FUSION OF SAR AND OPTICAL IMAGES USING PIXEL-BASED CNN

S.R. Bandi[*], *M. Anbarasan*[†], *D. Sheela*[‡]

Abstract: Sensors of different wavelengths in remote sensing field capture data. Each and every sensor has its own capabilities and limitations. Synthetic aperture radar (SAR) collects data that has a high spatial and radiometric resolution. The optical remote sensors capture images with good spectral information. Fused images from these sensors will have high information when implemented with a better algorithm resulting in the proper collection of data to predict weather forecasting, soil exploration, and crop classification. This work encompasses a fusion of optical and radar data of Sentinel series satellites using a deep learning-based convolutional neural network (CNN). The three-fold work of the image fusion approach is performed in CNN as layered architecture covering the image transform in the convolutional layer, followed by the activity level measurement in the max pooling layer. Finally, the decision-making is performed in the fully connected layer. The objective of the work is to show that the proposed deep learning-based CNN fusion approach overcomes some of the difficulties in the traditional image fusion approaches. To show the performance of the CNN-based image fusion, a good number of image quality assessment metrics are analyzed. The consequences demonstrate that the integration of spatial and spectral information is numerically evident in the output image and has high robustness. Finally, the objective assessment results outperform the state-of-the-art fusion methodologies.

Key words: *deep learning, image fusion, optical data, synthetic aperture radar, quality metrics*

Received: June 10, 2020
Revised and accepted: August 30, 2022

DOI: 10.14311/NNW.2022.32.012

1. Introduction

Image fusion is an important image processing procedure to augment various properties of image visual perceptions obtained from two or more images. There are

[*]Sudheer Reddy Bandi – Corresponding author; Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai-600069, Tamil Nadu, India, E-mail: sudheer653@gmail.com

[†]M. Anbarasan; Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai-600069, Tamil Nadu, India, E-mail: anbarasan.cse@gmail.com

[‡]D. Sheela; Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai-602105, Tamil Nadu, India, E-mail: sheela_rajesh@rediffmail.com

some good works of image fusion in various fields which includes health care [1], surveillance [2], and many more applications. In the same way, the image fusion or image integration approaches essentially improve the capability of remote sensing applications by enhancing the detail of fused images required in various applications of remote sensing like earth observations, ocean monitoring [15], target detection [28], controlling the emergencies like flood control [29], active fire detection [30] in forests, soil moisture detection [31], texture detection to differentiate vegetation and forest areas, and classification of urban areas [32]. Remote sensors are typical devices that take the energy from the earth's surface in the form of signals and convert them into a human-readable form in the form of images.

There are some drawbacks in both radar-based and optical-based remote sensing image information guiding the erroneous image interpretation mechanism. The following observations from the literature [33] list the drawbacks of both SAR and optical images. SAR image leads to speckle noise whereas Optical remote sensing data don't. Optical remote sensing gives high spectral information whereas SAR data gives a high spatial resolution. Optical adherent devices are helpless to haze and unpleasant climate conditions, which constrain the awareness of the Earth's surface affecting the optical remote sensing. On the contrary, synthetic aperture radar is self-sufficient in solar radiances and variable climatic conditions, and it is capable to afford beneficial imagery within a considerable period of time than optical sensors as demonstrated in [44, 45]. As optical remote sensing depends on solar illuminations, the image gives reflective and emissive characteristics of the objects, while the microwave remote sensing image gives information on spatial distribution, surface coarseness, and dielectric properties of the objects present on the earth's surface. To overcome the disadvantages mentioned above, the current methodology advances to generate balancing images from the integration of SAR and multispectral images.

CNN in computer vision mainly focuses on biometric recognition [3], change detection in remote sensing data [4], object detection and human behavior tracking [5], subject classification [6], and medical image analysis [7] fields. The work of the microwave and optical image fusion is restricted to the following: In [8], SAR and optical images were fused using Atrous wavelet transform and applied fusion rule in pixel to obtain the high spatial image missing boundary characteristics covering the regional properties. The major hindrance to the approach was not considering the pixels, leaving the fusion process at the edges. Decision-level image fusion using feature normalization was adopted by [9] in fusing the SAR and optical images of land use land cover classification but found difficulty in the fusion results of water bodies and shaded areas. the applied wavelet transform-based fusion approach concentrates actively on the spatial details of the SAR image and avoided the spectral analysis of the image. Another major complexity of the above algorithm was the coefficient to be considered for the wavelet transform. The effort in [11] suggested a block regression-based fusion approach for SAR and optical images reducing computation time but the evaluation criteria to judge the image quality was deficient. CNN has shown tremendous improvements in the fusion of multi-spectral and hyperspectral image fusion given in [12]. Further, the analysis based on the recent survey [34–36] suggests that the work proposed here is the origination of radar and multispectral image merging from different scene classes.

The few works of neural networks-based image integration aimed at multispectral and hyperspectral are given as follows: CNN in image fusion was proposed in [12] to fuse Landsat and MODIS images. The work has overcome some of the shortcomings like spatial and temporal dynamics but worked on fewer data. Some of the shortcomings discussed above have been overcome in the proposed work such as qualitative measurement of the image through pixel-by-pixel analysis and reducing computation time with the help of reduced kernel size.

2. Methodology of the proposed work

2.1 Overview

Synthetic aperture radar and optical images are taken from sentinel series of satellites. The microwave remote sensing images contain speckle noise that occurred due to the radar receiving data from multiple targets. The images are filtered with the advanced version of local means-based and patch-based filters adapted from [13,14]. The optical satellite images are converted to grayscale and enhanced with histogram equalization. The next step is to fuse the images with the convolutional neural network from the work obtained from [15] to obtain the focus map, binary segmented map refined with guided filtering [16], and fused map images in the subsequent steps of this work.

2.2 Preprocessing

More frequently raw remotely sensed images contain flaws or deficiencies and correction is required for prior processing. Therefore pre-processing is considered as a preparatory segment to improve the quality of the image from undesirable atmospheric interference, system noise, sensor motion, etc. For the current methodology, SAR and optical images of size 256×256 are considered for image fusion. Frequently, satellite images are degraded with the aid of noise during the process of image procurement and transmission technique. The most important motive of the noise reduction methods in radar images is to put off speckle noise adopted from [13,14] through the observance of essential characteristics of the images. The short notations for different classes used in the experiments are given as Desert-Dt, Grassland-Gl, Harbour-Hr, Rain slicks-Rs, Residential-Rd, Snowland-Sl, and Vegetation-Vg. The different scene classes of the remote sensing dataset considered for fusion are given in Fig. 1. From the definition, the pixels have been widespread leading to the higher spatial resolution in SAR images, whereas they are narrow for optical sensor images.

3. CNN for image fusion: background

Convolutional neural network (CNN) is derived from traditional feed-forward neural networks which are a specific type of artificial neural network (ANN) architecture. The main resolution of CNN is to extract low-level and high-level depth information by learning the essential hierarchical features with numerous generalization levels or (levels of abstraction) [17]. Some of the most popular applications

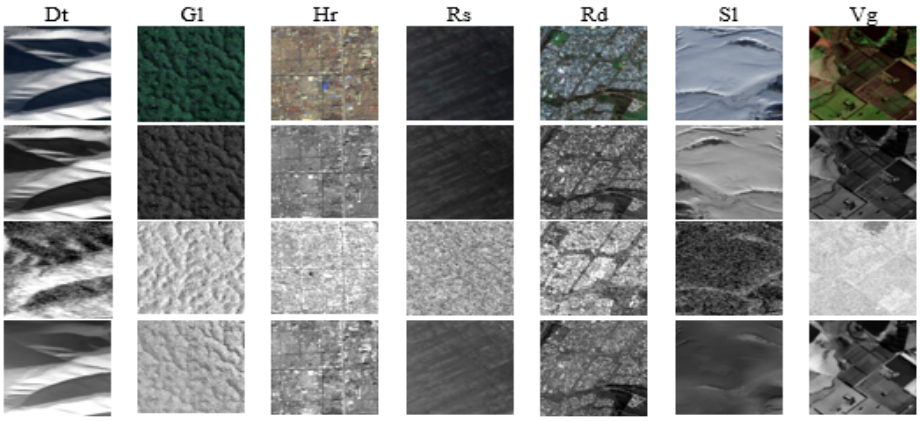


Fig. 1 *Row wise: (a) Optical images (b) Optical grayscale images (c) SAR speckle noise images (d) SAR speckle noise-free images.*

of CNN include image classification [37], video surveillance [38], exploring patterns from satellite images [39], target detection [40], medical applications [41], etc., CNN has proven very effective in the area of remote sensing-based image fusion. The procedure of Image integration involves the incorporation of two or more classes of data to obtain a composite image of high quality that helps in the consequent applications of computer vision like classification, recognition, and detection. The classes may be different sensors as in medical (CT and MRI), remote sensing (optical, thermal, microwave), multi-focus, and multi-temporal indicating different times. The success of the fusion method is mainly dependent on the image information extraction involving image transform, activity levels of the obtained image transform, and appropriate fusion principles. The performance of the image fusion is primarily reliant on the fusion result which involves spatial consistency, effectively representing features, and removal of artefacts and noise. For example, combining greater spatial statistics (SAR) in one band with greater spectral data (optical) in any other dataset to create 'synthetic' higher resolution multispectral datasets (fused images) eliminates the noise and reduces the conflicts between the radar-dependent spatial and optical-based spectral resolutions.

Furthermore, each feature map neuron is connected to the previous neurons of the neighboring feature maps. The neighboring future map in the former layer is referred to as the receptive field. The new feature maps are generated by taking the input of the learned filter and then applying the non-linear activation function as element-wise multiplication. To generate the remaining feature maps the spatial location in the input images is shared by different types of kernels. The derivation of feature pixel value at (x, y) location in the z-th feature map of n-th layer, $F^n_{x,y,z}$ is given by:

$$F^n_{x,y,z} = {W^n_z}^{\mathrm{T}} a^n_{x,y} + b^n_z, \tag{1}$$

where W^n_z and b^n_z are the weights with vector values and bias term of z-th filter of the n-th layer respectively.

The activation function presents a non-linear transformation of the network nodes to identify the non-linear features. Let A(·) represent the non-linear activation function. The activation value $A^n_{x,y,z}$ of convolution feature $F^n_{x,y,z}$ can be computed as:

$$A^n_{x,y,z} = A\left(F^n_{x,y,z}\right). \tag{2}$$

The most popular type of activation functions used in convolution layers is sigmoid, tanh, and ReLU functions. The second layer is known as the pooling layer. Max-pooling and average pooling are the two well-known operations in CNN for spatial input dimension. The pooling layer will help to downsize the convolved feature map generated from the previous layer of convolution. Generally, pooling layers are arranged in between the two layers of convolution or after/before the normalization. Let P(·) denote the pooling function. For each feature map $A^n_{:,:,z}$ the pooling function is updated as

$$Y^n_{x,y,z} = P\left(A^n_{m,n,z}\right), \forall\,(m,n) \in \mathcal{R}_{x,y}, \tag{3}$$

where $\mathcal{R}_{x,y}$ is a local neighborhood pixel values around the location (m,n). The various types of spatial pooling operations available are average pooling and max pooling. At first convolution layers kernels are used to detect low-level attributes like edges, shapes, curves, lines, etc., moving on to the next layer higher level features are learned by encoding more abstract features of hidden neurons. After a sequential batch of convolution and pooling layers, there may be a fully-connected layer with one or more layers. This development to form a single layer is called flattening. Finally, the flattened matrix with a single column is directed through a fully connected layer to classify the images. Therefore, the above operations (convolution, activation function, and pooling) furnish the steps of deep CNN by learning an extracted feature to escalate the correctness of fusion while measuring the dependency between SAR and optical features.

3.1 CNN model for SAR optical fusion

The detailed CNN model of SAR optical fusion is given in Fig. 2. The convolution layer uses small-sized patches to operate on images. Generally, a patch is a small

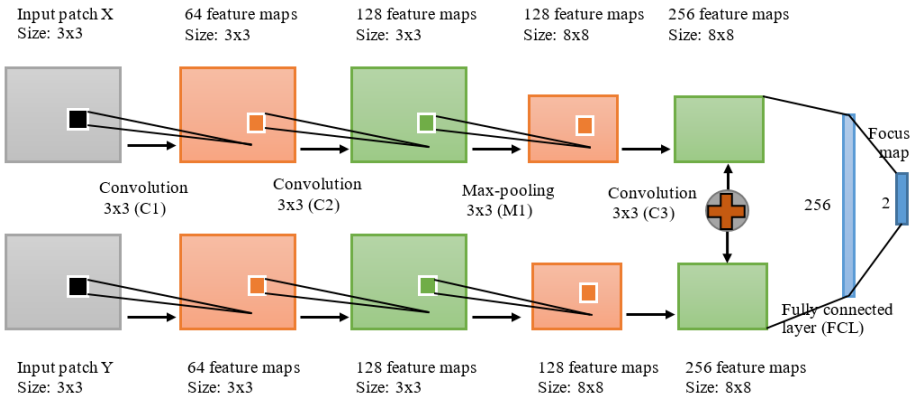


Fig. 2 *CNN model of SAR optical fusion.*

rectangular-sized image pixel or it can also be a part of an input image [18]. This paper uses a 3×3 patch (square of pixels) containing 256×256 size as the input image. Due to the diminished dimension, some of the computer vision algorithms such as denoising, super-resolution, etc. are less complicated to function on patches instead of working on the whole image itself. The process of CNN-based fusion can be given as follows in detail:

Step 1, the filters and patch sizes required for each and every layer are initialized in this step. The parameters are given in Tab. I.

Step 2, the CNN model has three convolutional layers acting as an image transform approach in which the first layer has a patch size of 3×3. This layer takes the grayscale images of radar and optical remote sensing images as input with the (C1) as input mentioned in Tab. I. The convolution process occurs in the first CNN layer for the two image inputs and the output for the final filter (64) is obtained for all the respective classes.

Step 3, the image transform from step 2 is taken as input to step 3 and the convolution process is applied with parameters given in (C2) in Tab. I. This layer takes the grayscale images of radar and optical remote sensing images as input with the (C1) as input mentioned in Tab. I. The feature selection in this layer is established on the optimal value i.e maximization of the filters, this type of feature selection improves the resolution of the image as well. The output for the last filter 128 for both SAR and optical images is obtained for all the classes. The work done in step 2 and step 3 are the same i.e. convolution is applied for all the radar and optical images with 64 filters for each patch with a size of 3×3 and 128 filters with a patch size of 3×3 respectively.

Step 4, the feature map obtained in step 3 acts as an input to the max pooling layer in step 4. Now, the 128 feature maps are down-sampled using a max pooling layer $M1$ of size 2×2 and with a stride of 2. The max pooling step is an important operation in the current work to perform the activity level measurement of the process. The maximum of the feature map is considered for every path size 3×3 in both satellite images. The output for the last filter 128 for both SAR and optical images is obtained.

Step 5, the third convolutional layer (C3) provides 256 filters of size 128×9×256. Similar to layer 1 and layer 2 convolution process is applied to input images from the max pooling layer. In addition to the convolution process, data is concatenated

Image Type	Layers (L)	Filter (f)	Feature maps	Pooling
SAR image &	C1	3×3×64/1	64	none
optical image	C2	64×9×128/1	128	none
Size: 256×256	M1	128×9×256/2	128	Max-pooling
	C3	128×9×256/1	256	none
Fusion	FCL	512×64×256	256	none
Fusion	FCL	256×2×2	2	none
Focus Map	FCL	1×1×1	1	none

Tab. I *Hyper parameters of each layer.*

Algorithm 1 CNN fusion.

Input: Two batch images of size 256×256 with patch size as 16×16
Output: Two classes of fused images
Procedure: Considered SAR and optical grayscale images for fusion as img1 and img2.
Basic steps: (pre-processing)

1. Compare the size (H×W) of two images
2. If both the sizes are equal then pass the images to CNN fusion

CNN layer steps:

1. Check the type of images (RGB/Grayscale. If type RGB is present convert it to grayscale.
2. Load the CNN model
3. Define convolution layers by passing the number of patchsize (p), filters (f), and size as weights (w) to networks.
   (a) For first convolution layer (C1) set p1=3×3, f1=64, w1=9×64
   (b) For the second convolution layer (C2) set channel (c) as the first dimension then set p and f, c2=64 (i.e. C1 filter), p2=9 (3×3), f2=128, w2=64×9×128
   (c) Define pooling layer as max pooling after C2 layer with kernel size as 3×3 and s=1
   (d) For third convolution layer (C3) set c3=128, p3=9 (3×3), f3=256, w3=128×9×256
   (e) For fourth convolution layer (C4) set c4=256, p3=64, f4=256, w4=256×64×256
   (f) For last convolution layer (C5) set c5=256, p5=1, f5=2, w5=256×1×2
4. The final convolution layer (C5) is passed to a softmax function which generates a probability of two fusion classes.

Fusion steps:

1. Generate focus map (pixel-level map)
2. Focus map is segmented into a binary map with a threshold set as 0.5
3. Refine the binary segmented map with two verification strategies
   (a) Small region removal
   (b) Guided image filtering
4. Generate final decision map
5. Display the fused image with final decision map using pixel-wise weighted-average strategy

i.e., the fusion rule involving the max rule strategy is applied to fuse the data and acquire the non-refined fused image.

Step 6, the fourth and fifth convolutional layers are similar to the previous layer with weights given as (C4) and (C5) in Tab. I. In the fourth layer, for the

last time, all 256 filters are applied and then reduced to two filters in the fifth layer as the number of images is from two different sensors. The work from (C5) to the last stage is the decision level in the fusion stage.

Step 7, the probability distribution over two fusion classes is given by the softmax layer. In this stage of the image fusion process, the neural network system takes two source images of random pixel size as a whole input to produce a dense score map. The two source images are of size H × W, the size of the output score map is $(\lceil H2/2\rceil - 8 + 1) \times (\lceil W/2\rceil - 8 + 1)$. Where $\lceil H/2\rceil$ and $\lceil W/2\rceil$ denotes the ceiling operation. Each coefficient in the score map preserves the output score of a pair of source image patches of size 3×3 going forward through the network. The output of the softmax layer is obtained.

Step 8, the focus map is generated from the input of the softmax layer. If the coefficients are closer in both sources, the quantities are covered with a stride of two pixels with a patch size of 16. Now, the binary segmented image is obtained with a threshold of 0.5 considering the effects of both sources of images. The pixels are in reverse order in a binary segmented map with the high spatial value being dark, and the less spatial value being bright. Unwanted boundaries and artefacts are reduced to obtain the initial decision map. The obtained map is filtered with a guided filter to reserve the contours and obtain the initial fused map and final decision maps. Finally, at the last stage of the proposed fusion, the average weighted fusion rule is applied to all the pixels to obtain the resultant image. All these steps are given in Algorithm 1.

4. Experimental results and analysis

This segment is divided into three sub-categories involving experimental setup describing the dataset and the system settings to run the work in the first. In the second sub-section, experimental results are briefly discussed. Finally, in the last section the comparison of results with other standard approaches.

4.1 Dataset preparation

The SAR and optical data are provided from the SENTINEL series as a part of Copernicus free data service. The first satellite launched by the European space agency as a part of the Copernicus programme satellite constellation is SENTINEL-1. Sentinel-1 is a radar satellite providing data at C-band in all weather conditions. The spatial resolution of the sensor is 5 m with a frequency of 5.405 GHz. The multispectral satellite of the Copernicus Programme satellite constellation is SENTINEL-2 with a spatial resolution of 10 m to 60 m. The images are mainly covering the areas of Europe in different seasons present in [19] is freely accessible. The image scene classes are biological slicks, desert, grassland, harbour, rain slicks, residential area, seashore, sea ice, snow land, and vegetation area. For every categorical class 3000 images are taken in SAR, and 3000 in optical resulting in a total of 30000 SAR images and 30000 optical images.

4.2 Comparison of CNN-based image fusion with traditional and conventional fusion methods

The fusion methods ranging from traditional methods like spatial-based, multi-scale transform, and sparse representation to conventional approaches like pulse code neural network approaches are compared to analyse the performance and efficiency of the modern-day deep learning-based CNNs. The fusion techniques used for comparison based on wavelet transforms are discrete wavelet transform (DWT) [20], curvelet transform [21], and laplacian pyramid image fusion [22] based on multiresolution decomposition, edge-based gradient image fusion (GRD) [23], non-subsampled contourlet transform (NSCT) based on saliency of an image [24], sparse representation (SR) based image fusion [25], fusion of Laplacian pyramid and sparse representation, and parameter adaptive pulse coupled neural network (PCN) [26]. The limitations of transform-based approaches are the convolution calculations, decomposition levels, and less spatial resolution in the resultant image. The disadvantages of spatial-based techniques especially in remote sensing-based fusion are spectral distortion resulting in more noise level and may introduce artefacts in the resultant image. Gradient-based fusion techniques concentrate more on edges than on texture with the given inputs in which sparse inputs are not accepted directly [41]. Additionally, edge efficiency is not more accurate with more interruption at all levels of methodology [42, 43]. The shortcomings of contourlet transformation are time-consuming, shift-invariant, and cannot be applied to complex structures. The drawbacks of sparse representation are the selection of fusion rule and dictionary construction.

Activity level measurement, image transform, and fusion rule are the three important steps in most image fusion approaches. The primary objective of the proposed work is to reduce the limitations in the image fusion approach with the assistance of a deep learning-based CNN image fusion approach. Generally, the classification and fusion problems in computer vision are divided into feature extraction-activity level measurement, feature scaling-image transform, and prediction-fusion rule. All three steps can be done through CNN layers pixel by pixel, overcoming the drawbacks of traditional image fusion approaches like a fusion of edges, homogeneous regions, and noise levels. The result of the image fusion process can be assessed by the quality of the image. The measurement of image quality is given as an image quality assessment. The unbiased image quality metrics are mainly classified as i) reference quality metrics, ii) no reference quality metrics, and iii) reduced reference quality metrics. In this study, the current methodology mainly implements the first two types of quality metrics. The reference quality metrics take both the original and resultant image to find the score. In no reference type of metric, the strength of only the fused image is utilized to find the quality score.

4.3 Inference from the reference-quality assessment metrics

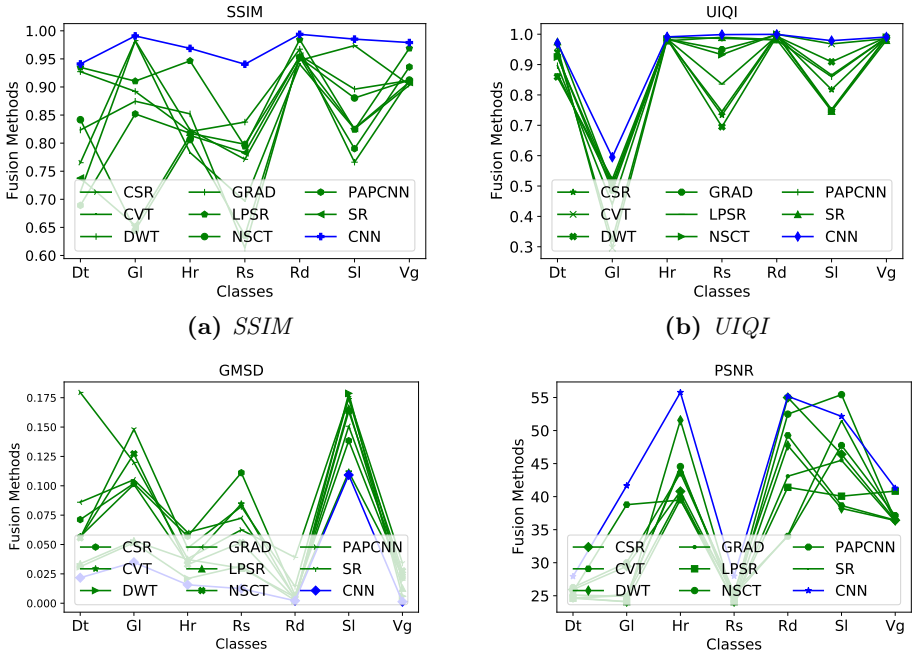
The fused results are compared with different types of image fusion techniques available such as traditional, conventional, multiscale transform, sparse-based representation, and multiresolution, etc. The various reference quality metrics used are structure similarity index measurement (SSIM), universal image quality index (UIQI), gradient magnitude similarity deviation (GMSD), and peak signal to noise

ratio (PSNR). The values are tabulated in Tab. II. SSIM is a full reference metric as the result depends completely on the reference image. The value of SSIM depends on PSNR and MSR (mean square error). The structure similarity is given by the statistical value of SSIM. PSNR is an extensively used quality metric to measure the image. This metric is calculated by finding the number of grey levels in the image distributed by the matching pixels in both the reference and the fused im-

Method	CSR	CVT	DWT	GRD	LSR	NCT	PCN	SR	CNN
Class					SSIM				
Dt	0.7652	0.7134	0.9273	0.8234	0.9349	0.8417	0.6892	0.7377	0.941
Gl	0.9826	0.9818	0.8921	0.8746	0.9104	0.645	0.852	0.6536	0.9909
Hr	0.8237	0.7832	0.8206	0.8521	0.9465	0.8064	0.8175	0.8138	0.9687
Rs	0.7711	0.6969	0.8375	0.6134	0.7942	0.6357	0.798	0.783	0.9405
Rd	0.9474	0.9405	0.9673	0.9563	0.9843	0.9522	0.9535	0.9548	0.9938
Sl	0.9737	0.8254	0.7661	0.8961	0.8242	0.8804	0.7902	0.8263	0.9852
Vg	0.9037	0.9031	0.9121	0.9117	0.9688	0.9123	0.9356	0.9084	0.9791
Class					UIQI				
Dt	0.9427	0.9568	0.8592	0.9284	0.8691	0.925	0.8964	0.9759	0.9704
Gl	0.3229	0.2961	0.5221	0.4859	0.5018	0.5081	0.4395	0.5026	0.595
Hr	0.986	0.9821	0.9853	0.9842	0.9844	0.9845	0.9896	0.9784	0.9912
Rs	0.7336	0.7456	0.6951	0.9495	0.8357	0.9327	0.9863	0.9897	0.9987
Rd	0.9971	0.9997	0.993	0.9973	0.9935	0.9997	0.9811	0.9823	0.9995
Sl	0.8176	0.9681	0.909	0.7483	0.8656	0.7521	0.86	0.7467	0.9788
Vg	0.9894	0.9853	0.99	0.9898	0.9884	0.99	0.9894	0.98	0.9907
Class					GMSD				
Dt	0.0712	0.0568	0.0311	0.086	0.0336	0.0557	0.1795	0.0557	0.0217
Gl	0.102	0.1012	0.0515	0.1052	0.0532	0.1272	0.1199	0.1482	0.0349
Hr	0.057	0.0341	0.021	0.0603	0.0372	0.0329	0.0379	0.0469	0.0157
Rs	0.1111	0.0846	0.0316	0.0724	0.0298	0.0517	0.0625	0.0823	0.0124
Rd	0.0067	0.0044	0.0024	0.0052	0.0051	0.004	0.039	0.0142	0.0021
Sl	0.1384	0.1118	0.1787	0.1507	0.166	0.1634	0.1743	0.175	0.1092
Vg	0.022	0.0237	0.006	0.0252	0.0128	0.0276	0.0195	0.0342	0.0013
Class					PSNR				
Dt	25.936	25.619	24.702	24.661	24.62	25.112	26.259	24.643	27.926
Gl	29.46	38.784	24.099	25.02	24.099	24.913	30.005	25.178	41.659
Hr	40.799	39.482	51.559	40.161	39.576	44.551	43.562	44.166	55.771
Rs	24.901	25.59	24.102	24.168	24.099	24.159	25.545	25.032	27.972
Rd	54.996	49.264	47.733	43.087	41.41	52.484	34	34.119	55.191
Sl	46.402	38.651	38.219	45.503	40.068	55.437	47.757	51.496	52.145
Vg	36.431	36.439	36.385	36.254	40.836	36.463	37.144	36.373	41.29

Tab. II *Mathematical values of the reference-quality assessment metrics for various fusion methods adopted for all the classes.*

ages. The high value of PSNR indicates both images are similar. Universal image quality index mainly depends on the structural distortion of the reference image. The three distortion factors which impact the quality of the fused image in UIQI are given as contrast, luminance and correlation. The value ranges from −1 to +1. A value near 1 gives the best quality of the complementary image. The deviation of the reference image and resultant image in the edges are measured by GMSD. The higher value indicates the higher distortion in the images and gives the lower image perceptual quality. The statistical value in the cell of SSIM represents the mean similarity of a single class from the 3000 images of the merged image with the mean value of the SAR image from 3000 images. The PSNR mean value for an individual fusion methodology is represented concerning the fused image and SAR image in the cell, likewise, the values for the other metrics UIQI and GMSD are denoted in the cell. The acquired mathematical results are compared with various fusion methods and the results are visualized in Fig. 3. Fig. 3, gives a clear picture that the blue line indicates the proposed work values which are better when compared with the values of other fusion approaches indicated in green lines.



(a) *SSIM* **(b)** *UIQI*

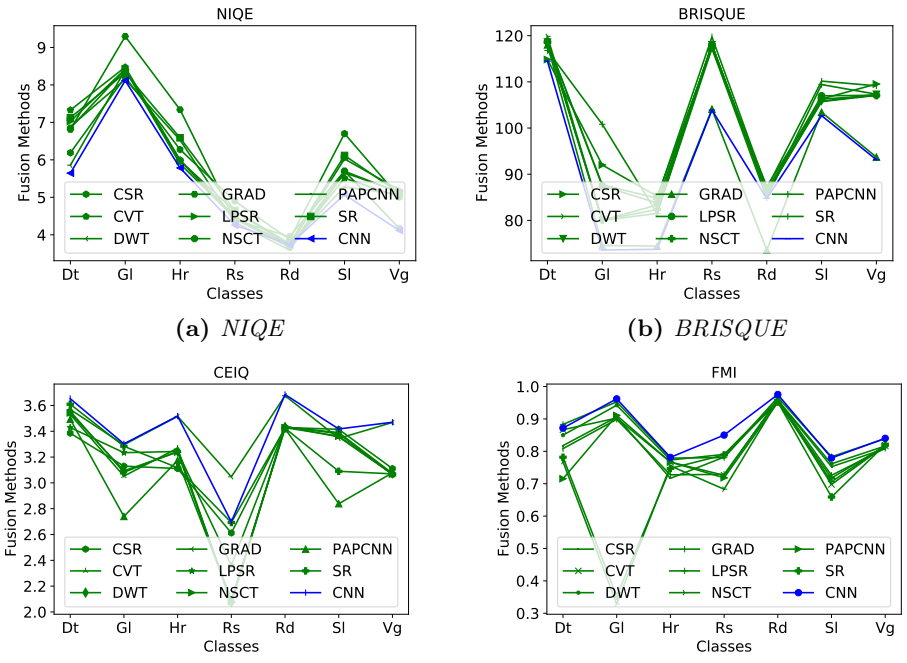
(c) *GMSD* **(d)** *PSNR*

Fig. 3 *Comparison of different reference quality assessment metrics with different fusion methods.*

4.4 Inference from the no reference quality metrics

The natural image quality evaluator (NIQE) is also known as blind image quality assessment as it depends only on the fused image. NIQE mainly operates on the

distortions of natural images. The statistical measure of NIQE depends on the spatial performance and features of natural scene statistics of information theory. Blind/referenceless image spatial quality evaluator (BRISQUE) is similar to NIQE which operates on the spatial scattering of the pixels in the image. The lower values of NIQE and BRISQUE indicate good perceptual quality of the fused image. Another conventional metric was introduced by [26], which analyzes various distortions like image compression, blur, non-uniform intensities, and white noise. The metric is opinion aware fused image quality analyzer that depends upon human judgments and implements the support vector regression to find the result. In contrast enhanced image quality (CEIQ) technique, the quality of the images is based on contrast enhancement. This approach is also based on information theory under natural scene statistics. The quality score in CEIQ is learned based on the SSIM, histogram based entropy, and cross-entropy. Finally, feature mutual information (FMI) metric is used to estimate the quality of the fused image, which results in more visual information in the images. The higher value of FMI gives the desired result. The cell values in the Tab. III represents the mean values for all the fusion images of different classes.



(a) *NIQE* **(b)** *BRISQUE*

(c) *CEIQ* **(d)** *FMI*

Fig. 4 *Comparison of different no reference quality assessment metrics with different fusion methods.*

The statistics of no reference quality assessment metrics are compared in Fig. 4. As explained in the above section, the blue line indicates the better values of the suggested CNN-based fusion approach when related to the green lines of other

Method	CSR	CVT	DWT	GRD	LPSR	NSCT	PCN	SR	CNN
Class					NIQE				
Dt	6.188	7.329	5.858	6.811	7.027	6.847	6.886	7.126	5.645
Gt	8.212	8.474	8.401	9.295	8.421	8.416	8.147	8.321	8.117
Hr	6.275	5.995	5.893	7.339	5.98	5.983	6.528	6.588	5.785
Rs	4.899	4.422	4.447	4.39	4.626	4.488	4.331	4.645	4.256
Rd	3.875	3.664	3.745	3.951	3.721	3.689	3.593	3.766	3.737
Sl	6.701	5.512	5.277	5.7	6.025	5.697	5.656	6.117	5.072
Vg	5.004	5.011	5.075	4.169	5.197	5.124	5.147	5.114	4.114
Class					BRISQUE				
Dt	114.84	119.85	118.62	118.01	118.7	118.55	116.8	118.69	114.49
Gt	92	79.99	80.03	74.57	87.44	80.08	100.81	87.7	73.59
Hr	85.21	81.51	83.35	74.41	83.76	82.35	81.51	84.81	73.73
Rs	117.31	117.99	118.03	104.08	117.92	118.02	117.24	119.72	103.9
Rd	87.36	85.82	85.94	73.58	86.42	85.82	84.83	86.92	84.75
Sl	106.3	105.73	105.78	103.42	106.98	106.27	109.41	110.16	102.68
Vg	109.53	107.07	107.32	93.7	107.1	107.03	107.33	109.14	93.14
Class					CEIQ				
Dt	3.385	3.53	3.556	3.569	3.429	3.542	3.493	3.61	3.653
Gt	3.13	3.052	3.082	3.295	3.234	3.094	2.74	3.284	3.303
Hr	3.111	3.269	3.245	3.515	3.244	3.239	3.174	3.118	3.518
Rs	2.612	2.063	2.081	3.047	2.361	2.083	2.095	2.693	2.698
Rd	3.429	3.438	3.433	3.676	3.435	3.428	3.419	3.423	3.685
Sl	3.415	3.356	3.39	3.345	3.381	3.362	2.839	3.09	3.418
Vg	3.111	3.068	3.074	3.468	3.064	3.068	3.078	3.07	3.47
Class					FMI				
Dt	0.818	0.867	0.85	0.808	0.883	0.768	0.715	0.782	0.872
Gt	0.906	0.903	0.942	0.899	0.953	0.329	0.911	0.353	0.962
Hr	0.716	0.767	0.773	0.727	0.779	0.755	0.767	0.747	0.781
Rs	0.782	0.726	0.79	0.729	0.781	0.683	0.718	0.789	0.85
Rd	0.957	0.96	0.958	0.958	0.967	0.956	0.951	0.952	0.975
Sl	0.752	0.698	0.761	0.726	0.783	0.706	0.718	0.659	0.779
Vg	0.808	0.817	0.816	0.809	0.839	0.818	0.816	0.821	0.84

Tab. III *Mathematical values of the no reference-quality assessment metrics for various fusion algorithms adopted for all the classes.*

fusion approaches. The statistical values with respect to CEIQ, and FMI should be higher whereas lower for NIQE and BRISQUE. The values of the proposed methodology are higher for grassland and lower for rain slicks when compared to other classes in the case of entropy. Similar results have been followed by the other fusion methodologies. NIQE and BRISQUE metrics have similar results in which the values are lower for residential areas and higher for grassland and desert

respectively for the proposed methodology, likewise other fusion techniques. A similar analysis is absorbed with other fusion techniques. Residential areas have given the best values for CEIQ and FMI in the occasion of the projected technique and lower in the instance of rain slicks and snow land areas. Here, the observations of grassland, rain slicks, residential, and desert are similar in most of the fusion methods. The proposed values of CNN-based fusion are given in the last column of both tables for all the quality metrics.

5. Conclusion and future work

The current study essentially emphasises the integration of microwave remote sensing and optical remote sensing images using deep learning-based CNN. The three important stages of the image fusion process are image transformation, activity level measurement, and fusion decision. In the first stage, the score map is generated with a patch size of 3×3, leading to the generation of the binary segmented image in the second level. Finally, the fusion is done at the patch level to generate the final decision map. Now, the demonstration of fused image quality is assessed with various reference and no reference quality metrics. Further, to enhance the current work is compared with various other fusion techniques. The results of CNN-based image fusion outperform traditional and conventional-based image fusion in all aspects.

Another important observation from the results is classes like desert, grassland, rain slicks, and residential areas play an important role in evaluating the fusion results when compared to other classes. The proposed work does not provide good results for snowland with respect to the PSNR value. This drawback can be taken as a future enhancement to overcome the problem and also work on more classes like a barren land, airports, and parking lots. The obtained results of the four classes mentioned above can be classified and analysed in future work. Also, to achieve better performance in the current study, the classification of the resultant fused images shall be compared with the existing classification procedures and will be considered as a future recommendation. In this manner, the present work leads to highly recommended future work. Despite many achievements, the complexity to learn deep learning-based models is very high along with the computational power.

The chief contributions and originality of the work proposed are given in four steps as follows: (1) A deep architecture of CNN is designed to work for image transformation, activity level measurement, and fusion rule in a unified way. (2) There is a direct mapping between the source images (SAR, optical) and the resultant image to show the relationship among them. (3) The very important aspect of the proposed work is that the algorithm is applied to a good number of scene classes of images. Different types of objective image quality assessment metrics like information theory-based, image structure, and feature-based metrics are compared with other fusion methods.

References

[1] LIU Y., CHEN X., WARD R.K., WANG Z.J. Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Processing Letters*. 2019, 26(3), pp. 485–489, doi: 10.1109/LSP.2019.2895749.

[2] PARAMANANDHAM N., RAJENDIRAN K. Multi sensor image fusion for surveillance applications using hybrid image fusion algorithm. *Multimedia Tools and Applications*. 2018, 77(10), pp. 12405–12436. doi: 10.1007/s11042-017-4895-3.

[3] NGUYEN K., FOOKES C., ROSS A., SRIDHARAN S. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access*. 2017, 6, pp. 18848–18855. doi: 10.1109/ACCESS.2017.2784352.

[4] ATASEVER U.H., KESIKOGLU M.H., OZKAN C. A new artificial intelligence optimization method for PCA based unsupervised change detection of remote sensing image data. *Neural Network World*. 2016, 26(2), pp. 141–154, doi: 10.14311/nnw.2016.26.008.

[5] KUANG P., MA T., LI F., CHEN Z. Real-time pedestrian detection using convolutional neural networks. *International Journal of Pattern Recognition and Artificial Intelligence*. 2018, 32(11), pp. 1856014–1856029, doi: 10.1142/S0218001418560141.

[6] LINDA G.M., THEMOZHI G., BANDI S.R. Color-mapped contour gait image for cross-view gait recognition using deep convolutional neural network. *International Journal of Wavelets, Multiresolution and Information Processing*. 2019, pp. 1941012–1941040, doi: 10.1142/S0219691319410121.

[7] ZAORALEK L., PLATOS J., SNASEL V. Patient-adapted and inter-patient ECG classification using neural network and gradient boosting. *Neural Network World*. 2018, 26(2), pp. 241–254, doi: 10.14311/nnw.2018.28.015.

[8] BYUN Y. A texture-based fusion scheme to integrate high-resolution satellite SAR and optical images. *Remote sensing letters*. 2014, 5(2), pp. 103–111, doi: 10.1080/2150704X.2014.880817.

[9] ZHANG H., LIN H., LI Y. Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification. *IEEE Geoscience and Remote Sensing Letters*. 2015, 12(5), pp. 1061–1065, doi: 10.1109/LGRS.2014.2377722.

[10] ZHANG W., XU M. Translate SAR data into optical image using IHS and wavelet transform integrated fusion. *Journal of the Indian Society of Remote Sensing*. 2019, 47(1), pp. 125–137, doi: 10.1007/s12524-018-0879-7.

[11] ZHANG J., YANG J., ZHAO Z., LI H., ZHANG Y. Block-regression based fusion of optical and SAR imagery for feature enhancement. *International Journal of Remote Sensing*. 2010, 31(9), pp. 2325–2345, doi: 10.1080/01431160902980324.

[12] SONG H., LIU Q., WANG G., HANG R., HUANG B. Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2018, 11(3), pp. 821–829. doi: 10.1109/JSTARS.2018.2797894.

[13] DELEDALLE C.A., DENIS L., TUPIN F. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*. 2009, 18(12), pp. 2661–2672. doi: 10.1109/TIP.2009.2029593.

[14] YU Z., WANG W., LI C., LIU W., YANG J. Speckle noise suppression in SAR images using a three-step algorithm. *Sensors*. 2018, 18(11), p.3643. doi: 10.3390/s18113643.

[15] LIU Y., CHEN X., PENG H., WANG Z. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*. 2017, 36, pp. 191–207. doi: 10.1016/j.inffus.2016.12.001

[16] HE K., SUN J., TANG X. Guided image filtering. *IEEE Transactions on Pattern Analysis an d Machine Intelligence*. 2012, 35, pp. 1397—1409. doi: 10.1109/TPAMI.2012.213.

[17] HERMESSI H., MOURALI O., ZAGROUBA E. Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain. *Neural Computing and Applications*. 2018, 30(7), pp. 2029–2045. doi: 10.1007/s00521-018-3441-1.

[18] SHARMA A., LIU X., YANG X., SHI D. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks* 2017, 95, pp. 19–28. doi: 10.1016/j.neunet.2017.07.017.

[19] SCHMITT., MICHAEL., HUGHES., LLOYD. mediatum1474000 Home Page. In: SCHMITT, ed. *SEN12MS – dataset* [online]. Technical University of Munich, 2017 [viewed 2019-08-13]. Available from: https://mediatum.ub.tum.de/1474000.

[20] ZHANG Z., BLUM R.S. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*. 1999, 87(8), pp. 1315–1326. doi: 10.1109/5.775414.

[21] NENCINI F., GARZELLI A., BARONTI S., ALPARONE L. Remote sensing image fusion using the curvelet transform. *Information fusion*. 2007, 8(2), pp. 143–156. doi: 10.1016/j.inffus.2006.02.001.

[22] BURT P., ADELSON E. The Laplacian pyramid as a compact image code. *IEEE Transactions on communications*. 1983, 31(4), pp. 532–540. doi: 10.1109/TCOM.1983.1095851.

[23] PAUL S., SEVCENCO I.S., AGATHOKLIS P. Multi-exposure and multi-focus image fusion in gradient domain. *Journal of Circuits, Systems and Computers*. 2016, 25(10), p.1650123. doi: 10.1142/S0218126616501231.

[24] ZHANG Q., GUO B.L. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal processing*. 2009, 89(7), pp. 1334–1346. doi: 10.1016/j.sigpro.2009.01.012.

[25] YANG B., LI S. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement* 2009, 59(4), pp. 884–892. doi: 10.1109/TIM.2009.2026612.

[26] YIN M., LIU X., LIU Y., CHEN X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Transactions on Instrumentation and Measurement*. 2018, 68(1), pp. 49–64. doi: 10.1109/TIM.2018.2838778.

[27] LIU M., DAI Y., ZHANG J., ZHANG X., MENG J., XIE Q. PCA-based sea-ice image fusion of optical data by HIS transform and SAR data by wavelet transform. *Acta Oceanologica Sinica*. 2015, 34(3), pp. 59–67. doi: 10.1007/s13131-015-0634-7.

[28] LIU J., CHEN H., WANG Y. Multi-Source Remote Sensing Image Fusion for Ship Target Detection and Recognition. *Remote Sensing*. 2021, 13(23), p.4852. doi: 10.3390/rs13234852.

[29] ANUSHA N., BHARATHI B. Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data. *The Egyptian Journal of Remote Sensing and Space Science*. 2020, 23(2), pp. 207–219. doi: 10.1016/j.ejrs.2019.01.001.

[30] ZHANG P., BAN Y., NASCETTI A. Learning U-Net without forgetting for near real-time wildfire monitoring by the fusion of SAR and optical time series. *Remote Sensing of Environment*. 2021, 261, p.112467. doi:

[31] CHEN N., CHENG B., ZHANG X., XING C. Surface soil moisture estimation at high spatial resolution by fusing synthetic aperture radar and optical remote sensing data. *Journal of Applied Remote Sensing*. 2020, 14(2), p.024508. doi: 10.1117/1.JRS.14.024508.

[32] PRABHAKAR K.R., NUKALA V.H., GUBBI J., PAL A., BALAMURALIDHAR P. Improving SAR and Optical Image Fusion for Lulc Classification with Domain Knowledge. *IEEE International Geoscience and Remote Sensing Symposium*. 2022, pp. 711–714. doi: 10.1109/IGARSS46834.2022.9884283.

[33] KARIMI D., AKBARIZADEH G., RANGZAN K., KABOLIZADEH M. Effective Supervised Multiple-Feature Learning for Fused Radar and Optical Data Classification. *IET Radar, Sonar and Navigation*. 2017, 11, pp. 768–777. doi: 10.1049/iet-rsn.2016.0346.

[34] MAHYOUBA S., FADILB A., MANSOUR E.M., RHINANEA H., AL-NAHMIA F. Fusing of optical and synthetic aperture radar (SAR) remote sensing data: A systematic literature review (SLR). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2019, 42(4/W12), pp. 127–138. doi: 10.5194/isprs-archives-XLII-4-W12-127-2019.

[35] DIAN R., LI S., SUN B., GUO A. Recent advances and new guidelines on hyperspectral and multispectral image fusion. *Information Fusion*. 2021, 69, pp. 40–51. doi: https://doi.org/10.1016/j.inffus.2020.11.001.

[36] ZHANG H., XU R. Exploring the optimal integration levels between SAR and optical data for better urban land cover mapping in the Pearl River Delta. *International journal of applied earth observation and geoinformation*. 2018, 64, pp. 87–95. doi: 10.1016/j.jag.2017.08.013

[37] MEHMOOD M., SHAHZAD A., ZAFAR B., SHABBIR A., ALI N. Remote sensing image classification: A comprehensive review and applications. *Mathematical Problems in Engineering*. 2022. doi: 10.1155/2022/5880959.

[38] SAPONARA S., ELHANASHI A., GAGLIARDI A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *Journal of Real-Time Image Processing*. 2021, 18(3), pp. 889–900. doi: 10.1007/s11554-020-01044-0.

[39] DU J., LU H., HU M., ZHANG L., SHEN, X. CNN-based infrared dim small target detection algorithm using target-oriented shallow-deep features and effective small anchor. *IET Image Processing*. 2021, 15(1), pp. 1–15. doi: 10.1049/ipr2.12001.

[40] EL-REWAIDY H., FAHMY A.S., PASHAKHANLOO F., CAI X., KUCUKSEYMEN S., CSECS I., NEISIUS U., HAJI-VALIZADEH H., MENZE B., NEZAFAT R. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. *Magnetic Resonance in Medicine*. 2021, 85(3), pp. 1195–1208. doi: 10.1002/mrm.28485.

[41] KAUR H., KOUNDAL D., KADYAN V. Image fusion techniques: a survey. *Archives of computational methods in Engineering*. 2021, 28(7), pp. 4425–4447. doi: 10.1007/s11831-021-09540-7.

[42] LI S., KANG X., FANG L., HU J., YIN H. Pixel-level image fusion: A survey of the state of the art. *information Fusion*. 2017, 33, pp. 100–112. doi: 10.1016/j.inffus.2016.05.004.

[43] ZHOU Y., YANG X., ZHANG R., LIU K., ANISETTI M., JEON G. Gradient-based multi-focus image fusion method using convolution neural network. *Computers and Electrical Engineering*. 2021, 92, p.107174. doi: 10.1016/j.compeleceng.2021.107174.

[44] SINHA S., JEGANATHAN C., SHARMA L.K., NATHAWAT M.S. A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology*. 2015, 12(5), pp. 1779–1792. doi: 10.1007/s13762-015-0750-0.

[45] ECKERSTORFER M., BUHLER Y., FRAUENFELDER R., MALNES E. Remote sensing of snow avalanches: Recent advances, potential, and limitations. *Cold Regions Science and Technology*. 2016, 121, pp. 126–140. doi: 10.1016/j.coldregions.2015.11.001.