Single Image Super Resolution for Underwater Images with ESRGAN and EDSR
Advanced Machine Learning
A.Y. 2020/2021

Nana Teukam Yves Gaetan
Matricola: 1741352

Simone Marretta
Matricola: 1911358

Daniele Trappolini
Matricola: 1710415

Luca Scofano
Matricola: 1762509

ABSTRACT

Image super-resolution (SR) problem, particularly single image super-resolution (SISR), has gained increasing research attention for decades. In fact some sciences like oceanography relies on the quality of the images to have statistically significant results. Our project is based on Single Image Super Resolution with and additional caveat, underwater images. The first part of our work is based on preprocessing the images with different techniques in order to reduce the noise produced by the water and the lack of brightness. In the second part of the project we used GANs and CNNs in order to obtain good result for different datasets. The first model we chose was EDSR and the best performance was when using the ULAP preprocessing-method, on the other hand the newer version of the SRGAN, the ESRGAN performed best with the original dataset.

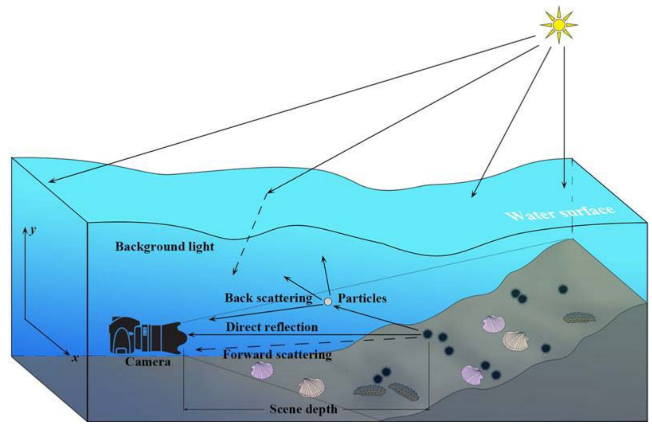


Figure 1: Schematic diagram of underwater optical imaging

INTRODUCTION

Oceanography is the study to analyse and interpret physical, biological, chemical, archaeological and geological data collected from the sea. Oceanographers use imaging technology for their study. However, there are limitations still prevailing in underwater surrounding while collecting and processing the underwater image. Naturally, underwater images are degraded by the adverse effects of light absorption and scattering due to particles in the water including micro phytoplankton colored dissolved organic matter and non-algal particles [1]. When the light propagates in an underwater scenario, the light received by a camera is mainly composed of three types of light: direct light, forward scattering light, and back scattering light. The direct light suffers from attenuation resulting in information loss of underwater images Figure 1. The forward scattering light has a negligible contribution to the blurring of the image features. The back scattering light reduces the contrast of underwater images and suppresses fine details and patterns. Additionally, the red light first disappears, followed by the green and blue lights (the wavelengths of the red, green and blue lights are 600nm, 525nm, and 475nm, respectively). As a result, most underwater images are dominated by a bluish or greenish tone.

Due to the complexity of the underwater imaging environment, the underwater image distortion is severe, and it is difficult to obtain clear and high quality images [2]. In order to solve this problem, high resolution images can be obtained by hardware and software. But as we all know, the hardware is relatively expensive and difficult to implement, so the super resolution (SR) technology of underwater images is a necessary job.

Notable progress has been made to improve the visual quality of underwater images in recent years. Existing underwater images sharpness methods can be classified into one of two broad categories: image enhancement methods and image restoration methods.

In this paper, we show two models that are well suitable for underwater single image super resolution: ESRGAN and EDSR & SRGAN. We also show that different image preprocessing methods can be used to increase the initial quality of the images in order to have better performances. The preprocessing methods that we have explored are based on image enhancement (CLAHE and RGHS) and image restoration (ULAP and DCP).

DATA PREPROCESSING

Dataset. Our dataset used is USR-248 [CITA SITO WEB], 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×, 4×, or 8× single image super resolution models. From this dataset, two set of datasets have been created - during the preprocessing step, for a total of 3 datasets for training and testing. For preprocessing, we started from the hr images then applied image enhancement and image resolution algorithms.

Underwater Image Enhancement (UIE) aim to produce a high-quality image that human favor from a single degraded input. These algorithms either increase the visibility or alleviate color casts by combating the light scattering and other ambient circumstance factors during capturing underwater scenes. According to the means of modeling imaging process [3]. Generally, UIE algorithms can be categorized into three types including model-free, model-based, and learning-based convolutional neural networks (CNNs).

- **Contrast Limited Adaptive Histogram Equalization (CLAHE)** is a variant of Adaptive histogram equalization (AHE) which takes care of over-amplification of the contrast. CLAHE operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' value. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image [4].
- **Relative Global Histogram Stretching (RGHS)** consists of two parts: contrast correction and color correction. The contrast correction in RGB color space firstly equalizes G and B channels and then re-distributes each R-G-B channel histogram with dynamic parameters that relate to the intensity distribution of original image and wavelength attenuation of different colors under the water. The color correction is performed by stretching the 'L' component and modifying 'a' and 'b' components in CIE-Lab color space [5].

Image Restoration Usually establishes an effective degradation model by analyzing the underwater imaging mechanism and the basic physics of light propagation, then deduces the key parameters of the constructed physical model via some prior knowledge, and finally recovers the restored image by reserving compensation processing [6].

- **Dark Channel Priority (DCP)** is widely used for image dehazing. Due to the similarities between a hazed outdoor image and an underwater image, the DCP-based dehazing method is widely applied to underwater image enhancement. The dark channel prior was based on the observation that clear day images contain some pixels which have very low intensities (close to zero) in at least one color channel. The traditional DCP was used to estimate the transmission maps (TM) of R channel, and then the TMs of GB channels were derived by considering the exponential relationship with the attenuation coefficient. While a recent way of performing DCP is by picking up the top 0.1% brightest pixels in dark channel and then selected the average value of the corresponding intensities in the input image as final background light.
- **Underwater Light Attenuation Prior (ULAP)** assumes the difference between the maximum value of G-B intensity and the value of R intensity in one pixel of the underwater image strongly related to the change of the scene depth. Based on the ULAP, a linear model was established to rapidly obtain scene depth map, which can be used to estimate the BL and transmission maps (TMs) for R-G-B channels are easily estimated to recover the true scene radiance under the water. which use instance data to extract some valuable feature vectors [7].

The lr images have been obtained through down-sampling hr images. The original lr have been created by down-sampling hr images using BiCubic interpolation with 20% Gaussian noise. Instead, we simply did the down-sampling and BiCubic interpolation without the Gaussian noise [DA CANCELLARE???]

We have decided to combine image enhancement and color restoration. Therefore, we used the combinations ULAP-RGHS and CLAHE-DCP. These combinations have been chosen to compaire recent algothims (ULAP-RGHS, 2018-2018) and old ones (CLAHE-DCP, 1994-2011). We also observed that these combinations gave the best result for image enhancement (INSERIRE FOTO)

Generally, DCP-CLAHE are used together, because if you compare the peak signal-to-noise ratio (PSNR) of images coming from DCP and CLAHE distinctly, you will find that the PSNR value is better when they are combined. [LOOK AT TABLE]. This happens because the combination of these two methods, resolves the problems they encounter individually.

There are major defects occurred in DCP. In any kind of image, object and the background. After DCP, where there is a removal of haze, there is very low contrast with the background and sometimes this may also darken the local regions of the image. This is because background may also be mixed with the haze. This problem over come by performing the CLAHE algorithm. While, one defect of CLAHE is that it does not perform dehazing.

Le miglior combinazioni sono ULAP-RGHS and CLAHE-DCP. INSERIRE FOTO PER CONFERMARE QUESTA TESI

MODELS

EDSR+SRGAN and ESRGAN. We picked two state of the art models in order to compare which one performed better. The key difference between the two is the model used as initialization for the generator, one is the EDSR and the other is the RRDBnet. The ideas behind this type of task is that if the downgrade function is unknown, supervised model training requires existing LR and HR image pairs to be available. Alternatively, unsupervised learning methods can be used to learn to approximate the downgrade function from unpaired LR and HR images as in our case.

EDSR and SRGAN

Residual Design. Super-resolution models therefore mainly learn the residuals between LR and HR images. Residual network designs are therefore of high importance: identity information is conveyed via skip connections whereas reconstruction of high frequency content is done on the main path of the network.

Local skip connections in residual blocks make the network easier to optimize and therefore support the construction of deeper networks.

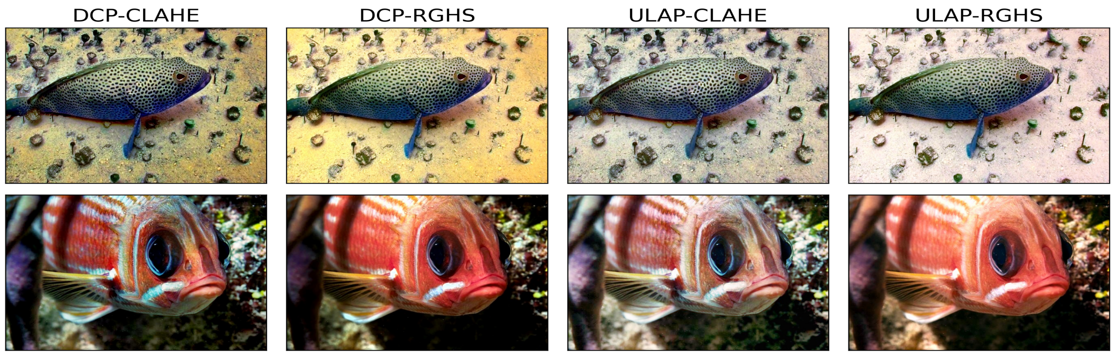


Figure 2: sample images of two fishes. Enhancement algorithm are combined with restoration images to improve the original images.

Upsampling. The upsampling layer used in this article is a subpixel convolution layer (Fig. 3).

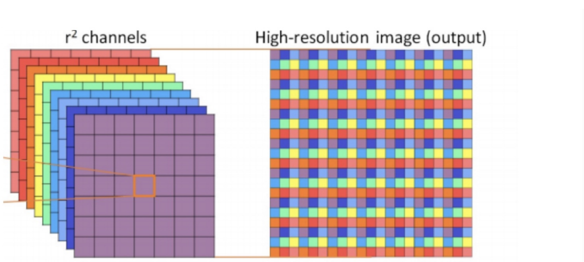


Figure 3: Subpixel convolution

This layer essentially uses regular convolutional layers followed by a specific type of image reshaping called a phase shift. In other words, instead of putting zeros in between pixels and having to do extra computation, they calculate more convolutions in lower resolution and resize the resulting map into an upscaled image.

Network

Network. It is a winner of the NTIRE 2017 super-resolution challenge. Here's an overview of the EDSR architecture:

Its residual block design differs from that of ResNet. Batch normalization layers have been removed together with the final ReLU activation.

As the authors of the paper say, batch normalization loses scale information of images and reduces the range flexibility of activations. Removal of batch normalization layers not only increases super-resolution performance but also reduces GPU memory up to 40 percent. (Fig. 4)

Data

Data. For data augmentation, random crops, flips and rotations are made to get a large number of different training images. Random crops are used because we want filters to process small pieces of the image in order to detect features (edges, etc). This also has a nice regularization property, since we're estimating a smaller number of parameters.

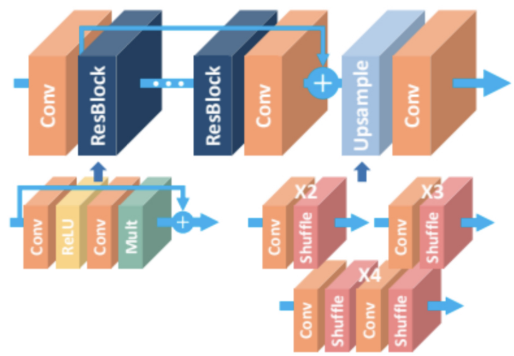


Figure 4: EDSR Architecture

Pixel Loss

Pixel Loss. The pixel-wise L2 loss and the pixel-wise L1 loss are frequently used loss functions for training super-resolution models.

The pixel-wise L2 loss directly optimizes PSNR, an evaluation metric often used in super-resolution competitions. Experiments have shown that the pixel-wise L1 loss can sometimes achieve even better performance and is therefore used for EDSR.

Generated SR images often lack high-frequency content, realistic textures and are perceived blurry. This problem is addressed with perceptual loss functions by using the SRGAN.

Perceptual Loss

Perceptual Loss. The authors use a perceptual loss function composed of a content loss and an adversarial loss.

They also train their super-resolution model as generator G in a generative adversarial network (GAN). The GAN descriminator D is optimized for descriminating SR from HR images whereas the generator is optimized for generating more realistic SR images in order to fool the discriminator.

Instead of training the super-resolution model the authors use the trained EDSR model.

The following example fine-tunes an EDSR baseline model that was pre-trained with a pixel-wise L1 loss.

Training

Training. We performed some parameter-tuning as well for both parts of the model.

EDSR

For the EDSR we tried with different combinations of *batch sizes, patch size, normalization, number of iterations and learning rate.*

Batch Size. First off the initial *batch size* was **16**,smaller batch sizes have been empirically shown to have faster convergence to “good” solutions. The downside of using a smaller batch size is that the model is not guaranteed to converge to the global optima. It will bounce around the global optima. this is why we used 32.

ULAP	16 Batch Size	32 Batch Size
Highest value	132.5	201.7
Lowest value	63.8	10.6

ORIGINAL	16 Batch Size	32 Batch Size
Highest value	50.6	62.8
Lowest value	63.8	60.1

CLAHE	16 Batch Size	32 Batch Size
Highest value	55.3	60.8
Lowest value	50.3	49.8

Number of iteration

After several trials we used **60.000** steps.

Patch size

So a patch is an area of a single image, like a convolutional kernel, but it doesn’t convolve.

They talk about adaptive patches, meaning that you need to select a pixel, then adapt the patch size used in order to include enough surrounding information to reproduce a homogenous patch as the output.

CNN kernels/filters only process one patch at a time, rather than the whole image. This is because we want filters to process small pieces of the image in order to detect features (edges, etc). This also has a nice regularization property, since we’re estimating a smaller number of parameters, and those parameters have to be "good" across many regions of each image, as well as many regions of all other training images.

The values we used are **24, 124, 192**. The best result we had was produced using the intermediate patch size **124**.

Learning Rate

When training deep neural networks, it is often useful to reduce learning rate as the training progresses. This can be done by using pre-defined learning rate schedules or adaptive learning rate methods.

We used a adaptive learning rate method, more in particular a step decay approach.

We set some boundaries at the *30.000* step where the learning rate goes from **0.0001 to 1e-4** and at step *45.000* we go from **1e-4 to 5e-5**.

(a) LR

(b) HR

(c) Pixel Loss

(d) Perceptual Loss

Figure 5: Results of EDSR SRGAN on ULAP dataset

ESRGAN

Components. The ESRGAN model is an improved version of the SRGAN. It won the first place in PIRM2018-SR challenge. This model has three key components. First,Residual-in-residual Dense Block(RRDB) without batch normalization is introduced. Moreover,the discriminator predicts relative realness instead of the absolute value. Finally,the perceptual loss is used by using the features before activation.

Network

Batch normalization. Batch normalization layers are removed from the model. This increased the performance of the model and reduced computational complexity of it,because batch normalization in SR tasks tends to introduce artifacts and limit the generalization ability of the model.

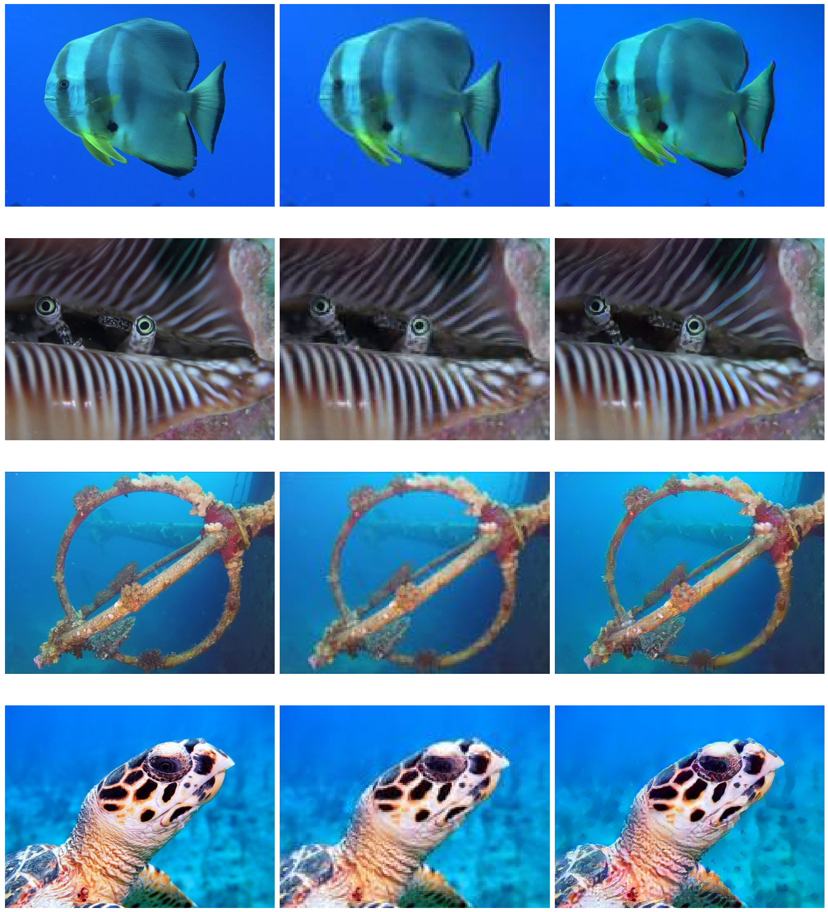
Dense block. The dense block,taken by the dense convolutional neural network(DenseNet),replace the original residual block of the SRGAN.

Perceptual Loss

SRGAN problems. In ESRGAN the features before the activation layer are used. In SRGAN we know that it was the opposite. Using the features after the activation cause two problems. First,the features are very sparse,especially in the case of a really deep neural network. Second,it can be shown that using activated features cause inconsistent reconstructed brightness compared to the ground-truth image.

Training

Training. The mini-batch size is set to 16. The spatial size of cropped HR patch is 128 × 128. For optimization, we use Adam with Beta 1 = 0.9, Beta2 = 0.999. We used Google Colab Pro to run our



(a) HR **(b) LR** **(c) ESRGAN**

Figure 6: Results of ESRGAN on original dataset

models and we used the official implementation of the model on the BasicSR repository.

Conclusion and Discussion

In this paper we explore two different methods for underwater single image super resolution - ESRGAN and EDSR & SRGAN. Before running the different models, the images have been preprocessed using ULAP-RGHS and DCP-CLAHE. These preprocessing methods, are both the result of the combination of image enhancement and image restoration algorithms. Using ESRGAN we had better performances with the original images respect to the preprocessed ones. This might come from the fact that the original images contained additional Gaussian noise. However, the result coming from all datasets were pretty good. However, the additional Gaussian noise of the original images do not bring to the same result when using EDSR & SRGAN. In fact the best performing dataset with this second models is ULAP. We hypothesize that ULAP perform better because of the images frequencies. The experimental results prove that preprocessing the images can be well-suitable for underwater image restoration under different scenarios, faster and more effective to improve the quality of underwater images, according to the best objective evaluations and the lowest running time.

REFERENCES

[1] C. Mobley, “Light and water: radiative transfer in natural waters,” *Academic Press*, 1994.
[2] S. Anwar, C. Li, and F. Porikli, “Deep underwater image enhancement,” *CoRR*, vol. abs/1807.03528, 2018. http://arxiv.org/abs/1807.03528.
[3] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: Challenges, benchmarks, and solutions,” 2019.
[4] G. Yadav, S. Maheshwari, and A. Agarwal, “Contrast limited adaptive histogram equalization based enhancement for real time video system,” in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2392–2397, 2014.
[5] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, “Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition,” in *MultiMedia Modeling* (K. Schoeffmann, T. H. Chalidabhongse, C. W. Ngo, S. Aramvith, N. E. O’Connor, Y.-S. Ho, M. Gabbouj, and A. Elgammal, eds.), (Cham), pp. 453–465, Springer International Publishing, 2018.
[6] H. Chang, C. Cheng, and C. Sung, “Single underwater image restoration based on depth estimation and transmission compensation,” *IEEE Journal of Oceanic Engineering*, vol. 44, no. 4, pp. 1130–1149, 2019.
[7] Y. Wang, W. Song, G. Fortino, L. Qi, W. Zhang, and A. Liotta, “An experimental-based review of image enhancement and image restoration methods for underwater imaging,” 2019.