An image enhancement algorithm of video surveillance scene based on deep learning

Wei-wei Shen | Lin Chen | Shuai Liu | Yu-Dong Zhang
1 Institute of Information Engineering, Suqian University, Suqian, China
2 Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, China
3 Hunan Xiangjiang Artificial Intelligence Academy, Hunan Normal University, Changsha, China
4 School of Informatics, University of Leicester, Leicester, UK

Correspondence
Shuai Liu, Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China. Email: liushuai@hunnu.edu.cn
Yu-Dong Zhang, School of Informatics, University of Leicester, Leicester LE1 7RH, UK. Email: yudongzhang@ieee.org

Abstract
Target enhancement is the most important task in a video surveillance system. In order to improve the accuracy and efficiency of target enhancement, and better deal with the subsequent recognition, tracking, behaviour understanding and other processing of targets, a deep learning-based image enhancement algorithm for video surveillance scenes is proposed. First, the super-resolution reconstruction of the image is carried out through the image super-resolution reconstruction method based on the hybrid deep convolutional network to improve the sharpness of the image. Then, for the reconstructed video surveillance scene image, the watershed image enhancement algorithm based on morphology and region merging is used to realize the enhancement of the video surveillance scene image. Deep learning algorithms can improve the accuracy of image enhancement through iterative calculations. Experimental results show that after image enhancement in daytime, night and noisy video surveillance scenes, the maximum enhancement difference rate is less than 0.5%, the cross-linking degree is close to 1, and the average image enhancement time is less than 1.3 s. It can realize image enhancement of video surveillance scenes and improve the image clarity of the video surveillance scene.

1 INTRODUCTION

In daily life, human beings get most information by using visual system, which belongs to image information, and so image is an irreplaceable information carrier for human perception of the world [1]. With the development of science and technology, humans use computer to simulate the process of human brain recognizing the target in the image, in order to get the method of image processing, and make the image more suitable for human eye observation or convenient for instrument detection [2]. In the field of computer vision application research, image enhancement is the first step, which is very important in the process of digital image processing. If the quality of image enhancement results is very good, it will be of great help to the next step of target recognition, image recognition, scene analysis and other work. Therefore, the research of image enhancement is of great significance for image processing, image analysis and other fields. The research of image enhancement algorithm has made great progress, but because of its complexity, there are still many problems to be solved. So far, there is no image enhancement method that can be applied to all fields, and there is no general evaluation standard [3]. Traditional image enhancement methods are difficult to extract features, and the enhancement effect is difficult to fully meet various requirements, so there is...
still a lot of room for improvement. People are still trying to combine the application of advanced algorithms into the field of image enhancement, hoping to produce better enhancement results. Therefore, further research on image enhancement algorithm is still of great practical significance [4].

In the research and application of image, people are often interested in some parts of the image, which are often called targets or prospects. They generally correspond to specific regions with unique properties in the image. In order to identify and analyse the objects in the image, it is necessary to separate them from the image [5]. On this basis, people can further measure the target and use the image. With the development of computer technology and digital image monitoring technology, the first level targets of military key alert have begun to use the digital image monitoring system to improve the monitoring efficiency [6]. In order to further improve the intelligent degree of the monitoring system, it is necessary to analyse the image effectively. Video surveillance scene image enhancement is an important image analysis technology [7].

Related scholars have done research on image enhancement algorithms for video surveillance scenes, such as signal-to-noise ratio (PSNR) [8] and structural similarity (SSIM) [9]. The main advantage of PSNR algorithm is that the algorithm does not need parameter control, and the enhancement accuracy is relatively high, but the algorithm needs to calculate the similarity between adjacent regions. As the most basic calculation unit, the SSIM is an index that measures the similarity of two images. It can improve the accuracy of image enhancement, but the algorithm takes a long time, resulting in low efficiency of the algorithm.

Deep learning is a branch of machine learning, which is a new direction in the recent decade. The essence of deep learning is bionics, which simulates the process of human brain perception and cognition. It is the research goal of deep learning to construct the neural network of human brain for analytical learning and imitate the mechanism of visual system to analyse image, sound and text information. Deep learning is developed on the basis of artificial neural network. At the end of the last century, neural network was once paid great attention to in the field of machine learning, but later, there appeared some problems, such as easy fitting, difficult to adjust parameters, slow training speed, and not more efficient than other methods when there were fewer levels, so it gradually faded out [10]. Until 2006, Geoffrey Hinton put forward the idea of deep learning. Subsequently, deep learning has been paid attention to by many scholars in academia and widely used in industry. It has made a breakthrough in the fields of speech recognition, image recognition and natural language processing. The difference between deep learning and traditional shallow learning is that deep learning requires deep network structure and attaches great importance to feature learning. Compared with the method of artificial feature construction, it is easier to obtain all kinds of connotative information by using massive data for training and learning features. At present, deep learning has made remarkable achievements in many fields, and has become a research hotspot that attracts the attention of scholars in the industry. Moreover, the upsurge of deep learning is still continuing, and continuous major breakthroughs have been made, which will have a significant impact on the field of machine learning.

In this paper, deep learning technology is used in the video surveillance scene image enhancement, and an image enhancement algorithm of video surveillance scene based on deep learning is proposed to realize the video surveillance scene image enhancement.

2 | IMAGE ENHANCEMENT ALGORITHM OF VIDEO SURVEILLANCE SCENE BASED ON DEEP LEARNING

This chapter is based on the super-resolution (SR) reconstruction of the image based on the hybrid deep convolutional network, which is based on the image background and target characterized by morphological opening, closing mixing operation and the Laplacian sharpening technology. The contour is then enhanced by the Laplacian operator. According to the brightness of the video surveillance scene image area, the object and background area of the video surveillance scene image are marked, and the modified gradient amplitude image watershed is segmented. Finally, RMMS is used to merge the segmented image regions, and to merge the missing labelled regions into the target or background regions until there is no new region merging algorithm. The above steps can effectively improve the clarity of the reconstructed image and improve the accuracy of the reconstruction, and are used to implement an image enhancement algorithm for a video surveillance scene.

2.1 | Image SR reconstruction method based on hybrid deep convolution network

Super resolution (SR) reconstruction aims to reconstruct high-resolution (HR) images with rich details by inputting one or more low-resolution (LR) images. As an ill posed inverse problem, reconstruction technology needs to collect and analyse as many adjacent pixels as possible to obtain more clues, which can be used to supplement the missing pixel information in the process of up-sampling. The signal image super resolution (SISR) reconstruction of video surveillance scene image is to use the rich information contained in a video surveillance scene image and the visual priors obtained from the sample image to identify important visual clues, fill in details, and present as faithfully and aesthetically as possible. In recent years, reconstruction technology has been widely used in medical image, satellite remote sensing, video surveillance and other fields [11]. This paper proposes an image SR reconstruction method based on hybrid deep convolution network, which is one of the core technologies in the field of deep learning. In this method, the codec structure constructed by convolution and deconvolution can remove the noise generated in the process of image reconstruction. Through the mixed use of different convolution methods, the end-to-end network is constructed, and the SR image in line with the original image is reconstructed [12].
In this paper, deep learning technology is used in the video surveillance scene image enhancement, and an image enhancement algorithm of video surveillance scene based on deep learning is proposed to realize the video surveillance scene image enhancement. Through the image SR reconstruction method based on the hybrid deep convolutional network, the image is SR reconstruction to improve the sharpness of the image. Then, for the reconstructed video surveillance scene image, the watershed image enhancement algorithm based on morphology and region merging is used to realize the enhancement of the video surveillance scene image.

2.1.1 Up-sampling

In the task of SR reconstruction of video surveillance scene image based on deep learning, the reconstruction model of high-resolution image (formula 1) is used. In the up-sampling module, the LR image of video surveillance scene is scaled to a sub LR image with the same number of pixels as the high-resolution image [13]. The simplest way of up-sampling is resampling and interpolation: the input image is rescaled to a desired size, and the pixels of each point are calculated at the same time. In this paper, a deconvolution network structure is used to sample the input LR image to the target image size [14]. In this part, a deconvolution layer with convolution kernel of $3 \times 3$ is used to sample the LR image of the video surveillance scene to the same scale of the target image as the input of the feature extraction layer.

$$I^H = D^{-1}I^L - S$$  (1)

where $I^H$ is the high-resolution image of video surveillance scene; $D^{-1}$ is the up-sampling operation; $I^L$ is the low-resolution image of video surveillance scene; $B^{-1}$ is the deblurring operation; $-S$ is the noise reduction operation.

2.1.2 Image feature extraction

In traditional image restoration, a strategy of image feature extraction is to extract image blocks intensively, and then use a set of pre-trained bases (such as principal component analysis (PCA), discrete cosine transform (DCT) etc.) to represent them. In convolution neural network, this part can be incorporated into the basic optimization of the network, and convolution operation can automatically extract video surveillance scene image features [15]. Formally, the sub network of feature extraction in this paper is represented as follows:

$$F_1(X) = \max (0, W_1 \otimes X + b_1)$$  (2)

where $Z$ is the LR image of the video surveillance scene after up-sampling; $W_1$ and $b_1$ represent the convolution weight and offset, respectively, where the size of $W_1$ is $3 \times 3 \times 64$; $\otimes$ is the convolution operation, 0 boundary is added and step size is 1, so that the input and output sizes are consistent and the boundary rank reduction is prevented.

2.1.3 Structure of codec denoising

In the feature denoising structure, convolution and deconvolution are cascaded to construct the encoding and decoding structure, which can eliminate the feature noise of video surveillance scene image to the maximum extent. Convolution layer retains the main video surveillance scene image content, while deconvolution layer is used to compensate for the details, so as to achieve good denoising effect and better retain the video surveillance scene image content. Formally, this layer is expressed as:

$$H_1(F_1) = \max (0, W_2 \otimes F_1 + b_2)$$  (3)

$$H_2(F_1) = \max (0, W_3 \otimes H_1 + b_3)$$  (4)

where the output in the image feature extraction stage of video surveillance scene is $F_1$; $W_2$ and $b_2$ are convolution weight and offset size. $W_3$ and $b_3$ are deconvolution weights and offsets; $H_1(F_1)$ and $H_2(F_1)$ are the image features of video surveillance scene extracted by deconvolution one operation and two operations in turn; $H_1$ is deconvolution one operation.

The convolution layer gradually reduces the size of the feature map, retains the main image information of the video surveillance scene, and obtains the abstract content of the image features of the video surveillance scene. The deconvolution layer gradually increases the size of the feature map, enlarges the size of the feature map, and restores the detail information of the image features of the video surveillance scene. At the same time, jump connection is used to speed up the training process, so as to ensure that the input and output sizes of the encoding and decoding structure are consistent. At the same time, it also ensures the test efficiency in the case of limited computing power of the mobile terminal, and obtains the image feature map of the video surveillance scene to remove the noise [16].

All 64 convolution cores of $3 \times 3 \times 64$ are used in this part of the convolution layer. However, the first half of the convolution operation does not add 0 boundary and the step size is 2, so that the output size of the feature becomes half of the input size. In the second part of deconvolution feature recovery, 0 boundary is added and the step size is 2. The video surveillance scene feature map is restored to the original size to ensure the image scale integrity.

2.1.4 Reconstruction of video surveillance scene image

In the process of video surveillance scene image reconstruction, the input of hidden state feature image $F_2$ and the output of SR reconstruction image can be regarded as the inverse operation of feature extraction stage. In network convolution, convolution
kernel $W^d_c$ is used as a reaction base coefficient, and each position of high-dimensional hidden state image feature is regarded as SR image domain to obtain SR reconstruction image. For this reason, a convolution layer is defined to generate the final SR image $I^{SR}$

$$I^{SR} = W^d_c \otimes F_4 + b_3 \tag{5}$$

where $W^d_c$ is $\epsilon$ convolutions with the size of $1 \times 1 \times 64$, $\epsilon$ is the number of video surveillance scene image channels, $b_3$ is the convolution bias, and $F_4$ is the convolution kernel after four convolutions.

The first four convolutions in this part use 64 convolutions with $3 \times 3 \times 64$ to check the image features of video surveillance scene for high-dimensional extraction. The convolution kernel convolution point spacing is 1, 1, 2 and 4, respectively, which can calculate the image features with a larger receptive field. Then, a convolution layer composed of 64 convolution cores with $3 \times 3 \times 64$ is used to sum the first extracted image features of video surveillance scene with the high-dimensional features to ensure the full utilization of the features. Finally, the required SR image is reconstructed by using the convolution layer of $\epsilon$ convolution cores with $1 \times 1 \times 64$.

2.1.5 Loss function and optimization

For any given training data set $\{X^{(i)}, Y^{(i)}\}_{i=0}^{n-1}$, the goal is to find an accurate mapping value $F(x)$, so as to minimize the mean square error (MSE) between the SR reconstructed image $I^{SR}$ and the real image $T$, and improve the image quality evaluation index—peak signal-to-noise ratio (PSNR). Although high PSNR does not represent absolute excellence of reconstructed image, satisfactory performance is still observed when alternative evaluation indexes are used to evaluate the model.

$$MSE = \frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (P(i, j) - T(i, j))^2 \tag{6}$$

where $P(i, j)$ and $T(i, j)$ represent the predicted image and the real image, respectively; $i$ and $j$ are the reconstruction times of the height and width of the video surveillance scene image; $H$ and $W$ are the height and width of the video surveillance scene image, respectively.

$$f_{het} = MSE + \beta \|\theta\|^2 \tag{7}$$

That is,

$$f_{het} = \frac{1}{N} \sum_{i=0}^{N-1}\|F(x) - Y\|^2 + \beta \|\theta\|^2 \tag{8}$$

where $f_{het}$ is the loss function; $\beta \|\theta\|^2$ is the multiplicative coefficient of weight attenuation; $N$ is the optimization times.

2.2 Watershed image enhancement algorithm based on morphology and region merging

2.2.1 Morphological open–close hybrid operation

Mathematical morphology is a tool for mathematical analysis of images based on structural elements. In the process of mathematical analysis, structural elements with certain morphological structure are used to measure and extract the corresponding shape in the image, so as to further achieve the purpose of image recognition and analysis. In the process of image processing, there are four basic operations of mathematical morphology: expansion, corrosion, open operation and closed operation.

Image etching can remove small protrusions or bursts of objects smaller than structural elements. By selecting structural elements of different sizes, objects of different sizes can be removed from the image. The expansion operation can enlarge the image and fill the holes in the object.

In morphology, structural element is the most basic concept. The function of structural elements in morphological transformation is equivalent to 'filter window' in signal processing. $B \in B(x)$ is used to represent the target structure element in the video surveillance scene image. For each point $x$ in the video surveillance scene image space $E$, the operations of corrosion $X$ and expansion $Y$ are defined as follows:

$$X = E \circ B = \{X : B(x) \subset E\} \tag{9}$$

$$Y = E \oplus B = \{Y : B(y) \cap E \neq \emptyset\} \tag{10}$$

Morphological open–close hybrid operation can achieve the goal of eliminating the hole of target details in the image on the premise of maintaining the integrity and position of target information. This step can eliminate the local extreme caused by irregular grey disturbance and noise in the gradient image before the subsequent processing, and preserve the object contour and information of the video surveillance scene image.

2.2.2 Laplacian sharpening

The edge and details of the video surveillance scene image are further emphasized, and the video surveillance scene image is sharpened to improve the contrast [17]. Laplacian sharpening image is related to the mutation degree of the surrounding pixels of a pixel to this pixel, and the transition degree of the pixel can be found according to the second differential. When the grey value of the centre pixel in the neighbourhood is lower than the average grey value of the neighbourhood, the grey value of the centre pixel should also decrease; when the grey value of the centre pixel is higher than the average grey value of the neighbourhood, the grey value of the centre pixel should increase. In
this way, image sharpening of video surveillance scene is realized [18].

The relationship between second-order differential and pixel is as follows:

\[
\begin{align*}
\frac{\partial^2 f}{\partial x^2} &= f(x, y) - f(x - 1, y) \\
\frac{\partial^2 f}{\partial y^2} &= f(x, y) - f(x, y - 1)
\end{align*}
\]  

(11)

where the input video surveillance scene image is \( f(x, y) \), and \( x \) and \( y \) represent the pixels of the image. The central mask coefficient of Laplace is as follows:

\[
\nabla^2 f = 4f(x, y) - f(x - 1, y) - f(x + 1, y) - f(x, y - 1) - f(x, y + 1)
\]

(12)

Four neighbourhood matrix templates are obtained.

\[
\begin{pmatrix}
0 & -1 & 0 \\
-1 & 4 & -1 \\
0 & -1 & 0
\end{pmatrix}
\]

(13)

Similarly, the sharpening process of eight neighbourhood of video surveillance scene image is as follows:

\[
\begin{align*}
\nabla^2 f &= 8f(x, y) - f(x - 1, y - 1) - f(x - 1, y + 1) - f(x - 1, y - 1) - f(x + 1, y - 1) - f(x + 1, y + 1) - f(x, y - 1) - f(x, y + 1) \\
\end{align*}
\]

(14)

The matrix template is:

\[
\begin{pmatrix}
-1 & -1 & -1 \\
-1 & 8 & -1 \\
-1 & -1 & -1
\end{pmatrix}
\]

(15)

By replacing the value at the original pixel \((x, y)\) with the obtained value, similar boundary can be obtained, and the sharpened video surveillance scene image can be obtained by the following formula:

\[
\delta(x, y) = \begin{cases} 
\sigma^2 f(x, y), & \sigma^2 f(x, y) < 0 \\
\sigma^2 f(x, y), & \sigma^2 f(x, y) \geq 0
\end{cases}
\]

(16)

2.2.3 Marking background and target

When looking for the mark, the target object is relatively bright, while the background corresponds to the dark area. If the background area of the video surveillance scene image is dark at this time, the image binarization can be used to find the corresponding area of the background, so as to ensure that the external markers can be included in the background. Because the target object in the video surveillance scene image is generally located in the prominent position of other objects in the image, the target object may be the largest area of the local area pixel value, which can be found by finding the local area extreme value. The large value is used as the marker of the target object [19]. The process of finding internal and external markers is shown in Figure 1.

First, the video surveillance scene image is continuously opened and closed to generate the maximum and minimum value of the image area, and the maximum value point is marked as the target. Then, the video surveillance scene image is binarized, and the pixel position information is transformed into the grey information of the video surveillance scene image by distance transformation. Then, the boundary between the adjacent areas of the video surveillance scene image is taken out to form the background marker, and the gradient amplitude image of the video surveillance scene image is modified.

2.2.4 Region merging based on maximum similarity (RMMS)

After the marking in the previous step, there are still some missing unmarked areas in the video surveillance scene image. This method can transform the merged region to the target region or background region. RMMS is divided into two stages. The first stage is sampling iteration method, which merges the unmarked regions which are more similar to the background set until there is no new region in the background set to merge with a certain region of the unmarked set. In each iteration step, the background set and unlabelled set are updated continuously, the background set expands and the unlabelled area set shrinks. The second stage focuses on the remaining regions after the completion of the first stage. In the unlabelled region set, regions within the set are merged according to the maximum similarity of regions. The merged regions belong to the target or background, and are handed over to the next round for completion. The two phases are implemented successively until there is no new regional merging. The unmarked area that is more similar to the background set is merged with the background set, until
there is no new area in the background set and a certain area of the unmarked set, the image is merged.

- $A$ is the neighbourhood of $B$, and $S_A = \{S^i\}_{i=1,2,...,q}$ is the neighbourhood of $A$.
- The degree of similarity $\rho(A, S^i)$ between $A$ and all its neighbours is calculated.
- When the similarity between region $A$ and region $B$ is $\rho(B, A) = \max \rho(A, S^i)$, region $B$ and region $A$ are merged.

2.2.5 Flow chart of the algorithm

Figure 2 shows the flow chart of the proposed algorithm. First, the obtained video surveillance scene image is transformed into grey image, and the grey image is opened and closed by morphology, and then the contour is enhanced by Laplacian [20]. Then, based on the brightness of the video surveillance scene image region, the object and background region of the video surveillance scene image are marked, and the modified gradient amplitude image watershed is segmented. Finally, RMMS is used to merge the segmented image regions, and the missing marked regions are merged into the target or background regions until there is no new region merging algorithm.

3 EXPERIMENTAL ANALYSIS

In order to further test the effect of the proposed algorithm on video surveillance scene image processing, the experiment is completed on a computer with Intel Core i7-7500u CPU and 8 GB RAM. The tool used in the experiment is MATLAB2015.

3.1 Super resolution reconstruction effect of video surveillance scene image

In order to test the enhancement effect of the proposed algorithm on video surveillance scene image, the video surveillance scene image in daytime, night and noisy video surveillance scene image, in Figure 3, are taken as examples to test the image quality change after processing by this algorithm.
In this paper, the visual effect of video surveillance scene image is processed by the algorithm. (a) Daytime video surveillance scene image, (b) night video surveillance scene image, (c) video surveillance scene image with noise.

The visual effect of video surveillance scene image in Figure 3 after SR reconstruction is shown in Figure 4. Comparing Figures 3 and 4, we can see that the clarity of video surveillance scene images of daytime, night and noisy types can be improved after the proposed algorithm is processed. The reason is that the proposed algorithm can directly realize the end-to-end mapping between LR and SR of video surveillance scene images, and adopt convolution and deconvolution cascade encoding and decoding method to improve the clarity of video surveillance scene images and eliminate noise, so as to solve the noise in the image features and optimize the image quality of video surveillance scene.

The above experimental results are used to test the effect of the algorithm from the visual angle, and the following two classical standard peak signal-to-noise ratio (PSNR) and SSIM are used to evaluate the SR reconstruction effect of the algorithm for video surveillance scene image.

\[
\text{PSNR} = 20 \log_{10} \left( \frac{255}{\text{RMSE}} \right)
\]

where the root MSE of video surveillance scene image reconstruction is RMSE and SSIM.

\[
\text{RMSE} = \sqrt{\frac{1}{pq} \| I_{SR} - T \|^2}
\]

where \( p \) is the number of rows of the video surveillance scene image; \( q \) is the number of columns of the video surveillance scene image; \( I_{SR} \) is the reconstructed video surveillance scene image; and \( T \) is the actual video surveillance scene image.

\[
\text{SSIM} (X, X_{ori}) = \left( \alpha_1 (l(X, X_{ori}))^{\alpha_2} (c(X, X_{ori}))^{\alpha_3} s(X, X_{ori})^{\alpha_3} \right)
\]

where \( \alpha_1, \alpha_2 \text{ and } \alpha_3 \text{ are weights; } \epsilon \text{ is constant; } l \text{ is the maximum pixel value of video surveillance scene image; and } s \text{ is the contrast of reconstructed video surveillance scene image.} \)

In general, it is considered that the higher the values of PSNR and SSIM the better the effectiveness of reconstructed video surveillance scene image and the accurateness of the structure.

The processing times of the video surveillance scene image is set to be processed to 150, 250, 350, 450 and 550 in turn. After testing the proposed algorithm, the peak signal-to-noise ratio and structure similarity of the three video surveillance scene images in Figure 3 are shown. The results are shown in Figure 5. It can be seen from the analysis of Figure 5 that after the algorithm is processed for many times, the peak signal-to-noise ratio of the three video surveillance scene images is greater than 36 dB; the SSIM is greater than 95%, and the SSIM is significant. It can be seen that after the three kinds of video surveillance scene image processing, the reconstructed image has good effect and accurate structure.

3.2 Image enhancement effect of video surveillance scene

When judging the performance of the proposed algorithm, the enhancement difference rate, the intersection and union ratio, and the enhancement efficiency are used. The enhancement difference rate is defined as:

\[
\frac{S_1 \cup S_2 - S_1 \cup S_2}{S_1 \cup S_2}
\]
It can be seen from the analysis of Figure 6 that after segmenting the video surveillance scene images of daytime, night and noisy types by using the proposed algorithm, the difference rate between the segmented video surveillance scene image and the original image is very low, and the maximum value is less than 0.5%, which indicates that the enhancement effect of the proposed algorithm is better.

The calculation method of intersection over-union (IOU) is as follows:

$$\text{IoU} = \frac{\text{area}(C) \cap \text{area}(G)}{\text{area}(C) \cup \text{area}(G)}$$  \hfill (21)

where \(\text{area}(C)\) and \(\text{area}(G)\) are the real background part of video surveillance scene image and the real background part in enhancement. The closer the intersection and union ratio is to 1, the higher the enhancement accuracy of video surveillance scene image.

The image processing times of the video surveillance scene are set to be processed to 150, 250, 350, 450 and 550 in turn, and the test results of the enhancement and cross parallel ratio are shown in Figure 7.

It can be seen from the analysis of Figure 7 that after the enhancement of video surveillance scene images of daytime, night and noisy types, the intersection and union ratio of the enhancement of the proposed algorithm is very close to 1, and the ratio between the real background part and the real background part in the enhancement is close to 1, indicating that this algorithm has high enhancement accuracy for video surveillance scene images.

The processing times of the video surveillance scene image are set to be processed to 150, 250, 350, 450 and 550 in turn, and the execution time of the proposed algorithm used in the enhancement of the three types of video surveillance scene images in Figure 3 is tested. The test results are shown in Table 1.

Analysis of Table 1 shows that before using the proposed algorithm to segment daytime, night and noisy video surveillance scene images, the average enhancement time of daytime, night and noisy video surveillance scene images is...
more than 12 s, and the enhancement efficiency is low; after using the proposed algorithm to segment daytime, night and noisy video surveillance scene images, the average time of scene image enhancement of daytime, night and noisy video surveillance scene images is less than 1.3 s, and the enhancement efficiency is higher. The proposed method uses a deep learning algorithm, through iterative learning, the optimal convergence state can be quickly obtained, and the image is efficiently reconstructed, which improves the efficiency of the algorithm.

4 | CONCLUSION

Image enhancement is the basis of image analysis, image recognition and image understanding. The quality of image enhancement is directly related to the effect of subsequent image processing, so image enhancement plays a very important role in the process of image processing. Image enhancement is the use of image grey, colour, texture, shape and other information in accordance with the standard, and the image has a special meaning of different regions separated, in order to ensure that the internal consistency of each region is satisfied, and the differences between regions are satisfied. Image enhancement can extract the objects that people are interested in from the complex scene, and then proceed to the next step. Video surveillance scene image enhancement can effectively divide the target and background into two parts, and help people effectively and quickly judge the target information in the video surveillance scene image. This paper takes the video surveillance scene image as the research goal, and proposes an image enhancement algorithm of video surveillance scene based on deep learning. In the experiment, the simulation test is carried out by MATLAB2016 software. After the test, the algorithm can segment various types of video surveillance scene images with high accuracy and high efficiency, and it is an effective image enhancement algorithm.

1. Through various video surveillance scene image enhancement crossover algorithms and comparative test results, it is shown that the designed algorithm can effectively improve the definition of video surveillance scene images.

2. The average enhancement time of the algorithm in this paper is less than 1.3 s, and the enhancement efficiency is high. This shows that the algorithm has better efficiency.

In the future research, the goal of applying the video surveillance scene image enhancement algorithm to the batch processing of video surveillance scene images should be to further improve the video surveillance scene image enhancement algorithm based on deep learning.

ACKNOWLEDGEMENTS

This work was supported by the Suqian Science and Technology Project (grant no. Z2019109); Natural Science Foundation of Hunan Province (grant no. 2020JJ4434); Key Scientific Research Projects of Department of Education of Hunan Province (grant no. 19A312); Hunan Provincial Science & Technology Project Foundation (2018TP1018, 2018RS3065); Innovation and Entrepreneurship Training Program of Hunan Xiangjiang Artificial Intelligence Academy, Educational Reform Project of Hunan Xiangjiang Artificial Intelligence Academy, Open Project Program of the State Key Lab of CAD&CG (grant no. A1926)

ORCID

Shuai Liu https://orcid.org/0000-0001-9909-0664

REFERENCES

1. Kotov, V.M., Averin, S.V.: Two-dimensional image edge enhancement using two orders of Bragg diffraction. Quantum Electron. 50(3), 305–308 (2020)
2. Yan, Z., Liu, S., Gu, H.: Fault image enhancement using a forward and backward diffusion method. Comput. Geosci. 131(8), 1–14 (2019)
3. Manju, V.N., Lenin, F.A.: AC coefficient and K-means cuckoo optimisation algorithm-based segmentation and compression of compound images. IET Image Proc. 12(2), 218–225 (2018)
4. Mao, J., et al.: GrabCut algorithm for dental X-ray images based on full threshold segmentation. IET Image Proc. 12(12), 2330–2335 (2018)
5. Florea, L., Florea, C.: Directed color transfer for low-light image enhancement. Digital Signal Process. 93, 1–12 (2019)
6. Fan, J., Wang, J.: A two-phase fuzzy clustering algorithm based on neurodynamic optimization with its application for PolSAR image segmentation. IEEE Trans. Fuzzy Syst. 26(1), 72–83 (2018)
7. Nongmeikapam, K., Kumar, W.K., Singh, A.D.: Fast and automatically adjustable GRBF kernel based fuzzy C-means for cluster-wise coloured
feature extraction and segmentation of MR images. IET Image Proc. 12(4), 513–524 (2018)
8. Chen, Y., et al.: Image super-resolution reconstruction based on feature map attention mechanism. Appl. Intell. 10(8), 1–14 (2021)
9. Fu, C., Duan, R., Kayacan, E.: Visual tracking with online structural similarity-based weighted multiple instance learning. Inf. Sci. 481(1), 292–310 (2019)
10. Vieira, R.B., Lambros, J.: Machine learning neural-network predictions for grain-boundary strain accumulation in a polycrystalline metal. Exp. Mech. 20(1), 1–13 (2021)
11. Abdoli, M., et al.: A quality assessment tool for performance measurement of image contrast enhancement methods. IET Image Proc. 13(5), 833–842 (2019)
12. Nazarinezhad, J., Dehghani, M.: A contextual-based segmentation of compact PolSAR images using Markov random field (MRF) model. Int. J. Remote Sens. 40(3-4), 985–1010 (2019)
13. Roy, S., Maji, P.: Medical image segmentation by partitioning spatially constrained fuzzy approximation spaces. IEEE Trans. Fuzzy Syst. 28(5), 965–977 (2020)
14. Kowal, M., et al.: Cell nuclei segmentation in cytological images using convolutional neural network and seeded watershed algorithm. J. Digit. Imaging 33(1), 231–242 (2020)
15. Haberl, M.G., et al.: CDeep3M—plug-and-play cloud-based deep learning for image segmentation. Nat. Methods 15(9), 677–680 (2018)
16. Sikandar, T., Ghazali, K.H., Rabbi, M.F.: ATM crime detection using image processing integrated video surveillance: A systematic review. Multimed. Syst. 25(3), 229–251 (2018)
17. Athanesious, J., et al.: Detecting abnormal events in traffic video surveillance using superorientation optical flow feature. IET Image Proc. 14(9), 1881–1891 (2020)
18. Luo, X., et al.: Comparative analysis on landsat image enhancement using fractional and integral differential operators. Computing 102(1), 247–261 (2020)
19. Kumar, N., et al.: Hyperspectral tissue image segmentation using semi-supervised NMF and hierarchical clustering. IEEE Trans. Med. Imaging 38(5), 1304–1313 (2019)
20. Srinivas, K., Bhandari, A.K.: Low light image enhancement with adaptive sigmoid transfer function. IET Image Proc. 14(4), 668–678 (2020)

How to cite this article: Shen, W., et al.: An image enhancement algorithm of video surveillance scene based on deep learning. IET Image Process. 1–10 (2021). https://doi.org/10.1049/ipr2.12286