


REGULAR PAPER

A CSR-based visible and infrared image fusion method in low illumination conditions for sense and avoid

N. Ma¹, Y. Cao¹, Z. Zhang¹, Y. Fan² and M. Ding³

¹College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China, ²Shenyang Aircraft Design & Research Institute, Aviation Industry Corporation of China, Shenyang, China and ³College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Corresponding author: Y. Cao; Email: cyfac@nuaa.edu.cn

Received: 6 July 2022; **Revised:** 21 April 2023; **Accepted:** 25 May 2023

Keywords: Image fusion; Unmanned aerial vehicles; Machine vision; Low illumination; Visible/infrared

Abstract

Machine vision has been extensively researched in the field of unmanned aerial vehicles (UAV) recently. However, the ability of Sense and Avoid (SAA) largely limited by environmental visibility, which brings hazards to flight safety in low illumination or nighttime conditions. In order to solve this critical problem, an approach of image enhancement is proposed in this paper to improve image qualities in low illumination conditions. Considering the complementarity of visible and infrared images, a visible and infrared image fusion method based on convolutional sparse representation (CSR) is a promising solution to improve the SAA ability of UAVs. Firstly, the source image is decomposed into a texture layer and structure layer since infrared images are good at characterising structural information, and visible images have richer texture information. Both the structure and the texture layers are transformed into the sparse convolutional domain through the CSR mechanism, and then CSR coefficient mapping are fused via activity level assessment. Finally, the image is synthesised through the reconstruction results of the fusion texture and structure layers. In the experimental simulation section, a series of visible and infrared registered images including aerial targets are adopted to evaluate the proposed algorithm. Experimental results demonstrates that the proposed method increases image qualities in low illumination conditions effectively and can enhance the object details, which has better performance than traditional methods.

Nomenclature

ADS-B	Automatic dependent surveillance-broadcast
CNN	Convolutional neural network
CSR	Convolutional sparse representation
GAN	Generative adversarial network
LiDAR	Light Detection and Ranging
SAA	Sense and Avoid
SR	Sparse representation
SWAP	Size, weight and power
TCAS	Traffic Alert and Collision Avoidance System
UAV	Unmanned aerial vehicle

Greek symbol

μ	Control texture smooth degree
σ	Control texture size
λ	Control weight of l_1 norm

1.0 Introduction

The capability of Sense and Avoid (SAA) is considered to be the most important component for the integration of unmanned aerial vehicles (UAV) into the National Aerospace System [1, 2]. SAA is composed of two crucial parts in general: (1) the sensing part, aimed at detecting all aerial targets threatening UAV flight safety with the help of on-board sensing devices; and (2) the avoiding part, aimed at eliminating the potential hazards based on the sensing result by trajectory re-planning and corresponding flight control [3].

The sensing part is the foundation of SAA. According to the working pattern of onboard sensing devices, SAA can be divided into two main categories: cooperative and non-cooperative. The sensing devices for cooperative SAA contain Traffic Alert and Collision Avoidance System (TCAS) [4] and Automatic Dependent Surveillance-Broadcast (ADS-B) [5], which have been widely installed on manned aircraft. The non-cooperative SAA is operated with onboard sensing devices free from information exchange. The sensing devices for non-cooperative SAA contain machine vision [6], acoustic system [7], and Light Detection and Ranging (LiDAR) [8]. Different from non-cooperative devices, the onboard sensing devices for cooperative SAA largely depend on information exchange with aerial targets.

Compared to all airborne sensing devices, machine vision shows great potential to enhance the capabilities of SAA for the following reasons [9]:

- Machine vision can detect all the dangerous flying targets without information exchange, which makes the perception of non-cooperative aerial targets possible.
- The information gathered by machine vision is abundant compared with other non-cooperative sensors. For example, the category of aerial targets can be acquired by image recognition algorithm, and proper collision avoidance maneuvers can be selected according to the target category.
- Machine vision outperforms other airborne sensor devices in terms of size, weight, and power (SWAP) [10], which makes the installation of machine vision on small UAVs possible.

The application of machine vision for SAA has been extensively researched, and a series of algorithms and systems based on vision have been developed. However, there are still some challenges in the perception part of SAA, and one of the most serious is the high demand for image quality. The factors that may affect the airborne image quality can be summarised as follows: (1) inadequate illumination in dark environments [11]; (2) blurred image caused by aircraft position and posture variation [12]; and (3) target occlusion caused by cloud and mist [13]. It is worth mentioning that previous SAA research work based on vision, including aerial target detection, tracking and pose estimation, are all proposed with the precondition that image quality is good enough. However, low image quality will influence the performance of these algorithms greatly in real application because it will directly lead to the loss of target texture information, which will bring great difficulties to feature extraction and target detection. Therefore, the low image quality will reduce the perception ability of the target, and the visual perception range will be reduced, which is extremely unfavourable for SAA [14]. Among these factors, low illumination is the most typical one, which will directly influence UAV's threat target perception ability in low illumination conditions. Therefore, the research motivation of this paper includes two points: restraining the attenuation of UAV's visual perception range in low illumination and enhancing the structure and texture information of the detected target for SAA application.

As shown in Fig. 1, the helicopter is not clear in visible image in low illumination conditions. But the infrared image can effectively capture the structure of helicopter, compared with the lack of texture information for visible images. Considering the complementarity of visible and infrared sensors, it is feasible to combine these two visual information to improve the image quality. As shown in Fig. 1(c), The fused image combines the advantages of both visible and infrared images effectively.

Therefore, a CSR-based visible and infrared image fusion method is proposed to enhance the SAA ability of UAVs in low illumination conditions in this paper. Firstly, the source image is decomposed into

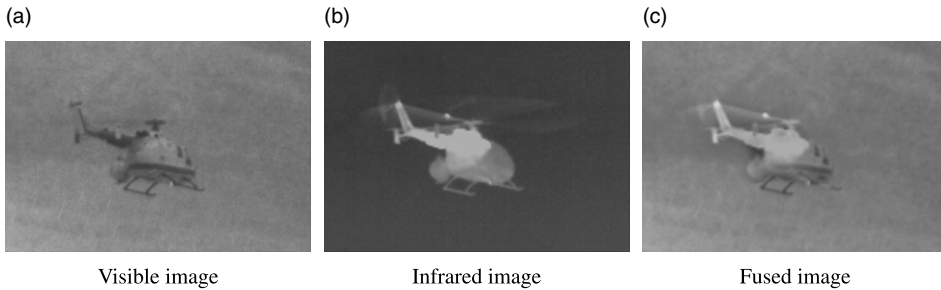


Figure 1. Visible, infrared and fused image containing aerial targets.

texture and structure layers since infrared images are good at characterising structural information and visible images have richer texture information. Then both the structure and texture layers are transformed into the sparse convolutional domain through the CSR mechanism, and the CSR coefficient mapping are fused by activity level assessment. Finally, the image is synthesised through the reconstruction results of the fusion structure and texture layers.

The main contribution of this paper can be summarised below:

- Aiming at shortcomings of semantic information loss and low tolerance of detail information caused by traditional methods, a visible and infrared image fusion method based on multi-layer CSR is proposed to enhance the visual perception of UAVs in low illumination.
- Different from the local transformation in traditional methods, the global modelling ability of the proposed method has obvious advantages under the condition of mismatch.
- Compared with the deep learning methods, the proposed method adopts an unsupervised learning mode, which does not require many labelled samples for training and is easier to implement.

The rest of this paper is organised as follows. Firstly, the application background of this method is introduced, and the significance and motivation of the research are explained in Section 1. Section 2 conducts an analysis of relevant literature reviews. Section 3 introduces the framework and mechanism of the proposed visible and infrared image fusion algorithm in this paper. In Section 4, the effectiveness of the proposed algorithm is verified by a series of experiments in three scenarios and compared and analysed with other algorithms. The whole paper is summarised in Section 5.

2.0 Literature review

Due to the advantages, including non-cooperative target perception, abundant information acquisition capability and good size, weight and power (SWAP) features, vision-based SAA has shown great potential for increasing UAV safety levels in recent years. The general framework of vision-based SAA consists of four key components: aerial target detection, tracking, relative pose and position estimation and avoidance [2]. The related research for each component can be concluded as follows:

- Aerial target detection [10, 15, 16, 17]. Aerial target detection is the first step of vision-based SAA, which aims at picking out targets with potential risk from images/videos. Research on aerial target detection can be classified into foreground modelling-based methods and background modelling-based methods, which utilises information from single image and consecutive frames respectively.
- Aerial target tracking [18]. After the detection of aerial targets, the detected bounding box should be tracked continuously by target tracking algorithms. Vision-based target tracking algorithms can be classified into generative and discriminative tracking, and both categories have been

applied to vision-based SAA. The main challenge of vision-based aerial target tracking is the adaptive scale transformation of the tracking bounding box.

- Relative state estimation [19, 20]. This component aims to obtain the relative position and attitude between host UAV and aerial targets with potential risk. Since the risk level is determined based on the estimated angle and range, this step is crucial for collision avoidance.
- Collision avoidance [3, 21]. Finally, the potential risk of aerial targets should be eliminated by trajectory re-planning and tracking control based on the estimated pose and position. The biggest challenge for vision based collision avoidance is that the range information can be hard to acquired in some cases, especially for monocular vision.

Generally the four key components concluded above are important, but the foundation of vision-based SAA is the high quality image. Therefore, the enhancement of image quality in some undesirable conditions is imperative. It is worth noting that the previous research related to vision-based SAA is carried out assuming that image quality is good enough. Several factors may deteriorate airborne image quality in real applications, and insufficient illumination in dark environments is the most crucial one. Since visible and infrared images containing aerial targets are complementary in dark environments, increasing image quality by visible and infrared image fusion is desirable.

In general, the algorithm designed for visible and image fusion can be concluded as three steps: image transformation, image fusion and image reconstruction. Among the three steps, the method for image transformation is the foundation of the whole algorithm. For this reason, the research of image fusion algorithms during the past decade mainly focuses on developing a more concise and effective transformation method. The most widely used transformation methods for image fusion are sparse representation (SR), convolutional sparse representation (CSR) and deep learning-based methods such as convolutional neural networks (CNN).

The application of SR-based image fusion has achieved great success in the past few years. However, due to the local representation nature of SR, the drawbacks of SR-based fusion algorithm can be concluded as follows [22, 23]: (1) The context information loss. Since SR-based fusion must firstly decompose source image into local patches, the context information within the source image is neglected. It is worth noting that context information is essential for vision understanding and analysis. (2) The high sensitivity to registration errors. As SR fuses all the image patches respectively, all the image patches need to be accurately registered. However, image registration itself is also a difficult task, and registration error may always exist. To overcome this problem, the fusion framework designed based on global representation algorithms has been proposed in recent years, and the most representative algorithms are CNN and CSR [24].

Deep learning has revealed the powerful potential for computer vision tasks recently. The advantage of deep learning-based image fusion methods is that the fusion strategy can be obtained through learning, and the fused result can be obtained without the artificial design of fusion rules [25]. As a supervised learning approach, the framework of CNN can be classified into two main categories, namely the regression CNN and classification CNN [26]. Both the regression CNN and classification CNN have been successfully applied to image fusion [27, 28]. Generative adversarial network (GAN), a novel deep learning model, can extract typical characteristics by using different network branches according to the modality of different image sources [29]. Because of its advantages in processing multi-modality information, GAN is the future development direction of image fusion methods, which is suitable for task-driven image fusion, such as target perception [30]. However, the restriction of CNN-based image fusion may come from the high demand for labelled training samples. CSR originated from the deconvolutional networks designed for unsupervised image feature analysis [31]. With applications to image fusion, CSR can be treated as a global image transformation approach. The advantages of CSR-based image fusion over SR and deep learning can be concluded as follows [23]: (1) the global modelling capability of CSR makes it free from image decomposition when applied to image fusion. For this reason, the above-mentioned deficiencies of SR based fusion, including context information loss and high sensitivity to misregistration caused by local transformation, is easy to solve; (2) the unsupervised learning

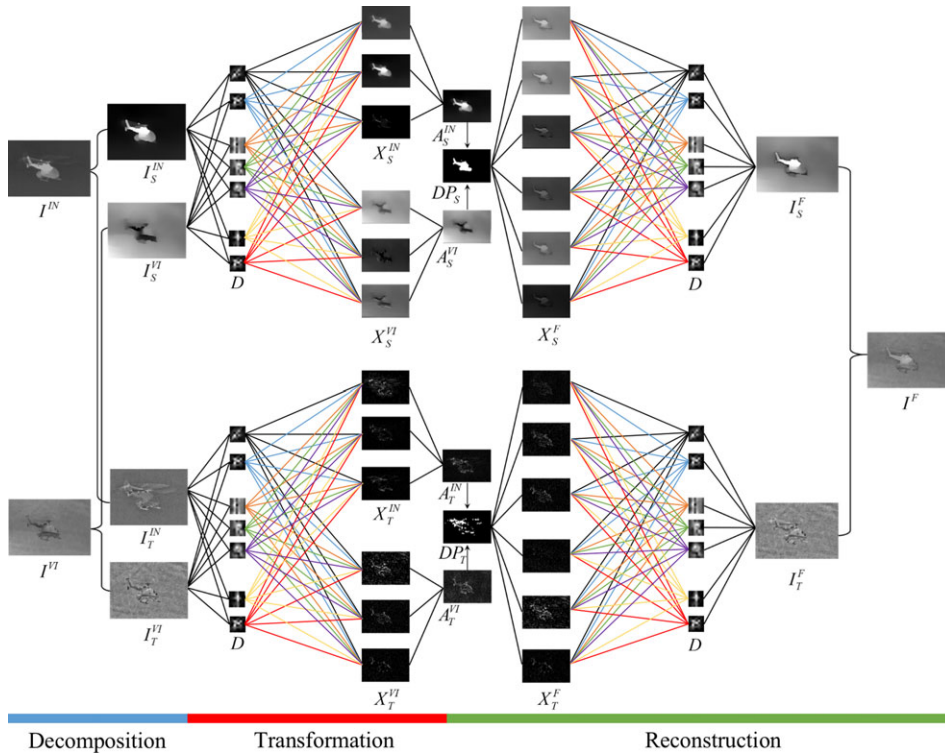


Figure 2. Framework of visible and infrared aerial targets image fusion.

nature of CSR makes it free from a large amount of labelled ground truth images. Therefore, CSR has revealed great potential for image fusion.

For this reason, the elastic-net regularisation based multi-layer CSR is adopted for image fusion in this paper. Instead of directly fusing source images, both visible and infrared images are decomposed into structure and texture layers. The image layers, after decomposition, are then transformed into the sparse convolutional domain for image fusion. Finally, the transformed sparse convolutional coefficient mapping corresponding to visible and infrared images are fused by activity level assessment, and the fused image is obtained by image reconstruction.

3.0 Visible and infrared image fusion method for SAA

The general framework of the image fusion method for SAA contains three parts, as shown in Fig. 2. Firstly, both the visible image I^{VI} and infrared image I^{IN} are decomposed into two layers, namely structure layers I_S^{IN}, I_S^{VI} and texture layers I_T^{IN}, I_T^{VI} . Secondly, the decomposed structure layers I_S^{IN}, I_S^{VI} and texture layers I_T^{IN}, I_T^{VI} are transformed into the sparse convolutional domain via CSR by pre-learned dictionary D , and the transformed sparse convolutional coefficient mapping corresponding to $I_S^{IN}, I_S^{VI}, I_T^{IN}$ and I_T^{VI} are $X_S^{IN}, X_S^{VI}, I_S^{IN}$ and I_S^{VI} respectively. Thirdly, by transforming the activity map of X_S^{IN} and X_S^{VI} into A_S^{IN}, A_S^{VI} , the decision map for structure layer DP_S is generated. Similarly, the decision map for texture layer DP_T is generated. Based on the decision map, the fused convolutional sparse coefficient map X_S^F and X_T^F for structure and texture layers are obtained. The fused result I_S^F and I_T^F for structure and texture layers are reconstructed by utilising sparse convolutional dictionary D . Finally, the fused image I^F is obtained by synthesising X_S^F and I_T^F . In this section, the method for image decomposition, image transformation and image reconstruction will be introduced in detail.

3.1 Image decomposition

Typically, image I is composed of two layers: the structure layer I_S and the texture layer I_T , as expressed in Equation (1). The structure layer usually represents the semantic information and captures salient objects inside the image, while the texture layer emphasises the preservation of the details of the image. And the semantically meaningful structure layers are usually covered by texture layers as shown in Fig. 2. As mentioned above, since infrared image and visible image are good at retaining structure and texture information respectively, so it is desirable to decompose these two layers for image fusion.

$$I = I_S + I_T \tag{1}$$

This paper adopts the relative total variation based image decomposition algorithm for image decomposition [32]. The objective function for image decomposition is expressed as Equation (2), where $I_S(i)$ and $I(i)$ are pixel values of the structure layer and original image at location i respectively, p is the total pixel of the input image, μ is the parameter controlling smooth degree, and ε is a small positive number to avoid denominator being zero. $V_x(i)$ and $V_y(i)$ from in Equations (3) and (4) are total variations in x and y direction for pixel i , where $R(i)$ is the rectangle region centred at i , g_{ij} is weighting function designed to avoid spatial affinity, ∂_x and ∂_y are partial derivatives in x and y direction respectively. The formulation of g_{ij} is expressed as Equation (5), where σ is the parameter controlling window size. The influence of the image decomposition parameters on image fusion will be analysed in Section 5.

$$\operatorname{argmin}_{I_S} \sum_{i=1}^p (I_S(i) - I(i))^2 + \mu \cdot \left(\frac{V_x(i)}{V_x(i) + \varepsilon} + \frac{V_y(i)}{V_y(i) + \varepsilon} \right) \tag{2}$$

$$V_x(i) = \left| \sum_{j \in R(i)} g_{ij} \cdot (\partial_x I_S)_j \right| \tag{3}$$

$$V_y(i) = \left| \sum_{j \in R(i)} g_{ij} \cdot (\partial_y I_S)_j \right| \tag{4}$$

$$g_{ij} \propto \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2}\right) \tag{5}$$

3.2 Image transformation

In this part, the structure layers I_S^{VI} , I_S^{IN} and texture layers I_T^{VI} , I_T^{IN} of visible and infrared images are transformed into the sparse convolutional domain by elastic net based CSR. In Equation (6), the basic idea of CSR is that an input image I can be represented by the sum of the convolutional product of equal-sized convolutional dictionary filters $D = \{d_1, d_2, \dots, d_m\}$ and sparse convolutional coefficient mapping $X = \{x_1, x_2, \dots, x_m\}$, where m is the number of convolutional dictionary filters.

$$I = \sum_{i=1}^m d_i * x_i. \tag{6}$$

For each input image, the convolutional dictionary is pre-learned. Therefore, the computation of sparse convolutional coefficient mapping X is essential for image transformation. Conventionally, the computation of X can be operated by l_1 norm regularisation, and the objective function can be expressed as Equation (7), where λ is the regularisation parameter.

$$\operatorname{arg min}_{x_i} \frac{1}{2} \left\| \sum_{i=1}^m d_i * x_i - I \right\|_2^2 + \lambda \sum_{i=1}^m \|x_i\|_1. \tag{7}$$

Since l_1 norm regularisation could not guarantee group selection when applied to image transformation, the elastic-net based regularisation is proposed in this paper to combine the advantages of l_1 norm

and l_2 norm regularisation. The objective function for elastic-net based regularisation can be expressed as Equation (8). The solution of Equation (8) can be acquired by the alternating direction method of multipliers (ADMM) [33].

$$\arg \min_{x_i} \frac{1}{2} \left\| \sum_{i=1}^m d_i * x_i - I \right\|_2^2 + \lambda \sum_{i=1}^m \|x_i\|_1 + (1 - \lambda) \sum_{i=1}^m \|x_i\|_2^2 \tag{8}$$

Moreover, the solution of the convolutional dictionary filters can be regarded as an optimisation problem, as shown in Equation (9). The solution of Equation (9) can be further divided into the optimisation of variables x_i and d_i . The solution of the former is the same as Equation (7). The solution of the latter can be regarded as the convolution form of the method of optimal directions (MOD) [34].

$$\arg \min_{d_i, x_i} \frac{1}{2} \left\| \sum_{i=1}^m d_i * x_i - I \right\|_2^2 + \lambda \sum_{i=1}^m \|x_i\|_1, \text{ s.t. } \|d_i\|_2 = 1 \tag{9}$$

Therefore, as presented in Equation (10), given the structure layers of I_S^{VI} , I_S^{IN} and texture layers I_S^{VI} , I_T^{IN} of visible and infrared images, the sparse convolutional coefficient mapping X_S^{IN} , X_S^{VI} , X_T^{IN} and X_S^{VI} can be estimated via elastic net regularisation-based CSR.

$$\arg \min_{x_{S,T}^{IN,VI}} \frac{1}{2} \left\| \sum_{i=1}^m d_i * x_{S,T}^{IN,VI} - I_{S,T}^{IN,VI} \right\|_2^2 + \lambda \sum_{i=1}^m \|x_{S,T}^{IN,VI}\|_1 + (1 - \lambda) \sum_{i=1}^m \|x_{S,T}^{IN,VI}\|_2^2 \tag{10}$$

3.3 Image reconstruction

The l_1 norm max strategy is adopted to fuse the sparse convolutional coefficient mapping of structure and texture layers after the computation of X_S^{IN} , X_S^{VI} , X_T^{IN} and X_S^{VI} in Equation (11), where $X_S(i, j)$ and $X_T(i, j)$ denotes the content of X_S and X_T at location (i, j) respectively.

$$X_{S,T}^F(i, j) = \begin{cases} X_{S,T}^{IN}(i, j) & \|X_{S,T}^{IN}(i, j)\|_1 > \|X_{S,T}^{VI}(i, j)\|_1 \\ X_{S,T}^{VI}(i, j) & \|X_{S,T}^{IN}(i, j)\|_1 < \|X_{S,T}^{VI}(i, j)\|_1 \end{cases} \tag{11}$$

Since the fusion of the structure layer is operated in the transformation domain, the fused results of structure and texture layers $X_{S,T}^F = \{x_{S,T_1}^F, x_{S,T_2}^F, \dots, x_{S,T_m}^F\}$ need to be transformed back to image domain. As presented in Equation (12), the reconstruction of the fused result of structure layer $X_{S,T}^F$ can be acquired by utilising the convolutional dictionary filter.

$$I_{S,T}^F = \sum_{i=1}^{m_1} d_i * x_{S,T_i}^F. \tag{12}$$

Finally, the fused image I^F can be obtained by superimposing the fused result of structure and texture layers, as presented in Equation (13).

$$I^F = I_S^F + I_T^F. \tag{13}$$

4.0 Experimental simulation and analysis

In this section, three scenes of visible and infrared images containing aerial targets are selected for the image fusion experiment as presented in Fig. 3, where the selected images obtained by different sensors are all complimentary.

To effectively evaluate the algorithm performance, objective and subjective metrics are adopted to function the quality of image fusions. Subjective metrics are operated by human eyes observation, which is intuitive and easy to operate. However, subjective metrics may fail to capture the slight differences

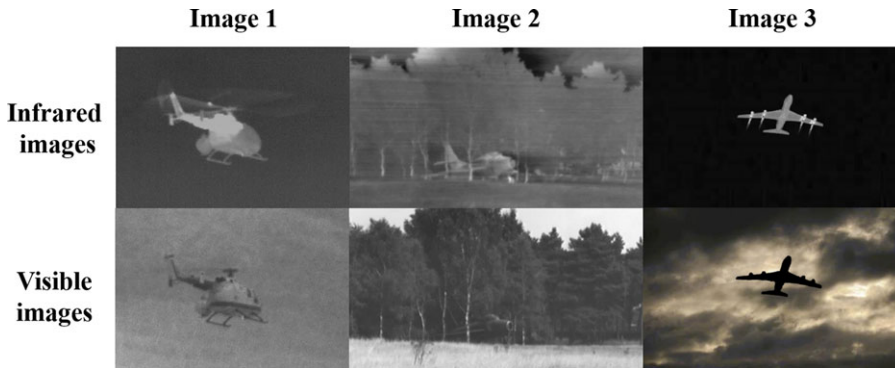


Figure 3. Different scenes of visible and infrared images containing aerial targets.

between fused results. Therefore, the objective metrics are adopted to evaluate the image fused results. The definition and function for each objective metric is presented as follows [35]:

- Q_{MI} . Objective Metric Q_{MI} is defined based on information theory, and a higher value of Q_{MI} indicates a better fused result.
- Q_M . Objective metric Q_M is defined based on a multi-scale scheme, and a higher value of Q_M indicates a better fused result.
- Q_S . Objective metric Q_S is defined based on image structural similarity, and a higher value of Q_S represents a better fused result.
- Q_{CB} . Objective metric Q_{CB} is defined based on the human perception system, and a higher value of Q_{CB} represents a better fused result.

Three important parameters have been recognised to bring some influence on fused results. The parameter's name, influences and value range are concluded as Table 1. The parameter μ affects the smoothness of the image decomposition texture. Some noises can be filtered by proper image smoothing. However, the effect of image fusion decreases significantly with the increase of parameter μ , because higher texture smoothness will lead to the loss of image details. We have found that when μ exceeds 0.5, the image quality can not achieve satisfactory results for any other parameters. Considering the above factors comprehensively, the range of value of μ should not be too large, so we choose (0, 0.05] as the range of value of μ , which can explain the changing trend. The parameter σ affects the texture size of image decomposition. In the experiment, the change of σ value has no obvious influence on the fusion effect, so we choose a larger range. Parameter λ indicates regularisation of convolution sparse representation, which must be in the range of (0, 1]. With the increase of λ , the effect of image fusion will decline because larger λ will bring greater reconstruction errors, eventually affecting the fusion effect. When λ is close to 1, the model is under fitted, while when λ tends to 0, the model is easily over-fitted. Therefore, in order to facilitate the adjustment of λ in order of magnitude, the range of value of λ should not be too large, we choose [0.0099, 0.99] as the range of value of λ .

4.1 Quantitative evaluation for μ

The value range of μ is presented in Table 1, and the values of σ and λ when evaluating μ are 3 and 0.001 respectively. The fused results with the variation of μ are presented in Fig. 4. The Q_{MI} , Q_M , Q_S , Q_{CB} for different scenes are presented in Fig. 5. It can be seen intuitively from Fig. 4 that with the increase of μ , the contours of the background and the target will be weakened in most cases, which

Table 1. Definition of three important parameters for evaluation

	Function	Value range
μ	Control texture smooth degree	(0, 0.05]
σ	Control texture size	(0, 6]
λ	Control weight of l_1 norm	[0.0099, 0.99]

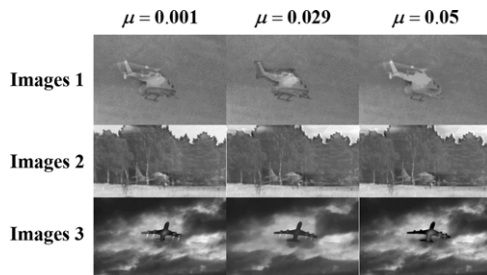


Figure 4. Image fused results with the variation of μ for different scenes.

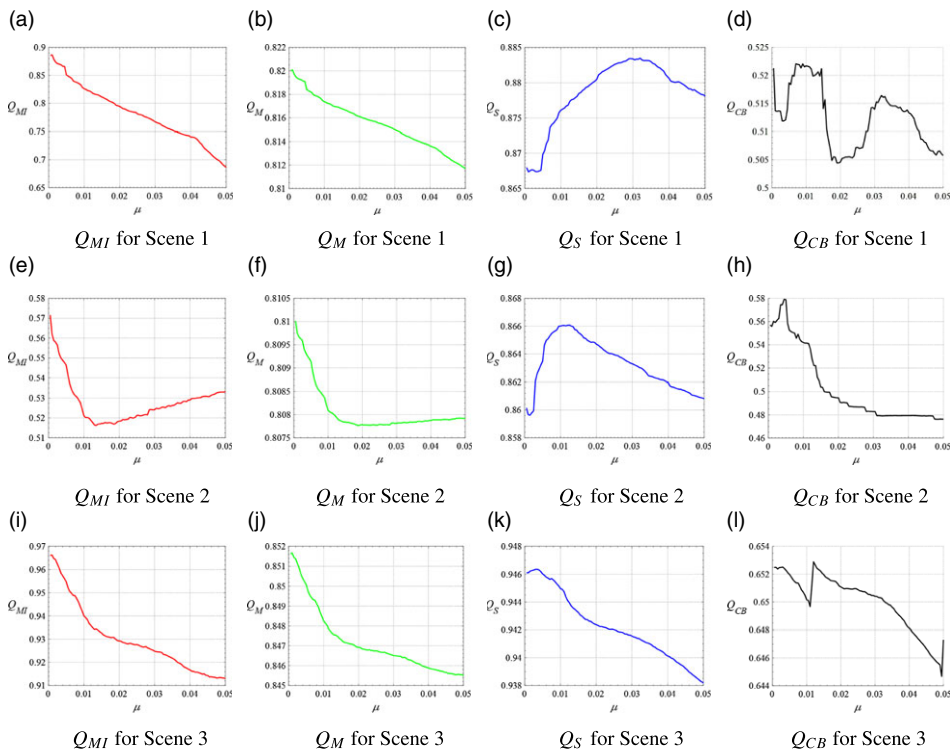


Figure 5. Q_{MI} , Q_M , Q_S , Q_{CB} with the variation of μ .

is not conducive to subsequent detection tasks. According to Fig. 5, it can be known that the effect of image fusion is negatively correlated with μ in most cases, but smoothing out background noise can bring some improvement in some specific scenes. The fusion effect decreases with the increase of μ , because the higher smooth degree may cause detail loss of the fused image.

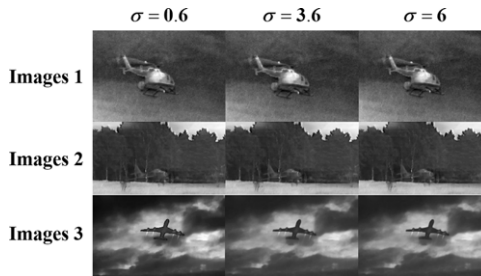


Figure 6. Image fused results with the variation of σ for different scenes.

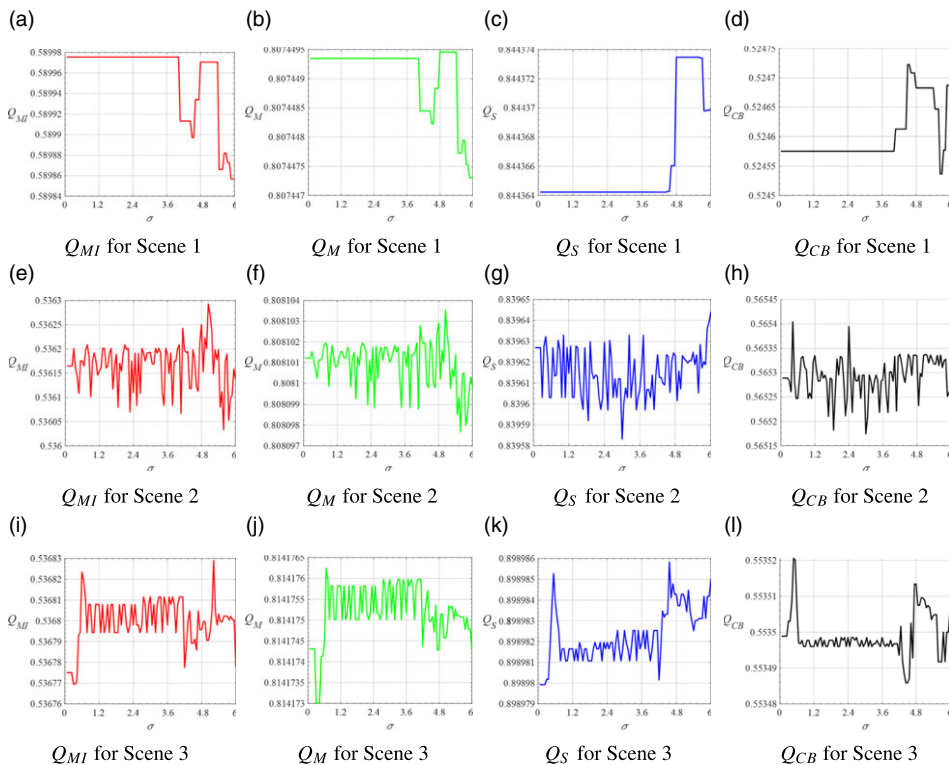


Figure 7. Q_{MI} , Q_M , Q_S , Q_{CB} with the variation of σ .

4.2 Quantitative evaluation for σ

The value range of σ is presented in Table 1, and the values of μ and λ when evaluating σ are 0.0015 and 0.001, respectively. The fused results with the variation of σ are presented in Fig. 6. The Q_{MI} , Q_M , Q_S , Q_{CB} for different scenes are presented in Fig. 7. It can be seen from Fig. 6 that there is no significant difference in the effect of image fusion with the change of σ value. As can be seen from Fig. 7, when σ is in the range of (2, 4), the evaluation index of image fusion fluctuates stably in a fixed interval. Experiment results reveal that the variation of σ dose not significantly influence fused results.

4.3 Quantitative evaluation for λ

The value range of λ is presented in Table 1, and the values of μ and σ when evaluating λ are 0.0015 and 3 respectively. The fused results with the variation of λ are presented in Fig. 8. The Q_{MI} , Q_M , Q_S ,

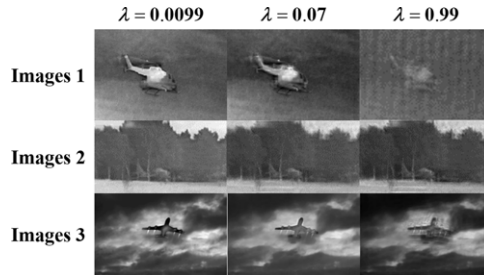


Figure 8. Image fused results with the variation of λ for different scenes.

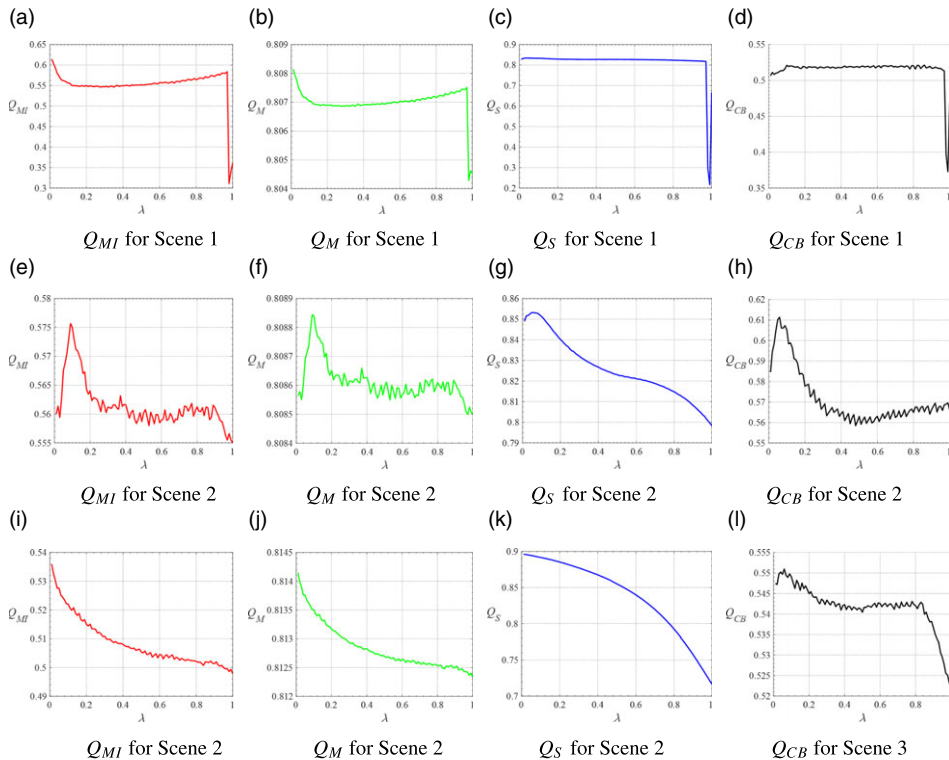


Figure 9. Q_{MI} , Q_M , Q_S , Q_{CB} with the variation of λ .

Q_{CB} for different scenes are presented in Fig. 9. As seen from Fig. 8, with the increase of λ , the target is blurred to produce ghosts, and the background noise is also increased. As seen from Fig. 9, the effect of image fusion is negatively correlated with the increase of λ , and the increase of λ is not conducive to subsequent target detection. Generally, the quality of the fused result decreases with the increase of λ . It is worth noting that the increase of λ will cause the coefficient map to be more sparse, finally indicating the decrease in the fused result.

4.4 Comparison experiments

To measure the effectiveness of the algorithm proposed in this paper more effectively, three image fusion algorithms, including SR [36], lasso-based CSR [23] and CNN [27], are selected to compare with the

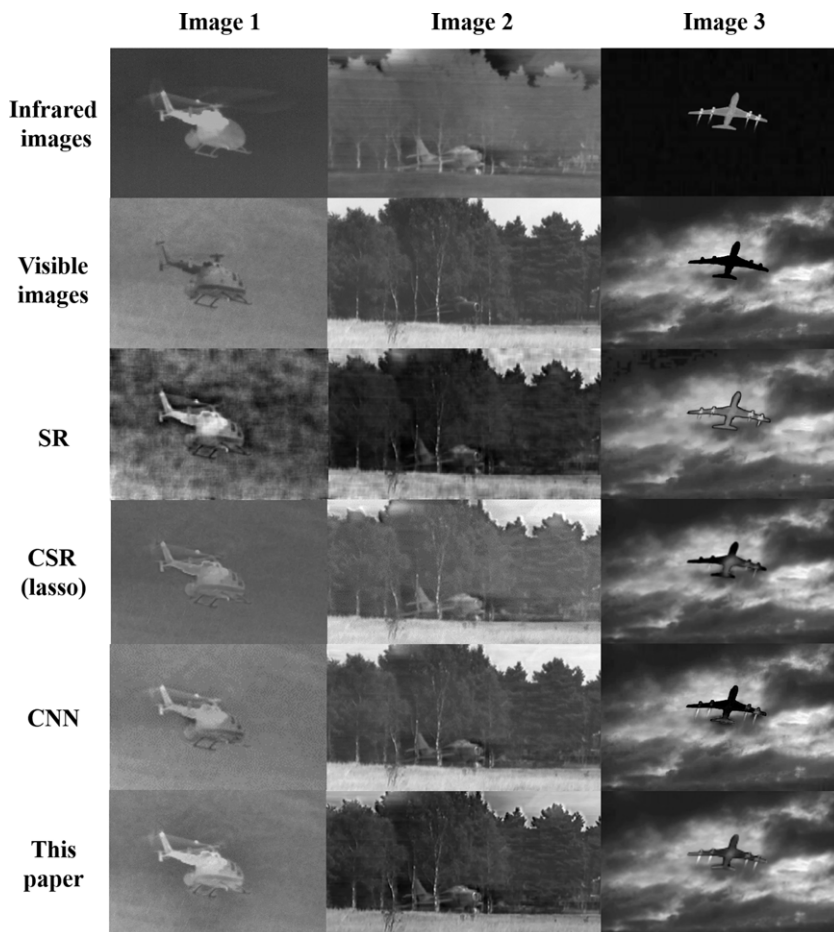


Figure 10. Comparison of fused results.

algorithm proposed in this paper. The fused results are compared in Fig. 10. It can be seen from Fig. 10 that there is too much noise in the fused results obtained by SR-method, which will bring many disadvantages. The fusion method based on CSR (lasso) is more robust than SR but loses structural information such as edges. The fusion method based on CNN is a compromise compared with the above two methods, but the texture features are poorly maintained compared with the proposed method. From the aspect of subjective measurement, the algorithm proposed in this paper is capable of preserving image details while strengthening the object. The comparison of objective measurements containing Q_{MI} , Q_M , Q_S , Q_{CB} are presented in Table 2. Obviously that the objective measurements of the algorithm proposed in this paper outperforms other algorithms in most cases. Based on the fused results based on SR, CSR and CNN in Table 2 (refer to the maximum value among the three), Q_{MI} is increased by nearly 19%, Q_M is increased by nearly 4%, Q_S is increased by nearly 2% and Q_{CB} is increased by nearly 4% on average. Although the proposed algorithm is slightly lower than the comparative method in several aspects, its performance is superior on average and has better robustness.

All the algorithms are implemented in MATLAB 2016b with a 2.4 GHz CPU and 8 GB RAM. The processing time of the four algorithms are counted by using the tic and toc command of MATLAB, and the results are shown in Table 3. As can be seen from Table 3, the algorithm proposed in this paper has obvious advantages in processing time compared with SR and CNN. However, compared with CSR, the processing time slightly increases due to the need for two convolution sparse operations.

Table 2. Comparison results of different methods by Q_{MI} , Q_M , Q_S , Q_{CB}

		SR	CSR(lasso)	CNN	Ours
Q_{MI}	Scene 1	2.412	2.915	2.615	3.212
	Scene 2	1.697	0.897	2.237	2.496
	Scene 3	1.013	1.634	2.254	2.793
Q_M	Scene 1	0.493	0.672	0.691	0.715
	Scene 2	0.587	0.657	0.684	0.728
	Scene 3	0.686	0.682	0.731	0.747
Q_S	Scene 1	2.233	2.637	2.614	2.713
	Scene 2	2.276	1.754	1.989	2.012
	Scene 3	2.348	2.521	2.409	2.706
Q_{CB}	Scene 1	0.349	0.417	0.429	0.413
	Scene 2	0.367	0.208	0.473	0.502
	Scene 3	0.384	0.408	0.415	0.428

Table 3. Processing time comparison of different image fusion methods (seconds)

	SR	CSR (lasso)	CNN	Ours
Scene 1	859	33	132	65
Scene 2	633	29	96	42
Scene 3	757	31	117	58

5.0 Conclusion

In this paper, an elastic net regularisation based CSR is presented for visible and infrared image fusion. Since visible images and infrared images are good at preserving texture and structure, the source images are first decomposed into texture layer and structure layer before image fusion. Then, the structure layers of the source images are fused by CSR using the pre-learned convolutional sparse dictionary filter. The texture layers are fused by utilising the decision map generated during the fusion process of structure layers. Finally, the fused results of the texture and structure layers are synthesised to acquire the fused image. To verify the effectiveness of the proposed algorithm, both subjective and objective measurements are selected to evaluate the effectiveness of the fusion algorithm proposed in this paper. The simulation results reveal that the proposed algorithm can preserve image details while strengthening objects, and it is superior to other image fusion methods in most cases.

Acknowledgements. This study is supported by the National Natural Science Foundation of China (No. 61673211), Open Project Funds of Ministry of Industry and Information Technology for the Key Laboratory of Space Photoelectric Detection and Perception (No. NJ2020021-01).

References

- [1] Yu, X. and Zhang, Y. Sense and avoid technologies with applications to unmanned aircraft systems: Review and prospects, *Progr. Aerosp. Sci.*, 2015, **74**, pp 152–166.
- [2] Mcfadyen, A. and Mejias, L. A survey of autonomous vision-based see and avoid for unmanned aircraft systems, *Progr. Aerosp. Sci.*, 2016, **80**, pp 1–17.
- [3] Fu, Y., Zhang, Y. and Yu, X. An advanced sense and collision avoidance strategy for unmanned aerial vehicles in landing phase, *IEEE Aerosp. Electron. Syst. Mag.*, 2016, **31**, (9), pp 40–52.

- [4] Lin, C.E., Lai, Y.-H. and Lee, F.-J. UAV collision avoidance using sector recognition in cooperative mission to helicopters, 2014 *Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings*, IEEE, 2014, pp F1–1.
- [5] Lin, C.E. and Lai, Y.-H. Quasi-ADS-B based UAV conflict detection and resolution to manned aircraft, *J. Electric. Comput. Eng.*, 2015, pp 1–12.
- [6] Zhang, Z., Cao, Y., Ding, M., Zhuang, L. and Yao, W. An intruder detection algorithm for vision based sense and avoid system, 2016 *International Conference on Unmanned Aircraft Systems(ICUAS)*, IEEE, 2016, pp 550–556.
- [7] Harvey, B. and O’Young, S. Acoustic detection of a fixed-wing UAV, *Drones*, 2018, **2**, (1), pp 4–22.
- [8] Sabatini, R., Gardi, A. and Ramasamy, S. A laser obstacle warning and avoidance system for unmanned aircraft sense-and-avoid, *Appl. Mech. Mater.*, 2014, **629**, pp 355–360.
- [9] Zhang, Z., Cao, Y., Ding, M., Zhuang, L., Yao, W., Zhong, P. and Li, H. Candidate regions extraction of intruder airplane under complex background for vision-based sense and avoid system, *IET Sci. Meas. Technol.*, 2017, **11**, (5), pp 571–580.
- [10] Zhang, Z., Cao, Y., Ding, M., Zhuang, L. and Wang, Z. Spatial and temporal context information fusion based flying objects detection for autonomous sense and avoid, 2018 *International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 2018, pp 569–578.
- [11] Wang, X., Zhou, Q., Liu, Q. and Qi, S. A method of airborne infrared and visible image matching based on hog feature, *Pattern Recognition and Computer Vision*, 2018.
- [12] Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D. and Matas, J. Deblurgan: Blind motion deblurring using conditional adversarial networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp 8183–8192.
- [13] Liu, R., Fan, X., Hou, M., Jiang, Z., Luo, Z. and Zhang, L. Learning aggregated transmission propagation networks for haze removal and beyond, *IEEE Trans. Neural Netw. Learn. Syst.*, 2018, pp 1–14.
- [14] James, J., Ford, J.J. and Molloy, T.L. Learning to detect aircraft for long-range vision-based sense-and-avoid systems, *IEEE Robot. Automat. Lett.*, 2018, **3**, (4), pp 4383–4390.
- [15] Lyu, Y., Pan, Q., Zhao, C., Zhang, Y. and Hu, J., Vision-based UAV collision avoidance with 2D dynamic safety envelope, *IEEE Aerosp. Electron. Syst. Mag.*, 2016, **31**, (7), pp 16–26.
- [16] Zhang, S., Guo, Y., Lu, Z., Wang, S. and Liu, Z. Cooperative detection based on the adaptive interacting multiple model-information filtering algorithm, *Aerosp. Sci. Technol.*, 2019, **93**, p 105310.
- [17] Yu, M., Li, S. and Leng, S. On-board passive-image based non-cooperative space object capture window estimation, *Aerosp. Sci. Technol.*, 2019, **84**, pp 953–965.
- [18] Yang, H., Yu, J., Wang, S. and Peng, X., Design of airborne target tracking accelerator based on KCF, *J. Eng.*
- [19] Lai, J., Ford, J.J., O’Shea, P. and Mejias, L. Vision-based estimation of airborne target pseudobearing rate using hidden Markov model filters, *IEEE Trans. Aerosp. Electron. Syst.*, 2013, **49**, (4), pp 2129–2145.
- [20] Vetrilla, A.R., Fasano, G. and Accardo, D. Attitude estimation for cooperating UAVs based on tight integration of GNSS and vision measurements, *Aerosp. Sci. Technol.*, 2019, **84**, pp 966–979.
- [21] Lee, K., Park, H., Park, C. and Park, S.-Y. Sub-optimal cooperative collision avoidance maneuvers of multiple active spacecraft via discrete-time generating functions, *Aerosp. Sci. Technol.*, 2019, **93**, pp 105298.1–105298.10.
- [22] Gu, S., Zuo, W., Xie, Q., Meng, D., Feng, X. and Zhang, L. Convolutional sparse coding for image super-resolution, *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp 1823–1831.
- [23] Liu, Y., Chen, X., Ward, R.K. and Wang, Z.J., Image fusion with convolutional sparse representation, *IEEE Sig. Process. Lett.*, 2016, **23**, (12), pp 1882–1886.
- [24] Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K. and Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects, *Inform. Fusion*, 2018, **42**, pp 158–173.
- [25] Liu, J., Fan, X., Jiang, J., Liu, R. and Luo, Z., Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion, *IEEE Trans. Circ. Syst. Video Technol.*, 2022, **32**, (1), pp 105–119.
- [26] Singh, S. and Anand, R. Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering, *IEEE Trans. Instrument. Meas.*, 2020, **69**, (6), pp 3855–3865.
- [27] Liu, Y., Chen, X., Cheng, J., Peng, H. and Wang, Z. Infrared and visible image fusion with convolutional neural networks, *Int. J. Wavelets Multiresol. Inf. Process.*, 2018, **16**, (03).
- [28] Liu, Y., Chen, X., Cheng, J. and Peng, H., A medical image fusion method based on convolutional neural networks, 2017 *20th International Conference on Information Fusion*, IEEE, 2017, pp 1–7.
- [29] Liu, R., Liu, J., Jiang, Z., Fan, X. and Luo, Z. A Bilevel integrated model with data-driven layer ensemble for multi-modality image fusion, *IEEE Trans. Image Process.*, 2021, **30**, pp 1261–1274.
- [30] Liu, J., Fan, X., Huang, Z., Wu, G., Liu, R., Zhong, W. and Luo, Z., Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp 5802–5811.
- [31] Zeiler, M.D., Taylor, G.W. and Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning, 2011 *International Conference on Computer Vision*, 2011, pp 2018–2025.
- [32] Xu, L., Yan, Q., Xia, Y. and Jia, J. Structure extraction from texture via relative total variation, *ACM Trans. Graph. (TOG)*, 2012, **31**, (6), pp 139.
- [33] Wohlberg, B. Efficient algorithms for convolutional sparse representations, *IEEE Trans. Image Process.*, 2015, **25**, (1), pp 301–315.
- [34] Madhuri, G. and Negi, A., Discriminative dictionary learning based on statistical methods, *arXiv e-prints*, 2021.

- [35] Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R. and Wu, W. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study, *IEEE Trans. Patt. Anal. Mach. Intell.*, 2011, **34**, (1), pp 94–109.
- [36] Yang, B. and Li, S. Multifocus image fusion and restoration with sparse representation, *IEEE Trans. Instrument. Meas.*, 2009, **59**, (4), pp 884–892.