An Improved Multi-Exposure Image Fusion Method for Intelligent Transportation System

Mingyu Gao, Junfan Wang, Yi Chen, Chenjie Du, Chao Chen and Yu Zeng

Abstract: In this paper, an improved multi-exposure image fusion method for intelligent transportation systems (ITS) is proposed. Further, a new multi-exposure image dataset for traffic signs, TrafficSign, is presented to verify the method. In the intelligent transportation system, as a type of important road information, traffic signs are fused by this method to obtain a fused image with moderate brightness and intact information. By estimating the degree of retention of different features in the source image, the fusion results have adaptive characteristics similar to that of the source image. Considering the weather factor and environmental noise, the source image is preprocessed by bilateral filtering and dehazing algorithm. Further, this paper uses adaptive optimization to improve the quality of the output image of the fusion model. The qualitative and quantitative experiments on the new dataset show that the multi-exposure image fusion algorithm proposed in this paper is effective and practical in the ITS.

Keywords: multi-exposure image fusion; intelligent transportation system; adaptive optimization

1. Introduction

With the rapid development of digital image technology, more and more digital image technologies will be applied to intelligent transportation systems [1]. As an important part of a smart city, intelligent transportation systems (ITS) are the effective comprehensive application of advanced science and technology in the field of transportation. It can strengthen the connection between vehicles, roads, and users, thereby forming a comprehensive transportation system that guarantees safety, improves efficiency, and saves energy. In the field of intelligent transportation, one of the very important pieces of information is the road signs taken by the camera. For different exposure environments, multi-exposure fusion technology [2] is very important to obtain high-quality road sign images.

Traditional fusion methods include three main steps: image transformation, activity level measurement, and fusion rule design [3]. However, it is time-consuming, expensive, and difficult to design the feature extraction and fusion rules. Gu et al. [4] proposed a new iterative correction method for gradient field by using quadratic average filtering and nonlinear compression. Mertens et al. [5] proposed a multi-resolution fusion method based on the Laplacian pyramid.

The trained network can automatically extract features from images and merge features without manual participation in transformation and activity level measurement. However, the ground-truth of supervised learning is also difficult to obtain in practical application. Therefore, this paper uses a unified unsupervised image fusion network, which does not need the ground-truth to generate fusion results and can fuse the source image adaptively [6].

At present, in the field of intelligent transportation, the light and the ambient noise will produce certain interference to the image. Considering the influence of extreme weather,
the irrelevant information in the image is eliminated by preprocessing, the detectability of important information is enhanced, and the reliability of subsequent feature extraction is improved. The fusion results are adaptively optimized to ensure that the output image can reflect the retained important information to the greatest extent. The main contributions of our work are summarized as follows:

1. By calculating the retention degree of different features in the source image, the fusion result is adaptively similar to the source image. The end-to-end unsupervised image fusion network is used to overcome the problems that the ground-truth cannot achieve and the reference metric is not available in most image fusion problems;

2. Multi-exposure image fusion is applied to the transportation field. The source images are preprocessed according to weather conditions and environmental noise, and the fusion result is adaptively optimized. The final generated image can reflect the source image information and has great practicability and effectiveness;

3. We have released a new multi-exposure image dataset, TrafficSign [7], which is aimed at the fusion of traffic signs in the intelligent transportation field, and provides a new option for image fusion benchmark evaluation.

The rest of the paper is arranged as follows: The current methods and applications of multi-exposure image fusion are described in Section 2. The proposed method for multi-exposure image fusion is introduced in Section 3. Additionally, the effectiveness of this method is verified through computer simulation in Section 4. Finally, the conclusion is described in Section 5.

2. Related Works

At present in intelligent transportation systems, intelligent vehicles make decisions by sensing the surrounding environment. The image taken by the camera is an important source of information. In recent years, the research on image fusion has developed rapidly and has received widespread attention. Li et al. [8] decomposed the source image into a basic part and a detailed part, and used the deep learning network to extract the features of the detailed part for fusion. Paul et al. [9] proposed to mix the gradient of the brightness component of the input image, utilized the maximum gradient amplitude at each pixel position, and obtained the fused brightness by using image reconstruction technique based on Haar wavelet. The above algorithms are mainly aimed at the fusion between infrared and visible images and multi-focus images [10–12]. In the camera shooting process, the phenomenon of over-exposure or under-exposure will occur due to the influence of light and environmental factors. Therefore, the road sign information cannot be extracted, which is not conducive to subsequent decision-making.

This paper proposes an improved multi-exposure image fusion algorithm and applies it to the processing of traffic sign images in the field of intelligent transportation [6]. Prabhakar et al. [13] proposed a deep unsupervised approach for exposure fusion (Deep-fuse), which uses a novel CNN (Convolutional Neural Networks) architecture that is trained to learn the fusion operation without reference ground-truth image. This fusion process will lose other key information, such as contrast and texture information. A new multi-exposure fusion method based on dense scale-invariant feature transform (DSIFT) is presented by Liu et al. [14]. The fusion effect of this method is largely based on the selection of reference images and the selection of specific scenes, and cannot be applied to the multi-change scenes under intelligent transportation. Yang et al. [15] proposed a strategy by producing virtual images. The downside of this approach is that it only deals with images that have equal exposure degrees. Paul et al. [9] presented multi-exposure and multi-focus image fusion in the gradient domain (GBM). This method is not suitable for multi-exposure images with too large of a gradient difference, and the fusion speed is slower than other methods, so it is not suitable for the intelligent transportation field that needs to process a large amount of information. Fu et al. proposed the fusions of visual and infrared images on the feature level and decision level. On the feature level, the feature fusion was realized by wavelet transformation. On the decision level, the fusion
method based on Dempster-Shafer evidence theory was proposed [16]. However, the wavelet transform for feature fusion will cause image distortion to a certain extent, and affect subsequent decision-making. Goshtasby et al. took non-overlapping blocks with the highest information from each image to obtain the fused result. This is prone to suffer from block artifacts [17].

All of the above works rely on hand-crafted features for image fusion. These methods are not robust in the sense that the parameters need to be varied for different input conditions say, linear and non-linear exposures, filter size depends on image sizes. At the same time, among the above-mentioned fusion methods, artificially designed extraction methods will make the fusion method more and more complex, thereby increasing the difficulty of fusion rule design. For different fusion tasks, the extraction method needs to be improved accordingly. In addition, attention needs to be paid to the appropriateness of the extraction method to ensure the integrity of the features.

This paper uses a unified unsupervised end-to-end image fusion network which is proposed by Xu et al. [6]. Through feature extraction and information measurement, an adaptive information preservation degree is given. Feature extraction is performed through the pre-trained VGGNet-16, and information is measured on the feature map to obtain the degree of information retention in the feature map, which is sent to the loss function. The method trains DenseNet to fuse the source images and minimizes the loss function to optimize the DenseNet network to achieve a better fusion effect. It has a high requirement for the quality of the input multi-exposure image. If noise or distortion occurs in the acquisition of the input image, the multi-exposure fusion model will amplify these problems. Based on this method, this paper utilizes the dehazing algorithm [18] and bilateral filtering to denoise the source image and reduce the influence of the weather environment. Further, Laplace operator and histogram equalization are applied to the proposed fusion network, which are utilized to keep the important information of the source image. Meanwhile, the Laplace operator is helpful to enhance the contrast and the entire brightness of the final image [19,20].

3. Proposed Method

In general, the photos taken by the camera are usually three-channel (RGB) images. We first convert them from RGB to YCbCr color space due to a large amount of calculation in processing RGB images. In the YCbCr space, Y is the luminance channel, and Cb and Cr are the density offset components of blue and red, respectively. When directly processing an RGB image, the brightness will also change when the color is adjusted, which is prone to color cast and is not beneficial to direct fusion. Focusing on the fusion of the Y channel, the detailed texture of the image is mainly represented by this channel [21]. The values of the Cb channel and Cr channel (chroma channel) adopt the traditional fusion method.

The fusion framework proposed in this paper is shown in Figure 1. Taking into account the camera shakes during shooting and the internal noise interference of the instrument, the input original image may have partial blur or more noise, which will have a greater impact on the subsequent multi-exposure fusion and may lead to express the information in the fusion image inaccurately. Therefore, we first preprocessed the image by using bilateral filtering and dehazing algorithm. After that, we merged the images that have been preprocessed. We used pre-trained VGGNet-16 to extract the shallow features and deep features of the multi-exposure image to estimate the amount of information. Subsequently, Laplace operator and histogram equalization were used to optimize the over-bright or over-dark images. The procedure is summarized in Algorithm 1.
Algorithm 1: The description of the training procedure

* Training Procedure of the Proposed Method

Parameter: \( \phi_{n}^{k} \) means the feature map of the \( k \)-th input image \( I_{n}^{k} \) before the \( n \)-th max-pooling layer. \( g_{l} \) means the information measurement of image \( I_{l} \). \( \theta \) denotes the parameters in DenseNet. \( D \) is the training dataset. \( a \) is set as 20. \( c \) is set as 3e3, 3.5e3, and 1e2.

Input: RGB images \( I_{n}^{1} \) and \( I_{n}^{2} \), \( n \) denotes the \( n \)-th pair images.

Output: the parameters in DenseNet \( \theta \).

1. \( I_{1}^{n} \leftarrow \) bilateral filtering and dehazing on \( I_{1}^{n} \).
2. \( I_{2}^{n} \leftarrow \) bilateral filtering and dehazing on \( I_{2}^{n} \).
3. Set the number of training iterations;
4. for the number of training images do
5. Feed the input images into VGGNet-16, and extract the feature maps: \( \{ \phi_{1}^{1}, \ldots, \phi_{c}^{1} \} \) and \( \{ \phi_{c}^{2}, \ldots, \phi_{C}^{2} \} \);
6. Compute the gradients \( g_{l} \) using \( g_{l} = \frac{1}{5} \sum_{j=1}^{D} \frac{1}{D} \sum_{k=1}^{D_{k}} \| \nabla \phi_{C}^{j}(l) \|^2_{F} \) to measure the information of input images;
7. Define two weights \( \omega_{1} \) and \( \omega_{2} \) as the information preservation degrees, which can compute using \( [\omega_{1}, \omega_{2}] = \text{softmax} \left( \left[ \frac{g_{1}}{\omega}, \frac{g_{2}}{\omega} \right] \right) \), the weights is to preserve the information in source images;
8. SSIM and MSE is used to obtain the \( L_{\text{sim}}(\theta, D) \), \( L_{\text{sim}}(\theta, D) \) can compute by \( L_{\text{sim}}(\theta, D) = E \left[ \omega_{1} \cdot \left( 1 - S_{I_{f}, I_{1}} \right) + \omega_{2} \cdot \left( 1 - S_{I_{f}, I_{2}} \right) \right] + E \left[ \omega_{1} \cdot \text{MSE}_{I_{f}, I_{1}} + \omega_{2} \cdot \text{MSE}_{I_{f}, I_{2}} \right] \);
9. Update \( \theta \);
10. The number of training iterations minus 1;
11. if the number of training iterations is 0 break;
12. end;
13. end;
14. return \( \theta \);

Figure 1. Multi-exposure image fusion framework.

3.1. Image Preprocessing

Bilateral filtering is an edge protection filtering method using a weighted-average strategy based on Gaussian distribution [22,23]. Bilateral filtering consists of two parts: the spatial matrix and range matrix. The spatial matrix is analogous to Gaussian filtering, which is used for fuzzy denoising; the range matrix is obtained according to the gray-scale similarity, and is used to protect edges. The specific formulas of the spatial matrix and range matrix are as follows:

\[
d(i, j, k, l) = e^{-\frac{(i-k)^2 + (j-l)^2}{2\sigma_x^2}}
\]

\[
r(i, j, k, l) = e^{-\frac{(r(i,j), k, l))^2}{2\sigma_y^2}}
\]

where \((i, j)\) is the coordinate of the center point of the filter window, and \((k, l)\) is any point in the field of the center point.

For Gaussian filtering, only the weight coefficient kernel of the spatial distance is used to convolve the image to determine the gray value of the center point, the closer the point to the center point, the larger the weight coefficient. The weight of gray information is added
to the bilateral filtering. In the field, the closer the gray value is to the center point, the greater the point weight [22]. This weight is determined by Equation (2). By multiplying Equations (1) and (2), the final convolution template is obtained.

Multiplying Equations (1) and (2) is the calculation formula of the bilateral filter weight matrix, the final weight matrix by multiplying the two weight matrices:

$$w(i,j,k,l) = e^{-\frac{(i-k)^2+(j-l)^2}{2\sigma^2}} + w(i,j)$$  

Finally, calculate the weighted average as the pixel value of the center point after filtering:

$$g(i,j) = \frac{\sum_{(k,l) \in S} f(i,j)w(i,j,k,l)}{\sum_{(k,l) \in S} w(i,j,k,l)}$$  

In the dehazing algorithm [18], firstly, the dark channel prior theory believes that in most non-sky local areas, certain pixels will always have at least one color channel with a very low value.

Given a mathematical definition of the dark channel, for any input image \(I\), the dark channel expression is as follows:

$$I_{\text{dark}}(x) = \min_{y \in \Omega(x)} \left( \min_{c \in \{r,g,b\}} I_{c}(y) \right)$$  

where \(I_{c}\) represents each channel of the color image, \(x\) and \(y\) represent the pixel. \(\Omega(x)\) represents a window centered on pixel \(x\). \(I_{\text{dark}}(x)\) is the dark primary color of the image in the \(\Omega(x)\) neighborhood. For clear and fog-free images, its value tends to 0.

In computer vision and computer graphics, the fog map formation model described by the following equation is widely used:

$$I(x) = I(x)t(x) + A(1 - t(x))$$  

where \(I(x)\) is the original image, \(I(x)\) is the image after defogging, \(A\) is the atmospheric light value, and \(t(x)\) is the transmittance.

The current known condition is \(I(x)\), and the target value \(f(x)\) is required. This is an equation with countless solutions. Therefore, some priors were needed.

The formula of the fog map formation model is processed and can be written by Equation (7).

$$\frac{I_c(x)}{A_c} = t(x) \frac{I_c(x)}{A_c} + 1 - t(x)$$  

where \(c\) represents the R/G/B three channels.

He et al. [24] assumed that the transmittance was constant in each window, which was defined as \(\tilde{t}(x)\). From the dark channel map, take the first 0.1% of pixels according to the brightness. In these positions, find the corresponding value of the point with the highest brightness in the original foggy image \(I\) as the value of \(A\). Therefore, Equation (7) can be transformed into:

$$\min_{y \in \Omega(x)} \left( \min_{c} \frac{I_{c}(y)}{A_{c}} \right) = \tilde{t}(x) \min_{y \in \Omega(x)} \left( \min_{c} \frac{I_{c}(y)}{A_{c}} \right) + 1 - \tilde{t}(x)$$  

According to the previous prior theory, the method to further deduce the estimated value of transmittance is as follows:

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left( \min_{c} I_{c}(x) / A_{c} \right)$$  

where \(\omega\) is the artificially introduced correction constant (generally 0.95), which is used to retain the fog at part of the perspective and maintain the variation of the depth of field.
We set the lower limit \( t_0 \) to limit \( t(x) \) in order to prevent the contrast from getting too large. When \( t \) is less than \( t_0 \), \( t \) will be set to \( t_0 \). Therefore, the final recovery Equation (10) is as follows:

\[
J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A
\]

(10)

where \( t_0 \) is set to 0.1.

In order to improve the speed of the dehazing algorithm and achieve real-time effects, when optimizing the image, the original image is first down-sampled. First, the image is reduced to one-quarter of the original image, the transmittance of the reduced image is calculated, and then the approximate transmittance of the original image is obtained by interpolation, which greatly improves the execution speed, while the effect is basically unchanged.

### 3.2. Fusion Module

The multi-exposure images fusion model mainly includes three aspects: fusion network, feature retention, and loss function. The specific structure is shown in Figure 2.

![Figure 2. Fusion model.](image)

The input multi-exposure images are represented as \( I_1^*, I_2^* \) and the fusion image is generated through DenseNet training in the fusion network. In the feature retention module, the outputs of feature extraction parts are feature maps \( \phi_{C_1}(I_1), \ldots, \phi_{C_5}(I_1) \) and \( \phi_{C_1}(I_2), \ldots, \phi_{C_5}(I_2) \). In information measurement, the amount of information extracted from the special graph is expressed as \( g_{I_1} \) and \( g_{I_2} \). Through subsequent processing, the degree of information retention in the final obtained source images is represented by \( \omega_1 \) and \( \omega_2 \). \( I_1^*, I_2^*, I_{\omega_1}^*, I_{\omega_2}^* \) are sent into the loss function without the need for ground-truth. During the training, the DenseNet was continuously optimized to minimize the loss function [25]. It was not necessary to measure \( \omega_1 \) and \( \omega_2 \) again in the process of testing, and the fusion speed was faster in practical application.

The DenseNet architecture in the fusion network consisted of 10 layers, each of which had a convolutional layer and an activation function [26]. Dense connections were applied inside each Dense Block, and a convolutional layer plus a pooling layer were utilized between adjacent Dense Block. The advantage of DenseNet is that the network is narrower, has fewer parameters, and reduces the phenomenon of gradient disappearance.
The activation function of the first nine layers is LeakReLU with a slope of 0.2 and the last layer is the tanh. The kernel size of all convolutional layers is set to $3 \times 3$, and the stride is set to 1 \cite{6}.

We extracted the features through the pre-trained VGGNet-16, which is shown in Figure 3. The convolutional layer output before the maximum pooling layer is the feature graph, which was used for subsequent information measurement \cite{27}. In the source image, the underexposed image has lower brightness. Therefore, the overexposed image contains more texture details or greater gradients than the underexposed image. The shallow features such as texture and details were extracted from $\phi_{C_1}(I)$ and $\phi_{C_2}(I)$, and the deep features such as structural content were extracted from the feature maps of the later layers.

![VGGNet-16](image)

**Figure 3.** VGGNet-16.

After the image gradient was estimated, the feature map information could be measured. Information measurement is defined as follows:

$$g_I = \frac{1}{5} \sum_{j=1}^{5} \frac{1}{H_j W_j D_j} \sum_{k=1}^{D_j} \| \nabla \phi_{C_1}(I) \|^2_F$$

(11)

where $\phi_{C_1}(I)$ is the feature map and $k$ denotes the feature map in the $k$-th channel of $D_j$ channels. $\| \cdot \|^2_F$ represents Frobenius norm and $\nabla$ represents a Laplace operator.

In this method, $\omega_1$ and $\omega_2$ were obtained through $g_{I_1}$ and $g_{I_2}$. As the difference between $g_{I_1}$ and $g_{I_2}$ is an absolute value and is too small compared with itself, therefore, we scaled them with a predefined positive number $c$ to better distribute the weight. Through the function, the expressions of $\omega_1$ and $\omega_2$ are as follows:

$$[\omega_1, \omega_2] = \text{softmax} \left( \left[ \frac{g_{I_1}}{c}, \frac{g_{I_2}}{c} \right] \right)$$

(12)

The loss function consists of two parts defined as follows:

$$L(\theta, D) = L_{\text{sim}}(\theta, D) + \alpha L_{\text{mse}}(\theta, D)$$

(13)

where $\theta$ represents the parameter in DenseNet and $D$ represents the training dataset. $L_{\text{sim}}(\theta, D)$ is the similarity loss between the fused image and the source multi-exposure image, and $L_{\text{mse}}(\theta, D)$ is the mean square error loss between the images. $\alpha$ is used to control the trade-off.

Structural similarity index measure (SSIM) \cite{28,29} is widely used in modeling distortion based on the similarity of light, contrast, and structural information. In this paper, SSIM was used to constrain the structural similarity between $I_{g_1}^*$, $I_{g_2}^*$ and $I_{g}^*$. The loss function under the SSIM framework is as follows:

$$L_{\text{sim}}(\theta, D) = \mathbb{E} \left[ \omega_1 \left( 1 - S_{I_{g_1}, I_{g_1}^*} \right) + \omega_2 \left( 1 - S_{I_{g_2}, I_{g_2}^*} \right) \right]$$

(14)

where $S_{x,y}$ denotes the value of SSIM between $x$ and $y$. 


Considering that SSIM only focuses on contrast and structure changes and it shows weaker constraints on the difference of the intensity distribution, we supplemented it by mean square error (MSE) between two images, and the loss function of this part was as follows:

\[
L_{mse}(\theta, D) = E \left[ \omega_1 \text{MSE}_{I_1, I_1} + \omega_2 \text{MSE}_{I_2, I_2} \right] 
\]  

(15)

### 3.3. Image Optimization

The fusion result should be a normal exposure under ideal conditions, which can reflect the road sign information. However, due to the limited training set and the small number of individual road signs, the exposure of the fusion result will still not reach the normal level. Therefore, we utilized histogram equalization and Laplacian to optimize the image with a poor fusion effect.

In actual applications, the fusion image was not what we expected due to the defects of our fusion algorithm. Although it retained the general characteristics of the under-exposure image and the over-exposure image, the overall brightness presented made ours feel inappropriate. Therefore, we introduced optimization algorithms next to adjust the brightness of the image without affecting the image quality.

When the brightness of the fusion image is lower than the pre-set value, the Laplace algorithm [30] is used to sharpen the image. When the value is higher than the pre-set value, Histogram equalization [31] is adopted to enhance the over-bright part of the image. The procedure is summarized in Algorithm 2.

**Algorithm 2: The description of the optimization**

* Process of Optimization

**Parameter:** \( B^*_f \) denotes the brightness of \( I^*_f \), \( M_{gray} \) means the gray average of an image.

**Input:** Fused image \( I^*_f \) from DenseNet, the high threshold \( T_h \), and the low threshold \( T_l \).

**Output:** Final image \( I_f \) after optimization.

1: Compute the Brightness \( B^*_f \) of \( I^*_f \) using \( M_{gray}/255.0 \);
2: if \( 1 > B^*_f > T_h \) then
3: \( I_f \leftarrow \) Brightness reduction on \( I^*_f \);
4: return \( I_f \);
5: else if \( 0 < B^*_f < T_l \) then
6: \( I_f \leftarrow \) Brightness enhancement on \( I^*_f \);
7: return \( I_f \);
8: else if \( T_l \leq B^*_f \leq T_h \) then
9: return \( I_f \);
10: else
11: return false;
12: end if

The function of image sharpening is to enhance the gray-scale contrast, so that the sharpness of the image is improved. The essence of image blur is that the image is subjected to averaging operation or integration operation. The Laplacian operation is a kind of differential operator; through the inverse operation of the image, it enhances the region of the gray mutation in the image, highlights the details of the image, and obtains a clear image.

An image that describes the gray level mutation is generated through the Laplacian operator to process the original images, then the Laplacian image is superimposed with the original image to produce a sharpened image. This principle is actually a convolution operation.
Laplacian operator is the simplest isotropic differential operator with rotation invariance. The Laplace transform of two-dimensional image function is the isotropic second derivative, which is defined as:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$  \hspace{1cm} (16)

In a two-dimensional function \(f(x,y)\), the second-order difference in \(x\) and \(y\) directions is as follows:

$$\frac{\partial^2 f}{\partial x^2} = f_{x+1,y} + f_{x-1,y} - 2f_{x,y}$$  \hspace{1cm} (17)

$$\frac{\partial^2 f}{\partial y^2} = f_{x,y+1} + f_{x,y-1} - 2f_{x,y}$$  \hspace{1cm} (18)

where \(f_{x,y}\) represents \(f(x,y)\).

In order to be more suitable for digital image processing, the equation is converted into a discrete form, which is as follows:

$$\nabla^2 f = [f_{x+1,y} + f_{x-1,y} + f_{x,y+1} + f_{x,y-1}] - 4f_{x,y}$$  \hspace{1cm} (19)

The basic method of Laplace operator enhanced image can be expressed as follows:

$$g(x, y) = f(x, y) + c[\nabla^2 f(x, y)]$$  \hspace{1cm} (20)

where \(f(x, y)\) and \(g(x, y)\) are the input image and the sharpened image, respectively. \(c\) is the coefficient, indicating how much detail is added.

This simple sharpening method can produce the effect of Laplacian sharpening while retaining the background information. The original image is superimposed on the processing result of the Laplace transform, so that the gray value in the image can be retained, and the contrast at the gray level mutation can be enhanced. The end result is to highlight small details in the image while preserving the background of the image.

The fusion result will have the phenomenon of partial exposure. In the case, histogram equalization was utilized to optimize the fusion image.

Histogram equalization is a simple and effective image enhancement technology that changes the grayscale of each pixel in the image by changing the histogram in the image. The gray levels of overexposed pictures are concentrated in the high brightness range. Through the histogram equalization, the gray value of the large number of pixels in the image is expanded, and the gray value of the small number of pixels is merged, and the histogram of the original image can be transformed into a uniform distribution (balanced) form. This increases the dynamic range of the gray value difference between pixels, thereby achieving the effect of enhancing the overall contrast of the image \([31,32]\).

The gray histogram of the image is a one-dimensional discrete function, which can be written as:

$$h(r_k) = n_k, k = 0, 1, \ldots, L - 1$$  \hspace{1cm} (21)

where \(n_k\) is the number of pixels whose gray level is \(r_k\) in the source image.

Based on the histogram, the relative frequency \(P_r(r_k)\) of gray level appearing in the normalized histogram is further defined, and the expression is as follows:

$$P_r(r_k) = n_k / N$$  \hspace{1cm} (22)

where \(N\) represents the total number of pixels in the source image \(I\).

Histogram equalization is to transform the gray value of pixels according to the histogram. \(r\) and \(s\) represent the normalized original image grayscale and the histogram equalized grayscale, respectively, and the values of them are between 0 and 1. For any \(r\) in
the interval of $[0, 1]$, a corresponding $s$ can be generated by the transformation function $T(r)$, the expression is as follows:

$$ s = T(r) \quad (23) $$

$T(r)$ is a monotonically increasing function to ensure that the gray level of the image after equalization does not change from black to white. At the same time, the range of $T(r)$ is also between 0 and 1, ensuring that the pixel gray of the equalized image is within the allowable range.

The inverse transformation of the above formula is as follows:

$$ r = T^{-1}(s) \quad (24) $$

It is known that the probability density function of $r$ is $p_r(r)$ and $s$ is the function of $r$. Therefore, the probability density $p_s(s)$ of $s$ can be obtained from $p_r(r)$. Further, because the probability density function is the derivative of the distribution function, the probability density function of $s$ is further obtained through the distribution function $F_s(s)$. The specific derivation process is as follows:

$$ F_s(s) = \int_{-\infty}^{s} p_s(s) ds = \int_{-\infty}^{r} p_r(r) dr \quad (25) $$

$$ p_s(s) = \frac{dF_s(s)}{ds} = \frac{d}{ds} \left[ \int_{-\infty}^{r} p_r(r) dr \right] = p_r(r) \frac{dr}{ds} = p_r(r) \frac{dr}{d[T(r)]} \quad (26) $$

Equation (26) shows that the probability density function $p_s(s)$ of the image gray level can be controlled by the transformation function $T(r)$, thereby improving the image gray level. Therefore, in the histogram equalization, $p_s(s)$ should be a uniformly distributed probability density function. As we have normalized the $r$, the value of the $p_r(s)$ is 1. Therefore, $ds = p_r(r) dr$, the integral on both sides of the formula can be obtained as follows:

$$ s = T(r) = \int_{0}^{r} p_r(r) dr \quad (27) $$

Equation (24) is the expression of the transformation function $T(r)$. It shows that when the transformation function $T(r)$ is the cumulative distribution probability of the original image histogram, the histogram can be equalized. For digital images with discrete gray levels, using frequency instead of probability, the discrete mindset of the transformation function $T(r_k)$ can be expressed as:

$$ s_k = T(r_k) = \sum_{i=0}^{k} p_r(r_i) = \sum_{i=0}^{k} \frac{n_i}{N}, \quad k = 0, 1, \ldots, L - 1 \quad (28) $$

$r_k$ is between 0 and 1, which represents the gray value after normalization, which is calculated by the quotient of $k$ and $L - 1$. $k$ represents the gray value before normalization. Equation (28) shows that the gray value $s_k$ of each pixel after equalization can be directly calculated from the histogram of the source image.

### 4. Experimental Result and Analysis

Since there are few multi-exposure datasets available, which can hinder our training for exposure fusion tasks. So we selected some images from the existing road-related datasets to make the multi-exposure dataset. We selected 1400 images of various traffic scenarios from the CCTSDB (CSUST Chinese Traffic Sign Detection Benchmark) as the dataset to be processed later and use ACDSee [33] to produce the multi-exposure images. We compared our approach with the other four representative methods, including DSIFT, FLER, GBM, and DeepFuse. Experiments were performed on an NVIDIA GeForce 920M and 2.4 GHz Intel(R) Core i7-5500U CPU. All of our training procedure is on NVIDIA GTX 1080Ti and 32 GB memory.
4.1. Qualitative Comparisons

We chose typical images as comparison pictures for qualitative analysis. The fusion results of this paper are compared with the other four fusion methods, which are shown in Figure 4a.

![Comparison Images](image_url)

(a) The separate from left to right in Figure 4a are source image, multi-exposure image pairs, DSIFT, FLER, GBM, DeepFuse, our method.

(b) The separate from left to right in Figure 4b are source image, DSIFT, FLER, GBM, DeepFuse, our method.

**Figure 4.** Qualitative comparison of our method with four methods on three typical multi-exposure image pairs in our dataset.

It can be seen from the figure that the fusion result of DSIFT and FLER will reduce the sharpness of the image, the degree of preservation of the details of the image is not high, and there will be obvious black areas in local areas. FLER also has the defect of poor fusion effect of large-scale exposure. The final fusion of GBM and DeepFuse has a natural visual effect but suffers from a lack of detail and texture. The fusion result of GBM also has the problem of low contrast, which makes it impossible to display the road sign information. The method proposed in this paper shows that the fusion result looks the best overall, as it has the highest definition, the edge contour is closer to the source image, and the information retention is more complete than other methods.

We further analyzed the experiment results. The overall image visual effects of the FLER and DeepFuse results are more consistent with human visual perception, but at traffic signs, their results also lose some detailed texture information. In contrast, the landmark in our result has rich texture information, and its contour information is closer to the real information.
We also placed an enlarged view of the road signs in one of the experimental results in Figure 4b. It can be seen that our method can retain the road sign information as much as possible. It can be seen from the results of GBM and DSIFT that their information is a bit vague, and it is not suitable for intelligent transportation systems and smart cities that require precise identification of road signs. The signpost information in our method is clearest, and the exposure level seems to be the most appropriate.

In the optimization process, the images with poor fusion effects are processed. The brightness of the image ranges from 0 to 1. Laplace algorithm and histogram equalization are applied to images with brightness values less than 0.3 and brightness values greater than 0.7, respectively, to ensure that the image brightness is within the normal range.

The effects of the two optimization methods are shown in Figure 5.

![Figure 5](image)

**Figure 5.** The results of optimizing the fusion images, \( I_{f1} \) denotes the fusion image with a brightness value lower than 0.3 and \( I_{o1} \) is a result image after processing by Laplace algorithm; \( I_{f2} \) denotes the fusion image with a brightness value higher than 0.7, and \( I_{o2} \) is a result image after processing by histogram equalization.

### 4.2. Quantitative Comparisons

We made a quantitative comparison of 16 pairs of multi-exposure images in the dataset, and the results are shown in Figures 6 and 7. In the figure, we also indicate the average value of each method’s corresponding indicator. In the test dataset, the method proposed in this paper achieves the optimal value in all the other five indexes except MI. Figure 6 compares spatial frequency (SF) [34], standard deviation (STD) [35] and entropy (EN) [36]. The larger these parameters, the better the retention of information. Table 1 shows the comparison results of five methods in six indicators. The average value of these indicators in our method is 12.990, 70.907, and 7.328. The results show that the method in this paper retains more information and has higher image quality. Figure 7 shows the comparison results of mutual information (MI) [37], SSIM and peak signal to noise ratio (PSNR) [38]. The larger these parameters, the better the retention of information. In addition, the average value of these indicators in the proposed method is 4.409, 0.923, and 24.720. It seems that MI is a suboptimal index. It can be seen that the fusion results of our method have a higher similarity with the source images.
Figure 7. Quantitative comparisons of the SSIM, MI, and PSNR on 16 image pairs from our dataset.

Table 1. Quantitative comparisons of the six metrics.

| Metrics   | SF   | STD  | EN   | SSIM | MI   | PSNR |
|-----------|------|------|------|------|------|------|
| DSIPT     | 8.415| 32.449| 6.597| 0.833| 1.644| 11.583|
| FLER      | 10.750| 54.242| 7.103| 0.861| 3.152| 14.283|
| GBM       | 8.877| 35.244| 6.420| 0.888| 2.407| 15.364|
| DeepFuse  | 9.295| 67.090| 6.884| 0.907| 5.716| 14.742|
| Ours      | 12.990| 70.907| 7.328| 0.923| 4.409| 21.720|

Note: There are six metrics in the table. SF is the spatial frequency, STD is the standard deviation, EN is the entropy, SSIM is the structural similarity index measure, MI is the mutual information, PSNR is the peak signal to noise ratio.

From the result, we see that our method can achieve comparable efficiency compared with the other four methods.

4.3. Expended Experiment

In order to better illustrate the application of multi-exposure image fusion in the field of intelligent transportation, we carried out experiments on the recognition and classification of traffic signs. We applied the multi-exposure image fusion method proposed in this paper to the identification of traffic signs, and the recognition accuracy of traffic signs has been improved. According to the traffic sign recognition method in [39], the fusion method proposed in this paper is compared with the other four fusion methods. The multiscale recognition method for traffic signs based on the Gaussian Mixture Model (GMM) and Category Quality Focal Loss (CQFL) to enhance recognition speed and recognition accuracy. Specifically, GMM is utilized to cluster the prior anchors, which are in favor of reducing the clustering error. Meanwhile, considering the most common issue in supervised learning (i.e., the imbalance of data set categories), the category proportion factor is introduced into Quality Focal Loss, which is referred to as CQFL. Furthermore, a five-scale recognition network with a prior anchor allocation strategy is designed for small target objects i.e., traffic sign recognition. We chose this algorithm due to its superior recognition accuracy and recognition speed. The classifier used can distinguish specific road signs in 30, which is shown in Table 2.
Table 2. Categories of traffic signs.

| Traffic Signs               | Over-Exposure | Under-Exposure | Fusion |
|----------------------------|---------------|----------------|--------|
| Red Light for Go Straight  | 0.4622        | 0.4618         | 0.4629 |
| Green Light for Go Straight| 0.5259        | 0.5333         | 0.5418 |
| Red Light for Left Turn    | 0.3867        | 0.3867         | 0.3991 |
| Green Light for Left Turn  | 0.4072        | 0.4627         | 0.4649 |
| Green Light for Right Turn | 0.8468        | 0.8509         | 0.8573 |
| Go Straight Slot           | 0.8494        | 0.8559         | 0.8588 |
| Left Turn Slot             | 0.9149        | 0.9153         | 0.9178 |
| Strictly No Parking         | 0.8669        | 0.8537         | 0.8726 |
| No left or Right Turn      | 0.8340        | 0.8335         | 0.8439 |
| School Crossing Ahead      | 0.7017        | 0.7017         | 0.7037 |
| Stop                       | 0.7977        | 0.7797         | 0.8117 |

We selected 1400 pairs of multi-exposure images from our dataset (TrafficSign) [7] for the traffic sign classification experiment. The classification results of fusion images from our method, and the over-exposure and under-exposure images are shown in Table 3. It can be seen from Table 3 that the fusion result has a certain influence on the recognition of traffic signs in the case of over-exposure and under-exposure. The last row of Table 3 shows the recognition accuracy of all classes of traffic signs.

Table 3. Classification accuracy of over-exposure, under-exposure and fusion images.

| Traffic Sign                     | Over-Exposure | Under-Exposure | Fusion  |
|----------------------------------|---------------|----------------|--------|
| Green Light for Go Straight      | 0.4622        | 0.4618         | 0.4629 |
| Red Light for Left Turn          | 0.5259        | 0.5333         | 0.5418 |
| Green Light for Left Turn        | 0.3867        | 0.3867         | 0.3991 |
| Green Light for Right Turn       | 0.4072        | 0.4627         | 0.4649 |
| Go Straight Slot                  | 0.8468        | 0.8509         | 0.8573 |
| Left Turn Slot                    | 0.8494        | 0.8559         | 0.8588 |
| Strictly No Parking               | 0.9149        | 0.9153         | 0.9178 |
| No left or Right Turn            | 0.8669        | 0.8537         | 0.8726 |
| School Crossing Ahead            | 0.8340        | 0.8335         | 0.8439 |
| Stop                             | 0.7017        | 0.7017         | 0.7037 |
| All Classes of Signs             | 0.7977        | 0.7797         | 0.8117 |

Since the dataset we use is a wide range of road images, the classification accuracy of many traffic signs is not high, but this will not affect our comparison of the final results. It can be proved in Table 3 that the accuracy of our fusion results in the classification of traffic signs is generally higher than that of over-exposure images and under-exposure images, which shows that our method can effectively avoid over-exposure and under-exposure situations that the traffic sign information is not accurate enough.

We not only compare the accuracy of traffic sign classification among the over-exposure image, the under-exposure image, and the fusion image, but also compare our method with the other four methods in the field of traffic sign recognition. Using the result image of these five methods for traffic sign recognition or traffic sign classification can be very effective and clear to know whether the fusion results of various methods are good.

We selected 1000 pairs of multi-exposure images from our multi-exposure dataset (TrafficSign) and obtained the fusion result images of each method. We conducted experiments on the classification of traffic signs on these fusion results and compared them. The results obtained through the experiment are shown in Table 4, which shows the accuracy of various methods for the classification of some traffic signs. Further, the last row of Table 4 shows the recognition accuracy of all classes of traffic signs.

From Table 4, compared with the other four methods, we can see that the classification accuracy obtained by our method is mostly the best. For example, for the traffic sign “Strictly No Parking”, our accuracy can reach 0.9178. Although the images obtained by our method are used for traffic sign recognition, the recognition effect of some traffic signs is not good, such as “Go Straight Slot”, but it will be too far that compared with the result with the highest accuracy and the result obtained by our method.
Table 4. Comparison of classification accuracy of other four methods and ours.

| Traffic Sign                        | DeepFuse | DSIFT | FLER | GBM | Ours  |
|-------------------------------------|----------|-------|------|-----|-------|
| Green Light for Go Straight         | 0.4118   | 0.4459 | 0.4514 | 0.4566 | 0.4629 |
| Red Light for Left Turn             | 0.5363   | 0.5326 | 0.5274 | 0.5303 | 0.5418 |
| Green Light for Right Turn          | 0.4496   | 0.4209 | 0.4072 | 0.4072 | 0.4649 |
| Go Straight Slot                    | 0.8550   | 0.8580 | 0.8567 | 0.8544 | 0.8573 |
| Strictly No Parking                 | 0.9137   | 0.9071 | 0.9076 | 0.9070 | 0.9178 |
| No Trucks                           | 0.8557   | 0.8574 | 0.8573 | 0.8596 | 0.8605 |
| No left or Right Turn               | 0.8221   | 0.8706 | 0.8702 | 0.8674 | 0.8726 |
| Weight Limit 40 Tons                | 0.8193   | 0.8191 | 0.8028 | 0.8140 | 0.8221 |
| School Crossing Ahead               | 0.8434   | 0.8436 | 0.8457 | 0.8391 | 0.8439 |
| Pedestrian Crossing Ahead           | 0.8503   | 0.8578 | 0.8569 | 0.8563 | 0.8574 |
| All Classes of Signs                | 0.8101   | 0.8115 | 0.8095 | 0.8059 | 0.8117 |

It is worth mentioning that because our dataset contains many traffic signs, and there are few specific traffic signs, such as traffic lights, the accuracy results of the classification test are generally low. However, this result will not affect our comparison of over-exposure images and under-exposure images and the images obtained by the other four methods.

5. Conclusions

In this paper, multi-exposure image fusion was applied to the identification of traffic signs. Fusion images were generated using a unified unsupervised end-to-end fusion network, U2Fusion. By obtaining the degree of information retention in the source image, the adaptive similarity between the fusion result and the source image was maintained. Considering the influence of weather and environmental noise, image preprocessing should be carried out before image fusion. At the same time, image optimization was carried out for fusion results. Through brightness judgment, the fusion results are adaptive adjusted, so that the final result diagram could reflect the information and facilitate subsequent analysis and decision. The qualitative and quantitative results show that the method was effective and practical in traffic signs. In addition, we have released a new multi-exposure image data set, which provides a new evaluation option for image fusion.

Author Contributions: Methodology, M.G.; software, J.W.; validation, Y.C. and C.D.; resources, C.C.; data curation, Y.Z.; writing—original draft preparation, M.G.; writing—review and editing, C.C. and J.W.; visualization, C.C.; supervision, C.D.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Program of Zhejiang Province (No. 2020C01110).

Acknowledgments: The authors would like to thank the editorial board and reviewers for the improvement of this paper.

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

References

1. Yang, Z.X.; Wang, Z.M.; Wu, J.; Yang, C.; Yu, Y.; He, Y.H. Image Fusion Scheme in Intelligent Transportation System. In Proceedings of the 2006 6th International Conference on ITS Telecommunications, Chengdu, China, 21–26 June 2006; pp. 935–938.
2. Kou, F.; Wei, Z.; Chen, W.; Wu, X.; Wen, C.; Li, Z. Intelligent Detail Enhancement for Exposure Fusion. IEEE Trans. Multimed. 2017, 20, 484–495. [CrossRef]
3. 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]
4. Gu, B.; Li, W.; Wong, J.; Zhu, M.; Wang, M. Gradient field multi-exposure images fusion for high dynamic range image visualization. J. Vis. Commun. Image Represent. 2012, 23, 604–610. [CrossRef]
5. Mertens, T.; Kautz, J.; Reeth, F.V. Exposure Fusion. In Proceedings of the 15th Pacific Conference on Computer Graphics and Applications (PG’07), Washington, DC, USA, 29 October–2 November 2007; pp. 382–390.
6. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. IEEE Trans.Pattern Anal. Mach. Intell. 2020, 1. [CrossRef] [PubMed]
7. Chen, Y. Dataset: TrafficSign. Available online: https://github.com/chenyi-real/TrafficSign (accessed on 27 December 2020).
8. Li, H.; Wu, X.; Kittler, J. Infrared and Visible Image Fusion using a Deep Learning Framework. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018, pp. 2705–2710.

9. Paul, S.; Sevcenco, I.S.; Agathoklis, P. Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain. J. Circuits Syst. Comput. 2016, 25, 25. [CrossRef]

10. Dong, Z.; Dai, C.S.; Qi, D.; Xu, Z.; Li, C.; Duan, S. A general memristor-based pulse coupled neural network with variable linking coefficient for multi-focus image fusion. Neurocomputing 2018, 308, 172–183. [CrossRef]

11. 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]

12. Qiu, X.; Li, M.; Zhang, L.; Yuan, X. Guided filter-based multi-focus image fusion through focus region detection. Signal Process. Image Commun. 2019, 72, 35–46. [CrossRef]

13. Prabhakar, K.; Srikar, V.; Babu, R. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, pp. 4724–4732.

14. Liu, Y.; Liu, S.; Wang, Z. Multi-focus image fusion with dense sft. Inf. Fusion 2015, 23, 139–155. [CrossRef]

15. Yang, Y.; Cao, W.; Wu, S.; Li, Z. Multi-Scale Fusion of Two Large-Exposure-Ratio Images. IEEE Signal Process. Lett. 2018, 25, 1885–1889. [CrossRef]

16. Fu, M.; Li, W.; Lian, F. The research of image fusion algorithms for ITS. In Proceedings of the 2010 International Conference on Mechatonic Automation and Control Engineering, Wuhan, China, 26–28 June 2010; pp. 2867–2870.

17. Goshtasby, A.A. Fusion of multi-exposure images. Image Vis. Comput. 2005, 23, 611–618. [CrossRef]

18. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 33, 2341–2353. [PubMed]

19. Dong, Z.; Lai, C.S.; He, Y.; Qi, D.; Duan, S. Hybrid dual-complementary metal–oxide–semiconductor/memristor synapse-based neural network with its applications in image super-resolution. IET Circuits Devices Syst. 2019, 13, 1241–1248. [CrossRef]

20. Dong, Z.; Du, C.; Lin, H.; Lai, C.S.; Hu, X.; Duan, S. Multi-channel Memristive Pulse Coupled Neural Network Based Multi-frame Images Super-resolution Reconstruction Algorithm. J. Electron. Inf. Technol. 2020, 42, 835–843.

21. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. IEEE Trans. Image Process. 2018, 27, 2049–2062. [CrossRef]

22. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, 4–7 January 1998; pp. 839–846.

23. Zhou, Z.; Wang, B.; Li, S.; Dong, M. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters. Inf. Fusion 2016, 30, 15–26. [CrossRef]

24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

25. Ma, J.; Liang, P.; Yu, W.; Chen, C.; Guo, X.; Wu, J.; Jiang, J. Infrared and visible image fusion via detail preserving adversarial learning. Inf. Fusion 2020, 54, 85–98. [CrossRef]

26. Li, H.; Wu, X.-J. DenseFuse: A Fusion Approach to Infrared and Visible Images. IEEE Trans. Image Process. 2019, 28, 2614–2623. [CrossRef]

27. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. Inf. Fusion 2017, 36, 191–207. [CrossRef]

28. Wang, Z.; Bovik, A. A universal image quality index. IEEE Signal Process. Lett. 2002, 9, 81–84. [CrossRef]

29. Ma, K.; Duanmu, Z.; Yeganeh, H.; Wang, Z. Multi-Exposure Image Fusion by Optimizing A Structural Similarity Index. IEEE Trans. Comput. Imaging 2017, 4, 60–72. [CrossRef]

30. Van Vliet, L.J.; Young, Y.T.; Beckers, G.L. A nonlinear Laplace operator as edge detector in noisy images. IEEE Trans. Image Process. 1995, 4, 167–195. [CrossRef]

31. Abdullah-Al-Wadud, M.; Kabir, H.; Dewan, M.A.A.; Chae, O. A Dynamic Hashing Equalization for Image Contrast Enhancement. IEEE Trans. Consum. Electron. 2007, 53, 593–600. [CrossRef]

32. Muniyappan, S.; Allirani, A.; Saraswathi, S. A novel approach for image enhancement by using contrast limited adaptive histogram equalization method. In Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 4–6 July 2013; pp. 1–6.

33. ACDSee. Available online: https://www.acdsee.cn/ (accessed on 29 October 2020).

34. Eskicioglu, A.; Fisher, P. Image quality measures and their performance. IEEE Trans. Commun. 1995, 43, 2959–2965. [CrossRef]

35. Rao, Y.J. In-fibre bragg grating sensors. Meas. Sci. Technol. 1997, 8, 355. [CrossRef]

36. Van Aardt, J.; Roberts, J.W.; Ahmed, F.B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. J. Appl. Remote. Sens. 2008, 2, 1–28. [CrossRef]

37. Vranjes, M.; Rimac-Drlje, S.; Grjic, K. Locally averaged PSNR as a simple objective Video Quality Metric. In Proceedings of the 2008 50th International Symposium ELMAR, Zadar, Croatia, 10–12 September 2008; pp. 17–20.
38. Hossain, M.A.; Jia, X.; Pickering, M. Improved feature selection based on a mutual information measure for hyperspectral image classification. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 June 2012; pp. 3058–3061.

39. Gao, M.; Chen, C.; Shi, J.; Lai, C.S.; Yang, Y.; Dong, Z. A Multiscale Recognition Method for the Optimization of Traffic Signs Using GMM and Category Quality Focal Loss. Sensors 2020, 20, 4850. [CrossRef]