

Article

A Comprehensive Image Quality Evaluation of Image Fusion Techniques Using X-Ray Images for Detonator Detection Tasks

Lynda Oulhissane ¹, Mostefa Merah ¹ , Simona Moldovanu ^{2,3}  and Luminita Moraru ^{3,4,5,*} 

¹ Laboratory of Signals and Systems (LSS), Faculty of Science and Technology, Abdelhamid Ibn Badis University of Mostaganem, Route Nationale N° 11 Kharouba, Mostaganem 27000, Algeria; lynda.oulhissane.etu@univ-mosta.dz (L.O.); mostefa.merah@univ-mosta.dz (M.M.)

² Department of Computer Science and Information Technology, Faculty of Automation, Computers, Electrical Engineering and Electronics, Dunărea de Jos University of Galați, 800008 Galați, Romania; simona.moldovanu@ugal.ro

³ Modelling & Simulation Laboratory (MSlab), Dunărea de Jos, University of Galați, 800008 Galați, Romania

⁴ Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment, Dunărea de Jos, University of Galați, 800008 Galați, Romania

⁵ Department of Physics, School of Science and Technology, Sefako Makgatho Health Sciences University, Medunsa, Pretoria 0204, South Africa

* Correspondence: luminita.moraru@ugal.ro

Abstract

Purpose: Luggage X-rays suffer from low contrast, material overlap, and noise; dual-energy imaging reduces ambiguity but creates colour biases that impair segmentation. This study aimed to (1) employ connotative fusion by embedding realistic detonator patches into real X-rays to simulate threats and enhance unattended detection without requiring ground-truth labels; (2) thoroughly evaluate fusion techniques in terms of balancing image quality, information content, contrast, and the preservation of meaningful features. **Methods:** A total of 1000 X-ray luggage images and 150 detonator images were used for fusion experiments based on deep learning, transform-based, and feature-driven methods. The proposed approach does not need ground truth supervision. Deep learning fusion techniques, including VGG, FusionNet, and AttentionFuse, enable the dynamic selection and combination of features from multiple input images. The transform-based fusion methods convert input images into different domains using mathematical transforms to enhance fine structures. The Nonsubsampled Contourlet Transform (NSCT), Curvelet Transform, and Laplacian Pyramid (LP) are employed. Feature-driven image fusion methods combine meaningful representations for easier interpretation. Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Random Forest (RF), and Local Binary Pattern (LBP) are used to capture and compare texture details across source images. Entropy (EN), Standard Deviation (SD), and Average Gradient (AG) assess factors such as spatial resolution, contrast preservation, and information retention and are used to evaluate the performance of the analysed methods. **Results:** The results highlight the strengths and limitations of the evaluated techniques, demonstrating their effectiveness in producing sharpened fused X-ray images with clearly emphasized targets and enhanced structural details. **Conclusions:** The Laplacian Pyramid fusion method emerges as the most versatile choice for applications demanding a balanced trade-off. This is evidenced by its overall multi-criteria balance, supported by a composite (geometric mean) score on normalised metrics. It consistently achieves high performance across all evaluated metrics, making it reliable for detecting concealed threats under diverse imaging conditions.



Academic Editor: Pedro Couto

Received: 3 September 2025

Revised: 3 October 2025

Accepted: 8 October 2025

Published: 13 October 2025

Citation: Oulhissane, L.; Merah, M.; Moldovanu, S.; Moraru, L. A Comprehensive Image Quality Evaluation of Image Fusion Techniques Using X-Ray Images for Detonator Detection Tasks. *Appl. Sci.* **2025**, *15*, 10987. <https://doi.org/10.3390/app152010987>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: X-ray images of luggage; detonator images; deep learning image fusion; transform-based image fusion; feature-level image fusion

1. Introduction

Raw luggage X-ray scans show low contrast, significant material overlap, and high noise levels, which make reliable interpretation difficult. Dual-energy acquisitions help distinguish materials but cause color inconsistencies that bias segmentation algorithms. Therefore, image fusion is essential for highlighting hidden threats, such as detonators. In this study, we use a connotative fusion approach that adds realistic detonator patches into genuine luggage X-rays to simulate believable threat scenarios, thereby improving automated detection without ground-truth annotations. Image fusion is a common technique that combines multiple images into a single, clearer, more informative image. Recently, many fusion algorithms and evaluation metrics have been developed. However, any new image fusion algorithm and its performance gains must be carefully evaluated within its specific application context. For better detonator detection, a multi-image fusion method is employed to enhance low-quality X-ray luggage images affected by poor lighting, low contrast, and potential color discrepancies. Poor lighting results in underexposed or overexposed areas, hiding small objects. Low contrast complicates the differentiation of object boundaries and materials. Color inconsistencies, caused by variable dual-energy imaging, lead to misleading material segmentation, which can hinder pattern recognition and detonator detection.

Our work focuses on connotative image fusion, which combines images to provide complementary information. Unlike spatial fusion, connotative fusion improves image quality and interpretability by preserving meaningful features. This study addresses image fusion in airport security, specifically integrating detonator patches into X-ray luggage images. Our goal is to simulate credible threat scenarios for automated threat detection system development. We use a dataset from HTDS (High Tech Detection Systems) with real radiographic luggage images and relevant object patches. This fusion approach creates visually and structurally coherent synthetic images to enhance the robustness of AI models in security screening. We evaluated the fusion performance and identified key challenges and unresolved issues.

The proposed methodology consists of two main stages. The first step involves the integration of detonator patches into X-ray baggage scans using a range of image fusion techniques. The second step involves evaluating the quality of the fused images.

The image fusion techniques include deep learning-based image fusion, transform-based fusion, and feature-level fusion. Deep learning-based image fusion has shown significant promise in multi-image fusion tasks, particularly for enhancing X-ray imagery by preserving critical structural and spectral features, which are key elements for accurately identifying concealed threat objects like detonators. Three major families of methods are generally explored for image fusion, beginning with deep learning-based approaches that leverage convolutional and generative neural networks such as DenseFuse [1,2], FusionGAN [3], and SwinFuse [4,5]. These models are designed to learn cross-modal representations, enabling the seamless and visually coherent fusion of patches into complex radiographic scenes. DenseFuse network preserves fine structural and edge details, works well on multi-spectral and dual-energy inputs, and its densely connected layers improve feature reuse, making fusion more efficient and informative. It enhances structural clarity in low-contrast or noisy X-ray images [1]. FusionGAN network works as a generator that fuses very different images (e.g., severely underexposed and/or noisy input) and as

a discriminator that enforces more realistic and informative outputs [3]. The generator learns to fuse images into a natural-looking, high-quality result while the discriminator tries to distinguish between real and fused images. It generates better texture, contrast, and sharpness preservation and can be trained with various custom fusion losses. The SwinFuse network uses a self-attention mechanism to extract long-range dependencies and fine features from input images [4,5]. The feature maps provided by Swin blocks are fused using a channel-spatial attention mechanism, and a transformer-based decoder then realizes the image reconstruction. It can capture both global and local contexts that are critical in cluttered X-ray scenes and can fuse X-ray inputs with non-aligned illumination levels or non-uniform exposure. It is important to note that DenseFuse is directly related to convolutional neural network (CNN) architectures, whereas FusionGAN employs a generative adversarial framework with a different backbone, but may incorporate VGG-based perceptual loss, making it only indirectly related to CNNs. In contrast, SwinFuse is grounded in a transformer-based architecture, representing a different design philosophy, and is likewise only indirectly connected to CNN models. DenseFuse has a moderate computational cost, FusionGAN is computationally expensive and less stable due to adversarial optimization, and SwinFuse can capture long-range dependencies, but at the cost of higher memory and compute requirements. To address these limitations, this study investigates the fusion capabilities of VGG-based networks, FusionNet, and AttentionFuse models, focusing on their ability to capture multi-scale features, effectively combine complementary information, and enhance image quality in complex X-ray baggage scans.

Transform-based methods form the second category, utilizing multi-resolution decomposition techniques, such as wavelet, contourlet, and shearlet transforms [6], to extract and fuse information across various levels of detail. These approaches are effective in preserving both coarse and fine image features, contributing to a more accurate and detailed fusion outcome.

Finally, feature-based methods employ local visual descriptors like SIFT, SURF, and gray-level co-occurrence matrices (GLCM) [7] to guide the fusion process based on the structural and textural characteristics of the source images. This feature-level guidance ensures that salient visual patterns are retained, supporting the realistic insertion of detonator patches within complex X-ray imagery.

The second step involves evaluating the quality of the fused images. The image quality assessment is carried out using three no-reference objective metrics. Entropy (EN) measures the informational richness of the fused image by assessing the diversity of intensity values, making it particularly useful for quantifying the preservation of fine details [8]. Standard deviation (SD) indicates global contrast and keeps detail from the image, a key factor in visually distinguishing the inserted detonator from the baggage background [9]. The average gradient (AG) assesses the sharpness of edges and how well edges are preserved in the fused image; a high AG suggests better definition of boundaries and intensity transitions, which is essential for segmentation and recognition tasks [10]. To ensure a rigorous and unbiased comparison across the different fusion methods, all evaluation metrics are normalized using the min–max normalization technique [11].

To summarize, we provide a comprehensive analysis and evaluation of various image fusion techniques designed to combine X-ray images of luggage with detonator images. Our goal is to identify the most effective methods for detonator detection. Notably, our framework works without the need for ground truth supervision, enhancing its applicability in real-world scenarios. The primary objective is to identify the fusion methods that best preserve and enhance the structural and textural details relevant to detonator components. The ideal fusion technique should produce output images that display more distinguishable features. Therefore, the main contributions of this paper are threefold:

- (1) to explore image fusion algorithms grounded in machine learning and deep learning theory, focusing on how these approaches enhance fusion performance through feature extraction, representation learning, and adaptive integration;
- (2) to evaluate eleven image fusion algorithms spanning the spatial domain, frequency domain, and deep learning paradigms, using objective image quality metrics such as EN, SD, and AG;
- (3) to analyze the strengths and limitations of each fusion method, offering insights and recommendations on their suitability based on the combined assessment of three no-reference objective metrics, EN, SD, and AG.

2. Related Work

Image fusion aims to combine data from multiple sources or modalities to create a single, more informative image that is easier for humans or algorithms to interpret. This technique is important in security, surveillance, threat detection, and medical analysis, where it improves visual perception while maintaining essential details. In the domain of baggage security, specialized image fusion techniques have been developed to enhance threat detection in complex and cluttered environments.

Liu et al. [12] provide a comprehensive overview of state-of-the-art image fusion techniques and recommend using statistical comparison tests to evaluate the performance of modern image fusion methods. Li et al. [13] present an extensive review study of multi-sensor image fusion, organized into three approaches: pixel-level, feature-level, and decision-level. They also discussed the fusion quality assessment metrics, fusion frameworks, and signal processing approaches, such as multiscale transforms, sparse representations, and neural-network or pulse-coupled- neural-network-based methods. Ma et al. [14] introduced a novel image fusion approach that combines gradient transfer with total variation minimization to merge infrared and visible images. The main idea was to preserve the salient structural features (edges and gradients) from both source images while suppressing noise and redundant information. They formulated the fusion task as an optimization problem and employed gradient transfer to retain strong edge information. The total variation regularization ensured smoothness and consistency in the fused image. Quantitative results demonstrated significant improvements over conventional methods. They reported an average EN increase of approximately 10%, reflecting a higher level of information content in the fused images. Additionally, the AG improved by nearly 20%, indicating superior preservation of structural and textural details. Zhang et al. [15] proposed IFCNN (Image Fusion Convolutional Neural Network), a deep learning-based framework designed to fuse infrared and visible images. Unlike traditional hand-crafted or transform-based fusion methods, IFCNN leverages the representation power of convolutional neural networks to learn and extract meaningful features from source images automatically. The architecture is trained to generate fused outputs that retain the complementary information from both modalities.

Li et al. [2] introduced DenseFuse, a deep learning-based image fusion model designed to merge infrared and visible images using a dense concatenation architecture. The key innovation of DenseFuse lies in its use of densely connected convolutional layers, which enhance feature propagation, encourage feature reuse, and improve gradient flow throughout the network. This enables the model to preserve both intensity from infrared images and fine-grained textures from visible images. It achieved a higher entropy compared to unimodal approaches, indicating a significant enhancement in the richness and utility of the fused images. Xie et al. [16] introduced an end-to-end deep learning model that incorporates 3D convolutional neural networks directly into the image fusion process. The novelty of this study lies in extending attention mechanisms, commonly used in 2D

image pair fusion networks, to 3D, and in introducing a comprehensive set of loss functions that enhance fusion quality and processing speed while avoiding image degradation. Park et al. [17] proposed CMTFusion, a cross-modal transformer-based fusion algorithm for infrared and visible image fusion. Unlike conventional CNN-only approaches that have difficulties capturing long-range dependencies, this hybrid architecture combines CNNs for local feature extraction with transformer modules for modeling global cross-modal interactions. The multiscale feature maps of infrared and visible images were extracted, and the fusion result was obtained based on the spatial-channel information in refined feature maps using a fusion block. This design achieves superior fusion quality and robustness in complex and challenging environments. Wen et al. [18] introduced a generator with separate encoders for infrared and visible images, utilizing multi-scale convolution blocks. Fusion happens via a Cross-Modal Differential Features Attention Module (CDAM), which employs spatial and channel attention paths to compute attention weights from differential features, thus reconstructing fused output progressively. In the specialized field of baggage security, Duan et al. [19] introduced RWSC-Fusion, a region-wise style-controlled fusion network designed for synthesizing and enhancing X-ray images used in security screening. This model focused on localized fusion by applying style control mechanisms to specific regions of interest, namely those containing suspicious or prohibited items, while preserving the overall image consistency. The luminance fidelity and sharp edge details in critical areas were kept, which are essential for accurate threat identification. Quantitative results demonstrated significant improvements in EN and AG within regions containing threat objects, highlighting the method's effectiveness in enhancing image quality where it matters most. Zhao et al. [20] proposed MMFuse, a multi-scale fusion algorithm that integrates morphological reconstruction with membership filtering to improve thermal-visible image fusion. By preserving structural details and enhancing contrast, MMFuse operates across multiple scales to effectively combine both high- and low-frequency components, resulting in more comprehensive and informative fused images. Experimental results demonstrated significant gains in EN and AG metrics, reflecting higher information content and sharper edges in the fused outputs.

Shafay et al. [21] introduced a temporal fusion-based multi-scale semantic segmentation framework designed specifically for detecting concealed threats in X-ray baggage imagery. Their method employed a residual encoder-decoder architecture that integrated features across sequential scan frames, allowing for improved visibility of partially occluded or hidden objects. They demonstrated that the model achieved superior detection accuracy and object localization compared to static image-based techniques. Wei et al. [22] proposed a manual segmentation method for dangerous objects in the SIXRay dataset. They also introduced a composition method for augmenting positive samples using affine transformations and Hue Saturation Value features. This resulted in an algorithm variant called SofterMask RCNN, which improved detection accuracy by 6.2% when used with an augmented image dataset. Dumagpi and Jeong [23] analysed the impact of employing Generative Adversarial Networks (GAN)-based image augmentation on the performance of threat object detection in X-ray images. They synthesised new X-ray security images by combining threat objects with background X-ray images, which were used to augment the dataset. The experiment results demonstrated that image synthesis is an effective approach to addressing the imbalance problem by reducing the false-positive rate (FPR) by up to 15.3%. Based on the real-world finding that items are randomly stacked in luggage scanning, causing penetration effects and heavily overlapping content, Miao et al. [24] proposed a classification approach called class-balanced hierarchical refinement. This approach uses mid-level features in a weakly-supervised manner to accurately localise prohibited items.

They report mean Average Precision improvements across all studied classes (gun, knife, wrench, pliers, scissors).

3. Methodology

This section outlines the framework for low-contrast detonator patch fusion into X-ray luggage images using deep learning, transform-based, and feature-driven methods, using data from the HTDS database. The block diagram of the experimental setup is schematically illustrated in Figure 1.

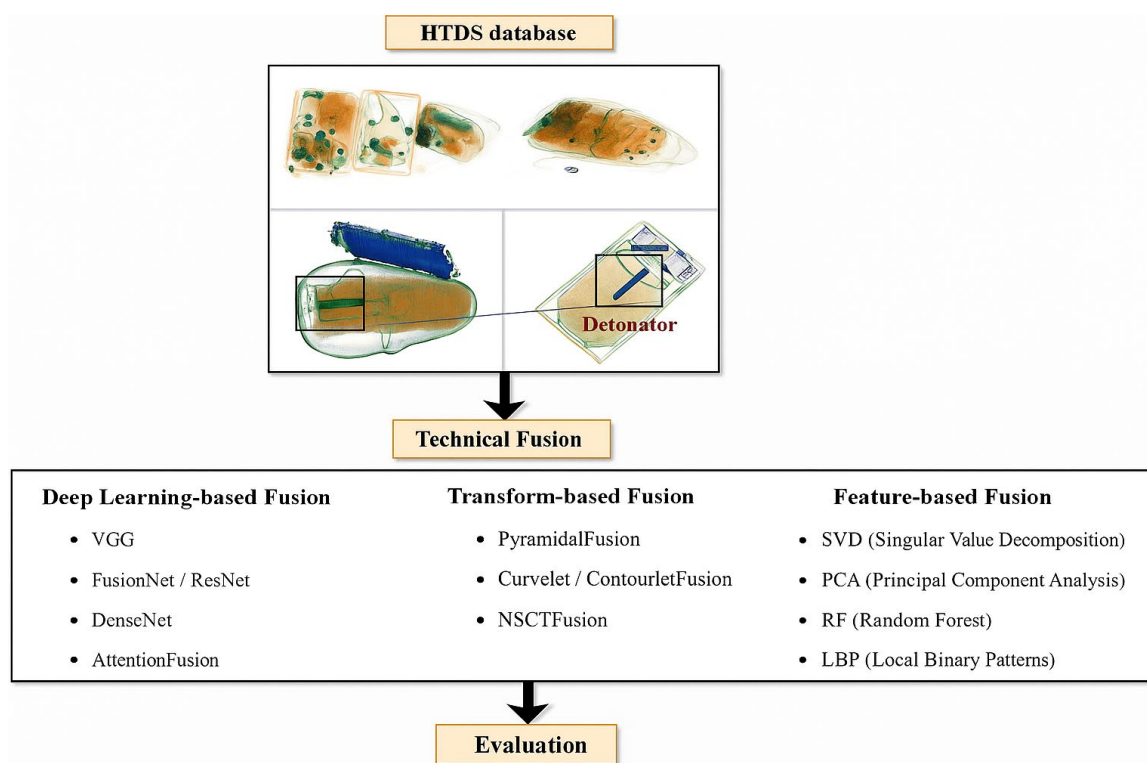


Figure 1. The block diagram of the experimental setup.

3.1. Database

The dataset used in this study was provided by HTDS (High Tech Detection Systems), a French company specializing in the distribution and maintenance of advanced security equipment for passenger screening, baggage and vehicle inspection, and cargo control [25]. The dataset is structured into two main categories: 1000 X-ray scans of luggage and 150 images of detonator objects, as illustrated in Figure 2.



Figure 2. Examples of images from the database.

The baggage scans represent a variety of real-world scenarios, including cluttered and non-threat items, simulating the conditions typically encountered in airport security checkpoints. The detonator images were extracted from radiographic scans and serve

as patch-level components to be inserted into background images for fusion and threat simulation tasks. All images are provided in standard formats (e.g., PNG or TIFF). This dataset supports the development and evaluation of image fusion techniques aimed at enhancing automated threat detection systems, particularly for generating synthetic yet realistic scenarios involving concealed explosive devices.

3.2. Deep Learning—Based Image Fusion

The existing deep learning-based image fusion techniques typically rely on CNN models due to their strong capability in hierarchical feature extraction and representation learning. These models have an architecture with three main stages: feature extraction, fusion, and reconstruction. The feature extraction module extracts deep representations from each input image. The feature fusion module merges the features into a single representation, while the reconstruction module decodes fused features into a final high-resolution image (Figure 3).

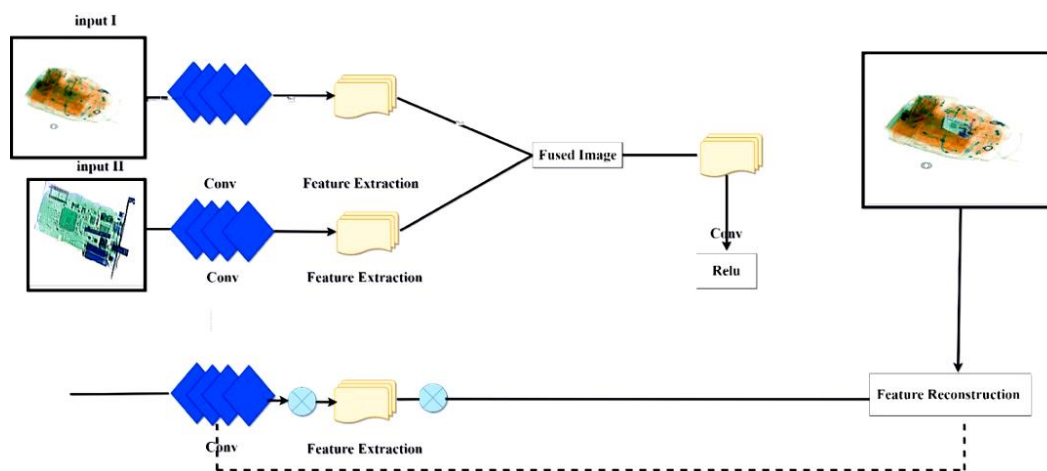


Figure 3. Deep learning-based image fusion.

The VGG-based deep learning image fusion technique uses pretrained CNNs (VGG-19) as a feature extractor. Features from source images are extracted from intermediate convolutional layers and are fused using rules like averaging, max-selection, or learned attention. The fused features are used to reconstruct the output image. The main advantage of using VGG networks for X-ray detonator detection lies in their ability to capture multi-level image details by combining low-level textures with high-level structural features [26]. This hierarchical representation is particularly effective for detecting small objects, such as detonators or wires, within complex baggage imagery. There are, however, certain limitations. For instance, VGG networks are not inherently optimized for the X-ray domain and often require domain-specific fine-tuning to achieve optimal performance. In addition, these approaches rely on a decoder for image reconstruction, which adds to the model complexity.

Although FusionNet is not usually recognized in the context of X-ray baggage inspection, it offers notable advantages over standard CNNs by extracting features from multiple modalities and employing a fusion strategy that simultaneously preserves spatial structures and intensity information. This results in a single high-quality fused image with richer detail and improved representation of critical objects. FusionNet is based on a CNN architecture block structure and performs end-to-end feature extraction using separate encoders for each input image, fusion by combining features from both encoders, and reconstruction with a decoder that rebuilds the fused image from the fused features [27]. It learns the entire process from scratch. FusionNet offers several advantages, including the

ability to function without pretraining, high adaptability to diverse image modalities, and strong performance in capturing fine-grained features such as small objects, edges, and textures. However, a key limitation is its limited semantic understanding, which may result in the loss of critical threat-related details, particularly in complex scenes where contextual interpretation is essential.

AttentionFuse is a neural network that uses attention mechanisms (channel, spatial, or both) to select and combine features from multiple input images, dynamically [28]. It improves upon traditional CNN-based fusion models (like FusionNet) by allowing the model to focus on more informative areas of the images. AttentionFuse introduces attention modules in the encoder or fusion stage to learn which features are important to retain and suppress irrelevant or noisy details. Thus, it can utilize spatial attention to enhance sharp regions (e.g., detonator boundaries), channel attention to assign more weight to fine textures, or cross-attention to link features across images. The main advantages include adaptive feature selection and robustness to noise and artifacts, achieved by learning to downweight irrelevant or redundant features. However, these benefits come at the cost of increased computational complexity compared to simpler CNN architectures, and the model requires careful tuning of hyperparameters to achieve optimal performance.

3.3. Transform-Based Image Fusion

Transform-based image fusion converts input images from the spatial domain to the frequency domain, in our case, using mathematical transforms. It decomposes images into multiple sub-bands, each capturing specific features such as edges, textures, or smooth regions at various scales and orientations. The corresponding sub-bands from the input images are then fused using appropriate fusion rules to retain the most informative features. Finally, an inverse transform is applied to reconstruct the fused image in the spatial domain, combining the complementary information from the sources into a single, enhanced output (Figure 4).

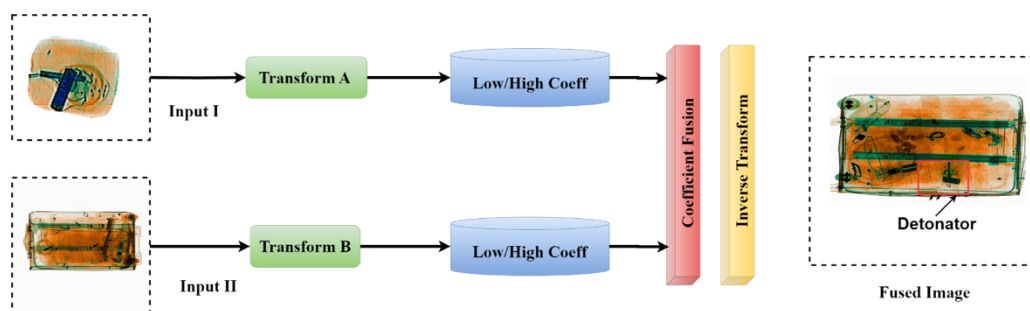


Figure 4. Transform-based image fusion.

The Non-subsampled Contourlet Transform (NSCT) is a multiscale and multidirectional approach used for capturing curved edges and complex textures better than wavelets. [29]. The Curvelet Transform also captures edge information along curves more effectively than traditional wavelets. Unlike wavelets, which are effective at capturing point singularities but struggle with curved discontinuities, curvelets provide sparse representations of objects with curved edges. This makes them particularly advantageous for identifying small and irregularly shaped items, like detonators, in complex backgrounds or cluttered scenes.

The Laplacian Pyramid (LP) transformation decomposes an image into a series of band-pass filtered images with multiple resolution levels, forming a hierarchical pyramid structure [30]. This multiscale representation enables effective separation of image details at different spatial frequencies. The fusion process is performed level-by-level, allowing

salient features from multiple input sources to be selectively integrated at each resolution. LP-based fusion is particularly effective for enhancing the visibility of important structures in poorly illuminated X-ray images, as it emphasizes fine details and edge information while preserving the overall image integrity.

3.4. Feature-Based Image Fusion

Feature-driven image fusion methods focus on combining salient feature representations extracted from the input images, rather than directly operating on raw pixel values. These methods extract specific features, such as edges, corners, gradients, textures, or even deep features, using hand-crafted or learned operators [31]. The extracted feature maps are fused according to predefined rules, highlighting complementary and informative structures. Subsequently, reconstruction is performed to generate a high-quality fused image. Feature-driven approaches are generally fast, lightweight, and yield interpretable intermediate representations, making them suitable for real-time applications and easier analysis compared to computationally intensive deep learning-based methods (Figure 5). Their main limitations lie in the reliance on hand-crafted fusion rules, which are often suboptimal and lack adaptability. Additionally, these methods tend to be less robust in the presence of extreme noise or low signal-to-noise ratios.

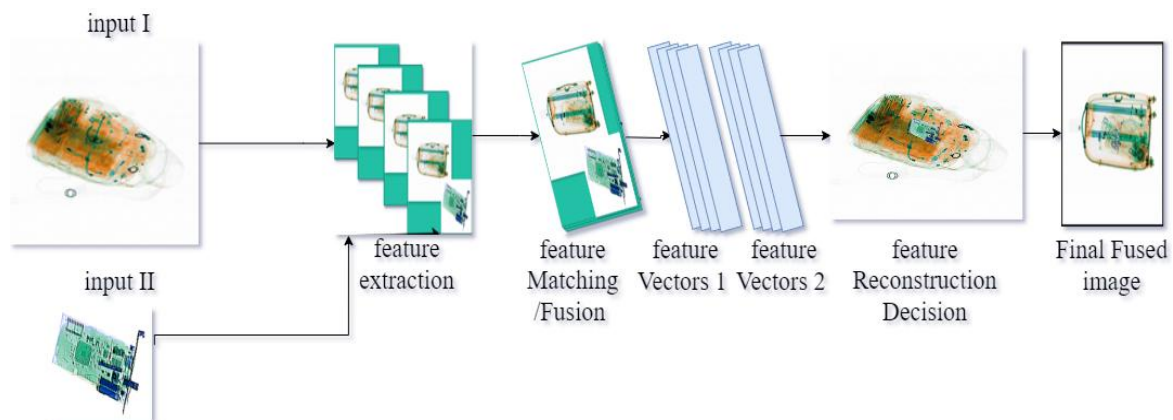


Figure 5. Feature-based image fusion.

Singular Value Decomposition (SVD)-based image fusion is an unsupervised approach that isolates significant structural and textural information by retaining dominant singular values [32]. This makes it well-suited for edge preservation and relatively resilient to noise, as smaller, noise-related singular values are typically discarded. However, since SVD is a global transformation applied to the entire image, it may overlook local variations and adaptive fusion opportunities. Furthermore, although it effectively eliminates minor noise, it remains susceptible to intense or structured noise, potentially compromising the quality of the fusion process.

Principal Component Analysis (PCA)-based image fusion transforms input images into a set of uncorrelated components, prioritizing those that capture the highest variance [33]. By reducing redundancy and discarding less informative data, PCA efficiently identifies the most significant features contributing to structural and textural differences between images. It is computationally lightweight, unsupervised, and automatically determines the most informative projection directions, making it suitable for real-time or resource-constrained applications.

Random Forest (RF)-based image fusion leverages ensemble learning to select and combine the most relevant features from multiple source images [34]. By constructing a large number of decision trees and aggregating their outputs, RF is capable of internally

ranking feature importance, thereby identifying and prioritizing the most informative attributes for fusion. This helps enhance the quality of the fused image by preserving critical structural and contextual information while suppressing redundant or irrelevant features. Additionally, RF's robustness to noise, non-linearity, and overfitting makes it well-suited for complex fusion tasks involving heterogeneous data sources.

LBP (Local Binary Pattern) in feature-based image fusion is used to extract and compare texture details from different source images before fusion [35]. By encoding local spatial patterns into binary representations, LBP enables the identification of texture-rich or feature-rich regions. The resulting fused image highlights detailed local textures and structural information. As an unsupervised method, LBP is computationally efficient and easy to implement. However, it operates on a local level and therefore lacks sensitivity to global structural context or semantic content, which may limit its performance in tasks that require high-level understanding.

3.5. Parameters Evaluations

Quality assessment of image fusion algorithms remains challenging, primarily due to the scarcity of benchmark datasets with reliable ground-truth fused images. Objective quantitative metrics are indispensable for systematically and rigorously comparing different fusion algorithms. These metrics assess factors such as spatial resolution, contrast preservation, and information retention, providing standardized benchmarks to guide the development and optimization of fusion methods (Table 1).

Table 1. Key Performance Metrics for Evaluating Fused Images.

Metric	Description	Purpose in Fusion Quality
Entropy (EN)	$EN = p_i \sum_{i=0}^{255} \log_2(p_i)$ p_i is the probability of intensity level i	Higher entropy indicates richer details and texture.
Standard Deviation (SD)	$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \mu)^2}$ μ is the mean of the image	Higher SD reflects better contrast and dynamic range.
Average Gradient (AG)	$AG = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\frac{[I(i+1, j) - I(i, j)]^2}{[I(i, j+1) - I(i, j)]^2}}$	Higher AG implies sharper edges and clearer details.

EN quantifies the amount of information present in an image. An increase in entropy after fusion indicates that the fused image contains more information and thus reflects improved fusion performance. SD reflects the spread of pixel intensities, including both signal and noise, and, in low-noise conditions, serves as a robust measure of image contrast. A fused image with higher contrast will exhibit a larger SD. AG evaluates the amount of edge and texture detail present in a fused image; a higher AG value signifies better preservation of fine texture information [36,37].

To facilitate direct comparison among EN, SD, and AG (which have different units and value ranges), each metric was normalized using the min–max normalization method:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the raw metric value, x_{max} and x_{min} are the maximum and minimum values of that metric across all fused images [11]. After normalization, each metric is scaled to a range between 0 (lowest performance) and 1 (highest performance), enabling fair and direct comparison. The resulting heatmap of normalized evaluation metrics provides an intuitive visual overview, allowing for quick identification of top-performing methods.

To move beyond a simple metrics analysis and avoid the “best on every single metric” approach, the geometric mean of normalised EN, SD, and AG is used to summarise the overall balance.

4. Results and Discussion

The proposed approach conducts a comprehensive cross-method evaluation on spatial resolution, contrast preservation, and information retention in fused images.

We utilised MATLAB 2019 for generating the derived images and employed PyTorch 1.5.0 as the machine learning and deep learning framework for model development. The Google Collaboratory environment, equipped with the A100 GPU hardware accelerator, was utilised to minimise runtime. Pre-trained CNNs were trained and tested in the software environment using Python (3.0), scikit-learn (1.5.2), TensorFlow (2.18.0), and Keras (3.6.0).

To verify the significance of the results from the analyzed method, we conducted a significance test along with a quantitative evaluation of the outcomes. Table 2 shows the performance of different fusion techniques used to combine detonator images into dual-energy X-ray scans. Higher values indicate better fusion quality regarding detail preservation, contrast, and sharpness.

Table 2. Results of Fusion of Detonator Images into Dual-Energy X-ray Baggage Scans, grouped by technique used: Deep learning, Transform-based, and Feature-driven methods.

	Methods	Entropy (EN)	Standard Deviation (SD)	Average Gradient (AG)
Deep learning	VGG19	5.4017	54.7072	7.4611
	FusionNet/ResNet	5.1210	51.8748	7.3522
	DenseFuse	5.1315	52.1295	7.3215
	AttentionFuse	5.1599	51.2348	7.0257
Transform-based	Fusion curvelet/contourlet	5.1255	51.0655	7.3263
	NSCT	5.1666	51.4534	7.3816
	LP	5.0128	59.5764	12.1236
Feature-driven	SVD	5.1852	57.9742	8.3136
	PCA	5.1031	75.4174	7.2482
	RF	4.8939	56.4910	8.5976
	LBP	4.9844	58.5407	14.2265

The highest EN is observed in the VGG-based fusion (5.4017), indicating superior retention of global detail, followed by NSCT (5.1666) and Attention Fusion (5.1599). In contrast, RF records the lowest EN (4.8939), suggesting potential information loss. PCA achieves the highest SD (75.4174), implying the most pronounced intensity differentiation between the detonator patch and its background. It is followed by LP (59.5764) and LBP (58.5407), while curvelet/contourlet-based fusion reports the lowest contrast (51.0655). Also, LBP leads significantly with an AG of 14.2265, which is over 16% higher than the next best method and LP (12.1236), highlighting its superior edge fidelity around the detonator. Classical approaches such as PCA and AttentionFuse exhibit more modest AG values (7.2482 and 7.0257, respectively).

Collectively, these results demonstrate that no single fusion algorithm excels across all evaluation metrics. VGG-based fusion is optimal for maximizing information content, Fusion PCA is best suited for enhancing contrast, and Fusion LBP excels in preserving edge

and texture detail. However, for applications requiring a balanced trade-off, such as reliable detection of concealed threats under diverse imaging conditions, Fusion LP offers the most versatile performance, with strong results in EN (5.0128), SD (59.5764), and AG (12.1236).

Figure 6 presents a normalized heatmap of the objective fusion metrics across eleven fusion algorithms. The heatmap visualizes the relative performance of various image fusion algorithms, while normalization facilitates direct comparison across metrics with different scales. EN peaks for the VGG-based method (normalized = 1.00), indicating superior retention of global details. SD, a proxy for contrast, is maximized by Fusion PCA (1.00), while Fusion LBP achieves the highest AG score (1.00), reflecting its exceptional preservation of fine textures and edges.

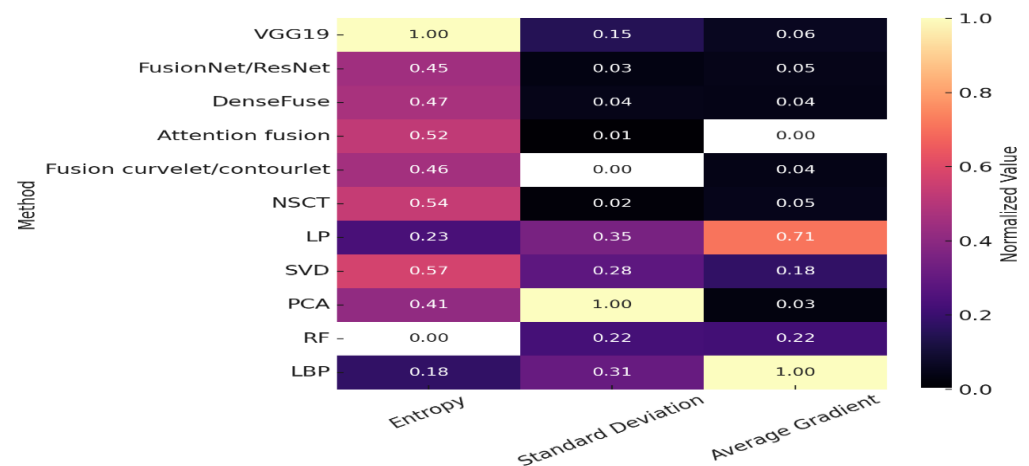


Figure 6. Heatmap of normalized evaluation metrics (Entropy, Standard Deviation, Average Gradient) across different image fusion methods.

These results indicate that Transform-based approaches provide a balanced performance across metrics. In the normalized heatmap, the fusion LP consistently ranks among the top three methods, with normalized scores of 0.23 for EN, 0.35 for SD, and 0.71 for AG and a geometric mean of 0.387, consistent with its strong cross-metric balance. In contrast, NSCT performs well in information retention (high EN) but falls short in both contrast and edge detail; the geometric mean is 0.047. Feature-based methods tend to specialize in specific aspects: PCA emphasizes contrast (high SD) but sacrifices fine detail, as indicated by its low AG score (0.03); its geometric mean is 0.230. Conversely, LBP prioritizes texture preservation with a leading AG score (1.00) but captures less global information (0.18 EN); its geometric mean is 0.382. Deep learning-based models, including VGG, FusionNet/ResNet, DenseFuse, and AttentionFuse, consistently achieve high entropy values (≥ 0.45). This indicates strong global feature retention. However, they generally underperform in contrast and edge detail, with normalized SD and AG scores typically ≤ 0.07 . All methods that perform exceptionally well on a single metric underperform on the composite geometric mean of the three min–max normalised metrics due to trade-offs with the other metrics.

The analysis reveals that VGG-based fusion is best suited for applications where maximizing information content is the primary objective. Fusion PCA is ideal for scenarios requiring enhanced contrast, while Fusion LBP and Fusion LP are preferable when preserving edge and texture details is critical. Among all methods, Fusion LP delivers the most balanced performance across entropy, contrast, and gradient metrics, making it the recommended choice for applications that demand a well-rounded fusion of global information, local contrast, and fine structural detail.

To nuance and enhance the comparison across diverse fusion strategies, we visualize the overall distribution, central tendency, and variability of each evaluation metric across all eleven image fusion methods (Figure 7). These boxplots provide a comprehensive view of performance spread, highlighting not only the median and interquartile range but also potential outliers and the relative consistency of each method.

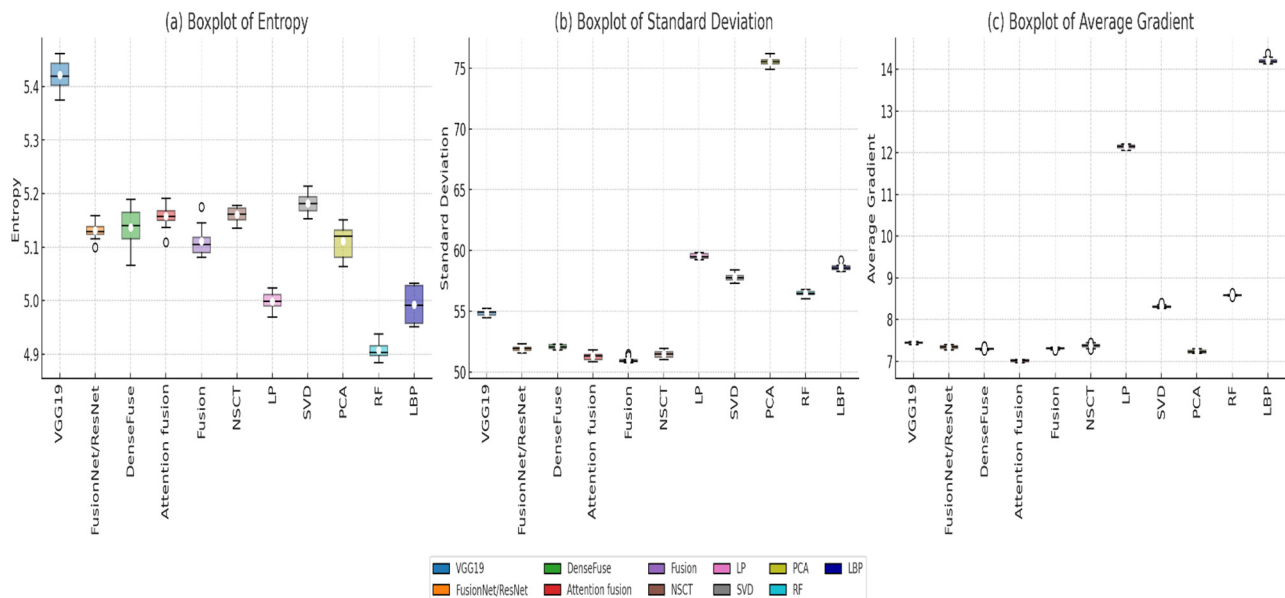


Figure 7. Boxplots of (a) Entropy, (b) Standard Deviation, and (c) Average Gradient across all fusion methods.

The average EN distribution (Figure 7a) is tightly concentrated, with a median of approximately 5.125 and an interquartile range (IQR) of [5.058, 5.163], indicating consistent performance across most methods in terms of information preservation. Correlated to data in Table 2, VGG-based fusion (5.402) emerges as a clear high-end outlier, confirming its dominant performance in retaining global information and structural richness. On the opposite end, Fusion RF (4.894) falls below the lower whisker, suggesting it persistently underperforms in preserving entropy, likely due to its limited capacity to capture diverse image features.

The SD boxplot (Figure 7b) shows a broader performance spread, with an average median near 54.707 and an IQR of [51.664, 58.257]. This metric reflects the algorithms' ability to enhance intensity contrast between the fused image elements. Correlated to data in Table 2, Fusion PCA (75.417) is a pronounced upper outlier, reinforcing its strength in boosting contrast far beyond the other techniques, especially beneficial for applications requiring high visual differentiation between foreground and background. In contrast, Curvelet/Contourlet and NSCT-based fusion methods cluster near the lower whisker (≈ 51), suggesting a more conservative approach to contrast enhancement, which may lead to smoother but less visually distinct outputs.

AG values are moderately dispersed, with a median of 7.382 and an IQR of [7.324, 8.456], representing variation in how well fine textures and edges are preserved (Figure 7c). Correlated to data in Table 2, Fusion LBP (14.226) stands out as a significant high-end outlier, highlighting its superior ability to retain edge sharpness and micro-textural details, that is critical in applications where feature clarity is essential, such as object detection. Conversely, Attention Fusion (7.026) lies at the bottom whisker, indicating it delivers the least gradient sharpness, which could be a limiting factor in edge-sensitive tasks.

Thus, VGG excels in information content, PCA dominates contrast, and LBP leads in texture and edge preservation. Fusion Pyramidal, while not an outlier in any single metric, consistently ranks near the upper quartile across all three, further supporting its role as the most balanced and versatile fusion method among those evaluated.

5. Conclusions

This study focuses on airport security screening and employs real luggage X-rays alongside detonator patches. Importantly, fusion is performed without ground-truth supervision, which facilitates practical deployment by eliminating the annotation bottleneck. The study explores various methods, including feature-driven, transform-based, and deep learning approaches, offering integrators a range of complexity-performance trade-offs. To comprehensively assess the effectiveness of each fusion method, we evaluate the performance across three critical image quality criteria that jointly determine suitability for detonator detection tasks. By analyzing the trade-offs between these metrics, we identify the relative strengths and weaknesses of each fusion technique and offer targeted recommendations based on their ability to balance these competing objectives. Taken together, the qualitative and quantitative analyses reveal that no single algorithm consistently outperforms across all three criteria. VGG-based fusion is most effective at maximizing information content (EN), Fusion PCA delivers superior contrast enhancement (SD), while Fusion LBP excels in preserving edge sharpness and texture detail (AG). The Laplacian Pyramid fusion method emerges as the most versatile choice for applications demanding a balanced trade-off. This is evidenced by its overall multi-criteria balance, supported by a composite (geometric mean) score on normalised metrics. It consistently achieves high performance across all evaluated metrics, making it reliable for detecting concealed threats under diverse imaging conditions. Moreover, deep learning-based fusion algorithms face several challenges, including the generation of suitable training datasets and the selection of appropriate architectures tailored to specific fusion tasks. Ensuring that the chosen deep learning models align with the unique requirements of each fusion problem remains a key concern in advancing this technology. Classical methods are already efficient. Future studies will focus on optimising deep fusion methods through techniques such as quantization, pruning, and GPU acceleration. Furthermore, large-scale open datasets are required for robust training and certification.

Author Contributions: Conceptualization, L.O., M.M. and L.M.; methodology, L.O., M.M. and L.M.; software, L.O., M.M. and S.M.; validation, L.O., M.M., S.M. and L.M.; formal analysis, L.M.; investigation M.M. and L.M.; writing—original draft preparation, L.O., M.M. and L.M.; writing—review and editing, L.M. and S.M.; supervision, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. Access to the source scans can be granted upon request, subject to a confidentiality agreement with High-Tech Defence Systems (HTDS).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Gong, J.; Sun, D.; Xu, L.; Peng, R. YOLO-DenseFuse: A Visible-Infrared Feature Fusion Framework for Intelligent Camouflage Target Detection in Complex Environments. In Proceedings of the AISNS '24: Proceedings of the 2024 2nd International Conference on Artificial Intelligence, Systems and Network Security, Mianyang, China, 20–22 December 2024; pp. 207–214. [\[CrossRef\]](#)
- Li, H.; Wu, X.J.; Kittler, J. DenseFuse: A fusion approach for infrared and visible images. *IEEE Trans. Image Process.* **2018**, *28*, 2614–2623. [\[CrossRef\]](#)
- Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [\[CrossRef\]](#)
- Li, L.; Shi, Y.; Lv, M.; Jia, Z.; Liu, M.; Zhao, X.; Zhang, X.; Ma, H. Infrared and Visible Image Fusion via Sparse Representation and Guided Filtering in Laplacian Pyramid Domain. *Remote Sens.* **2024**, *16*, 3804. [\[CrossRef\]](#)
- Wang, Z.; Chen, Y.; Shao, W.; Li, H.; Zhang, L. SwinFuse: A Residual Swin Transformer Fusion Network for Infrared and Visible Images. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5016412. [\[CrossRef\]](#)
- Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [\[CrossRef\]](#)
- Yashika, B.L.; Durdi, V.B. Image Fusion in Multi-View Videos Using SURF Algorithm. Lecture Notes in Networks and Systems. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS 2020), Jaipur, India, 11–12 December 2020; Kaiser, M.S., Xie, J., Rathore, V.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 190, pp. 1061–1071. [\[CrossRef\]](#)
- Piella, G. A general framework for multiresolution image fusion. *Inf. Fusion* **2003**, *4*, 259–280. [\[CrossRef\]](#)
- Yang, G.; Li, J.; Lei, H.; Gao, X. A multi-scale information integration framework for infrared and visible image fusion. *Neurocomputing* **2024**, *600*, 128116. [\[CrossRef\]](#)
- Rao, Y.; Gao, Y.; Huang, Y. TGFuse: Transformer and GAN for Image Fusion. *arXiv* **2022**. [\[CrossRef\]](#)
- Komal, H.; Chander, K. Analysis of Different Fusion and Normalization Techniques. *IJARCCCE Int. J. Adv. Res. Comput. Commun. Eng.* **2018**, *7*, 363–368.
- Liu, Z.; Blasch, E.; Bhatnagar, G.; John, V.; Wu, W.; Blum, R.S. Fusing synergistic information from multi-sensor images: An overview from implementation to performance assessment. *Inf. Fusion* **2018**, *42*, 127–145. [\[CrossRef\]](#)
- Li, B.; Xian, Y.; Zhang, D.; Su, J.; Hu, X.; Guo, W. Multi-Sensor Image Fusion: A Survey of the State of the Art. *J. Comput. Commun.* **2021**, *9*, 73–108. [\[CrossRef\]](#)
- Ma, J.; Chen, C.; Li, C.; Juang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [\[CrossRef\]](#)
- Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [\[CrossRef\]](#)
- Xie, X.; Qingyan, J.; Chen, D.; Guo, B.; Li, P.; Zhou, S. StackMFF: End-to-end multi-focus image stack fusion network. *Appl. Intell.* **2025**, *55*, 503. [\[CrossRef\]](#)
- Park, S.; Vien, A.G.; Lee, C. Cross-Modal Transformers for Infrared and Visible Image Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 770–785. [\[CrossRef\]](#)
- Wen, Y.; Liu, W. DAGANFuse: Infrared and Visible Image Fusion Based on Differential Features Attention Generative Adversarial Networks. *Appl. Sci.* **2025**, *15*, 4560. [\[CrossRef\]](#)
- Duan, L.; Wu, M.; Mao, L.; Yin, J.; Xiong, J.; Li, X. RWSC-Fusion: Region-Wise Style-Controlled Fusion Network for the Prohibited X-Ray Security Image Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 22398–22407. [\[CrossRef\]](#)
- Zhao, L.; Yang, H.; Dong, L.; Zheng, L.; Asiya, M.; Zheng, F. MMFuse: A multi-scale infrared and visible images fusion algorithm based on morphological reconstruction and membership filtering. *IET Image Process.* **2023**, *17*, 1126–1148. [\[CrossRef\]](#)
- Shafay, M.; Hassan, T.; Damiani, E.; Werghi, N. Temporal Fusion Based Mutli-scale Semantic Segmentation for Detecting Concealed Baggage Threats. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–21 October 2021; pp. 232–237. [\[CrossRef\]](#)
- Wei, Q.; Ma, S.; Tang, S.; Li, B.; Shen, J.; Xu, Y.; Fan, J. A deep learning-based recognition for dangerous objects imaged in X-ray security inspection device. *J. X-Ray Sci. Technol.* **2022**, *31*, 13–26. [\[CrossRef\]](#)
- Dumagpi, J.K.; Jeong, Y.-J. Evaluating GAN-Based Image Augmentation for Threat Detection in Large-Scale Xray Security Images. *Appl. Sci.* **2021**, *11*, 36. [\[CrossRef\]](#)
- Miao, C.; Xie, L.; Wan, F.; Su, C.; Liu, H.; Jiao, J.; Ye, Q. SIXray: A Large-Scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2114–2123. [\[CrossRef\]](#)

25. Oulhissane, L.; Merah, M.; Moldovanu, S.; Moraru, L. Enhanced detonators detection in X-ray baggage inspection by image manipulation and Deep Convolutional Neural Networks. *Sci. Rep.* **2023**, *13*, 14262. [\[CrossRef\]](#)
26. Andriyanov, N. Deep Learning for Detecting Dangerous Objects in X-rays of Luggage. *Eng. Proc.* **2023**, *33*, 20. [\[CrossRef\]](#)
27. Minh, Q.T.; Grant Colburn, H.D.; Jeong Won-Ki, J. FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics. *Front. Comput. Sci.* **2021**, *3*, 34. [\[CrossRef\]](#)
28. Dai, Y.; Gieseke, F.; Oehmcke, S.; Wu, Y.; Barnard, K. Attentional Feature Fusion. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 3559–3568. [\[CrossRef\]](#)
29. Goyal, S.; Singh, V.; Rani, A.; Yadav, N. Multimodal image fusion and denoising in NSCT domain using CNN and FOTGV. *Biomed. Signal Process. Control.* **2022**, *71*, 103214. [\[CrossRef\]](#)
30. Li, L.; Lv, M.; Jia, Z.; Jin, Q.; Liu, M.; Chen, L.; Ma, H. An Effective Infrared and Visible Image Fusion Approach via Rolling Guidance Filtering and Gradient Saliency Map. *Remote Sens.* **2023**, *15*, 2486. [\[CrossRef\]](#)
31. Tikariha, D.; Krishnan, P.T.; Kumar, T.S. Feature-Driven Fusion of Thermal and Visible Images for Face Analysis. In Proceedings of the 4th International Conference on Artificial Intelligence and Signal Processing (AISP), Vijayawada, India, 22–24 November 2024; pp. 01–05. [\[CrossRef\]](#)
32. Pandey, J.P.; Singh Umrao, L. Digital Image Processing using Singular Value Decomposition. In Proceedings of the 2nd International Conference on Advanced Computing and Software Engineering (ICACSE), Sultanpur, India, 8–9 February 2019. [\[CrossRef\]](#)
33. Kumar, B.; Rajput, S.S. PCA-IFNNR: Principal component analysis based image fusion directed nearest neighbor representation for dark face image super resolution. *Multimed. Tools Appl.* **2025**, *84*, 20587–20605. [\[CrossRef\]](#)
34. Kausar, N.; Majid, A. Random forest-based scheme using feature and decision levels information for multi-focus image fusion. *Pattern Anal. Appl.* **2016**, *19*, 221–236. [\[CrossRef\]](#)
35. Yin, W.; Zhao, W.; You, D.; Wang, D. Local binary pattern metric-based multi-focus image fusion. *Opt. Laser Technol.* **2019**, *110*, 62–68. [\[CrossRef\]](#)
36. Shama, S.; Padmalatha, L. Performance comparison of image fusion using singular value decomposition. *Int. J. Innov. Res. Sci. Eng. Technol.* **2015**, *4*, 8065–8073. [\[CrossRef\]](#)
37. Yang, K.; Xiang, W.; Chen, Z.; Zhang, J.; Liu, Y. A review on infrared and visible image fusion algorithms based on neural networks. *J. Vis. Commun. Image Rep.* **2024**, *101*, 104179. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.