



Super-Resolution Reconstruction of Weak Targets on Water Surfaces: A Generative Adversarial Network Approach Based on Implicit Neural Representation

Qilin Bi¹, Zhiqiang Lin^{1,2}, Boren Chen^{1,3*}, Minlin Lai^{1,3}, Yanyao Guo^{1,2}, Youjie Lv¹, Yali Tang¹, Chuxin Huang⁴

¹ School of Marine Engineering, Guangzhou Maritime University, Guangzhou 510725, China

² School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

³ School of Physics and Optoelectronic Engineering, Guangdong University of Technology, Guangzhou 510006, China

⁴ Yours Technology Co., Ltd., Jieyang 522095, China

Corresponding Author Email: 2112115045@mail2.gdut.edu.cn

Copyright: ©2023 IIETA. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.400630>

ABSTRACT

Received: 10 June 2023

Revised: 30 September 2023

Accepted: 17 October 2023

Available online: 30 December 2023

Keywords:

generative adversarial networks (GAN), super-resolution (SR) imaging, implicit neural representation, water surface target detection, environmental interference management

In the realm of super-resolution reconstruction, challenges are posed by the interference of various weather conditions, such as rain and fog, as well as complex environmental backgrounds, notably water surfaces. This research addresses the critical issue of feature information loss, lack of edge detail, and inconsistent lighting in the reconstruction of weak targets on water surfaces. The study introduces a novel approach employing a generative adversarial network (GAN) based on implicit neural representation. This method is specifically tailored for enhancing the clarity and detail of small targets on water surfaces. The methodology involves constructing a super-resolution (SR) image generation model that leverages the implicit neural representation of images. This model adeptly handles the nuances of small water surface targets. A comprehensive evaluation framework is developed, incorporating network weights, deviation coefficients, edge loss, and balance loss as key indicators. This aids in the formulation of an adaptive loss function for the SR image generation model, significantly improving the model's performance in challenging conditions. To validate the efficacy of the proposed approach, datasets of low-resolution (LR) and high-resolution (HR) images of weak targets on water surfaces were compiled. These datasets were created using simultaneous imaging of the target with both HR and LR cameras. Comparative analysis with existing popular algorithms demonstrates the superiority of the proposed method in SR reconstruction of weak targets in complex water surface environments. The results high-light the model's enhanced ability to identify and classify weak targets with high reliability and accuracy, even under challenging weather and environmental conditions.

1. INTRODUCTION

In environments characterized by complex water bodies such as ice regions, islands, and shallow seas, the collection of images via intelligent detection devices, including unmanned ships and drones, is significantly challenged. These devices often yield images of suboptimal quality and resolution, thereby impeding the acquisition of low-cost, high-quality imagery. In the context of such complex aquatic environments, there is an increasing focus on the investigation of SR image reconstruction, as evidenced in studies [1, 2]. Single Image Super-Resolution (SISR) is also garnering attention [3, 4]. SR reconstruction techniques have found extensive applications in diverse fields, such as face recognition [5, 6], remote sensing imaging [7], medical imaging [8], and satellite imaging [9].

The evolution of SR methodologies has traversed various stages, including classical interpolation and amplification [10-12], degenerate models [13-15], and classical machine learning [16, 17]. More recently, the incorporation of deep learning has marked a significant advancement in the

enhancement of SR image quality [18]. Dong et al. [19] pioneered the Super-Resolution Convolutional Neural Network (SRCNN) project to address SR challenges, demonstrating substantial improvements over traditional learning-based methods. Following this, Dong et al. [20] introduced the Fast Super-Resolution Convolutional Neural Network (FSRCNN), further refining the quality of SR images. Despite these advancements, the FSRCNN, with its deeper layers, incurs considerable computational overhead and still exhibits notable differences from the original images. Addressing the convergence speed, Kim et al. [21] developed the Very Deep Super-Resolution (VDSR) method, integrating a residual structure and achieving notable results in terms of reconstruction efficacy and computational efficiency. Lim et al. [22] proposed the Enhanced Deep Super-Resolution (EDSR) network, removing the Batch Normalization (BN) layer from SRResNet to expedite network convergence. Additionally, Ledig et al. [23] approached the SISR problem through GAN, introducing the SRGAN, which significantly enhanced the Mean Opinion Score (MOS) for the quality of

highest SR images.

In the realm of SISR, researchers typically initiate by generating a LR image. This process involves reducing the resolution and introducing artificial noise, such as Gaussian blur, into a HR image within a controlled setting. The resulting dataset serves as the foundation for both training and evaluating the neural network. Subsequently, the SR model is developed through this training process.

However, several factors impact the quality of image acquisition, including installation errors, strong light interference, fog occlusion, and the use of low-resolution cameras in acquisition equipment. These elements contribute to the pathological nature of the SR problem, wherein the transformation of a given LR image into an HR image leads to multiple possible solutions. Consequently, the approach of artificially degrading HR images to create LR counterparts fails to accurately mimic the low-quality LR images captured in challenging maritime environments. Models trained on such datasets are likely to fall short in delivering satisfactory reconstruction effects. This limitation underscores the instability of SISR technology in complex water environments and highlights significant constraints in the current SR algorithms for industrial applications.

In response, this research introduces an innovative approach incorporating an adaptive loss neural network method based on implicit neural representation and GAN. This method focuses on enhancing network accuracy by rapidly removing the normalized BN layer from the residual structure. The Local Implicit Image Function (LIIF) is employed to represent both natural and complex images effectively. This function assists in reconstructing SR images, magnifying the resolution of LR

images, invoking the adaptive robustness loss function, and dynamically selecting the loss function during the training phase. The culmination of this research involves applying the proposed method to train and reconstruct studies using independently constructed heterogeneous source image datasets. These efforts are aimed at assessing the quality of SR images. The results demonstrate a high evaluation in terms of image quality, with the reconstruction effect showing marked superiority over other prevalent methods. These outcomes reveal a significant enhancement in human eye perception and the image quality rating index.

2. MODELLING

The application SRGAN for the reconstruction of LR images substantially enhances the detection accuracy and precision of weak targets on water, contributing to an increase in sensory realism [24]. SRGAN's architecture comprises two primary components: the generator network (G) and the discriminator network (D). The generator network employs deep residual structures for extracting image information, followed by convolution operations to amplify the picture and generate SR images. In contrast, the discriminator network, utilizing a multi-layer convolutional network, captures image information and computes the differences between SR and HR images. It provides feedback on SR images both globally, covering the entire image, and locally, on a pixel-by-pixel basis. This process is illustrated in Figure 1, showcasing the operational mechanism of the SRGAN network.

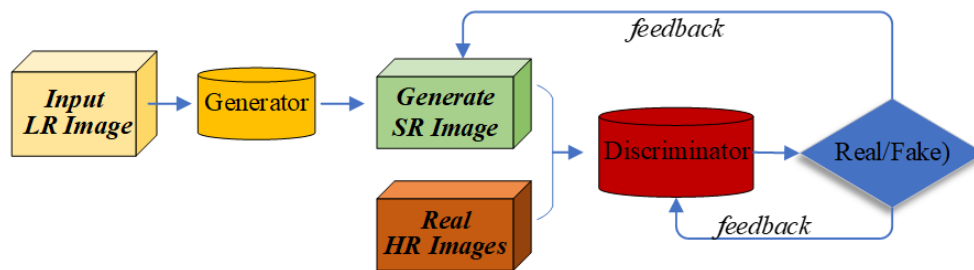


Figure 1. SRGAN gaming procedure

The operational sequence begins with feeding LR images into the generator (G) to construct SR images. Subsequently, both HR and SR images are inputted into the discriminator (D). This initiates a continuous discrimination process by both G and D, based on their respective parameters, fostering the learning process of the entire network. The generator (G) strives to produce SR images that deceive the discriminator (D), while the discriminator discriminates between the real and generated images. The interaction between G and D can be conceptualized as a game, represented by the following objective function:

$$\min_{\theta_G} \max_{\theta_D} E_{I^{HR} \sim p_{\min}(I^{HR})} \left[\log D_{\theta_D}(I^{HR}) \right] + E_{I^{LR} \sim p_G(I^{LR})} \left[\log \left(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})) \right) \right] \quad (1)$$

The discriminator (D) aims to classify the result $D(x)$ as close to 1 as possible, while the discriminator's output for $D[G(z)]$ should be closer to 0. In this context, ILR, IHR, and ISR represent the LR image, HR image, and SR image,

respectively. The training objective for the generator (G) is to produce outcomes for $D[G(z)]$ that are closer to 1.

Existing GAN algorithms face significant limitations in the HR reconstruction of weak targets on water. Challenges arise particularly when reconstructing SR images at 4x or higher resolutions, where environmental noise can lead to edge blurring and artifacts. This issue often results in substantial distortion and reduced quality of SR images. To overcome this problem, this research introduces an Adaptive Implicit Generative Adversarial Network (AIGAN) approach, based on implicit neural representation. This method is designed to mitigate the impact of environmental noise and enhance the clarity and quality of SR images.

2.1 Development of the AIGAN algorithm for SR reconstruction of targets on water surface

In the SR reconstruction process, a common approach involves the use of a HR image dataset. Deep learning techniques are then applied to generate a corresponding LR dataset. When GANs and their enhanced models are employed

for model training and parameter optimization, the SR reconstruction of the HR image can be accomplished. Effective results are typically achieved when the reconstructed HR and SR images are of the same scale. However, reconstructing the SR image at a higher scale presents challenges such as pronounced noise and a lack of detailed image information, leading to reduced correlation between the pixels of the reconstructed image. To address these challenges, this study introduces the AIGAN model, specifically designed for the super-resolution reconstruction of targets on water surfaces, as illustrated in Figure 2.

The traditional convolution approach for resolution scaling often results in lower pixel correlation in the reconstructed images, manifesting in excessive noise and lost edge detail, especially at higher magnification scales. To counter this, a linear SR network is developed that learns the implicit neural representation function of the image. This approach enables the prediction of an image with a higher resolution than the original LR image. The SR images reconstructed through this

method surpass the conventional magnification limitations, allowing for effectively infinite image magnification. Traditional neural network designs typically employ a fixed loss function, which can significantly impede the learning rate and precision of the network. To enhance network learning efficiency and reconstruction accuracy, the model incorporates an adaptive robustness loss function. This allows for the creation of a unique form of loss function during each training iteration, enabling the application of different loss functions based on the specific needs of each training regression.

In this research, the BN layer was removed from the generator network, building upon the SRGAN framework. This addresses the issue wherein the use of the BN layer in SR problems leads to extended training times, instability, and the loss of the original contrast information in the image. The network model, depicting this innovative approach, is illustrated in Figure 2.

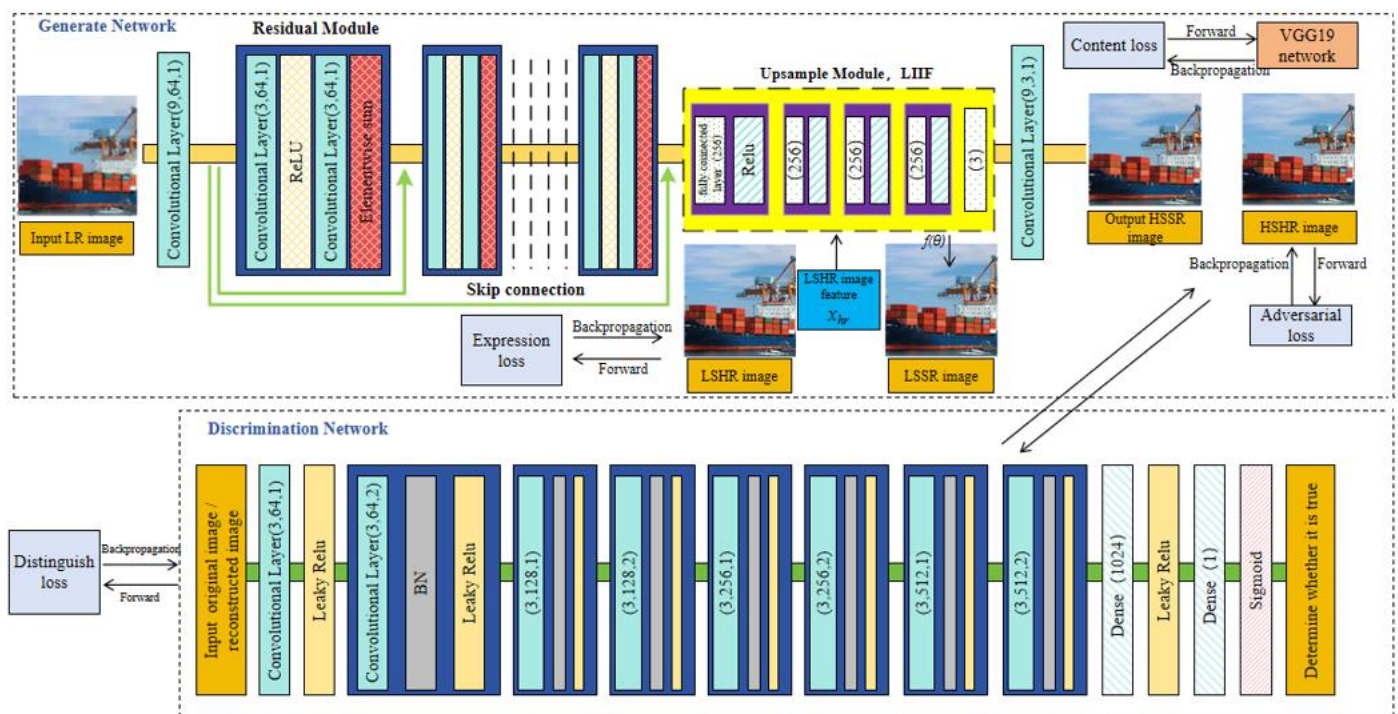


Figure 2. Model of SRGAN

Data Compilation Process: To develop a comprehensive and heterogeneous dataset for SR reconstruction, a series of images capturing water surface targets are acquired using multiple cameras. These images vary in pixel densities and are taken under different environmental conditions. This approach ensures the creation of a diverse dataset, encompassing a wide range of scenarios encountered in aquatic environments. The dataset is then categorized into two distinct groups: the HR dataset and the LR image dataset, based on the pixel density of the images. For the purpose of training and testing the SR model, the LR image dataset is divided into two subsets: LRtrain and LRtest. Similarly, the HR image dataset is split into HRtrain and HRtest subsets. This segregation results in two pairs of datasets: [LRtrain, HRtrain] for training and [LRtest, HRtest] for testing the SR model.

The images within these datasets are characterized using the implicit neural representation approach [25]. This method facilitates the continuous expression of images. A linear super-

resolution network is specifically trained to learn the local implicit image function of each image.

Generator Network: The generator network in this model is structured in several key steps. Initially, the input LR image is randomly segmented into small blocks. Each of these blocks is then represented as a shallow feature of the image, converted into a corresponding high-dimensional vector. Following this, the high-dimensional vector obtained is transformed back into a HR image block. This transformation is achieved by projecting the input shallow feature onto a subsequent high-dimensional vector through successive residual blocks. In the final stage, the LIIF replaces the up-sampling phase typically used in SRGAN. During the training process, the Multi-Layer Perceptron (MLP) determines the RGB or grayscale value of the SR image based on the HR image feature.

Discriminator Network: The discriminator network is designed to be binary, considering only two inputs: the SR image and the HR image. This design is premised on the

objective that the SR image should closely resemble the HR image. Given that the images in the dataset are grayscale, the binary discriminator network incorporates eight convolutional layers. This configuration is chosen to reduce the number of convolutional operations and enhance training efficiency. The Leaky ReLU function is selected as the activation function to prevent neuronal death. Each convolutional layer, except the first, is followed by a BN layer to streamline computations, prevent gradient vanishing, accelerate the convergence of the SR model, and improve the stability of the training process. Convolution operations are utilized to decrease image resolution as the number of features increases over time, resulting in two dense feature map layers. This approach enhances the discriminative network's convolution kernel size, further improving training efficiency. Once the structure of the discriminative network is complete, a fully connected layer is added, along with a sigmoid activation function. This combination allows for the connection of image features extracted in the earlier stages of the network, facilitating the classification of complex water image samples with probabilities ranging from 0 to 1.

Loss Function Assembly: The network uniquely incorporates three types of input images: LR, SR, and HR, distinguishing this approach from the conventional SRGAN technique. The loss function is composed of content loss, adversarial loss, discriminant loss, and expression loss.

2.2 Implicit neural expression (INR) of water surface target images

The representation of images for water surface targets draws upon the concept of implicit functions, commonly used in 3D reconstruction. Traditionally, images are represented as a 2D discrete dot matrix. However, this research proposes a continuous representation, suggesting that images can be conceptualized as continuous functions. In this approach, each image $I^{(i)}$ is represented in a 2D dot matrix form $M^{(i)} \in \mathbb{R}^{H \times W \times D}$, expressed through a MLP with the following expression:

$$s = f(x, z) \quad (2)$$

where, z is a vector representing the depth feature value of the image - the hidden feature - generated by the preceding network segment. The symbol $x \in X$ denotes the coordinates within the continuous image domain, corresponding to the pixel values of the LR image. The value s in the set S represents the pixel value of the SR image, which could be an RGB value for three-channel images or grayscale for single-channel images. The function f is the shared function for the full image representation. Figure 3 illustrates the concept of the LIIF. In the 2D space of the continuous image domain, the feature vectors $H \times W$ of $M(i)$ are uniformly distributed. The expected value $I^{(i)}(x_q)$ at point x_q for each image is represented as follows, given subsequent points assigned 2D coordinates:

$$I^{(i)}(x_q) = f(z^*, x_q - v^*) \quad (3)$$

This equation implies that the discrete hidden features are uniformly extended into the continuous 2D image domain. Here, $I^{(i)}$ is the continuous image, f is the shared function of all images using neural network representation, and z^* denotes the 2D feature vector distance from the hidden feature at position x_q . It is impractical to predict the hidden feature of x_q

using only a single point. Since z_{11}^* is the closest hidden feature to x_q , f will utilize it and its coordinates $x_q - v^*$ to calculate $I^{(i)}(x_q)$. To enable prediction of $I^{(i)}(x_q)$ using multiple points, the formula incorporates the four surrounding points, applying bilinear interpolation as follows:

$$I^{(i)}(x_q) = \sum_{\substack{m \in (0,1) \\ n \in (0,1)}} \frac{S_{mn}}{S} f(z_{mn}^*, x_q - v_{mn}^*) \quad (4)$$

For z_{mn}^* and x_q to form a bond, the area of the diagonal region is denoted by S . The MLP structure of LIIF, which combines the ReLU activation function with four fully connected layers, concludes with an additional fully connected layer, as depicted in Figure 3.

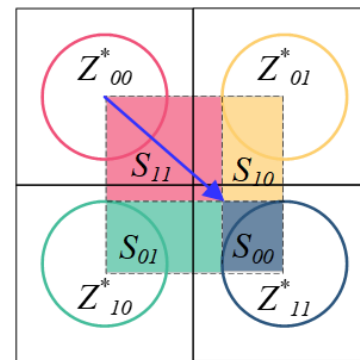


Figure 3. LIIF principle

2.3 Development of an adaptive loss function model

In the traditional SR methodology, L1 loss, L2 loss, and perceptual loss are commonly utilized as loss functions [26]. Although L2 loss is effective in addressing SR problems, it is hindered by derivative discontinuity and low sensitivity to large errors, leading to inefficiencies in solution finding. In super-resolution contexts, this can also result in disproportionately large gradients for very small loss values. The primary drawback of L1 loss is its constant uniqueness in the gradient update value. Utilizing L2 loss to guide the learning process can lead to a loss of detailed image information, and perceptual loss, essentially being a computation of L2 loss, presents similar challenges. Given the need for at least one effective loss function for each specific problem, and the impracticality of manually testing the robustness of each loss function, an adaptive robust perceptual loss function, ISR, has been formulated. This function incorporates robustness as a parameter, following the suggestion of Barron [27]:

$$l_{SR} = l_{L_2}^{HRSR} + \lambda_p l_{VGG}^{HSSR} + \lambda_g l_G^{HSSR} + \lambda_r l_R^{LSSR} \quad (5)$$

where, $l_{L_2}^{HRSR}$ represents the pixel loss function between HSSR and HSHR images; l_{VGG}^{HSSR} represents the pixel loss function of deep features of HR and SR images, and λ_p represents its weighting coefficient; l_G^{HSSR} represents the adversarial loss function, λ_r is its weighting coefficient; the weighting coefficients λ_p , λ_g , and λ_r are used to bring the L2 loss and VGG loss to the same level, generally taking $\lambda_p = 0.006$, $\lambda_g = 0.001$, $\lambda_r = 0.006$. The method for calculating the pixel loss function $l_{L_2}^{HRSR}$ between HR and SR images is:

$$I_{L_2}^{HRSR} = \frac{1}{S^2WH} \sum_{x=1}^{SW} \sum_{y=1}^{SH} [I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y}]^2 \quad (6)$$

For the LR image, W and H represent the width and height, respectively, S denotes the magnification factor, $I_{x,y}^{HR}$ represents the pixel value of the HR image at point (x, y) , and the SR image's pixel value at (x, y) is given by $G_{\theta_G}(I^{LR})_{x,y}$. The deep features of the HR and SR images are evaluated using the following formula to get the pixel loss function, I_{VGG}^{HSSR} .

$$I_{VGG}^{HSSR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left\{ \varphi_{i,j}(I^{HSSR})_{x,y} - \varphi_{i,j}[G_{\theta_G}(I^{LR})]_{x,y} \right\}^2 \quad (7)$$

where, $\varphi_{i,j}(I^{HSSR})_{x,y}$, $\varphi_{i,j}[G_{\theta_G}(I^{LR})]_{x,y}$ are the eigenvalues of the deep feature map of HR and SR images at point (x, y) , respectively. The adversarial loss function I_G^{HSSR} can be described as follows.

$$I_G^{HSSR} = \sum_{n=1}^N -\log D_{\theta_D}[G_{\theta_G}(I^{LR})] \quad (8)$$

The probability that the reconstructed image belongs to the HR image is represented by $D_{\theta_D}[G_{\theta_G}(I^{LR})]$. The following is the expression for the image implicit expression loss function, I_{LR}^{LSSR} .

$$I_{LR}^{LSSR} = I_{L_1}^{LSSR} = \frac{1}{SWH} \sum_{x=1}^{SW} \sum_{y=1}^{SH} |I_{x,y}^{LSHR} - G_{\theta_G}(I^{LR})_{x,y}| \quad (9)$$

The LSHR image's pixel value at (x, y) point is represented by $I_{x,y}^{LSHR}$, while the LSSR image's pixel value at point (x, y) generated by inputting the LR image is represented by $G_{\theta_G}(I^{LR})_{x,y}$.

3. EXPERIMENTAL METHOD AND RESULT ANALYSIS

3.1 Database construction

In the experimental phase, a critical step involves constructing a reliable database for the water surface target imaging. This is achieved by simultaneously capturing images of a weak target on water using two different cameras: a low-resolution camera (sionyx Aurora) and a high-resolution camera (PONY ES-AHD 1080PTZ). The concurrent use of these cameras aids in enhancing the generalization capability of the SR reconstruction model.

Directly utilizing images sourced from these distinct cameras presents challenges due to factors such as variations in camera installation positions, image distortion, and discrepancies in camera resolution. To effectively address these issues, a comprehensive calibration and registration process for imaging of varying sizes is employed. This involves utilizing a binocular camera calibration method for both the LR and HR cameras, as detailed in reference [28]. The implementation of this method proceeds as follows:

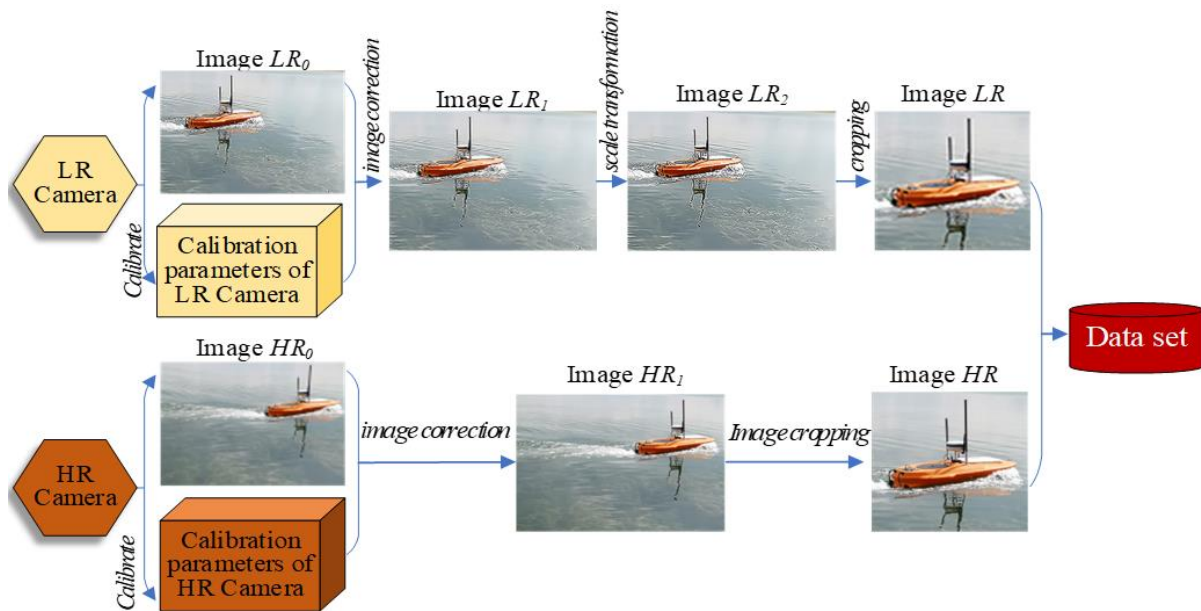


Figure 4. Method of water surface target data set construction

Initially, the experiment entailed calibrating the camera as described in Reference [28]. Subsequently, the calibrated parameters were applied to correct the initial LR₀ and HR₀ images, resulting in the formation of the adjusted image pairs LR₁ and HR₁. Following this, the LR₁ image underwent a linear interpolation process to match the resolution of the HR₁ image, thereby creating the LR₂ image. Finally, LR₂ was processed in accordance with the algorithm detailed in

literature [29], focusing on the target characteristics within both LR₂ and HR₁ images. As depicted in Figure 4, the final low-resolution and high-resolution images of the sea surface target were produced by segmenting the image into a(pixel) * b(pixel) blocks, centering the target in each segment.

In the experimental setup, two distinct sources of samples were utilized. The initial dataset is sourced from a low-resolution camera, which forms the basis for generating the LR

image dataset. Correspondingly, the HR image dataset was created by extracting matching target images from the LR dataset, captured by a HR camera. For the purpose of training the SR image reconstruction model, a total of 22,700 images were collected and cropped to a size of 512×512 pixels. The training dataset comprises 7,200 LR and 7,200 HR images of associated sea surface targets. Additionally, the test dataset includes 1,050 LR and 1,050 HR images of similar sea surface targets, ensuring that the categories of images in the test dataset are in line with those in the training dataset. For the experimental evaluation, 400 HR image test samples and 400 LR image test samples, both depicting water surface targets, were randomly selected. The final phase of the experiment involves the SR reconstruction of the LR images to obtain the matching SR images. These SR images, when paired with their corresponding HR counterparts, facilitate the derivation of evaluation parameters for the reconstructed SR images.

Table 1. Network training pseudo-code

Discriminant Network Training:

Prepare LR image set $LR\{i_1, i_2, \dots, i_m\}$ and corresponding heterogeneous image HR image set $HR\{I_1, I_2, \dots, I_m\}$ to the dataset;

for epoch do
for k-th do
Take out HR images from the HR image set IHR and LR image ILR from the LR image set according to a uniform distribution
Generate SR datasets $SR\{S_1, S_2, \dots, S_m\}$.
Update the discriminant network through the adaptive discriminant loss function.
end for
SR images are generated by extracting small leather LR images from the LR image set according to uniform distribution;
Update the generator with an adaptive loss function;
end for

3.2 Evaluation criteria

The evaluation of image quality, especially in the context of SR image reconstruction, often begins with visual observation to assess the detailed information and visual perception of the images. However, to more objectively reflect the quality of the resulting images, specific algorithms and models are employed. Standard methods for evaluating images include the Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index Measure (SSIM), and Peak Signal to Noise Ratio (PSNR) [30].

(1) PSNR: it serves as an objective assessment of image quality, primarily based on error sensitivity. It is calculated by measuring the error between corresponding pixels in the images. A higher PSNR value indicates lower image distortion following super-resolution reconstruction. The Mean Squared Error (MSE) is calculated from the following expression and is instrumental in determining PSNR:

$$MSE = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M (f_{ij} - f'_{ij})^2 \quad (10)$$

where, M and N represent the pixel size of the image, i and j denote the pixel position, and f is the pixel value. PSNR, which

represents the ratio of the signal's maximum power to its noise power, can be expressed as:

$$PSNR = 10 \lg\left(\frac{MAX_I^2}{MSE}\right) \quad (11)$$

where, MAX_I is the image's maximum PSNR pixel value expressed in decibels (dB).

(2) SSIM: it is a metric used to compare the similarity of two images, with values ranging from 0 to 1. SSIM evaluates three key components of an image: luminance, which is calculated using the mean values; contrast, determined by the standard deviation; and structure, assessed through covariance. A higher SSIM value indicates lower image distortion, reflecting better image quality. The SSIM formula, which gauges the similarity between two images, is as follows:

$$SSIM(x, y) = \frac{(2u_x u_y + c_2)(\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (12)$$

where, x and y represent the pixel values of the HR and SR images, respectively. u_x and u_y are their mean values, σ_x and σ_y are the standard deviations, and σ_{xy} is the covariance. Constants c_1 and c_2 are included to stabilize the division. Unlike PSNR, which quantifies absolute error, SSIM is a perceptual model that provides a score between 0 and 1, with higher values indicating less distortion and thus superior image quality.

(3) LPIPS: also known as perceptual loss, measures the difference between two images. It reflects the disparity between image pairs, with a lower LPIPS value indicating a smaller disparity. The LPIPS formula can be expressed as follows:

$$d(x, x_0) = \sum_1 \frac{1}{H_1 W_1} \sum_{h,w} \|w_1 \odot (\hat{y}_{hw}^1 - \hat{y}_{0hw}^1)\|_2^2 \quad (13)$$

where, d represents the distance between x and x_0 . The feature stack, extracted from the L layer, undergoes normalization in the channel dimension (channel-wise unit normalization). The number of excitation channels is then scaled using the vector W , followed by the computation of the L2 distance. The process concludes with a spatial average and a channel sum.

3.3 Performance comparison

The SR reconstruction of sea surface targets is enhanced through an advanced adversarial network model, which significantly improves the effectiveness and quality of the reconstruction. A common challenge in the SR reconstruction of sea surface targets is the limited number of identifiable targets and the predominance of single background features. Traditional SR reconstruction algorithms, which process the target and background with equal priority, often result in issues such as high noise, blurred textures, and distorted geometric features of the target. By employing image implicit neural representation, the adversarial network model elevates the reconstruction's effectiveness and accentuates the features of the image's weak targets. Therefore, an adversarial network model based on image implicit neural expression is introduced specifically for the SR reconstruction of complex, dim sea surface targets. This model is compared with various established SR methods, including Bicubic [31], SRGAN [23],

and EDSR [22]:











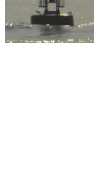










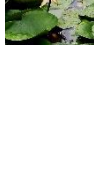
- Bicubic SR [31]: This method utilizes three interpolations based on the gray values of 16 surrounding points to achieve a closer-to-high-resolution magnification effect. The interpolation takes into account the rate of change in gray values between adjacent spots and the influence of the four nearest spots. The BiCubic function [32] serves as the basic function for this experiment.

- SRGAN [23]: The standard SR technique in SRGAN typically operates at a lower magnification, with excessive smoothness occurring at magnifications over four times, resulting in a non-photorealistic appearance. SRGAN generates detailed images through the network terminal design of GAN. The residual block in the generated network uses two 3×3 convolution kernels with 64 output channels. Post-convolution, a batch-normalized BN layer and a Parametric-ReLU activation layer are applied. The image's resolution is enhanced via two learned sub-pixel convolution layers, and the discriminative network employs the Leaky ReLU activation function ($\alpha = 0.2$) to avoid maximal pooling.

- EDSR [22]: This method leverages the residual learning-based mechanism of the ResNet network. The input image is divided into two paths by one convolution layer. One path undergoes additional convolution after passing through the n-layer ResBlock, while the other proceeds directly to the convolution output results and up-sampling processing, culminating in weighted summation at the intersection.

For testing these identification approaches, an image workstation equipped with a 2-GPU Tesla V100-32G and a 2-CPU Xeon (R) E5-2678W was utilized. The experiments were conducted on the PyTorch platform. Moreover, the training of the generative and discriminative networks was conducted separately. Table 2 presents a summary of the results of SR reconstruction for various SR models at different magnifications of water surface targets. The algorithm in this article is divided into training generative networks and discriminative networks. The training process is shown in Table 1, where the learning rate starts at 1e-4 and is halved every 100 iterations, over a total of 1000 training iterations.

Table 2. Comparison of experimental results

LR Image	Title	SR Image								HR Image
		Bicubic	SRGAN	EDSR	Our Method					
	2X									
	4X									
	8X									
	16X									

At a 16X magnification, the SR images reconstructed using the Bicubic, SRGAN, and EDSR methods show clear deformations. In contrast, the method introduced in this paper yields the most effective reconstruction, especially in maintaining uneven lighting and restoring textural details. This superiority is readily apparent to the naked eye, underscoring the proposed approach's capability in producing more accurate and visually compelling SR images.

Table 3 displays the evaluation results for different SR methods based on established criteria. In the case of the Bicubic technique used for SR image reconstruction, a noticeable decline in PSNR/SSIM values and an increase in LPIPS value are observed. This indicates that the algorithm's reconstruction quality is significantly affected by

environmental noise factors like rain and fog. Particularly, at a 16X magnification in noisy conditions, the PSNR/SSIM/LPIPS values deteriorate to 18.360/0.531/0.867, resulting in a relatively fuzzy SR image. Moreover, as the magnification rate for SR image reconstruction escalates, the performance efficiency of the Bicubic algorithm tends to decrease slightly. Notably, even at a 6X reconstruction rate, it maintains a computational speed of just 2 frames per second.

As the SRGAN algorithm is applied to higher magnification reconstructions, there is an increase in the LPIPS value and a decrease in PSNR/SSIM values. Notably, at an 8X magnification, these changes become more pronounced, although the algorithm's performance remains relatively unaffected by environmental factors such as rain and fog. At

this level, the SRGAN algorithm successfully maintains the texture features of the image, achieving PSNR/SSIM/LPIPS values of 21.96/0.736/0.402. However, at a 16X magnification, the algorithm's efficiency is limited to only two frames per second, resulting in a blurrier SR image with PSNR/SSIM/LPIPS values dropping to 20.48/0.606/0.486. This indicates that SRGAN faces challenges in producing high-quality SR images at magnifications beyond 8X, and its lower effectiveness in SR reconstruction poses difficulties for real-time image processing.

With the EDSR algorithm, as the magnification for super-resolution reconstruction increases, there is an observable rise in LPIPS values, alongside a decline in PSNR/SSIM metrics. At a 4X magnification, these values show significant changes, but external factors like rain and fog have minimal impact on the PSNR/SSIM/LPIPS readings. In this scenario, the PSNR/SSIM/LPIPS values reach 22.82/0.791/0.272. Comparing the super-resolution images produced by the SRGAN and EDSR algorithms, it is evident that SRGAN retains more textural details. However, at an 8X reconstruction rate, the EDSR algorithm yields PSNR/SSIM/LPIPS values of 22.36/0.783/0.476. This suggests that EDSR tends to produce blurrier images than SRGAN, especially at higher magnifications, with a reconstruction speed nearing three frames per second. Although EDSR shows improved efficiency, the image quality in high-magnification SR reconstructions of complex water surface environments is less effective compared to the results of the SRGAN method.

Table 3. Comparison of PSNR and SSIM values

Scale		Tested Algorithms			
		Bicubic	SRGAN	EDSR	Ours Method
2X	PSNR	19.013	23.03	0.823	21.291
	SSIM	0.628	0.759	0.795	0.828
	LPIPS	0.387	0.282	0.225	0.232
	Time(s)	0.001	1.052	0.552	0.442
4X	PSNR	18.951	23.01	22.82	23.087
	SSIM	0.626	0.756	0.791	0.821
	LPIPS	0.486	0.332	0.272	0.281
	Time(s)	0.002	1.512	1.421	0.601
8X	PSNR	18.932	21.96	22.36	22.012
	SSIM	0.615	0.736	0.783	0.791
	LPIPS	0.654	0.402	0.476	0.412
	Time(s)	0.003	2.068	2.296	0.896
16X	PSNR	18.360	20.48	20.92	20.361
	SSIM	0.531	0.606	0.661	0.628
	LPIPS	0.867	0.486	0.552	0.458
	Time(s)	0.005	3.026	1.963	1.132

In this study, the application of the proposed method for SR image reconstruction at magnification rates ranging from 2X to 16X results in a decrease in PSNR/SSIM values and an increase in LPIPS value. However, these differences are not statistically significant. Notably, environmental interferences such as rain and fog do not have a marked impact on the PSNR, SSIM, or LPIPS readings. The SR image generated using this method demonstrates better assessment indices, particularly in replicating images at a 16X magnification, compared to those produced using the SRGAN and EDSR algorithms. This indicates that the proposed method retains superior texture and quality even at high magnifications. Additionally, at a 4X magnification, the reconstruction speed of the proposed method is approximately 1 frame per second, aligning with the efficiency observed in the EDSR technique.

Table 3 presents a quantitative comparison of PSNR and SSIM values for the dataset, utilizing three distinct super-resolution techniques: our proposed algorithm, SRGAN, and EDSR.

4. CONCLUSION

This study introduces a novel super-resolution reconstruction method tailored for small and weak targets in complex aquatic environments. Utilizing datasets comprised of LR and HR images, captured by low- and high-resolution cameras respectively, the research focuses on model training about weak targets on the water surface. Central to this approach is the construction of a GAN-based SR model, incorporating implicit neural representation specifically designed for small and weak targets. The model encompasses a meticulously developed network and an adaptive loss function, facilitating the SR image reconstruction of high magnification targets on water in intricate environments.

Experimental results indicate that the proposed method adeptly addresses challenges related to blurring and the texture of high magnification SR image reconstruction of dim targets on water surfaces, even in the presence of complicating factors such as rain and fog. When compared to the SRGAN method, the proposed technique demonstrates a 62.59% improvement in SR image reconstruction efficiency at a 16X magnification rate. Moreover, there is a 3.63% increase in the SSIM values of the evaluation index, although a slight decrease of 0.58% in PSNR and 5.82% in LPIPS values is observed. In contrast with the EDSR method, the proposed approach shows a reduction in the PSNR, SSIM, and LPIPS values of the evaluation index by 2.67%, 4.99%, and 17.02% respectively, yet yields a 42.33% enhancement in SR image reconstruction efficiency.

The implications of this research are significant for marine navigation, particularly in enhancing collision detection capabilities. Given the challenges radar systems face in detecting small and weak targets at close range, the utilization of vision-based techniques for visualizing and extracting features of minute and weak objects on the ocean surface has emerged as a pivotal area of research. By achieving higher rates of synthetic radar image generation of such targets on water surfaces, the retrieval of small and weak target characteristics can be rendered more accurate, stable, and reliable. These advancements in visual recognition can substantially benefit various applications in sea navigation, marking a substantial stride in the field.

ACKNOWLEDGMENT

This research was funded by the National Natural Science Foundation of Guangdong Province, China, (Grant No.: 2021A1515010533); Science and Technology Program of Guangzhou, China, (Grant No.: 202201011735 and 202201011603); Characteristic innovation project of general universities in Guangdong province, China, (Grant No.: 2021KTSC097); Guangdong "Climbing" Program, (Grant No.: pdjh2022b0398 and pdjh2023b0397).

REFERENCES

- [1] Lu, H.T., Luo, M.K. (2022). Survey on new progresses

- of deep learning based computer vision. *Journal of Data Acquisition and Processing*, 2022(02): 247-278. <https://doi.org/10.16337/j.1004-9037.2022.02.001>
- [2] Wang, Z., Chen, J., Hoi, S.C. (2020). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3365-3387. <https://doi.org/10.1109/TPAMI.2020.2982166>
- [3] Soufi, O., Belouadha, F.Z. (2022). Study of deep learning-based models for single image super-resolution. *Revue d'Intelligence Artificielle*, 36(6): 939-952. <https://doi.org/10.18280/ria.360616>
- [4] Soufi, O., Belouadha, F.Z. (2023). FRSRI: New deep learning-based approach for super-resolution of multispectral satellite images. *Ingénierie des Systèmes d'Information*, 28(1): 113-132. <https://doi.org/10.18280/isi.280112>
- [5] Xu, S.H., Qi, M.M., Wang, X.M., Zhao, H.L., Hu, Z.Y., Sun, H.Y. (2022). A positive-unlabeled generative adversarial network for super-resolution image reconstruction using a Charbonnier loss. *Traitement du Signal*, 39(3): 1061-1069. <https://doi.org/10.18280/ts.390333>
- [6] Senalp, F.M., Ceylan, M. (2021). Deep learning based super resolution and classification applications for neonatal thermal images. *Traitement du Signal*, 38(5): 1361-1368. <https://doi.org/10.18280/ts.380511>
- [7] Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146: 305-319. <https://doi.org/10.1016/j.isprsjprs.2018.09.018>
- [8] Peng, C., Lin, W.A., Liao, H., Chellappa, R., Zhou, S.K. (2020). Saint: Spatially aware interpolation network for medical slice synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* pp. 7750-7759.
- [9] Lu, T., Wang, J., Zhang, Y., Wang, Z., Jiang, J. (2019). Satellite image super-resolution via multi-scale residual deep neural network. *Remote Sensing*, 11(13): 1588. <https://doi.org/10.3390/rs11131589>
- [10] Lee, Y.J., Yoon, J. (2010). Nonlinear image upsampling method based on radial basis function interpolation. *IEEE Transactions on Image Processing*, 19(10): 2682-2692. <https://doi.org/10.1109/TIP.2010.2050108>
- [11] Li, X., Orchard, M.T. (2001). New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10(10): 1521-1527. <https://doi.org/10.1109/83.951537>
- [12] Fattal, R. (2007). Image upsampling via imposed edge statistics. In *ACM SIGGRAPH 2007 papers*, San Diego California, pp. 95-es. <https://doi.org/10.1145/1275808.1276496>
- [13] Bishop, C. M., Blake, A., Marthi, B. (2003). Super-resolution enhancement of video. In *International Workshop on Artificial Intelligence and Statistics*, pp. 25-32
- [14] Tipping, M., Bishop, C. (2002). Bayesian image super-resolution. *Advances in Neural Information Processing Systems*, 15: 1303-1310.
- [15] Freeman, W.T., Jones, T.R., Pasztor, E.C. (2002). Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2): 56-65. <https://doi.org/10.1109/38.988747>
- [16] He, L., Qi, H., Zaretski, R. (2013). Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 345-352
- [17] Yang, J., Wright, J., Huang, T., Ma, Y. (2008). Image super-resolution as sparse representation of raw image patches. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, pp. 1-8. <https://doi.org/10.1109/CVPR.2008.4587647>
- [18] Su, H., Zhou, J., Zhang, Z.H. (2013). Survey of super-resolution image reconstruction methods. *Acta Automatica Sinica*, 39(8): 1202-1213.
- [19] Dong, C., Loy, C.C., He, K., Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295-307. <https://doi.org/10.1109/TPAMI.2015.2439281>
- [20] Dong, C., Loy, C.C., Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision—ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, pp. 391-407. https://doi.org/10.1007/978-3-319-46475-6_25
- [21] Kim, J., Lee, J.K., Lee, K.M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654.
- [22] Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 136-144.
- [23] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681-4690.
- [24] Anwar, S., Khan, S., Barnes, N. (2020). A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3): 1-34. <https://doi.org/10.1145/3390462>
- [25] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33: 7462-7473.
- [26] Janocha, K., Czarnecki, W.M. (2017). On loss functions for deep neural networks in classification. *arXiv preprint arXiv:1702.05659*. <https://doi.org/10.48550/arXiv.1702.05659>
- [27] Barron, J.T. (2019). A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Hubei, China, pp. 4331-4339.
- [28] Xu, G.Y., Chen, L.P., Gao, F. (2011). Study on binocular stereo camera calibration method. In *2011 International Conference on Image Analysis and Signal Processing*, pp. 133-137. <https://doi.org/10.1109/IASP.2011.6109013>
- [29] Zhang, Y.J. (2023). Camera calibration. In *3-D Computer Vision: Principles, Algorithms and Applications*, pp. 37-65. https://doi.org/10.1007/978-981-19-7580-6_2
- [30] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O. (2018). The unreasonable effectiveness of deep

- features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-595.
- [31] Han, D. (2013). Comparison of commonly used image interpolation methods. In Conference of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), pp. 1556-1559.
- <https://doi.org/10.2991/iccsee.2013.391>
- [32] Siu, W.C., Hung, K.W. (2012). Review of image interpolation and super-resolution. In Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, CA, USA, pp. 1-10.