Underwater Image Super-Resolution using Deep Residual Multipliers

Md Jahidul Islam[1], Sadman Sakib Enan[2], Peigen Luo[3], and Junaed Sattar[4]

Abstract— We present a deep residual network-based generative model for single image super-resolution (SISR) of underwater imagery for use by autonomous underwater robots. We also provide an adversarial training pipeline for learning SISR from paired data. In order to supervise the training, we formulate an objective function that evaluates the *perceptual quality* of an image based on its global content, color, and local style information. Additionally, we present *USR-248*, a large-scale dataset of three sets of underwater images of 'high' (640×480) and 'low' (80×60, 160×120, and 320×240) resolution. USR-248 contains paired instances for supervised training of $2\times$, $4\times$, or $8\times$ SISR models. Furthermore, we validate the effectiveness of our proposed model through qualitative and quantitative experiments and compare the results with several state-of-the-art models' performances. We also analyze its practical feasibility for applications such as scene understanding and attention modeling in noisy visual conditions.

I. INTRODUCTION

Visually-guided underwater robots require image synthesis and scene understanding in many important applications such as the monitoring of marine species and coral reefs [1], inspection of submarine cables and wreckage [2], human-robot collaboration [3], and more. Autonomous Underwater Vehicles (AUVs) and Remotely Operated Vehicles (ROVs) are widely used in these applications, where they harness the synthesized images for visual attention modeling to make navigation decisions such as 'where to look or go next', 'which snapshots should be recorded', etc. However, despite often using high-end cameras, underwater images are often greatly affected [4] by poor visibility, absorption, and scattering. Consequently, the objects of interest may appear blurred as the images lack important details. This problem exacerbates when the camera (*i.e.*, robot) cannot get close to the objects to get a closer view, *e.g.*, while following a fast-moving target, or surveying distant coral reefs or seabed. Fast and accurate techniques for Single Image Super-Resolution (SISR) can alleviate these problems by restoring the perceptual and statistical qualities of the low-resolution image patches.

The existing literature based on deep Convolutional Neural Networks (CNNs) provides good solutions for automatic SISR [5], [6]. In particular, several Generative Adversarial Network (GAN)-based models provide state-of-the-art (SOTA) performance [7], [8] in learning to enhance image resolution from a large collection of paired or unpaired data [9]. However, there are a few challenges involved in

The authors are with the Interactive Robotics and Vision Laboratory, Department of Computer Science and Engineering, Minnesota Robotics Institute, University of Minnesota- Twin Cities, US.
E-mail:{[1]islam034, [2]enan0001, [3]luo00034, [4]junaed}@umn.edu

(a) Zoom-in capability: HR image generation from LR image patches
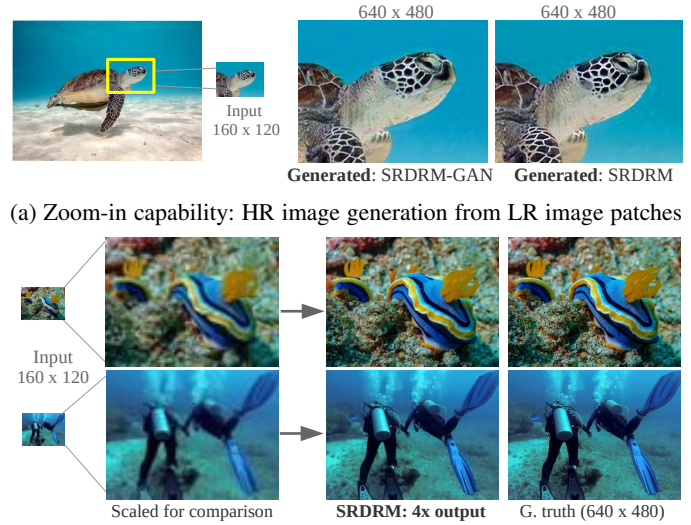
(b) Realistic HR image generation: comparison with the ground truths

Fig. 1: Demonstration of underwater image super-resolution using our proposed models: SRDRM and SRDRM-GAN.

adopting such models for underwater imagery. First, the underwater images suffer from a set of unique distortions. For instance, they tend to have a dominating green or blue hue because the red wavelengths get absorbed in deep water [10]. Other factors such as the lighting variations in different depths, amount of particles in the water, and scattering cause irregular non-linear distortions which result in low-contrast and blurry images [4]. Consequently, the off-the-shelf SISR models trained on arbitrary images fail to generate realistic higher resolution underwater images. Secondly, the lack of large-scale underwater dataset restricts extensive research attempts for the training and performance evaluation of SISR models on underwater images. Because of the high costs and difficulties associated with acquiring real-world underwater data, the existing datasets (that were originally proposed for training object detection and image enhancement models) often contain synthetic images [10] and/or their resolution are typically limited to 256×256 [4]. Due to these challenges, designing SISR models for underwater imagery and investigating their applicability in real-world underwater robotic applications have not been explored in-depth in the literature.

We attempt to address these challenges by designing a novel SISR model that can learn to generate $2\times$, $4\times$, or $8\times$ higher resolution (HR) underwater images from the respective low-resolution (LR) inputs. We also present a large-scale underwater dataset that provides the three sets of LR-HR pairs of images used to train the proposed model.

In addition, we perform thorough experimental evaluations of the proposed model and demonstrate its effectiveness compared to several existing SOTA models. Specifically, we make the following contributions in this paper:

(a) We present a fully-convolutional deep residual network-based generative model for underwater SISR, which we refer to as **SRDRM**. We also formulate an adversarial training pipeline (*i.e.*, **SRDM-GAN**) by designing a multi-modal objective function that evaluates the perceptual image quality based on its global content, color, and local style information. In our implementation, SRDRM and SRDM-GAN can learn to generate 640×480 images from respective inputs of size 320×240, 80×60, or 160×120. The model and associated training pipelines are available at https://github.com/xahidbuffon/srdrm.

(b) In addition, we present **USR-248**, a collection of over 1.5K samples (*i.e.*, paired HR-LR images) that facilitate large-scale SISR training. It has another 248 test images for benchmark evaluation. These images are rigorously collected during oceanic explorations and field experiments, and also from a few publicly available online resources. We make this available at http://irvlab.cs.umn.edu/resources/usr-248-dataset.

(c) Furthermore, we perform a number of qualitative and quantitative experiments that validate that the proposed model can learn to enhance underwater image resolution from both traditional and adversarial training. We also analyze its feasibility and effectiveness for improving visual perception in underwater robotic applications; a few sample demonstrations are highlighted in Fig. 1.

II. RELATED WORK

A. Single Image Super-resolution (SISR)

SISR has been studied [11], [12], [13] for nearly two decades in the area of signal processing and computer vision. Some of the classical SISR methods include statistical methods [14], [15], [16], patch-based methods [17], [18], [19], sparse representation-based methods [20], random forest-based method [21], etc. In recent years, with the rapid development of deep learning-based techniques, this area of research has been making incredible progress. In the pioneering work, Dong *et al.* [5] proposed a three-layer CNN-based end-to-end model named SRCNN, that can learn a non-linear LR-HR mapping without requiring any hand-crafted features. Soon after, Johnson *et al.* [22] showed that replacing the per-pixel loss with a perceptual loss (that quantifies image quality) gives better results for CNN-based SISR models. On the other hand, Kim *et al.* proposed deeper networks such as VDSR [23], DRCN [24] and used contemporary techniques such as gradient clipping, skip connection, and recursive-supervision in order to improve the training further. Moreover, the sparse coding-based networks [25], residual block-based networks (*e.g.*, EDSR [6], DRRN [26]), and other CNN-based models [27], [28] have been proposed that outperform SRCNN for SISR. These methods, however, have

rather complex training pipelines, and are often prone to poor performance for large scaling factors (*i.e.*, $4\times$ and higher). Thus far, researchers have been trying to address these issues by using Laplacian pyramid-based networks (LapSRN) [29], dense skip connections (SRDenseNet) [30], deep residual networks (RDN) [31], etc.

The CNN-based SISR models learn a sequence of non-linear filters from a large number of training images. This end-to-end learning of LR-HR mapping provide significantly better performance [32] compared to using hand-crafted filters, or traditional methods based on bicubic interpolation. On the other hand, Generative Adversarial Networks (GANs) [33] employ a two-player min-max game where the 'generator' tries to fool the 'discriminator' by generating *fake* images that appear to be *real* (*i.e.*, sampled from the HR distribution). Simultaneously, the discriminator tries to get better at discarding fake images and eventually (in equilibrium) the generator learns the underlying LR-HR mapping. GANs are known to provide SOTA performance for style transfer [34] and image-to-image translation [35] problems in general. As for SISR, the GAN-based models can recover finer texture details [36], [37] while super-resolving at large up-scaling factors. For instance, Ledig *et al.* showed that SRGAN [8] can reconstruct high-frequency details for an up-scaling factor of 4. Moreover, ESRGAN [7] incorporates a residual-in-residual dense block that improves the SISR performance. Furthermore, DeblurGAN [38] uses conditional GANs [39] that allow constraining the generator to learn a pixel-to-pixel mapping [35] within the LR-HR domain. Recently, inspired by the success of CycleGAN [40] and DualGAN [41], Yuan *et al.* [9] proposed a cycle-in-cycle GAN-based model that can be trained using unpaired data. However, such unpaired training of GAN-based SISR models are prone to instability and often produce inconsistent results.

B. SISR for Underwater Imagery

SISR techniques for underwater imagery, on the other hand, are significantly less studied. As mentioned in the previous section, this is mostly due to the lack of large-scale datasets that capture the distribution of the unique distortions prevalent in underwater imagery. The existing datasets are only suitable for underwater object detection [3] and image enhancement [4] tasks, as their image resolution is typically limited to 256×256, and they often contain synthetic images [10]. Consequently, the performance and applicability of existing and novel SISR models for underwater imagery have not been explored in depth.

Nevertheless, a few research attempts have been made for underwater SISR which primarily focus on reconstructing better quality underwater images from their noisy or blurred counterparts [42], [43], [44]. Other similar approaches have used SISR models to enhance underwater image sequence [45], and to improve fish recognition performance [46]. Although these models perform reasonably well for the respective applications, there is still significant room for improvement to match the SOTA performance. We attempt to address these aspects in this paper.
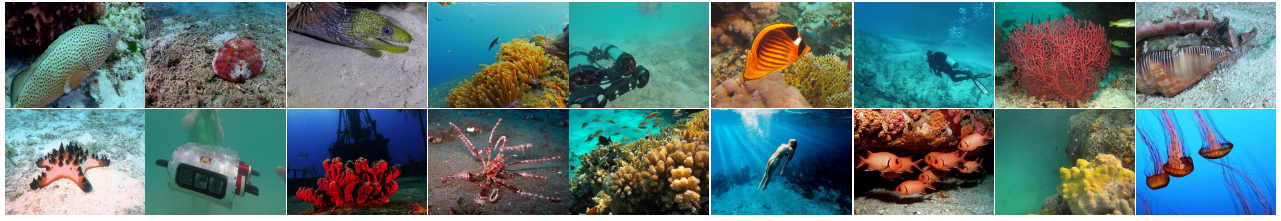
(a) A few instances sampled from the HR set; the HR images are of size 640×480.

HR image (640 x 480)   LR image (scaled for comparison)   LR images
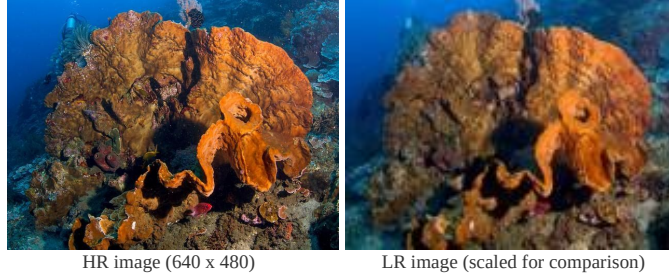
(b) A particular HR ground truth image and its corresponding LR images are shown.
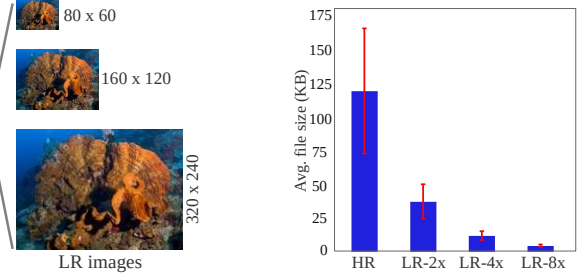
(c) Comparison of files sizes.

Fig. 2: The proposed USR-248 dataset has one HR set and three corresponding LR sets of images; hence, there are three possible combinations (*i.e.*, $2\times$, $4\times$ and $8\times$) for supervised training of SISR models.

III. USR-248 DATASET

The USR-248 dataset contains a large collection of HR underwater images and their respective LR pairs. As mentioned earlier, there are three sets of LR images of size 80×60, 160×120, and 320×240; whereas, the HR images are of size 640×480. Each set has 1060 RGB images for training and validation; another 248 test images are provided for benchmark evaluation. A few sample images from the dataset are provided in Fig. 2.

To prepare the dataset, we collected HR underwater images: *i)* during various oceanic explorations and field experiments, and *ii)* from publicly available Flickr[TM] images and YouTube[TM] videos. The field experiments are performed in a number of different locations over a diverse set of visibility conditions. Multiple GoPros [47], Aqua AUV's uEye cameras [48], low-light USB cameras [49], and Trident ROV's HD camera [50] are used to collect HR images during the experiments. We also compiled HR underwater images containing natural scenes from Flickr[TM], YouTube[TM], and other online resources[1]. We avoided multiple instances of similar scenes and made sure they contain different objects of interest (*e.g.*, coral reefs, fish, divers, wrecks/ruins, etc.) in a variety of backgrounds. Fig. 3 shows the modality in the data in terms of object categories. Once the HR images are selected and resized to 640×480, three sets of LR images are generated by compressing and then gradually downsizing the images to 320×240, 160×120, and 80×60; a comparison of the average file sizes for these image sets are shown in Fig. 2c. Overall, USR-248 provides large-scale paired data for training $2\times$, $4\times$, and $8\times$ underwater SISR models. It also includes the respective validation and test sets that are used to evaluate our proposed model.
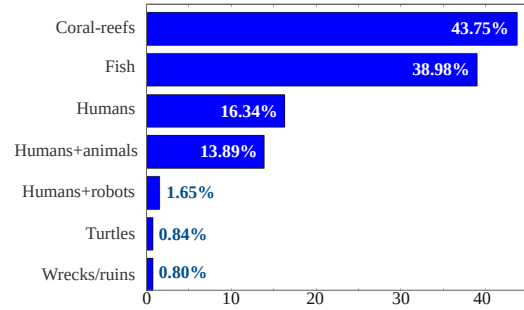
Fig. 3: Modality in the USR-248 dataset based on major objects of interest in the scene.

IV. SRDRM AND SRDRM-GAN MODEL

A. Deep Residual Multiplier (DRM)

The core element of the proposed model is a fully-convolutional deep residual block, designed to learn $2\times$ interpolation in the RGB image space. We denote this building block as Deep Residual Multiplier (DRM) as it scales the input features' spatial dimensions by a factor of two. As illustrated in Figure 4a, DRM consists of a convolutional (conv) layer, followed by 8 repeated residual layers, then another conv layer, and finally a de-convolutional (*i.e.*, deconv) layer for up-scaling. Each of the repeated residual layers (consisting of two conv layers) is designed by following the principles outlined in the EDSR model [6]. Several choices of hyper-parameters, *e.g.*, the number of filters in each layer, the use of ReLU non-linearity [51], and/or Batch Normalization (BN) [52] are annotated in Fig. 4a. As a whole, DRM is a 10 layer residual network that learns to scale up the spatial dimension of input features by a factor of two. It uses a series of 2D convolutions of size 3×3 (in repeated residual block) and 4×4 (in the rest of the network) to learn this spatial interpolation from paired training data.

(a) Architecture of a deep residual multiplier (DRM) block

(b) Generator: SRDRM model with multiple DRM blocks

(c) Discriminator: a Markovian PatchGAN [35] with nine layers and a patch-size of 40×30
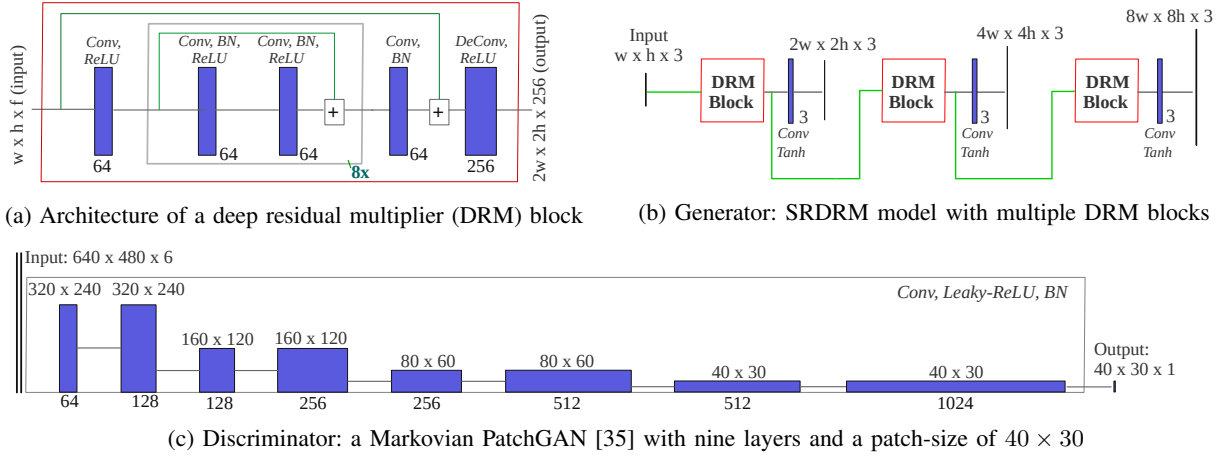
Fig. 4: Network architecture of the proposed model.

B. SRDRM Architecture

As Fig. 4b demonstrates, the SRDRM makes use of $n \in \{1, 2, 3\}$ DRM blocks in order to learn to generate $2^n \times$ HR outputs. An additional `conv` layer with `tanh` non-linearity [53] is added after the final DRM block in order to reshape the output features to the desired shape. Specifically, it generates a $2^n w \times 2^n h \times 3$ output for an input of size $w \times h \times 3$.

C. SRDRM-GAN Architecture

For adversarial training, we use the same SRDRM model as the *generator* and employ a Markovian PatchGAN [35]-based model for the *discriminator*. As illustrated by Fig. 4c, nine `conv` layers are used to transform a $640 \times 480 \times 6$ input (real and generated image) to a $40 \times 30 \times 1$ output that represents the averaged *validity* responses of the discriminator. At each layer, 3×3 convolutional filters are used with a stride size of 2, followed by a `Leaky-ReLU` non-linearity [54] and BN. Although traditionally PatchGANs use 70×70 patches [35], [41], we use a patch-size of 40×30 as our input/output image-shapes are of 4:3.

D. Objective Function Formulation

At first, we define the SISR problem as learning a function or mapping $G : \{X\} \to Y$, where X (Y) represents the LR (HR) image domain. Then, we formulate an objective function that evaluates the following properties of $G(X)$ compared to Y:

1) **Global similarity and perceptual loss**: existing methods have shown that adding an L_1 (L_2) loss to the objective function enables the generator to learn to sample from a globally similar space in an L_1 (L_2) sense [35]. In our implementation, we measure the *global similarity* loss as: $\mathcal{L}_2(G) = \mathbb{E}_{X,Y}\left[\left|\left|Y - G(X)\right|\right|_2\right]$. Additionally, as suggested in [55], we define a *perceptual loss* function based on the per-channel disparity between $G(X)$ and Y as:

$$\mathcal{L}_P(G) = \mathbb{E}_{X,Y}\left[\left|\left|(512 + \bar{\mathbf{r}})\mathbf{r}^2 + 4\mathbf{g}^2 + (767 - \bar{\mathbf{r}})\mathbf{b}^2\right|\right|_2\right].$$

Here, $\mathbf{r}$, $\mathbf{g}$, and $\mathbf{b}$ denote the normalized numeric differences of the red, green, and blue channels between $G(X)$ and Y, respectively; whereas $\bar{\mathbf{r}}$ is the mean of their red channels.

2) **Image content loss**: being inspired by the success of existing SISR models [32], we also formulate the *content loss* as:

$$\mathcal{L}_C(G) = \mathbb{E}_{X,Y}\left[\left|\left|\Phi(Y) - \Phi(G(X))\right|\right|_2\right].$$

Here, the function $\Phi(\cdot)$ denotes the high-level features extracted by the `block5_conv4` layer of a pre-trained `VGG-19` network.

Finally, we formulate the multi-modal objective function for the generator as: $\mathcal{L}_\mathcal{G}(G) = \lambda_c\mathcal{L}_C(G) + \lambda_p\mathcal{L}_P(G) + \lambda_2\mathcal{L}_2(G)$. Here, λ_c, λ_p, and λ_2 are scalars that are empirically tuned as hyper-parameters. Therefore, the generator G needs to solve the following minimization problem:

$$G^* = \arg\min_G \mathcal{L}_\mathcal{G}(G). \tag{1}$$

On the other hand, adversarial training requires a two-player min-max game [33] between the generator G and discriminator D, which is expressed as:

$$\mathcal{L}(G, D) = \mathbb{E}_{X,Y}\left[\log D(Y)\right] + \mathbb{E}_{X,Y}\left[\log(1 - D(X, G(X)))\right]. \tag{2}$$

Here, the generator tries to minimize $\mathcal{L}(G, D)$ while the discriminator tries to maximize it. Therefore, the optimization problem for adversarial training becomes:

$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_\mathcal{G}(G). \tag{3}$$

E. Implementation

We use TensorFlow libraries [56] to implement the proposed SRDRM and SRDRM-GAN models. We trained both the models on the USR-248 dataset up to 20 epochs with a batch-size of 4, using two NVIDIA™ GeForce GTX 1080 graphics cards. We also implement a number of SOTA generative and adversarial models for performance comparison in the same setup. Specifically, we consider three generative models named SRCNN [5], SRResNet [8], [32], and DSRCNN [57], and three adversarial models named SRGAN [8], ESRGAN [7], and EDSRGAN [6]. We already provided a brief discussion on the SOTA SISR models in Section II. Next, we present the experimental results based on qualitative analysis and quantitative evaluations in terms of standard metrics.

V. EXPERIMENTAL RESULTS

A. Qualitative Evaluations

At first, we analyze the sharpness and color consistency in the generated images of SRDRM and SRDRM-GAN. As Fig. 5 suggests, both models generate images that are comparable to the ground truths for $4\times$ SISR. We observe even better results for $2\times$ SISR, as it is a relatively less challenging problem. We demonstrate this relative performance margins at various scales in Fig. 6. This comparison shows that the global contrast and texture is mostly recovered in the $2\times$ and $4\times$ HR images generated by SRDRM and SRDRM-GAN. On the other hand, the $8\times$ HR images miss the finer details and lack the sharpness in high-texture regions. The state-of-the-art SISR models have also reported such difficulties beyond the $4\times$ scale [32].

Next, in Fig. 7, we provide a qualitative performance comparison with the state-of-the-art models for $4\times$ SISR. We select multiple 160×120 patches on the test images containing interesting textures and objects in contrasting background. Then, we apply all the SISR models (trained on $4\times$ USR-248 data) to generate respective HR images of size 640×480. In the evaluation, we observe that SRDRM performs at least as well as and often better compared to the generative models, *i.e.*, SRResNet, SRCNN, and DSRCNN. Moreover, SRResNet and SRGAN are prone to inconsistent coloring and over-saturation in bright regions. On the other hand, ESRGAN and EDSRGAN often fail to restore the sharpness and global contrast. Furthermore, SRDRM-GAN generates sharper images and does a better texture recovery than SRDRM (and other generative models) in general. We postulate that the PatchGAN-based discriminator contributes to this, as it forces the generator to learn high-frequency local texture and style information [35].

B. Quantitative Evaluation

We consider two standard metrics [58], [4] named Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) in order to quantitatively compare the SISR models' performances. The PSNR approximates the reconstruction quality of generated images compared to their respective ground truth, whereas the SSIM [59] compares the image patches based on three properties: luminance, contrast, and
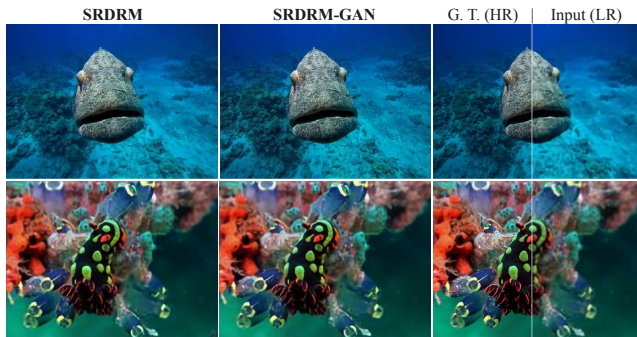
Fig. 5: Color consistency and sharpness of the generated $4\times$ HR images compared to the respective ground truth.

TABLE I: Comparison of average PSNR, SSIM, and UIQM scores for $2\times/4\times/8\times$ SISR on USR-248 test set.

| Model	$PSNR$ $(G(\mathbf{x}), \mathbf{y})$	$SSIM$ $(G(\mathbf{x}), \mathbf{y})$	$UIQM$ $(G(\mathbf{x}))$
SRResNet	25.98/24.15/19.26	0.72/0.66/0.55	2.68/2.23/1.95
SRCNN	26.81/23.38/19.97	0.76/0.67/0.57	2.74/2.38/2.01
DSRCNN	27.14/23.61/20.14	0.77/0.67/0.56	2.71/2.36/2.04
SRDRM	28.36/24.64/21.20	0.80/0.68/0.60	2.78/2.46/2.18
SRDRM-GAN	28.55/24.62/20.25	0.81/0.69/0.61	2.77/2.48/2.17
ESRGAN	26.66/23.79/19.75	0.75/0.66/0.58	2.70/2.38/2.05
EDSRGAN	27.12/21.65/19.87	0.77/0.65/0.58	2.67/2.40/2.12
SRGAN	28.05/24.76/20.14	0.78/0.69/0.60	2.74/2.42/2.10
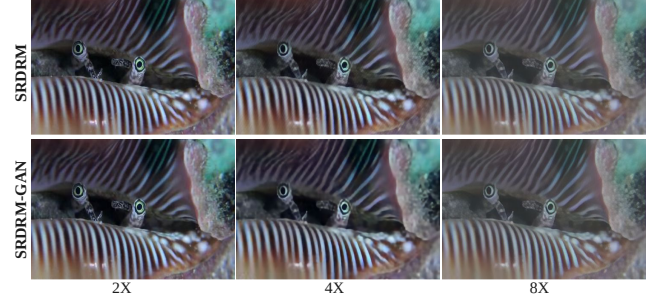
Fig. 6: Global contrast and texture recovery by SRDRM and SRDRM-GAN for $2\times$, $4\times$, and $8\times$ SISR.

structure. In addition, we consider Underwater Image Quality Measure (UIQM) [60], which quantifies underwater image colorfulness, sharpness, and contrast. We evaluate all the SISR models on USR-248 test images, and compare their performance in Table I. The results indicate that SRDRM-GAN, SRDRM, SRGAN, and SRResNet produce comparable values for PSNR and SSIM, and perform better than other models. SRDRM and SRDRM-GAN also produce higher UIQM scores than other models in comparison. These statistics are consistent with our qualitative analysis.

C. Practical Feasibility

The qualitative and quantitative results suggest that SRDRM and SRDRM-GAN provide good quality HR visualizations for LR image patches, which is potentially useful in tracking fast-moving targets, attention modeling, and detailed understanding of underwater scenes. Therefore, AUVs and ROVs can use this to *zoom in* a particular region of interest (RoI) for detailed and improved visual perception. One operational consideration for using such deep learning-based models in embedded robotic platforms is the computational complexity. As we demonstrate in Table II, the memory requirement for the proposed model is only 3.5-12 MB and it runs at 4-7 fps on NVIDIA$^{\text{TM}}$ Jetson TX2. Therefore, it essentially takes about 140-246 milliseconds for a robot to take a closer look at a LR RoI. These results validate the feasibility of using the proposed model for improving real-time perception of visually-guided underwater robots.

VI. CONCLUSION

In this paper, we present a fully-convolutional deep residual network-based model for underwater image super-resolution at $2\times$, $4\times$, and $8\times$ scales. We also provide generative and adversarial training pipelines driven by a
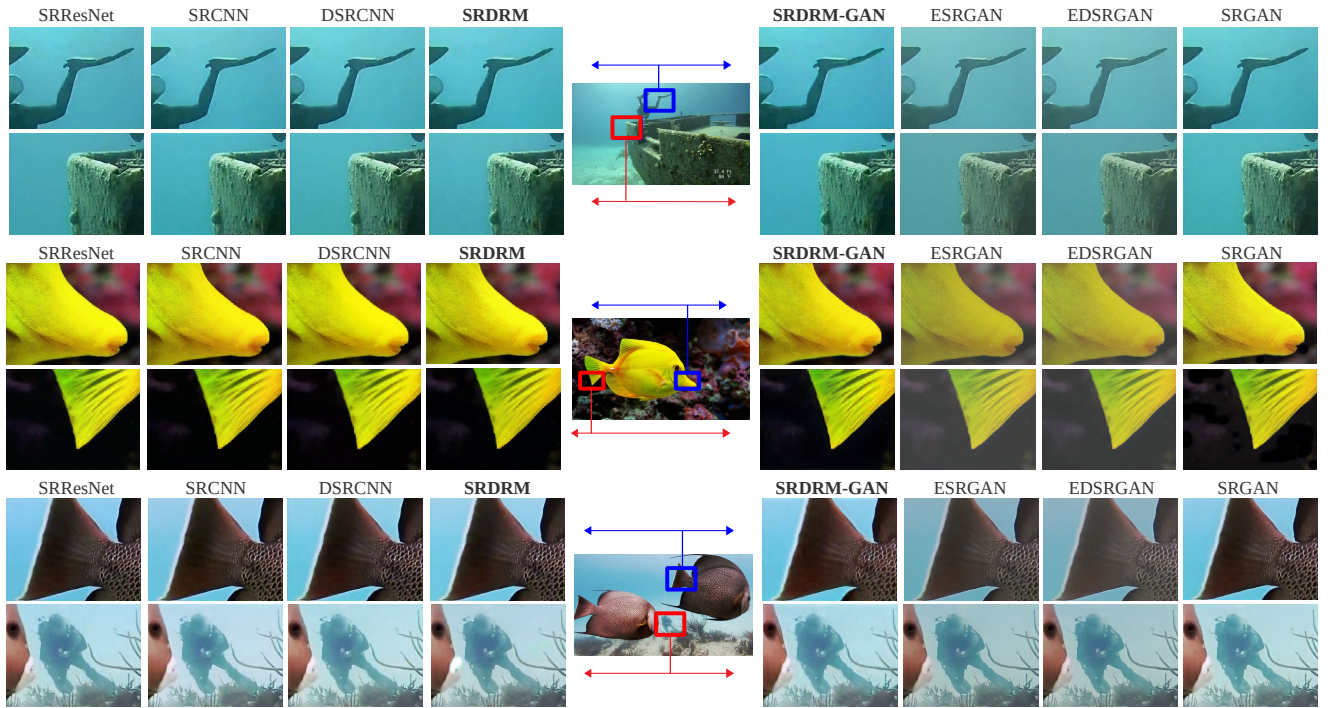
Fig. 7: Qualitative performance comparison of SRDRM and SRDRM-GAN with SRCNN [5], SRResNet [8], [32], DSRCNN [57], SRGAN [8], ESRGAN [7], and EDSRGAN [6]. (Best viewed at 400% zoom)

TABLE II: Run-time and memory requirement of SRDRM (same as SRDRM-GAN) on NVIDIA™ Jetson TX2 (optimized graph).

Model	$2\times$	$4\times$	$8\times$
Inference-time (ms)	140.6 ms	145.7 ms	245.7 ms
Frames per second (fps)	7.11 fps	6.86 fps	4.07 fps
Model-size	3.5 MB	8 MB	12 MB

multi-modal objective function, which is designed to evaluate image quality based on its content, color, and texture information. In addition, we present a large-scale dataset named USR-248 which contains paired underwater images of various resolutions for supervised training of SISR models. Furthermore, we perform thorough qualitative and quantitative evaluations which suggest that the proposed model can learn to restore image qualities at a higher resolution for an improved visual perception. In the future, we seek to improve its performance for $8\times$ SISR, and plan to further investigate its applicability in other underwater robotic applications.

REFERENCES

[1] O. Hoegh-Guldberg, P. J. Mumby, A. J. Hooten, R. S. Steneck, P. Greenfield, E. Gomez *et al.*, "Coral Reefs under Rapid Climate Change and Ocean Acidification," *Science*, vol. 318, no. 5857, pp. 1737–1742, 2007.

[2] B. Bingham, B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice *et al.*, "Robotic Tools for Deep Water Archaeology: Surveying an Ancient Shipwreck with an Autonomous Underwater Vehicle," *Journal of Field Robotics (JFR)*, vol. 27, no. 6, pp. 702–717, 2010.

[3] M. J. Islam, M. Ho, and J. Sattar, "Understanding Human Motion and Gestures for Underwater Human-Robot Collaboration," *Journal of Field Robotics (JFR)*, pp. 1–23, 2018.

[4] M. J. Islam, Y. Xia, and J. Sattar, "Fast Underwater Image Enhancement for Improved Visual Perception," *arXiv preprint arXiv:1903.09766*, 2019.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-resolution using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[6] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-resolution," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2017, pp. 136–144.

[7] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced Super-resolution Generative Adversarial Networks," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic Single Image Super-resolution using a Generative Adversarial Network," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690.

[9] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised Image Super-resolution using Cycle-in-cycle Generative Adversarial Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 701–710.

[10] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing Underwater Imagery using Generative Adversarial Networks," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7159–7165.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based Super-resolution," *IEEE Computer Graphics and Applications*, no. 2, pp. 56–65, 2002.

[12] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution Through Neighbor Embedding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2004, pp. I–I.

[13] D. O. Melville and R. J. Blaikie, "Super-resolution Imaging through a Planar Silver Layer," *Optics Express*, vol. 13, no. 6, pp. 2127–2134, 2005.

[14] J. Sun, Z. Xu, and H.-Y. Shum, "Image Super-resolution using Gradient Profile Prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[15] K. I. Kim and Y. Kwon, "Single-image Super-resolution using Sparse Regression and Natural Image Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.

[16] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the Nonlocal-means to Super-resolution Reconstruction," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 36–51, 2008.

[17] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a Sin-

gle Image," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2009, pp. 349–356.

[18] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled Dictionary Training for Image Super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, 3467–3478, 2012.

[19] J.-B. Huang, A. Singh, and N. Ahuja, "Single Image Super-resolution from Transformed Self-exemplars," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5197–5206.

[20] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image Super-resolution via Sparse Representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[21] S. Schulter, C. Leistner, and H. Bischof, "Fast and Accurate Image Upscaling with Super-resolution Forests," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3791–3799.

[22] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-time Style Transfer and Super-resolution," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 694–711.

[23] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate Image Super-resolution using Very Deep Convolutional Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.

[24] ——, "Deeply-recursive Convolutional Network for Image Super-resolution," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1637–1645.

[25] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust Single Image Super-resolution via Deep Networks with Sparse Prior," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3194–3207, 2016.

[26] Y. Tai, J. Yang, and X. Liu, "Image Super-resolution via Deep Recursive Residual Network," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3147–3155.

[27] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time Single Image and Video Super-resolution using an Efficient Sub-pixel Convolutional Neural Network," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.

[28] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-resolution Convolutional Neural Network," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 391–407.

[29] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-resolution," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 624–632.

[30] T. Tong, G. Li, X. Liu, and Q. Gao, "Image Super-resolution using Dense Skip Connections," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4799–4807.

[31] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual Dense Network for Image Super-resolution," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2472–2481.

[32] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for Single Image Super-resolution: A Brief Review," *IEEE Transactions on Multimedia*, 2019.

[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 2672–2680.

[34] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer using Convolutional Neural Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414–2423.

[35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image Translation with Conditional Adversarial Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1125–1134.

[36] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár, "Amortised Map Inference for Image Super-resolution," *arXiv preprint arXiv:1610.04490*, 2016.

[37] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and Accurate MRI Super-resolution using a Generative Adversarial Network and 3D Multi-level Densely Connected Net-

work," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 91–99.

[38] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind Motion Deblurring using Conditional Adversarial Networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8183–8192.

[39] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv preprint arXiv:1411.1784*, 2014.

[40] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-image Translation using Cycle-consistent Adversarial Networks," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2223–2232.

[41] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-image Translation," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2849–2857.

[42] Y. Chen, B. Yang, M. Xia, W. Li, K. Yang, and X. Zhang, "Model-based Super-resolution Reconstruction Techniques for Underwater Imaging," in *Photonics and Optoelectronics Meetings (POEM): Opto-electronic Sensing and Imaging*, vol. 8332. International Society for Optics and Photonics, 2012, p. 83320G.

[43] F. Fan, K. Yang, B. Fu, M. Xia, and W. Zhang, "Application of Blind Deconvolution Approach with Image Quality Metric in Underwater Image Restoration," in *International Conference on Image Analysis and Signal Processing*. IEEE, 2010, pp. 236–239.

[44] Y. Yu and F. Liu, "System of Remote-operated-vehicle-based Underwater Blurred Image Restoration," *Optical Engineering*, vol. 46, no. 11, p. 116002, 2007.

[45] E. Quevedo, E. Delory, G. Callicó, F. Tobajas, and R. Sarmiento, "Underwater Video Enhancement using Multi-camera Super-resolution," *Optics Communications*, vol. 404, pp. 94–102, 2017.

[46] X. Sun, J. Shi, J. Dong, and X. Wang, "Fish Recognition from Low-resolution Underwater Images," in *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2016, pp. 471–476.

[47] GoPro, "GoPro Hero 5," https://gopro.com/, 2016, accessed: 8-15-2019.

[48] G. Dudek, P. Giguere, C. Prahacs, S. Saunderson, J. Sattar, L.-A. Torres-Mendez, Jenkin *et al.*, "Aqua: An Amphibious Autonomous Robot," *Computer*, vol. 40, no. 1, pp. 46–53, 2007.

[49] BlueRobotics, "Low-light HD USB Camera," https://www.bluerobotics.com/, 2016, accessed: 3-15-2019.

[50] OpenROV, "TRIDENT," https://www.openrov.com/, 2017, accessed: 8-15-2019.

[51] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proc. of the International Conference on Machine Learning (ICML)*, 2010, pp. 807–814.

[52] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *CoRR, abs/1502.03167*, 2015.

[53] T. Raiko, H. Valpola, and Y. LeCun, "Deep Learning Made Easier by Linear Transformations in Perceptrons," in *Artificial Intelligence and Statistics*, 2012, pp. 924–932.

[54] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," in *International Conference on Machine Learning (ICML)*, vol. 30, no. 1, 2013, p. 3.

[55] CompuPhase, "Perceptual Color Metric," https://www.compuphase.com/cmetric.htm, 2019, accessed: 12-12-2019.

[56] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean *et al.*, "TensorFlow: A System for Large-scale Machine Learning," in *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.

[57] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image Restoration using Convolutional Auto-encoders with Symmetric Skip Connections," *arXiv preprint arXiv:1606.08921*, 2016.

[58] A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," in *International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[59] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[60] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired Underwater Image Quality Measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016.