Enhanced Deep Learning Super-Resolution for Bathymetry Data

Xingyan Li
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA
xingyanli@umbc.edu

Jian Li
NASA Goddard Space Flight Center
Greenbelt, MD, USA
jian.li@nasa.gov

Zachary Williams
NASA Goddard Space Flight Center
Greenbelt, MD, USA
zachary.w.williams@nasa.gov

Xin Huang
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA
xinh1@umbc.edu

Mark Carroll
NASA Goddard Space Flight Center
Greenbelt, MD, USA
mark.carroll@nasa.gov

Jianwu Wang
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA
jianwu@umbc.edu

***Abstract*—Spatial resolution is critical for observing and monitoring environmental phenomena. Acquiring high-resolution bathymetry data directly from satellites is not always feasible due to limitations on equipment, so spatial data scientists and researchers turn to single image super-resolution (SISR) methods that utilize deep learning techniques as an alternative method to increase pixel density. While super resolution residual networks (e.g., SR-ResNet) are promising for this purpose, several challenges still need to be addressed: (1) Earth data such as bathymetry is expensive to obtain and relatively limited in its data record amount; (2) certain domain knowledge needs to be complied with during model training; (3) certain areas of interest require more accurate measurements than other areas. To address these challenges, following the transfer learning principle, we study how to leverage an existing pre-trained super-resolution deep learning model, namely SR-ResNet, for high-resolution bathymetry data generation. We further enhance the SR-ResNet model to add corresponding loss functions based on domain knowledge. To let the model perform better for certain spatial areas, we add additional loss functions to increase the penalty of the areas of interest. Our experiments show our approaches achieve higher accuracy than most baseline models when evaluating using metrics including MSE, PSNR, and SSIM.**

***Index Terms*—deep learning, super-resolution, bathymetry data, transfer learning**

I. Introduction

Accurate determination of water depth is one of the most basic yet important Earth information products, especially bathymetry, and its spatial and temporal variation near coastal areas are essential for much thematic processing of remote sensing data. Most physical and optical environmental quantities derived from satellite observations rely on some prior knowledge of surface topography and depiction of the land and water, without such reliable information there will be areas of water to which terrestrial algorithms will be applied and conversely areas of land to which water algorithms are applied. Some retrieval algorithms utilize information on water extent to promote accuracies in the physical measurements, such as land surface temperature, aerosol retrievals, cloud detection, etc. [1], [2]. Applying bathymetry to a land-water mask as an additional decision criterion also aids to screen out any invalid retrievals and improve the quality of environmental products [3].

In the recent few years, remote sensing techniques on Earth's data with water-land masks, especially in coastal areas, have provided important information for economic development and ecological restoration. Remote sensing images with higher resolution feature higher pixel density, and thus capture more details and are more beneficial for further analysis. However high-resolution images with no blur are not always available due to limitations on equipment, therefore spatial data scientists and researchers turn to deep learning based computer vision techniques to generate high-resolution images from low-resolution ones. However, there are still three challenges remaining for the current super-resolution model to be used on bathymetry data. First, Earth data such as bathymetry is expensive to generate and relatively limited in terms of data records. We need a way to train a good deep learning model without requesting too much training data. Second, because Earth systems and data are governed by physics laws, domain knowledge should be complied with during model training. Third, the accuracy of bathymetry data for coastal areas is more important than that for deep ocean areas because of a lot of human activities in coastal areas, which requires studying how to adjust a super-resolution deep learning model to have better performance for certain areas.

In the light of above related research, we introduce a bathymetry data super-resolution network based on Super-Resolution-ResNet (SR-ResNet) to recover a global terrain image with under-ice topography information of both ocean and land. This study aims to train deep learning based residual neural network utilizing bathymetry data to generate high-resolution images with comparatively high accuracy in coastal areas, and evaluate deep learning based super-resolution for

bathymetry datasets. We have open-sourced our work at github[1]. Our major contributions are as follows.

- To deal with the data record shortage challenge, following transfer learning principle, we propose to leverage an existing pre-trained super-resolution deep learning models such as SR-ResNet for bathymetry data. Our experiments show leveraging a pre-trained model can greatly improve model performance.
- To comply with domain knowledge, we enhance the SR-ResNet model to add corresponding loss functions. Our experiments show these knowledge-based loss functions can help improve some model performance.
- To train the model to pay more attention to certain areas, we add additional loss functions to increase the penalty of the areas of interest. Our experiments show these knowledge-based loss functions can help improve model performance for these areas.

The rest of the paper is organized as follows. Background and related work is introduced in Section 2 and Section 3, respectively. Section 4 explains the datasets we work with. Our proposed method is shown in Section 5 and its evaluation is described in Section 6. Section 7 concludes the paper.

II. Background

This section introduces the background on super-resolution, transfer learning and knowledge-based loss function. As mentioned in the previous section, most of the current SISR models are trained by digital image databases containing RGB images. However, there are no such benchmark datasets of bathymetry for both training and testing. The ordinary real-world RGB images have pixel values ranging from **0 to 255**, while the bathymetry data we use has a global pixel range from **-10802 to 6787**. Although theoretically, any digital image database (e.g., ImageNet [4], LSUN [5], MC COCO [6]) can be used to train a SISR model, we cannot ignore the differences between ordinary RGB images and bathymetry on pixel value ranges and surface texture. The lack of training samples means there will not be enough knowledge to train our model by bathymetric images. Therefore transfer learning is used in this study to deal with limited training samples.

A. Super Resolution

Single image super-resolution (SISR) aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) observation [7]. Assume that LR image is represented by Y, HR image is represented by X, then the relationship between LR image and HR image can be represented by a degradation function $D(\cdot)$as below

$$Y = D(X, \theta_D)$$

Reversely, in the problem of SISR the HR image is the target. Assume that $\hat{X}$ represents the prediction of HR image, $R(\cdot)$ represents the SR function and is the set of parameters, the SISR problem can be defined as

$$\hat{X} = R(Y, \theta_R)$$

The super-resolution process $R(\cdot)$ can be divided into two categories including reconstruction-based and learning-based algorithms (Siu, et al., 2012) [8]. Some early works adopt the former methods, which means using prior knowledge to invent data fidelity algorithms to implement on LR images directly. The examples include example-based methods, sparse representation-based method, and regression-based method [9].

Reconstruction-based SISR features low computational intensity on each image pixel. However the learning-based super-resolution methods with knowledge from training samples are proved to be more effective and robust, especially those utilizing deep learning techniques. LeCun et al. [10] show in their work how convolutional neural network architecture downscales images to establish non-linearity between outputs and hidden layer outcomes, and extract feature maps from images. These show a promising approach to designing an end-to-end mapping function from LR image and HR image, and SR-CNN [11], which is constructed on top of this architecture, achieves higher PSNR and SSIM scores on benchmarking datasets Set5, Set14 and BSD200. Later, Ledig et al. proposed the Super-Resolution Generative Adversarial Network (SRGAN) [12] by adopting a GAN (Goodfellow., 2014) [13] pipeline to produce finer images. The proposed model SR-GAN has a perceptual loss function including a VGG-based content loss for the generator and an adversarial loss for the discriminator. In the evaluation results, incidental model without using the adversarial module SR-ResNet achieves higher PSNR and SSIM than interpolation method, SR-CNN, and even SR-GAN, and moreover, cost less training time and computational resources. However SR-GAN has better experiment results evaluated by MOS, and thus the authors believe it outperforms other models. In the same year, the Enhanced Deep Residual for Single Image Super-Resolution [14] is proposed by deleting unnecessary modules in the residual networks [15] and the proposed models EDSR and MDSR produce results with higher PSNR and SSIM than interpolation algorithms, SR-CNN and other super-resolution methods. These results prove that residual network is an effective basis to build SR models on.

B. Transfer Learning

Transfer learning is a machine learning technique that focuses on learning knowledge from one domain (task) and applying it to a different but related domain (task). For example, knowledge learned from training to recognize pedestrians in a large-scale simulated environment could apply when trying to recognize pedestrians on real street scenes in self-driving. This area of research bears some belief in reusing or transferring information from previously learned tasks for the learning of new tasks as the relationship or mapping structure in the previous task persists in the new task.

One of the challenges in learning remote sensing data is the limitation of label information for newly collected RS data.

[1]https://github.com/big-data-lab-umbc/bathymetry_super_resolution

It is very difficult to construct a large-scale well-annotated dataset due to the expense of data acquisition and the costly annotation. With the development of modern satellite sensors and the collection of new remote sensing data, processing such a large amount of data becomes even more challenging. A straightforward approach is to transfer existing knowledge to help understand unknown new data. To achieve this purpose, deep transfer learning-based frameworks can be used to overcome the semantic gap between different datasets. In recent years, many deep transfer learning models have all been successfully applied to analyzing remote sensing data [16]–[18]. Using pre-trained models (features) and fine-tuning on the target dataset is the most widely used way to exploit transfer learning. A recent study found fine-tuning pre-trained CNN models achieves the best performance for optical remote sensing change detection [19].

C. Knowledge-Based Loss Functions for Deep Learning

Loss function is a mathematical representation of the cost of algorithm output and the ground truth, and the loss is minimized to optimize the algorithm. Customizing loss functions is one of the methods to frame the problem of model optimizer into a special target. For super-resolution problems, the loss functions of the model optimizer should gauge the model performance by calculating the error between generated images and ground truth. In addition to that, we also specialize our problem by focusing more on coastal areas. Therefore, knowledge-based functions are necessary to exert a bigger penalty on wrongly-predicted pixels in coastal areas than in deep ocean area or land area. For environmental science applications, there is a guide to custom loss function for neural networks [20].

III. Related Work

Our study focuses on the application of super-resolution on single-band bathymetry data, which is different from three-channel RGB images commonly used as benchmarking datasets in most of the state-of-the-art models (bathymetry data is introduced in Section IV). Thus, the theoretical and experimental guidelines on similar Earth data or environment data would be essential. This section introduces the related works of customizing loss functions for deep learning and applying super-resolution models on Earth data.

In recent years, research on deep learning based super-resolution for spatial science has never stopped because it is a promising method to gain high-quality data. Vandal et al. proposed DeepSD [21] based on SR-CNN to generate high-resolution local-scale climate projections. The model is trained by historical climate observations and shows better evaluation results on Bias, Correlation and RMSE than SR-CNN and other deep learning models. This work proves that deep learning based super-resolution models can adopt earth data for training and generate reasonable results. Tao and Muller introduce a novel MSA-FAST-CNN-GPT-GAN (MAGiGAN) [22] super-resolution system based on a photogrammetric restoration approach as well as CNN to analyze Terra Multi-angle Imaging SpectroRadiometer (MISR) red band images. In 2020, Leong and Horgan proposed DeepBedMap [23] based on EDSR to generate high-resolution images on bed topography trained by scattered regions in Antarctica. The DeepBedMap is evaluated based on the digital elevation model (DEM) and surface roughness, and the model makes predictions on fine-scaled details but high-gradient areas due to lack of training data. In 2021, Han and Zhang et al. proposed a novel modified generative adversarial nets with a decoder (DeGAN) [24] combined with stochastic simulation. This study aims to super-resolution in reservoir simulation. Although this work does not analyze real-world images, it demonstrates the effectiveness of balancing the contributions of different neural networks to generate high-quality results.

There are also studies on how to do SISR for bathymetry data without using deep learning based approaches. For instance, Yutani et al. [25] improved linear sparse coding super-resolution (ScSR) to study ocean bathymetric maps and use visualization methods including residual images and distribution of clusters to show the efficiency of the model.

From the research above, we can see that deep learning based SISR methods for Earth data are constructed based on two neural networks: convolutional neural networks (CNN) and generative adversarial networks (GAN). On the other hand, the current study on SISR models for bathymetric data has not given rise to much attention, and there are no deep learning based approaches published yet. These works successfully adopt knowledge of Earth data into super-resolution approaches and show a promising research area of super-resolution specialized for spatial science.

However, there are still three limitations and challenges remaining to study super-resolution methods for bathymetry. 1) Although there have been a few studies on bathymetric data using conventional sparse representation-based methods, we still lack experience in deep learning based super-resolution models or improvements for bathymetric data. 2) The study of super-resolution for bathymetric data would be a multi-disciplinary study of spatial science and computer vision, so there is no conventional evaluation and visualization metric specialized for bathymetry. 3) Some of the studies are local-focused, which means the model is tested partially or only on images representing the same areas used as training data. This may cause a large bias when the model is tested in other areas.

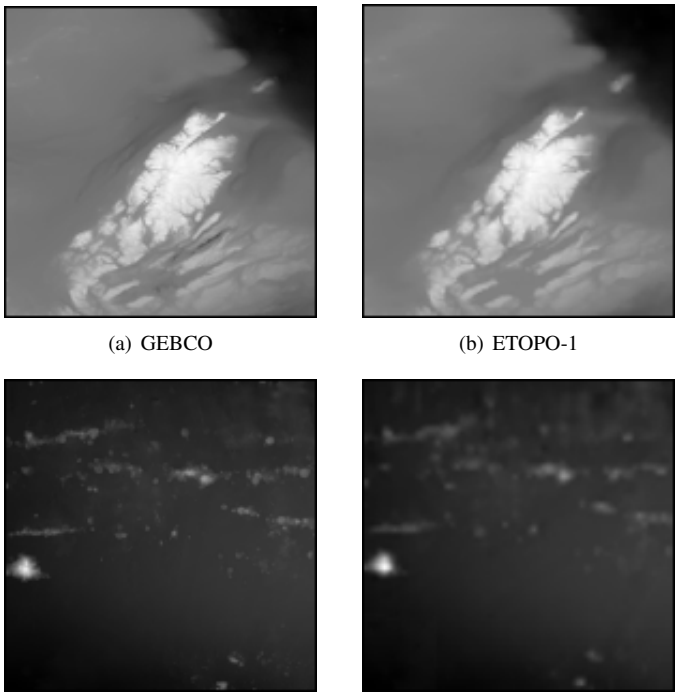
IV. Data

In this study, we created 2000 data tiles for training and testing, where each tile composes a stack of one high-resolution image, low-resolution image, and a corresponding binary water mask. High-resolution tiles are 15-arc seconds spatial resolution, derived from the General Bathymetric Chart of the Oceans (GEBCO) Grid, supported by the International Hydrographic Organization (IHO) and the Intergovernmental Oceanographic Commission (IOC) published in July 2021 and accessed in September 2021 [26]. Low-resolution tiles are 1-arc-minute spatial resolution and derived from ETOPO-1

Global Relief Model, published in June 2011 by the National Oceanic and Atmospheric Administration (NOAA) and accessed in September 2021 [27].

The bathymetric data contained within both GEBCO and ETOPO1 are amalgamations of the best available data sources compiled from both US and International agencies and academic institutions. Data collections were then re-gridded to common vertical and horizontal datums to create seamless, continuous global terrain models. Where no direct observations were available, interpolation was used. Pixel values within the GEBCO and ETOPO1 datasets represent 16-bit signed integer values depicting elevation ranging from -7,818 to 5,840 meters within the GEBCO and -10,898 to 8271 meters within ETOPO-1. Both datasets feature global coverage (180W, 180E) (90N, 90S) within the WGS 1984 geographic coordinate system. The corresponding water masks were derived from the MOD44W C5 global binary water mask [28], reprojected and downsampled to match the horizontal resolution of GEBCO.

The training and testing tiles were generated using a GIS approach. Coastlines were determined from ETOPO-1 data and buffered 1 degree on either side. This 2-degree global coastal region ensures all data tiles will contain both land and water pixels while capturing a diverse suite of bathymetric profiles (i.e. canyons, island chains, and bays), including several deeper water environments (greater than 300m in depth). 2000 random points were then generated within this buffered area, with a minimum distance of 1.4 degrees between points to minimize tile overlap, and used as center points to define an extraction window. Each tile then represents an area of 2.13 decimal degrees on each side, equivalent to 512x512 pixels in the GEBCO data and water mask, and 128x128 pixels for ETOPO-1 (Figure 1).



(a) GEBCO (b) ETOPO-1

(c) GEBCO (d) ETOPO-1

Fig. 1: Examples of high-resolution GEBCO (left) and low-resolution ETOPO-1 (right) data tiles used in this study. The top row demonstrates a tile stack within a complex coastal environment (near the mouth of the St. Lawrence River centered about 46.8 N, 59.9 W) and the bottom row demonstrates a deep water environment (seamount chain north of Hawaii centered about 23.5 N, 157.4 W).

V. Methods

A. Overall Architecture

The bathymetry super-resolution model is based on a pre-trained SR-ResNet model to generate remote-sensing images with higher resolution than the original ones with a limited training dataset. First we pre-train a SR-ResNet model by Large-scale CelebFaces Attributes (CelebA) [29] Dataset which contains 202,599 JPEG images. Then we use transfer learning to adapt our bathymetry dataset composed of 2000 Tiff images, with fine tuning on the last layer of the model. Finally, the model is used to generate high-resolution images for any Tiff inputs. Both the pre-trained model and bathymetry model are developed using python API and libraries of PyTorch. Figure 2 illustrated the overall architecture of our bathymetry super-resolution model.

B. Bathymetry Data Preprocessing

Transfer learning provides a solution to handle the shortage of training datasets to train the SR-ResNet model directly. However, most of the accessible images are in the format of JPEG or PNG, which are different from TIFF images in the number of bands and range of pixel values. Therefore we need to preprocess our bathymetry data before feeding it into the saved SR-ResNet model.

To feed our data into the SR-ResNet model which is pre-trained by JPEG images, we separate the geographic coordinates from bathymetry information layers and reshape the training single-banded images from N×N to three-dimensional images of N×N×3 by concatenating, where N represents 512 for high-resolution images and 128 for low-resolution images. Then we customize Min-Max normalization on the pixel values based on the maximum pixel values and minimum pixel values of all training tiles. Finally, the pixel data type should be transferred from integers to floats. The geographic coordinates are attached to the generated image during output.

C. Transfer Learning

We propose a new method to enhance the resolution of Bathymetry images by transferring the knowledge from the large-scale CelebFaces Attributes dataset. The mapping between low/high-resolution CelebFaces images is similar to that between low/high-resolution Bathymetry images. In our proposed framework, the structure and relationship between low and high-resolution images used in large-scale training can be learned by deep convolutional neural networks (SR-ResNet)

TABLE I: A comparison between ETOPO-1 and GEBCO

Data	Vertical Datum	Spatial Resolution	Pixel value data type
ETOPO-1	Sea Level	1 arc-minute	16-bit signed integer
GEBCO Global Grid	GSea Level	15 arc-seconds	16-bit signed integer

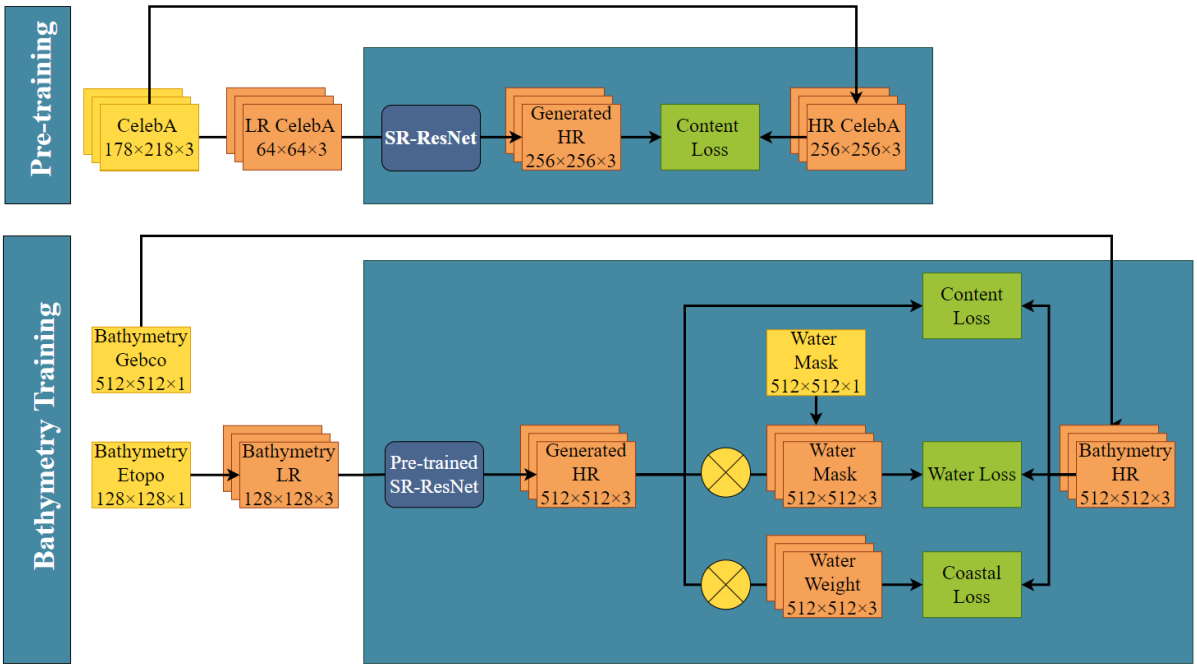


Fig. 2: Overall architecture of the bathymetry super-resolution model. In the pre-training model, a benchmark RGB-banded dataset CelebA is utilized. The original images are HR training tiles and they are downscaled by bicubic interpolation to be LR training counterparts. During training, the content loss from SRResNet penalty on errors. In our bathymetry training model, our datasets are pre-processed to be training tiles. This bathymetry model is then trained based on the pre-trained model with knowledge-based loss functions (water loss, weighted loss and edge loss) and content loss. The trained model is saved to generate HR images for any LR inputs.

and be transferred to Bathymetry images by exploiting the idea of transfer learning. To implement the transfer learning based super-resolution for Bathymetry data, we use a pre-training and fine-tuning approach. First, the SR-ResNet network is pre-trained on a Large-scale CelebFaces Attributes (CelebA) Dataset containing 202,599 JPEG images, and then the pre-trained model with updated parameters is used as an input into our model to give the model some 'knowledge' of training. In particular, transfer learning with the help of a large-scale CelebFaces Attributes dataset is performed for the external learning of backbone structure and mapping. We propose a new method to enhance the resolution of Bathymetry images by transferring the knowledge from the readily available dataset CelebA with a different task. The weights and parameters of convolutional layers in SR-ResNet are optimized during pre-training, and the outputs of pre-training are used as starting parameters to train on a similar network that requires a smaller dataset. This smaller network only needs to learn the relations and differences for our specific problem having already learned about patterns in the data from the pre-trained model.

D. Loss Functions

In our study, we use two categories of loss functions: content loss functions and knowledge-based loss functions.

The content loss represents the error between image content features of generated images and targeted images, and it is calculated by Euclidean norm, or L2 norm [30]. For super-resolution problems, the content loss can be calculated based on the images' pixel values directly, or high-level image features extracted from the original images. In the SR-ResNet model proposed by Ledig et al., the content loss function is based on image features [31] extracted by a VGG neural network. Although a pixel-wise loss function can be an alternative, the author argues in the paper that pixel-wise losses may give rise to a lack of high-frequency content. However, in our study, we use both pixel-wise losses and VGG-based loss in our content loss function $l_{Content}$,.

The knowledge-based loss functions are combined with the content loss to improve accuracy based on different areas in Earth images. To be more specific, two pixel-wise and VGG-

based knowledge-based loss functions are implemented in the bathymetry SR model and the loss functions in different experiments are their combinations. The different loss functions are water loss l_{Water} and coastal loss $l_{weighted}$. In the above equations, l_{MSE} is a pixel-wise loss function of mean square error, while l_{VGG} is an MSE loss function based on feature maps extracted by a VGG neural network.

Assume that r is a scaling ratio (the value is 4 in this study) W and H are the width and height of input images, $I^{HR}_{x,y}$ is the pixel located at (x, y) in a high-resolution image, $G_{\theta_G}(I^{LR})_{x,y}$ is the pixel at the same location in a generated image, and $\phi(I)$ is a function of VGG-19 feature extractor. Content loss is $l_{Content} = l_{MSE} + l_{VGG}$, then the components of the content loss function are calculated as below.

$$l_{MSE} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l_{VGG} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

Water loss is used to penalize the generated images only on ocean areas to minimize wrong predictions on ocean area (where the pixel values are negative) and land area (where the pixel values are positive). Since bathymetric data is used in conjunction with water masks, it is essential to maintain information about water masks in the generated images, and mistakes in predictions of ocean and land areas may be fatal in future applications of the generated high-resolution images. The water loss is calculated similarly to content loss, but adding an binary water mask E on top of the generated image before calculating loss. In other words, the high-resolution image $I^{HR}_{x,y}$ will be $E \cdot I^{HR}_{x,y}$, and the generated image $G_{\theta_G}(I^{LR})_{x,y}$ will be $E \cdot G_{\theta_G}(I^{LR})_{x,y}$. Water loss is a summation of a pixel-wise loss and a VGG based loss: $l_{Water} = l_{MSE_{water}} + l_{VGG_{water}}$. The components are as below.

$$l_{MSE_{water}} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(E \cdot I^{HR}_{x,y} - E \cdot G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l_{VGG_{water}} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\phi_{i,j}((\phi_{i,j}(E \cdot I^{HR})_{x,y} - \phi_{i,j}(E \cdot G_{\theta_G}(I^{LR}))_{x,y})^2$$

Coastal loss is used to give higher penalties for wrong predictions on pixels in coastal areas (where water depth is larger than -160 meters). A weight map M is used to weigh each pixel in the loss function according to the pixel value. Pixels with smaller magnitude values have higher weights when reciprocal pixel values are utilized, and thus pixels with smaller magnitude in the ocean area (coastal area) gain higher penalties. The weight map is calculated by $M_{i,j} = \frac{1}{||I^{HR}||} \cdot E_{i,j}$. Coastal loss is $l_{weighted} = l_{MSE_{weighted}} + l_{VGG_{weighted}}$, and the components are as below.

$$l_{MSE_{weighted}} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}M_{i,j} \cdot E_{i,j}(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_x, y)^2$$

$$l_{VGG_{weighted}} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\phi_{i,j}(M_{i,j} \cdot E_{i,j})_{x,y} \cdot (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

We conduct different experiments using models constructed by different combinations of the loss functions and compare the experimental results with three baseline models. Section VI discusses the description and results of baseline methods and our experiment models.

VI. Experiments

The experiment consists of two processes: training and testing. The training process refers to pre-training with CelebA dataset and training with our dataset consisting of 1,600 tiles of high-resolution and low-resolution images. Over the testing process, the trained mode generates 400 images from the test set, and we evaluate the generated images by evaluation metrics. We refer to the evaluation results as experiment results.

In this section, we first investigate the experiment results of different methods. Then we analyze the performance of knowledge-based loss functions and inspect whether they increase prediction accuracy in coastal areas. Finally, the proposed model is selected by performance and compared with baseline models. The experiment results are generated from 400 tiles of low-resolution images in the test dataset and are evaluated by frequently used metrics including mean squared error (MSE), peaks signal to noise ratio (PSNR) and structural similarity index metric (SSIM). The performance of knowledge-based loss functions is presented by changes in evaluation metrics on different areas: whole image, ocean areas and coastal areas. To be more specific, the evaluation on ocean areas represents the evaluation on generated images applied by binary water masks; the evaluation on coastal areas represents the evaluation on pixels with pixel values larger than -160.

As mentioned above, there are three baseline experiments to compare with our models. In Table II, baseline 0 model is bicubic-interpolation, baseline 1 is directly testing pre-trained SR-ResNet with our data and baseline 2 is directly training SR-ResNet using our data.

A. Evaluation Metrics

In this study, model performance is evaluated by three frequently used metrics including mean squared error (MSE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and one novel metric introduced by ourselves, wrong pixel ratio (WPR). This section concludes the mathematical fundamentals of these metrics.

TABLE II: Three baseline models and corresponding evaluation results by MSE, PSNR and SSIM on the whole area, ocean area and coastal area, respectively

	Evaluation on whole area			Evaluation on ocean area			Evaluation on coastal area			
Models	**MSE**	**PSNR**	**SSIM**	**MSE**	**PSNR**	**SSIM**	**MSE**	**PSNR**	**SSIM**	**WPR**
Bi-cubic interpolation	**1,303**	**49.654**	**0.988**	**7,853**	**47.401**	**0.989**	**1,771**	**52.996**	**0.995**	**0.011**
Pre-training	58,574,684	-0.846	0.004	57,881,520	-0.7663	0.220	37,101,514	1.817	0.481	0.284
Direct training	59,753,964	-1.027	0.023	60,407,267	-1.0263	0.240	40,019,392	1.440	0.484	0.302
Content loss	**62,357**	34.096	0.782	**62,137**	34.136	0.808	24,248	37.825	0.867	0.135
Content + water loss	612,187	**38.994**	**0.927**	611,307	**39.613**	**0.946**	31,449	**44.752**	**0.978**	**0.084**
Water loss	4,081,509	13.841	0.236	3,813,084	14.786	0.511	2,090,382	17.871	0.621	0.354
Water + coastal loss	3,310,707	14.3676	0.264	2,992,009	15.448	0.544	1,589,616	18.736	0.642	0.347
Coastal loss	**142,761**	36.039	0.898	102,043	37.294	**0.940**	**21,555**	**42.317**	**0.963**	0.133

MSE: Mean squared error is a commonly used metric in supervised machine learning and deep learning areas. For spatial data or images, it compares differences between generated images and the test data on a pixel basis. Assume I is an original image (ground truth), K is a generated image, m and n are the height and width of the images, respectively; then the MSE value should be:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

PSNR: Peak signal-to-noise ratio is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Mathematically it is proportional to the reciprocal of MSE, and thus larger PNSR indicates better results.

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

SSIM: The structural similarity index is used for measuring the similarity between two images based on image luminance, contrast and structure, and a larger value of SSIM indicates better results. Assume that where x,y are samples; σ is the standard deviation of pixel values; μ is the mean of pixel values.

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

WPR: The wrong pixel ratio is a knowledge-based evaluation metric. WPR is introduced to calculate the percentage of generated pixels with an opposite sign to the ground truth. In other words, if a pixel sign changes from positive to negative, or negative to positive after prediction, then it is collected as a pixel with the wrong sign. This evaluation metric is invented because the positive or negative value of pixels represents land or water area in the bathymetry images, and the correct water mask is essential for gaining geographic information from the images. Therefore it would be problematic to have wrong signed pixels in generated images. $\sum_{i=0}^{j=0} N_K(i,j)$ is the total number of pixels in the generated image, and function g equals 1 if the pixels $I(i,j)$ and $K(i,j)$ have opposite signs.

$$WPR = \frac{\sum_{i=0}^{j=0} g(I(i,j), K(i,j)}{\sum_{i=0}^{j=0} N_K(i,j)}$$

B. Models and Experiment Results

Based on our proposed model in Figure 2, the models are trained starting from a pre-trained model to recover low-resolution images directly from a size of (128,128) to high resolution images with height and width (512, 512) by upsampling. Different models under this architecture utilize different loss functions. There are 7 different combinations of loss functions implemented in the models. Among the seven experiments, Model 6 and Model 7 produce pixels with extraordinarily large values and the image patterns totally change from the ground truth. Since these two models cannot produce reasonable images, we exclude them from Tabel II and only keep records of the other five experiment models.

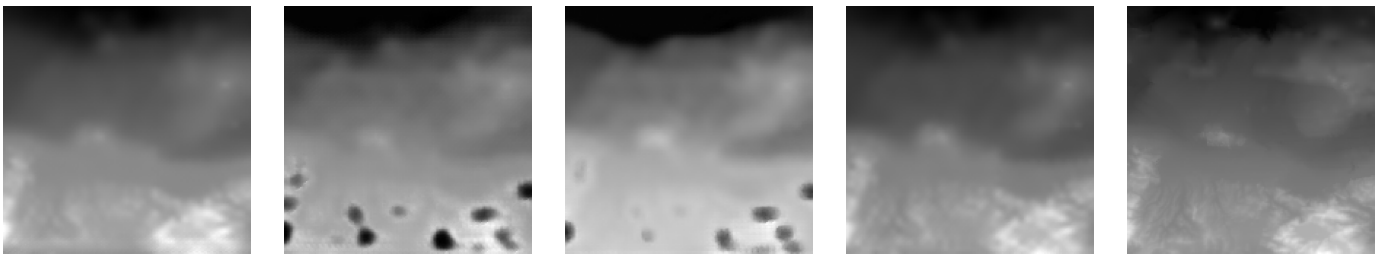
- Model 1: content loss
- Model 2: content loss + water loss
- Model 3: water loss
- Model 4: water loss + coastal loss
- Model 5: coastal loss
- Model 6: content loss + water loss + coastal loss
- Model 7: content loss + coastal loss

The evaluation results of three baseline models and five experiment models are shown in Table II and the visual comparison is shown in Figure 3. We evaluate the whole area, ocean area and coastal area respectively to verify whether our knowledge-based loss functions improve accuracy in the coastal area. By observing this table and comparing different models, we analyze model performance and draw conclusions as below.

By comparing the pre-trained model and model 1 (the experiment model with content losses), we verify our hypothesis that the knowledge of ordinary RGB images do not provide SR-ResNet with adequate information to reconstruct high-resolution bathymetric images. Both of the two models utilize content loss functions, while model 1 undergoes another 200 epochs of training by bathymetric images based on the pre-trained model. Even though SR-ResNet has already been trained by 202,599 images over 200 epochs, the generated bathymetric images from the pre-trained model have high MSE values and bad image patterns. Therefore it is essential to train the SISR model on datasets of the same image genre as the target.

By comparing the directly-trained model and model 1,

(a) LR image (b) Interpolation (c) Pre-training (d) Direct training (e) Content loss



(f) Content + water loss (g) Water loss (h) Water + coastal loss (i) Coastal loss (j) HR image

Fig. 3: The visual comparison among different experiments and ground truth at the same location. From (a) to (j) are original low-resolution image, bi-cubic interpolation (baseline 1), pre-trained model (baseline 2), directly-trained model (baseline 2), model 1 (content loss), model 2 (content + water loss), model 3 (water loss). model 4 (water + coastal loss), model 5 (coastal loss) and original high-resolution image (ground truth). Among all the experiments, model 5 (coastal loss) is the best model we have. The area in white color has positive pixel values and represents the land area, the black area represents the deep ocean and the area in grey around land area represents the coastal area.

we demonstrate that transfer learning improves model performance significantly. The directly-trained model is an SR-ResNet model directly trained by our bathymetric dataset, while model 1 transfers pre-trained knowledge from the pre-trained model. From the changes in values of four evaluation metrics, we can safely conclude that transfer learning leads to improvement in the performance of the SR-ResNet model to be implemented on bathymetry.

We can also indicate from a comparison among five experiment models that model 2 (utilizing content loss and water loss) and model 5 (utilizing coastal loss) perform the best. Predictions from Model 1 show the lowest MSE results on the whole area and ocean area, however Model 2 is the most accurate according to values of PSNR, SSIM and WPR. Besides model 2, we cannot ignore that model 3 also produce results with relatively low MSE, high PSNR, high SSIM and low WPR for all the selected areas.

Since model 2 and model 5 have been chosen as the best models, we move to analyze these two models by comparing different areas (whole area, ocean area and coastal area) to inspect the performance of knowledge-based loss functions. From Table II we can see the evaluation results of model 2 and model 5 experience a steady decrease in MSE and increase in PSNR and SSIM when the evaluation area changes from the whole to deep ocean and coastal. Model 5 with coastal loss implemented shows the best results in coastal area because the weight map assigns higher weights to coastal pixels than deep ocean pixels, and thus drive the model to achieve higher accuracy on the coastal area of generated images.

By comparing the bi-cubic interpolation method and our experiment models, we have to admit that our models based on SR-ResNet cannot exceed interpolation methods on evaluation metrics including MSE, PSNR, SSIM and WPR. However, these quantitative evaluation results are not fully reliable for super-resolution, which is also a conclusion drawn by Ledig et al. in SRGAN [12]. Actually, if we compare the generated images (b) and (i) in Figure 3, our experiment models with coastal loss distinguish more details in the coastal area (grey-colored area). For our deep learning based models, to achieve higher accuracy in quantitative results, a pre-training dataset with pixels of deeper bit depth (e.g. 16-bit) will help instead of RGB images (8-bit). As mentioned in Section II, most of the current SISR models utilize RGB images whose range of pixel values (0 to 255) are far different from bathymetry images (-10802 to 6787), so a pre-trained model using images with deeper bit depth (16-bit signed integer) would likely yield higher fidelity in the output maps. Moreover, the surface textures of real-world RGB images and bathymetry have different features. In RGB images, there are sharp edges to distinguish different items, while bathymetry values always change continuously and follow the change of topography. In other words, if we use the interpolation method, the generated pixel values are calculated using values of adjacent pixels, and thus the predictions cannot be far from the ground truth. Oppositely, the generated pixels value by a deep learning model is random and regulated by loss functions, so some

of them may differ from the ground truth. However, since we use VGG neural network to extract high-level features from images, the deep learning based model can capture image patterns better than the interpolation method. Hence our model shows better visual results while the interpolation method produces better evaluation results.

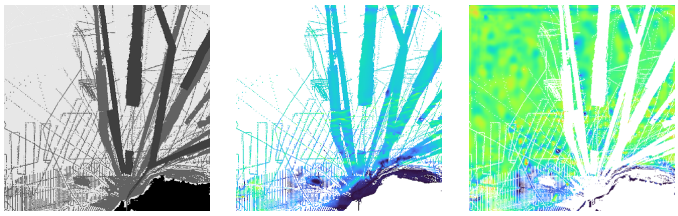
To summarize, our conclusions drawn from the analysis of experiment results are: 1) Transfer learning improves the performance of SR-ResNet in dealing with limitations in bathymetry data. 2) Our knowledge-based loss functions improve prediction accuracy in coastal areas. 3) Among our five experiment models, the model using coastal loss and the model using content loss with water loss is the best according to evaluation metrics and visual comparison. 4) We enhance SR-ResNet models for bathymetry data. The models fail to outperform the interpolation method by evaluation metrics, although our best model (the model using coastal loss) generates better visual results than the interpolation method does. The reason is that bathymetry images have a very special range of pixel values and image textures.

C. Spatial Model Performance Analysis: Direct Observations vs. Interpolated input data

The bathymetric data contained within the GEBCO Grid is an amalgamation of several direct and indirect measurement sources, including interpolated data (source). In order to accurately present the accuracy results of our model, we must demonstrate that model results do not perform better over interpolated data than direct observations. We use the GEBCO Type Identifier grid to parse GEBCO bathymetric data from five validation tiles into regions of direct observation and interpolated data and take the absolute difference between GEBCO pixel values and modeled pixels (Figure 4). We then compared the normalized mean and standard deviation of these areas to determine if there is preferential performance within the model between either pixel population. Similar mean and standard deviation values for the direct observations (0.48 and 0.12) and the interpolated data (0.41 and 0.10) indicate that the model performed similarly across all data inputs.

VII. Conclusion

The Enhanced Bathymetry Super-Resolution model presents a data-driven approach to reconstructing the ocean bathymetric data using SR-ResNet-based model trained by GEBCO and ETOPO-1 dataset. We enhance the SR-ResNet by knowledge-based loss functions (water loss and coastal loss) and transfer learning to train the model with limited bathymetry data and generate high-spatial-resolution bathymetry that recovers more accurate details in the coastal area where pixel values are larger than -160. The generated images are adaptable for future studies on water masks. Unlike other current bathymetry super-resolution models relying on sparse coding algorithms, the model uses a deep learning based approach to find suitable neural network parameters via an iterative error minimization approach. From the visual comparison of generated images, we can see this makes the resulting model properly recover



(a) (b) (c)

Fig. 4: An example of the GEBCO Tile Identifier grid (a) where dark colors represent direct observations, such as ship tracks, and light gray regions represent regions of interpolated data. The black region in the bottom right corner represents land and is excluded. The absolute difference between GEBCO data and model results for either region is shown for the direct (b) and interpolated (c) regions.

image details and the generated images show better image patterns than the interpolation method. This is because a VGG neural network is used to extract features from bathymetry data. However, the results of experiment models evaluated by MSE, PSNR, SSIM and a self-introduced metric WPR show that our best experiment model cannot beat interpolation. By analysis, we find the reason is that bathymetry has a very huge pixel value range and special surface texture features.

There are still limitations to our study. This study shows that the evaluation of bathymetric data cannot only relies on metrics PSNR and SSIM, although these two methods are conventional evaluation metrics in super-resolution studies [7]. In future studies, we will work on reliable evaluation metrics for bathymetry super-resolution. Furthermore, we should enlarge the dataset by collecting more samples for training and testing.

Acknowledgement

This work is partially supported by NSF grants: CAREER: Big Data Climate Causality (OAC-1942714) and HDR Institute: HARP - Harnessing Data and Model Revolution in the Polar Regions (OAC-2118285).

References

[1] C. Justice, J. Townshend, E. Vermote, E. Masuoka, R. Wolfe, N. Saleous, D. Roy, and J. Morisette, "An overview of MODIS land data processing and product status," *Remote Sensing of Environment*, vol. 83, no. 1-2, pp. 3–15, Nov. 2002. [Online]. Available: https://doi.org/10.1016/s0034-4257(02)00084-6

[2] E. Masuoka, D. Roy, R. Wolfe, J. Morisette, S. Sinno, M. Teague, N. Saleous, S. Devadiga, C. O. Justice, and J. Nickeson, "MODIS land data products: Generation, quality assurance and validation," in *Land Remote Sensing and Global Environmental Change*. Springer New York, 2010, pp. 509–531. [Online]. Available: https://doi.org/10.1007/978-1-4419-6749-7-22

[3] M. Wang, X. Liu, L. Tan, L. Jiang, S. Son, W. Shi, K. Rausch, and K. Voss, "Impacts of viirs sdr performance on ocean color products," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 18, pp. 10–347, 2013.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2009. [Online]. Available: https://doi.org/10.1109/cvpr.2009.5206848

[5] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," *arXiv preprint arXiv:1506.03365*, 2015.

[6] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, pp. 740–755. [Online]. Available: https://doi.org/10.1007/978-3-319-10602-1-48

[7] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu, "Real-world single image super-resolution: A brief review," *Information Fusion*, vol. 79, pp. 124–145, 2022.

[8] W.-C. Siu and K.-W. Hung, "Review of image interpolation and super-resolution," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2012, pp. 1–10.

[9] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002. [Online]. Available: https://doi.org/10.1109/38.988747

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] 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.

[12] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jul. 2017. [Online]. Available: https://doi.org/10.1109/cvpr.2017.19

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Jul. 2017. [Online]. Available: https://doi.org/10.1109/cvprw.2017.151

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2016. [Online]. Available: https://doi.org/10.1109/cvpr.2016.90

[16] J. N. B. M. L. D. . E. S. Xie, M., "Transfer learning from deep features for remote sensing and poverty mapping," in *In Proceedings of the thirtieth AAAI conference on artificial intelligence, AAAI-16*. AAAI Press, 2016, pp. 3929–3935.

[17] X. Huang, S. Ali, C. Wang, Z. Ning, S. Purushotham, J. Wang, and Z. Zhang, "Deep Domain Adaptation based Cloud Type Detection using Active and Passive Satellite Data," in *In: 2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 1330–1337.

[18] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Deep transfer learning for few-shot SAR image classification," *Remote Sensing*, vol. 11, no. 11, p. 1374, Jun. 2019. [Online]. Available: https://doi.org/10.3390/rs11111374

[19] M. E. Amin Larabi, S. Chaib, K. Bakhti, and M. S. Karoui, "Transfer learning for changes detection in optical remote sensing imagery," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 1582–1585.

[20] I. Ebert-Uphoff, R. Lagerquist, K. Hilburn, Y. Lee, K. Haynes, J. Stock, C. Kumler, and J. Q. Stewart, "Cira guide to custom loss functions for neural networks in environmental sciences–version 1," *arXiv preprint arXiv:2106.09757*, 2021.

[21] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "DeepSD," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2017. [Online]. Available: https://doi.org/10.1145/3097983.3098004

[22] Y. Tao and J.-P. Muller, "Super-resolution restoration of MISR images using the UCL MAGiGAN system," *Remote Sensing*, vol. 11, no. 1, p. 52, Dec. 2018. [Online]. Available: https://doi.org/10.3390/rs11010052

[23] W. J. Leong and H. J. Horgan, "DeepBedMap: a deep neural network for resolving the bed topography of antarctica," *The Cryosphere*, vol. 14, no. 11, pp. 3687–3705, Nov. 2020. [Online]. Available: https://doi.org/10.5194/tc-14-3687-2020

[24] F. Han, H. Zhang, S. Chatterjee, Q. Guo, and S. Wan, "A modified generative adversarial nets integrated with stochastic approach for realizing super-resolution reservoir simulation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1325–1336, Feb. 2020. [Online]. Available: https://doi.org/10.1109/tgrs.2019.2945946

[25] T. Yutani, O. Yono, T. Kuwatani, D. Matsuoka, J. Kaneko, M. Hidaka, T. Kasaya, Y. Kido, Y. Ishikawa, T. Ueki, and E. Kikawa, "Super-resolution and feature extraction for ocean bathymetric maps using sparse coding," *Sensors*, vol. 22, no. 9, p. 3198, Apr. 2022. [Online]. Available: https://doi.org/10.3390/s22093198

[26] GEBCO Bathymetric Compilation Group 2021, "The gebco-2021 grid - a continuous terrain model of the global oceans and land." 2021. [Online]. Available: https://www.bodc.ac.uk/data/published-data-library/catalogue/10.5285/c6612cbe-50b3-0cff-e053-6c86abc09f8f/

[27] C. Amante, "Etopo1 1 arc-minute global relief model: Procedures, data sources and analysis," 2009. [Online]. Available: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.dem:316

[28] M. L. Carroll, C. M. DiMiceli, J. R. G. Townshend, R. A. Sohlberg, A. I. Elders, S. Devadiga, A. M. Sayer, and R. C. Levy, "Development of an operational land water mask for MODIS collection 6, and influence on downstream data products," *International Journal of Digital Earth*, vol. 10, no. 2, pp. 207–218, Oct. 2016. [Online]. Available: https://doi.org/10.1080/17538947.2016.1232756

[29] Z. Liu, P. Luo, X. Wang, and X. Tang, "Large-scale celebfaces attributes (celeba) dataset," *Retrieved August*, vol. 15, no. 2018, p. 11, 2018.

[30] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels," *arXiv preprint arXiv:1205.2653*, 2012.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.