Sensitivity) is a metric to find the proportion of actual positives that were predicted correctly.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

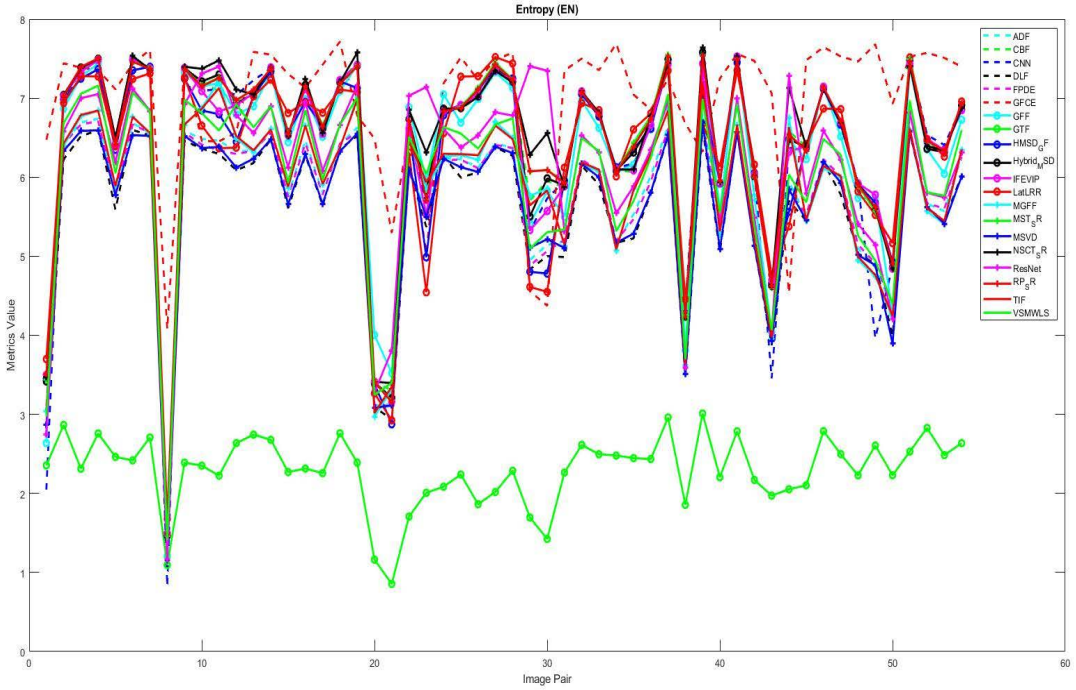


FIGURE 10. Quantitative comparisons of EN metric of 20 methods.

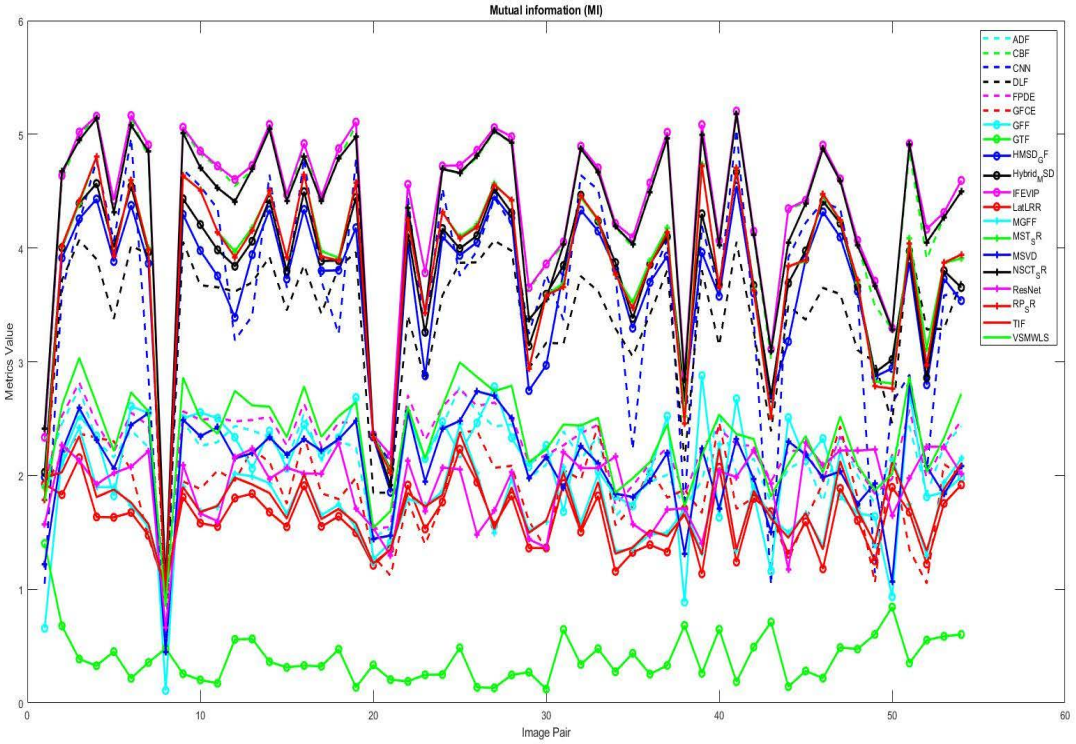


FIGURE 11. Quantitative comparisons of MI metric of 20 methods.

3) Precision (positive predictive value) is a metric to find the proportion of correct positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

4) F1 score is a metric to summarize precision and recall into a single metric.

$$F1score = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

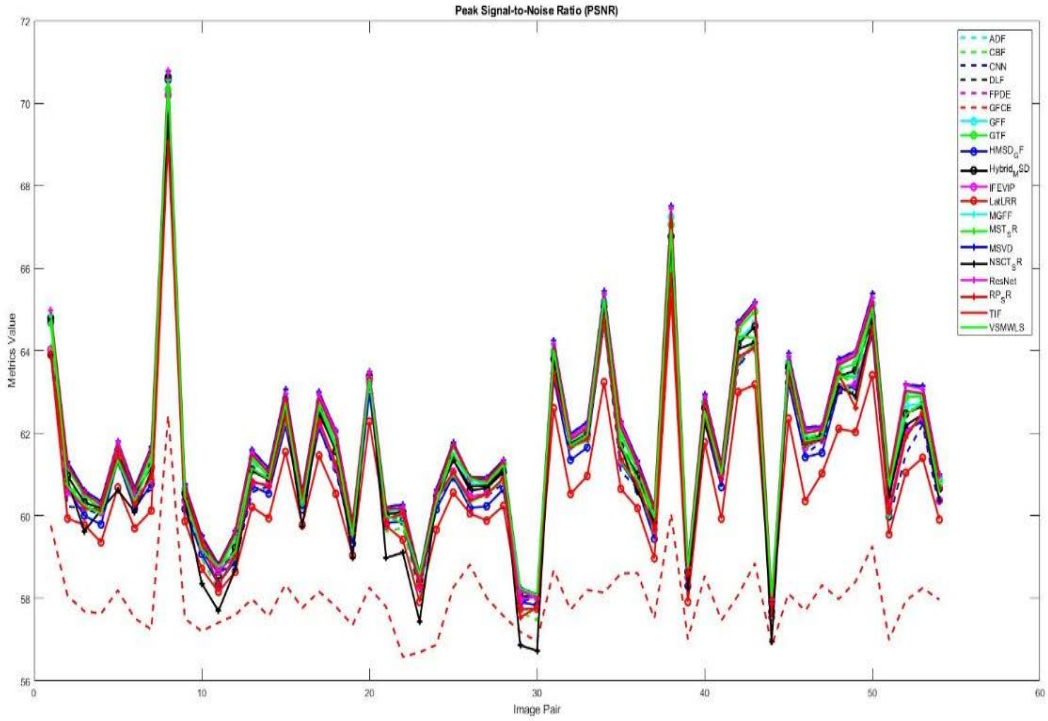


FIGURE 12. Quantitative comparisons of PSNR metric of 20 methods.

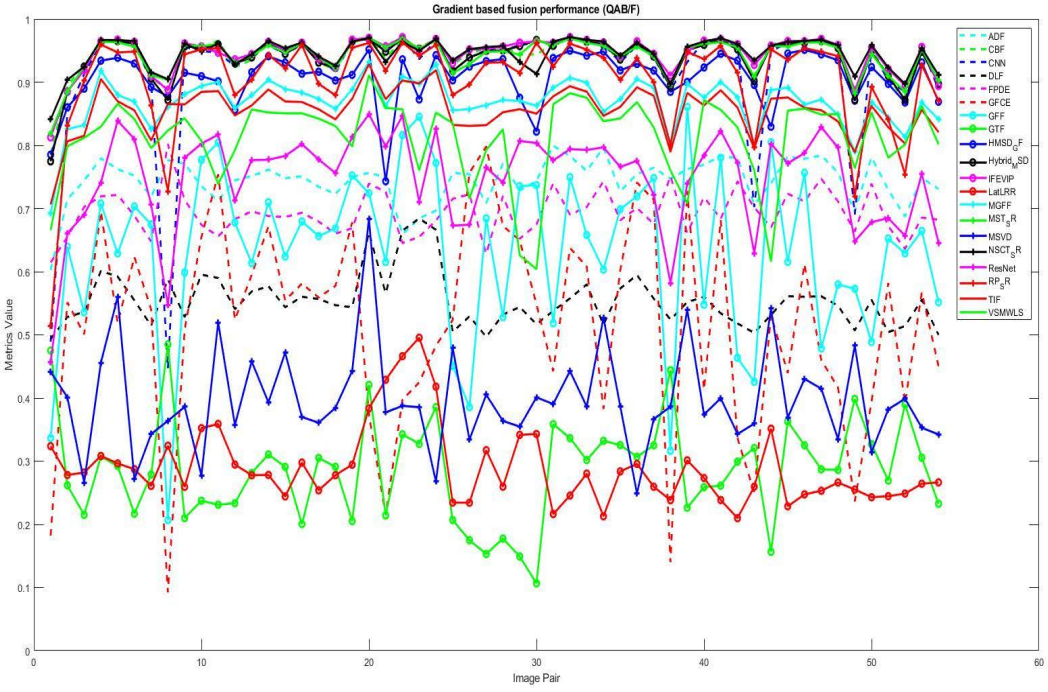


FIGURE 13. Quantitative comparisons of QAB/F metric of 20 methods.

where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

The three techniques of ensemble learning including bagging, boosting, and stacking were compared as shown in Figure 20. The confusion matrix for each technique is illustrated in Figure 20. As can be seen, the values of the main diagonal in confusion matrix of stacking model are higher than other ensemble learning models which leads to better classification accuracy.

The classification report that indicates the accuracy, recall, precision, and F1 score for each class of each ensemble learning method including bagging, boosting, and stacking was demonstrated in Table 12, Table 13, and Table 14.

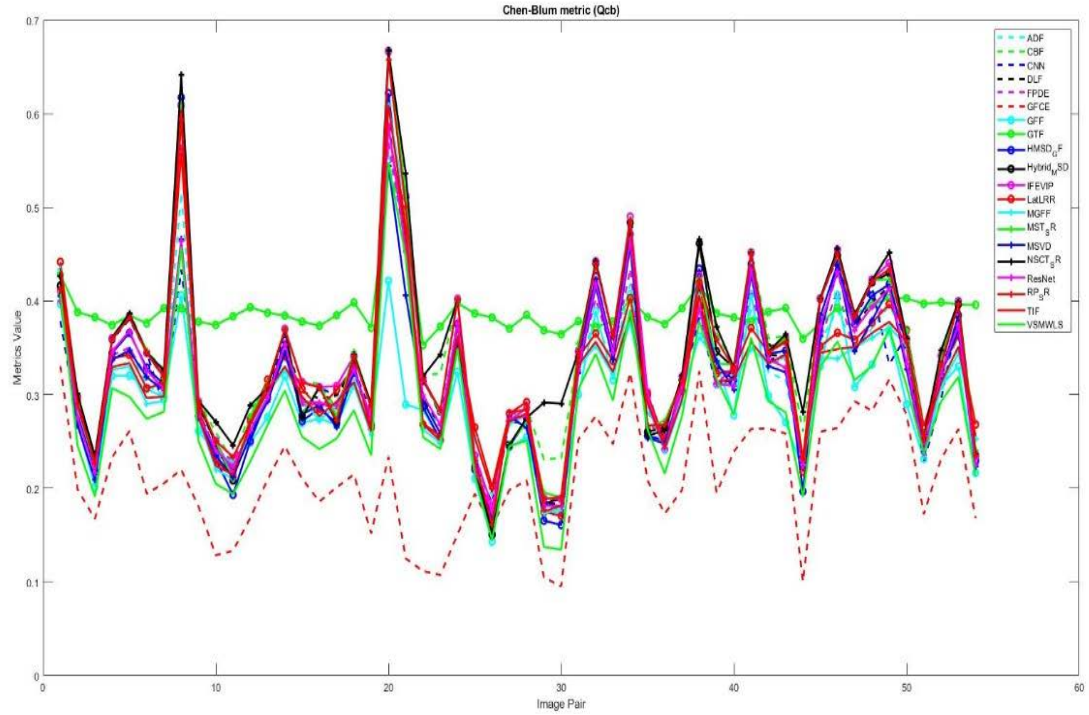


FIGURE 14. Quantitative comparisons of Qcb metric of 20 methods.

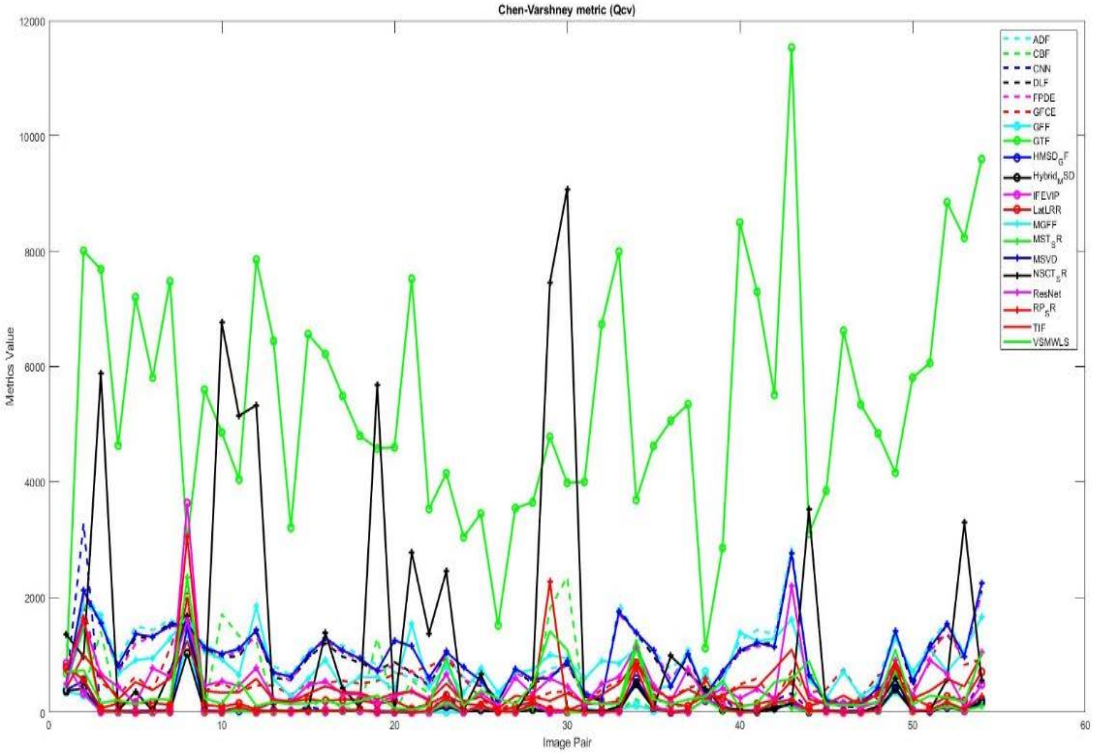


FIGURE 15. Quantitative comparisons of Qcv metric of 20 methods.

The three techniques of ensemble learning including bagging, boosting, and stacking were compared in term of average accuracy, average recall, average precision, and average F1 score as shown in Table 15. It was found that stacking model was able to outperform other ensemble learning models such as bagging and boosting in terms of accuracy, recall,

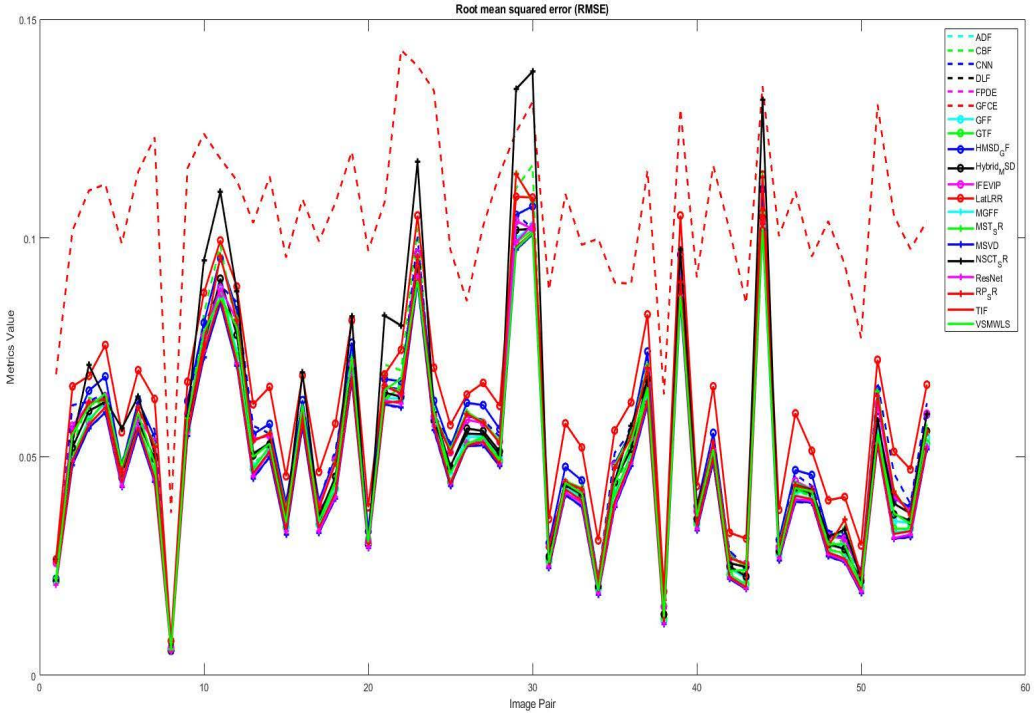


FIGURE 16. Quantitative comparisons of RMSE metric of 20 methods.

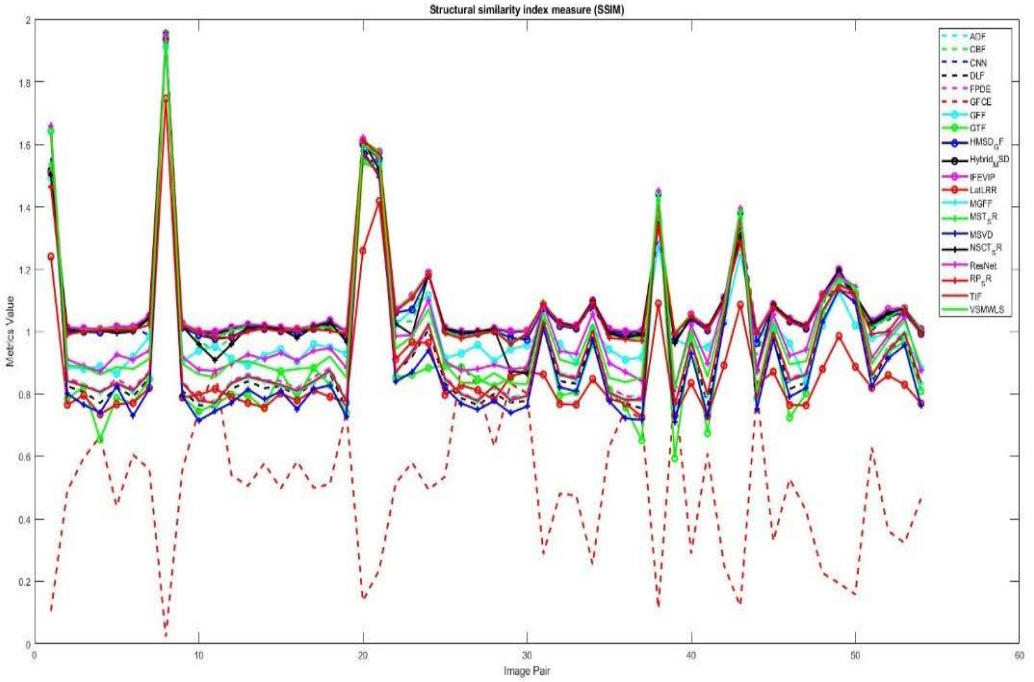


FIGURE 17. Quantitative comparisons of SSIM metric of 20 methods.

precision, and F1 score. Stacking was able to give the highest average accuracy of 89 % and the highest average F1 score oof 89 %. On the other hand, boosting gave the lowest average accuracy of 83 % and the lowest average F1 score of 82 %. The highest values are shown in bold font.

In summary, in this paper, an interesting challenge of space object recognition was addressed using combination of image fusion and ensemble learning models. The contributions of this work are summarized as follows:

1) RGB and Depth images were fused using twenty image fusion methods. The fusion methods were evaluated and compared in terms of 13 different evaluation metrics.

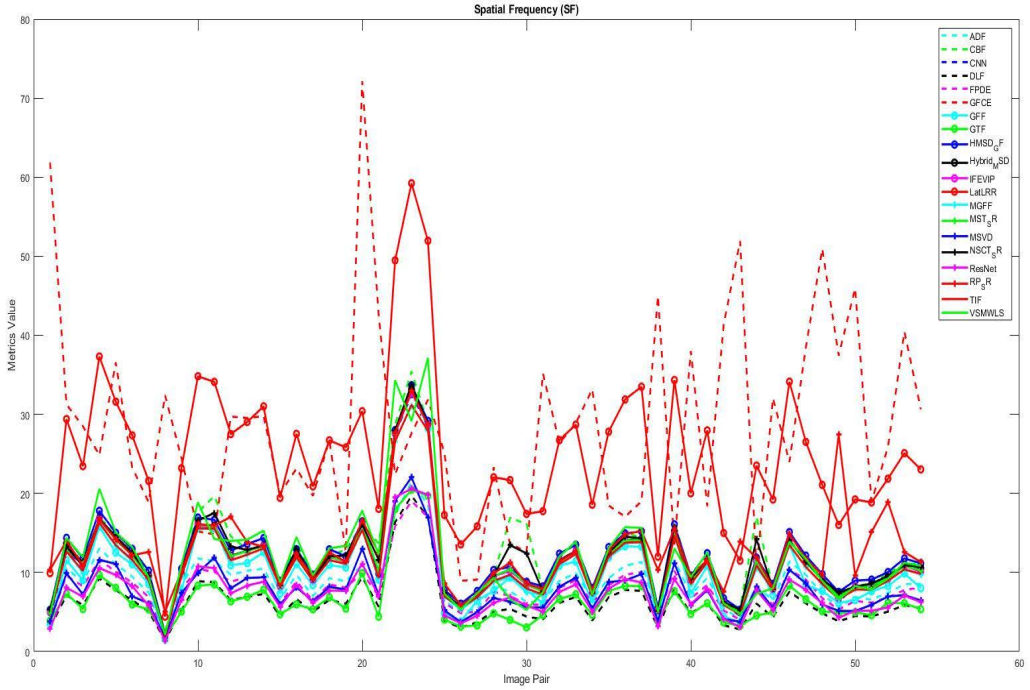


FIGURE 18. Quantitative comparisons of SF metric of 20 methods.

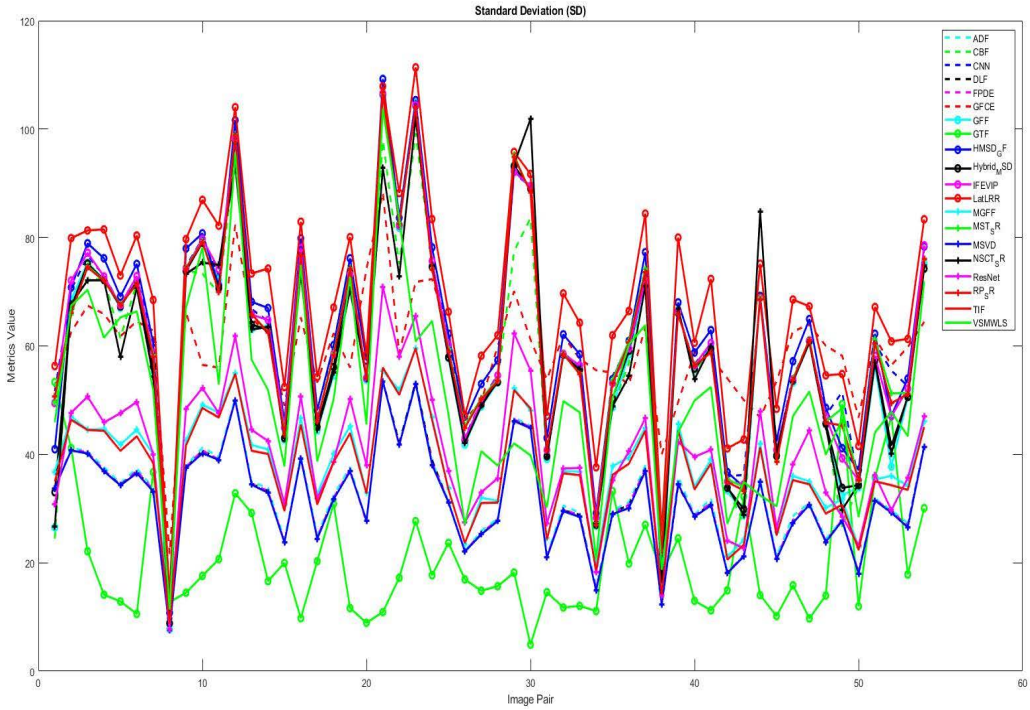


FIGURE 19. Quantitative comparisons of SD metric of 20 methods.

2) The fused images produced informative contents and comprehensive features which can enhance the task of space object recognition.
3) We proposed an ensemble of CoAtNets (Convolution and Attention Networks) using ensemble learning such as bagging, bosting, and stacking to classify space objects into ten categories.

V. CONCLUSION AND FUTURE WORK

In this work, various image fusion algorithms have been evaluated for task of RGB-D image fusion and they showed superior performance to explore more information that are not available in single images. We explored numerous fusion methods to fuse RGB and depth images that include space objects for recognition purposes. The experiments

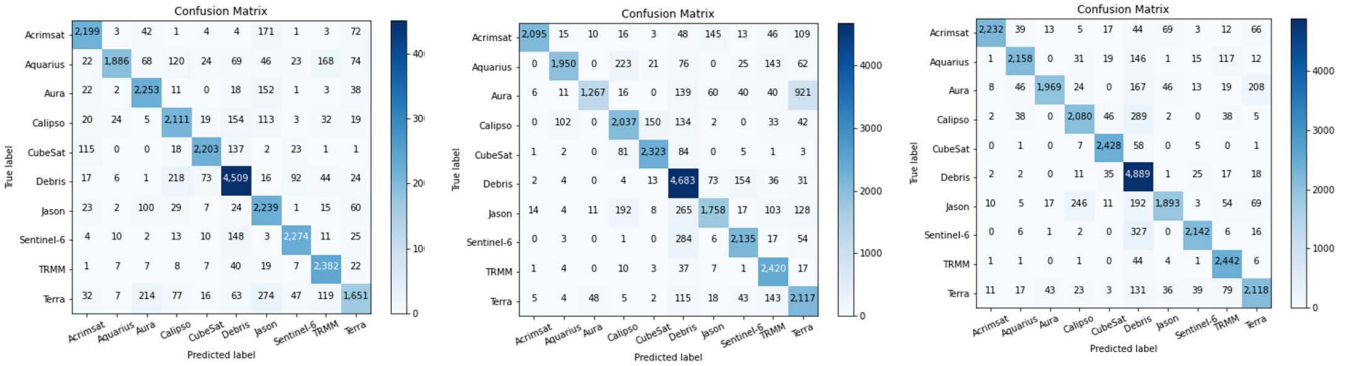


FIGURE 20. Confusion Matrix of Ensemble Learning Using bagging (on the left), boosting (in the middle), and stacking (on the right) techniques.

TABLE 15. Comparison between bagging, boosting, and sacking techniques.

Metric | Methods	Bagging (baseline)	Boosting (baseline)	Stacking (proposed)
Average Accuracy	86	83	**89**
Average Precision	87	85	**91**
Average Recall	86	82	**88**
Average F1-score	86	82	**89**

were carried out, and the performance was evaluated using 13 fusion performance metrics. It was found that the GFCE outperformed other image fusion methods in terms of average gradient (8.2593), spatial frequency (28.4114), and entropy (6.9486). Furthermore, LatLRR was able to produce the best cross entropy (2.0937), edge intensity (78.0973), and standard deviation (68.1338). Additionally, IFEVIP has the best mutual information (4.3275) and structural similarity index measure (1.0954). On the other hand, DLF was found to outperform others in terms of peak signal-to-noise ratio (61.9097) and root mean squared error (0.0476). However, due to its ability to balance between good performance and inference speed (11.41 second), GFCE was selected to fuse RGB and depth images before feature extraction and classification of space objects existing in images.

The outcome of fusion method is fused images that were used to train deep learning model to classify space objects into ten categories. The deep ensemble methods including bagging, boosting, and stacking contain several CoAtNets models that were trained using SPARK dataset. We utilized CoAtNet to enjoy the strengths of both ConvNets and Transformers. These ensemble learning methods were evaluated and compared for classification purposes. It was found that stacking of CoAtNets was able to outperform other ensemble learning models in terms of accuracy, recall, precision, and F1 score. Stacking was able to give the highest average accuracy of 89 % and the highest average F1 score of 89 %.

Finally, it was found that combination of fusion and stacking was able to enhance the space object recognition performance largely which helps to improve the performance of space situational awareness system.

Hence, we intend to enhance the performance of recognition by using object detection method after image fusion method. We expect that this can improve the detection performance in comparison with previous works that targeted RGB images only for detection purposes [81]. Additionally, an ensemble of object detectors may also contribute to enhance the detection performance in general and the recognition specifically. Therefore, it can be a good bonus to our future works.

ACKNOWLEDGMENT

The authors would like to thank the University of Luxembourg and LMO for sharing their dataset. The SPARK dataset used in this work was proposed in ICIP2021 challenge. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONTRIBUTIONS

The authors would like to acknowledge the contributions of each author. Conceptualization was done by Nouar Aldahoul, Mhd Adel Momo; data curation was done by Nouar Aldahoul; formal analysis was done by Nouar Aldahoul; funding acquisition was done by Hezerul Abdul Karim; investigation was done by Nouar Aldahoul, Myles Joshua Toledo Tan; methodology was done by Nouar Aldahoul, Mhd Adel Momo; project administration was done by Hezerul Abdul Karim; software was done by Nouar Aldahoul, Mhd Adel Momo; validation was done by Nouar Aldahoul; visualization was done by Nouar Aldahoul, Mhd Adel Momo; writing—original draft preparation was done by Nouar Aldahoul, Mhd Adel Momo, Francesca Isabelle Flores Escobar; and writing—review & editing was done by Nouar Aldahoul, Myles Joshua Toledo Tan, Hezerul Abdul Karim.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] (2009). *About Space Debris*. Esa.int. [Online]. Available: https://www.esa.int/Safety_Security/Space_Debris/About_space_debris

[2] J. Lei and L. Kong, "Fundamentals of big data in radio astronomy," *Big Data Astron.*, pp. 29–58, Jan. 2020, doi: 10.1016/b978-0-12-819084-5.00010-9.

[3] J. A. Dennerley, "State liability for space object collisions: The proper interpretation of 'fault' for the purposes of international space law," *Eur. J. Int. Law*, vol. 29, no. 1, pp. 281–301, May 2018, doi: 10.1093/ejil/chy003.

[4] V. Kopal, "Some remarks on issues relating to legal definitions of 'space object,' 'space debris,' and 'astronaut,'" *Int. Inst. Space Law*, vol. 37, pp. 99–108, 1994.

[5] (2020). *The Current State of Space Debris*. [Online]. Available: https://www.esa.int/Safety_Security/Space_Debris/The_current_state_of_space_debris

[6] *Near-Earth Objects*. Accessed: Sep. 1, 2022. [Online]. Available: https://www.unoosa.org/oosa/en/ourwork/topics/neos/index.html

[7] *Near-Earth Objects—NEO Segment*. Accessed: Sep. 1, 2022. [Online]. Available: https://www.esa.int/Safety_Security/Near-Earth_Objects_-_NEO_Segment

[8] (SSA). Accessed: Sep. 1, 2022. [Online]. Available: https://www.satcen.europa.eu/page/ssa

[9] J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. S. Pedersen, and C. Igel, "Big universe, big data: Machine learning and image analysis for astronomy," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 16–22, Mar. 2017, doi: 10.1109/MIS.2017.40.

[10] S. Sen, S. Agarwal, P. Chakraborty, and K. P. Singh, "Astronomical big data processing using machine learning: A comprehensive review," *Exp. Astron.*, vol. 53, no. 1, pp. 1–43, Jan. 2022, doi: 10.1007/s10686-021-09827-4.

[11] S. Angra and S. Ahuja, "Machine learning and its applications: A review," in *Proc. Int. Conf. Big Data Anal. Comput. Intell. (ICBDAC)*, vol. 1, Mar. 2017, pp. 57–60. [Online]. Available: https://ieeexplore.ieee.org/document/8070809

[12] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–22, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[13] *Artificial Intelligence in Space*. Accessed: Sep. 1, 2022. [Online]. Available: https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/Artificial_intelligence_in_space

[14] M. Khalil, E. Fantino, and P. Liatsis, "Evaluation of oversampling strategies in machine learning for space debris detection," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Dec. 2019, pp. 1–6, doi: 10.1109/IST48021.2019.9010217.

[15] J. D. Hefele, F. Bortolussi, and S. P. Zwart, "Identifying Earth-impacting asteroids using an artificial neural network," *Astron. Astrophys.*, vol. 634, p. A45, Feb. 2020, doi: 10.1051/0004-6361/201935983.

[16] M. Reza, "Galaxy morphology classification using automated machine learning," *Astron. Comput.*, vol. 37, Oct. 2021, Art. no. 100492, doi: 10.1016/j.ascom.2021.100492.

[17] D. Carrasco, L. F. Barrientos, K. Pichara, T. Anguita, D. N. A. Murphy, D. G. Gilbank, M. D. Gladders, H. K. C. Yee, B. C. Hsieh, and S. López, "Photometric classification of quasars from RCS-2 using random forest," *Astron. Astrophys.*, vol. 584, p. A44, Nov. 2015, doi: 10.1051/0004-6361/201525752.

[18] H. Klimczak, W. Kotłowski, D. Oszkiewicz, F. DeMeo, A. Kryszczyńska, E. Wilawer, and B. Carry, "Predicting asteroid types: Importance of individual and combined features," *Frontiers Astron. Space Sci.*, vol. 8, p. 216, Dec. 2021, doi: 10.3389/fspas.2021.767885.

[19] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202, doi: 10.1098/rsta.2015.0202.

[20] M. D. Perez, M. A. M. Ali, A. G. Sanchez, E. Ghorbel, K. Al Ismaeil, P. Le Henaff, and D. Aouada, "Detection & identification of on-orbit objects using machine learning," in *Proc. 8th Eur. Conf. Space Debris*, vol. 8, no. 1, Darmstadt, Germany, 2021, pp. 1–10.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[22] H. A. Dung, B. Chen, and T.-J. Chin, "A spacecraft dataset for detection, segmentation and parts recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 2012–2019, doi: 10.1109/CVPRW53098.2021.00229.

[23] P. F. Proenca and Y. Gao, "Deep learning for spacecraft pose estimation from photorealistic rendering," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 6007–6013, doi: 10.1109/ICRA40945.2020.9197244.

[24] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Martens, and S. D'Amico, "Satellite pose estimation challenge: Dataset, competition design, and results," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 4083–4098, Oct. 2020, doi: 10.1109/TAES.2020.2989063.

[25] S. Sharma, T. H. Park, and S. D'Amico, "Spacecraft pose estimation dataset (SPEED)," Stanford Digit. Repository, Stanford Univ., Stanford, CA, USA, Tech. Rep., 2019, doi: 10.25740/dz692fn7184.

[26] M. A. Musallam, K. Al Ismaeil, O. Oyedotun, M. D. Perez, M. Poucet, and D. Aouada, "SPARK: SPAcecraft recognition leveraging knowledge of space environment," 2021, *arXiv:2104.05978*.

[27] R. Afshar and S. Lu, "Classification and recognition of space debris and its pose estimation based on deep learning of CNNs," in *HCI International 2020—Posters* (Communications in Computer and Information Science). Cham, Switzerland: Springer, 2020, pp. 605–613, doi: 10.1007/978-3-030-50726-8_79.

[28] A. C. Rabeendran and L. Denneau, "A two-stage deep learning detection classifier for the ATLAS asteroid survey," *Publications Astronomical Soc. Pacific*, vol. 133, no. 1021, Feb. 2021, Art. no. 034501, doi: 10.1088/1538-3873/abc900.

[29] P. Jia, Q. Liu, and Y. Sun, "Detection and classification of astronomical targets with deep neural networks in wide-field small aperture telescopes," *Astronomical J.*, vol. 159, no. 5, p. 212, Apr. 2020, doi: 10.3847/1538-3881/ab800a.

[30] J. Xi, Y. Xiang, O. K. Ersoy, M. Cong, X. Wei, and J. Gu, "Space debris detection using feature learning of candidate regions in optical image sequences," *IEEE Access*, vol. 8, pp. 150864–150877, 2020, doi: 10.1109/ACCESS.2020.3016761.

[31] W. Liu, H. Xiao, and B. Chengchao, "Spatial multi-object recognition based on deep learning," in *Proc. IEEE Int. Conf. Unmanned Syst. (ICUS)*, Oct. 2019, pp. 736–741, doi: 10.1109/ICUS48101.2019.8995980.

[32] M. Gao, J. Jiang, G. Zou, V. John, and Z. Liu, "RGB-D-based object recognition using multimodal convolutional neural networks: A survey," *IEEE Access*, vol. 7, pp. 43110–43136, 2019, doi: 10.1109/ACCESS.2019.2907071.

[33] N. AlDahoul, H. A. Karim, and M. A. Momo, "RGB-D based multimodal convolutional neural networks for spacecraft recognition," in *Proc. IEEE Int. Conf. Image Process. Challenges (ICIPC)*, Sep. 2021, pp. 1–5, doi: 10.1109/ICIPC53495.2021.9620192.

[34] N. AlDahoul, H. A. Karim, and M. A. Momo, "RGB-D based multi-modal deep learning for spacecraft and debris recognition," *Sci. Rep.*, vol. 12, no. 1, pp. 1–18, Mar. 2022, doi: 10.1038/s41598-022-07846-5.

[35] H. Kaur, D. Koundal, and V. Kadyan, "Image fusion techniques: A survey," *Arch. Comput. Methods Eng.*, vol. 28, no. 7, pp. 4425–4447, Jan. 2021, doi: 10.1007/s11831-021-09540-7.

[36] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," 2018, *arXiv:1804.06992*.

[37] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, May 2018, Art. no. 2783902, doi: 10.1109/TGRS.2017.2783902.

[38] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, Mar. 2019, doi: 10.1109/LSP.2019.2895749.

[39] G. Sun, H. Huang, A. Zhang, F. Li, H. Zhao, and H. Fu, "Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images," *Remote Sens.*, vol. 11, no. 3, p. 227, Jan. 2019, doi: 10.3390/rs11030227.

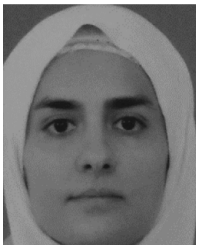
[40] Z. Duan, T. Zhang, J. Tan, and X. Luo, "Non-local multi-focus image fusion with recurrent neural networks," *IEEE Access*, vol. 8, pp. 135284–135295, 2020, doi: 10.1109/ACCESS.2020.3010542.

[41] B. Xiao, B. Xu, X. Bi, and W. Li, "Global-feature encoding U-Net (GEU-Net) for multi-focus image fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 163–175, 2021, doi: 10.1109/TIP.2020.3033158.

[42] T. Pan, J. Jiang, J. Yao, B. Wang, and B. Tan, "A novel multi-focus image fusion network with U-shape structure," *Sensors*, vol. 20, no. 14, p. 3901, Jul. 2020, doi: 10.3390/s20143901.

[43] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019, doi: 10.1016/j.inffus.2018.09.004.

[44] Z. Yang, Y. Chen, Z. Le, and Y. Ma, "GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 6133–6145, Jun. 2021, doi: 10.1007/s00521-020-05387-4.
[45] C. Yuan, C. Q. Sun, X. Y. Tang, and R. F. Liu, "FLGC-fusion GAN: An enhanced fusion GAN model by importing fully learnable group convolution," *Math. Problems Eng.*, vol. 2020, Art. no. 6384831, Oct. 2020, doi: 10.1155/2020/6384831.
[46] X. Zhang, P. Ye, and G. Xiao, "VIFB: A visible and infrared image fusion benchmark," 2020, *arXiv:2002.03322*.
[47] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.
[48] M. Kearns, "Thoughts on hypothesis boosting," Tech. Rep., 1988.
[49] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/s0893-6080(05)80023-1.
[50] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying convolution and attention for all data sizes," 2021, *arXiv:2106.04803*.
[51] *SPARK Challenge*. Accessed: Nov. 7, 2022. [Online]. Available: https://www.2021.ieeeicip.org/www.2021.ieeeicip.org/ChallengeSessions.html
[52] *SPARK Sponser*. Accessed: Nov. 7, 2022. [Online]. Available: https://www.lmo.space
[53] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.
[54] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, May 2018, Art. no. 1850018.
[55] D. P. Bavirisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen–Loeve transform," *IEEE Sensors J.*, vol. 16, no. 1, pp. 203–209, Jan. 2016.
[56] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image Video Process.*, vol. 9, no. 5, pp. 1193–1204, 2015.
[57] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–9.
[58] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Appl. Opt.*, vol. 55, no. 23, pp. 6480–6490, Aug. 2016.
[59] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[60] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Inf. Fusion*, vol. 30, pp. 15–26, Jul. 2016.
[61] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visual image fusion through infrared feature extraction and visual information preservation," *Infr. Phys. Technol.*, vol. 83, pp. 227–237, Jun. 2017.
[62] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*.
[63] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, "Multi-scale guided image and video fusion: A fast and efficient approach," *Circuits, Syst., Signal Process.*, vol. 38, no. 12, pp. 5576–5605, Dec. 2019.
[64] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.
[65] V. P. S. Naidu, "Image fusion technique using multi-resolution singular value decomposition," *Defence Sci. J.*, vol. 61, no. 5, pp. 479–484, 2011.
[66] H. Li, X. J. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infr. Phys. Technol.*, vol. 102, Mar. 2019, Art. no. 103039.
[67] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infr. Phys. Technol.*, vol. 76, pp. 52–64, May 2016.
[68] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.
[69] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.
[70] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Inf. Fusion*, vol. 8, no. 2, pp. 193–207, Apr. 2007.
[71] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Opt. Commun.*, vol. 341, pp. 199–209, Apr. 2015.
[72] B. Rajalingam and R. Priya, "Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis," *Int. J. Eng. Sci. Invention*, vol. 2, pp. 52–60, Jan. 2018.
[73] Y.-J. Rao, "In-fibre Bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, p. 355, 1997.
[74] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.
[75] C. S. Xydeas and P. V. V., "Objective image fusion performance measure," *Mil. Tech. Courier*, vol. 36, no. 4, pp. 308–309, 2000.
[76] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Proc.*, vol. 4, pp. 133–142, Jan. 2015.
[77] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[78] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.
[79] D. M. Bulanon, T. F. Burks, and V. Alchanatis, "Image fusion of visible and thermal images for fruit detection," *Biosyst. Eng.*, vol. 103, no. 1, pp. 12–22, 2009.
[80] J. Van Aardt, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, May 2008, Art. no. 023522.
[81] N. AlDahoul, H. A. Karim, A. De Castro, and M. J. T. Tan, "Localization and classification of space objects using efficientdet detector for space situational awareness," *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, Dec. 2022.
[82] M. A. Hearst, B. Scholkopf, S. Dumais, E. Osuna, and J. Platt, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.
[83] 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. (CVPR)*, Jun. 2009, pp. 248–255.



NOUAR ALDAHOUL received the B.Eng. and M.Eng. degrees in computer engineering from Damascus University, in 2008 and 2012, respectively, and the Ph.D. degree in machine learning from International Islamic University Malaysia, in 2019. She is currently a Researcher with the Faculty of Engineering, Multimedia University, Malaysia. She is also a Postdoctoral Associate with New York University Abu Dhabi. Her research interests include the area of deep learning, computer vision, and the Internet of Things. She was a recipient of several awards, such as the ICIP2020 Challenge Award and the Best Conference Paper Award.

HEZERUL ABDUL KARIM (Senior Member, IEEE) received the B.Eng. degree in electronics with communications from the University of Wales Swansea, U.K., in 1998, the M.Eng. degree in science from Multimedia University, Malaysia, in 2003, and the Ph.D. degree from the University of Surrey, U.K., in 2008. He is currently an Associate Professor with the Faculty of Engineering, Multimedia University. His research interests include telemetry, error resilience and multiple description video coding for 2D/3D image/video coding and transmission, and content-based image/video recognition. He is currently serving as a Treasurer in the IEEE Signal Processing Society Malaysia Chapter.

MHD ADEL MOMO received the B.Eng. degree in software engineering from Yarmouk Private University, in 2020. He is currently working as a Research and Development Software Engineer at FMS Tech. His research interests include deep learning, computer vision, natural language processing, and edge AI. He was a recipient of several awards, such as the ICIP2020 Challenge, ICIP2021 Challenge, and Malaysia Technology Expo 2022.

FRANCESCA ISABELLE FLORES ESCOBARA received the B.Sc. degree in biology from the University of Saint La Salle, Bacolod, Philippines, in 2022. Her research interests include computer vision, deep learning, and natural sciences.

MYLES JOSHUA TOLEDO TAN (Member, IEEE) was born in Bacolod, Philippines, in 1996. He received the B.S. degree (summa cum laude) in biomedical engineering from the University at Buffalo, The State University of New York, in 2017, and the M.S. degree in applied biomedical engineering from Johns Hopkins University, Baltimore, MD, USA, in 2018. He has been an Assistant Professor in chemical engineering with the University of Saint La Salle (USLS), since 2018, where he was also appointed as an Assistant Professor in natural sciences, in 2020. He has been actively involved in the education and training of students at the Department of Electronics Engineering and the Department of Electrical Engineering, USLS. He also leads the Tan Research Group. His research interests include biomedical signal processing, medical imaging, deep learning, and engineering and mathematics education. He is a member of the Institute of Physics, U.K., and Tau Beta Pi—The Engineering Honor Society (USA) and an Associate Member of the Institute of Mathematics and Its Applications, U.K. He was a recipient of the Tau Beta Pi Engineering Honor Society Record Scholarship and the Grace Capen Award.