Improved Recoloring of SAR Satellite Images

Sukrut Oak   Young Chol Song   Da Sun

Stanford University

{sukrut,youngcs,dasun}@stanford.edu

Abstract

Synthetic Aperture Radar (SAR) images, while invaluable for various remote sensing applications, lack the intuitive visual information provided by color. Our project seeks to enhance the quality of SAR image colorization, a process of assigning colors to grayscale SAR images, by employing a combination of image classification, segmentation, and coloring techniques. By utilizing an ensemble of colorization methods and selecting the optimal approach based on an improved understanding of scene structure and the distinct performance strengths of different models, we demonstrate that our method achieves more accurate and realistic colorization of SAR images. Our results underscore the potential of our approach in incorporating more context for SAR imaging in the enhanced visual interpretation and application of various environmental and surveillance tasks.

We utilize a subset of the SEN12MS-CR dateset, focusing on the San Francisco Bay Area. We first classify the SAR images into five terrain categories by using an instruction refined GPT-4o model. We then evaluate four recolorization models (cGAN, CNN, NL, and LR) via NRMSE, SAM, and Q4 metrics on their performance for each terrain category. When testing our terrain-based approach, we use the optimal model for the terrain category and compare the results of our terrain-based approach to the best performing model presented in Shen et al [14].

1. Introduction

As the costs of accessing space have decreased dramatically in the past decade, more technologies are being launched onboard the proliferating number of satellites in space. The SAR concept was invented in 1951 by the Goodyear Aircraft Company, but has since resurged in popularity during the past decade as one of the forefront remote sensing technologies to dominate the Earth observation market.

SAR imagery enables the capturing of features in the Earth's surface under various atmospheric conditions, including cloud cover and other visual obstructions not pen-
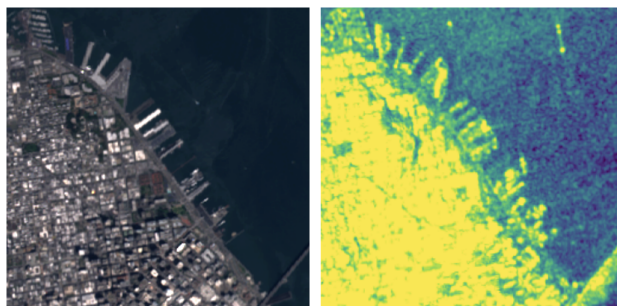

Figure 1: San Francisco from the SEN12MS-CR dataset (Left: Sentinel-2 RGB, Right: Sentinel-1 SAR Band 1)

etrable by standard spectral imaging techniques. SAR even works at night, making it a powerful technology capable of Earth observation under any condition. Despite these advantages, SAR images by themselves can sometimes be hard to analyze, and providing better interpretability through fusion of SAR and color imagery can be useful in examining a particular scene.

Our project focuses on improving the spectral-based baselines for SAR colorization presented in Shen *et al* [14] through terrain classification and segmentation, showing a quantitative improvement in colorization accuracy metrics compared with baseline methods, which leads to a qualitative improvement in the visual realism of the colorized images. The input to our method is a SAR satellite image. We then use an ensemble of colorization models (cGAN, CNN, NL, LR) to output a colorized SAR image.

2. Related Work

Acquiring simultaneous SAR and clear optical images can often prove difficult due to varying conditions on the surface [10]. An alternative approach is to train a model that learns to automatically colorize SAR images from readily available SAR and optical image data pairs, providing a way to augment SAR images with colorization techniques that simulate the appearance of natural colors, improving their utility and interpretability.

One such approach is to directly generate an optical im-

```
Driver: GTiff/GeoTIFF
Files: ROIs1868_summer_s1_123_p837.tif
Size is 256, 256
Coordinate System is:
PROJCRS["WGS 84 / UTM zone 10N",
[...]
        AREA["Between 126°W and 120°W, northern hemisphere between
equator and 84°N, onshore and offshore. Canada — British Columbia (BC);
Northwest Territories (NWT); Nunavut; Yukon. United States (USA) —
Alaska (AK)."],
[...]
Corner Coordinates:
Upper Left  (  551726.633, 4185072.779) (122d24'44.44"W, 37d48'41.70"N)
Lower Left  (  551726.633, 4182512.779) (122d24'45.09"W, 37d47'18.64"N)
Upper Right (  554286.633, 4185072.779) (122d22'59.74"W, 37d48'41.17"N)
Lower Right (  554286.633, 4182512.779) (122d23' 0.43"W, 37d47'18.10"N)
Center      (  553006.633, 4183792.779) (122d23'52.43"W, 37d47'59.91"N)
Band 1 Block=256x4 Type=Float32, ColorInterp=Gray
Band 2 Block=256x4 Type=Float32, ColorInterp=Undefined
```

Figure 2: Truncated metadata extracted from `gdalinfo`

Dataset	Info	Training	Testing
SEN12MS-CR	Patches	471	250
	SAR Size	256x256x1	256x256x1
	MS Size	256x256x3	256x256x3
	GT Size	256x256x3	256x256x3

Table 1: Dataset Summary of the SEN12MS-CR Subset

3. Data

We use a subset of the SEN12MS-CR dataset [2], a dataset covering 175 global distributed regions of interest in the world for multi-modal cloud removal. Our subset focuses on areas near the San Francisco Bay Area, containing observations conducted in the Summer of 2018. The dataset consists of a 2-band Sentinel-1 radar measurement (SAR), and corresponding multi-spectral Sentinel-2 observations (MS), which include those in the visible spectrum. For training spectral models, a separate pre-defined set of 20 SAR image pairs are used from the same dataset (Seed 1158 / Spring).

A total of 721 256×256 Sentinel-1 and 2 patches cover the North Bay, northern San Francisco and parts of Oakland. This region offers a diverse topography that includes mountains, hills, urban areas, and water bodies such as the Bay and the Pacific Ocean. This variety enables us to validate our methods across different terrains and environments. One patch of corresponding MS (Sentinel-2) and SAR (Sentinel-1) images are shown in Figure 1. Patches also include metadata (Figure 2), with supplemental information on location, exact corner coordinates and a text description.

For Sentinel-1 patches, we extract out Band 1 (polarimetric VV) for use as our SAR image. For Sentinel-2 patches, bands 4, 3 and 2, corresponding to R, G, B respectively are used for the multispectral (MS) image. The raw Sentinel-1 and Sentinel-2 images consist of unnormalized sensor values. To preprocess the SAR images, we apply the following equation to scale the values to a range between 0 and k:

$$\mathbf{SAR}_{adj} = \frac{\mathbf{SAR} - min(\mathbf{SAR})}{max(\mathbf{SAR} - min(\mathbf{SAR}))} \cdot k$$

$k = 2047$ when the image is input to the colorization network, and $k = 255$ when the image is visualized. Visualization of Sentinel-2 images are also similarly conducted by setting $k = 255$.

4. Methods

To enhance the results beyond the baseline, we incorporate context from the image. We start by classifying the

age from SAR data. This is called SAR-to-optical image translation [4, 8, 9, 7]. Unfortunately, this approach removes any SAR specific radar-specific effects, which may be extremely useful for terrain or scene analysis [11]. For example, speckle patterns allow for the determination of surface texture, all of which would be removed in the translation process of this approach.

Our chosen approach is creating an intermediate ground truth that preserves the characteristics of a SAR image, while also enabling coloring the image for better interpretability [13, 14]. This retains the textural and intensity details of the SAR image, while providing a more accurate and realistic color representation of the area. This is done by merging the SAR bands with the optical RGB bands to form a fused image. This method is called Fast IHS (Intensity-Hue-Saturation) [15], and is described using the following process (as described in Shen *et al.* [14]):

$$\mathbf{I} = T(\mathbf{MS})$$
$$\mathbf{SAR'} = (\mathbf{SAR} - \mu_{\mathbf{SAR}}) \cdot \sigma_{\mathbf{I}}/\sigma_{\mathbf{SAR}} + \mu_{\mathbf{I}}$$
$$\mathbf{D} = \mathbf{SAR'} - \mathbf{I}$$
$$\mathbf{GT} = \mathbf{MS} + \mathbf{D}$$

MS is the multispectral (RGB) image, **I** is the extracted average intensity component, **SAR** is the SAR image and **GT** is the ground truth (fused) image. An example of a generated **GT** image and its corresponding **MS** image is shown in Figure 4 (top-right and top-left, respectively).

The ground truth can be generated from the underlying SAR image using a variety of methods. Schmitt *et al.* [13] introduces a deep learning SAR colorization method using a combination of a variational autoencoder (VAE) [5] and mixed density network (MDN) [1]. Shen *et al.* [14] formalizes the evaluation of these models by introducing multiple evaluation metrics (NRMSE, SAM, Q4) and additional deep learning methods which use convolutional neural networks (CNN) and conditional generative adversarial networks (cGAN) [12] similar to Isola *et al.* [3]

image into its specific category. Following this, we apply recolorization using each available method. The ideal recolorization technique is then selected based on the image's classification.

4.1. SAR Data Classification

The first step in this process is to classify the SAR images into one of several categories based on the geography and scope of the image. We chose to create five distinct categories that could sufficiently summarize all the SAR images of the Bay Area in the training set: mountainous areas, urban areas, bodies of water, mix of urban areas with water, and mix of mountainous areas with water. For the classification task, we utilized GPT-4o [1], OpenAI's latest flagship model, because it can reason through vision tasks and is able to understand the basics of SAR imagery structure. In the prompt, our main instruction is to "Please classify the following SAR satellite image into one of the following categories: Mountain, Urban, Water, Water-Urban or Water-Mountain," and we provide further context and instruction on what features to focus on when deciding a category.

We found that GPT-4o did not fully understand how to interpret SAR imagery and often misclassified the water as an urban area or mountainous area. We improved the results by adding additional instructions like the following: "You often misclassify water. When the image is featureless and relatively uniformly dark and hazy, make sure your response is not Urban or Mountain but instead the correct answer of Water." Adding details on what features to pay attention to helped the model to improve its accuracy on the training images. After classifying the entire training set, we visually inspected a random sample of 100 of the SAR images and found 6 misclassifications. Thus, For the entire training set, we expect to see about 94% of the images to be classified correctly and 6% to be misclassified.

4.2. Recolorization Models

For spectral models, we trained our models using the same methodology specified by Shen et al. [14]. A separate subset of the SEN12MS-CR dataset [2] is used for training the spectral models [2]. For non-spectral models, we utilized the pretrained models provided by the authors to conduct SAR colorization. We evaluated five different methods for colorization:

1. **NoColSAR** This method simply replicates the SAR band over all three RGB channels. It served as a baseline in Shen *et al.* [14] However, it did not provide any benefit in our experiments and was therefore excluded from our final evaluations.

(a) Mountain Class Examples

(b) Urban Class Examples

(c) Water Class Examples

(d) Water-Urban Class Examples
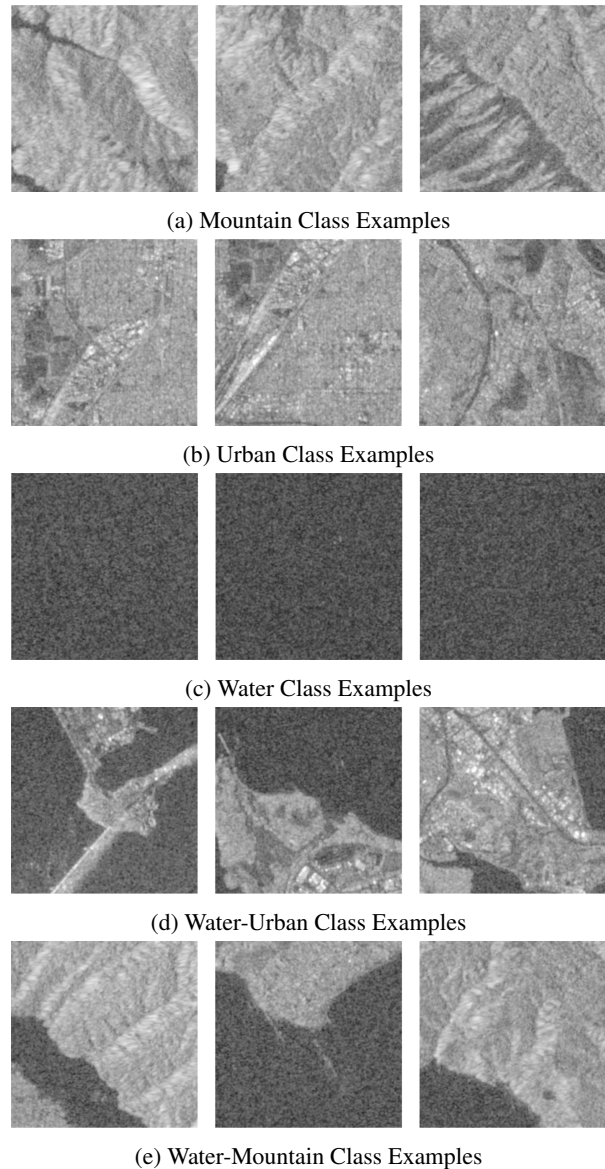

(e) Water-Mountain Class Examples

Figure 3: Terrain Class Examples

2. **LR4ColSAR** (**LR**) This spectral-based method employs linear regression. The SAR image is flattened and modeled using a linear regression approach, with weights and biases optimized using a mean squared error (MSE) loss function.

3. **NL4ColSAR** (**NL**) This is the second spectral-based method, incorporating a nonlinear element by implementing a single hidden layer neural network. The nonlinearity used in this model is a tan-sigmoid function as shown in Equation 1, using the Levenberg-
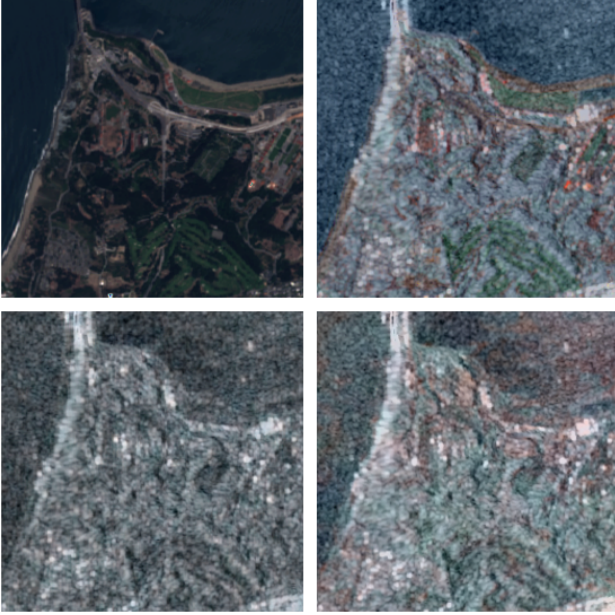
Figure 4: Model SAR colorization output examples (Top-left: RGB, Top-right: Ground Truth, Bottom-left: CNN4ColSAR, Bottom-right: cGAN4ColSAR)

Marquardt algorithm with a l_2-norm loss function.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \qquad (1)$$

4. **CNN4ColSAR (CNN)** We use the pretrained model [3] provided by Shen *et al*. [14], and specific details regarding parameters and setup are detailed in the paper. CNN4ColSAR utilizes a convolutional neural network with four convolutional layers having kernel sizes of 9, 5, 1, and 5, and numbers of kernels 64, 32, 32, and 3, respectively.

5. **cGAN4ColSAR (cGAN)** Like CNN4ColSAR, we used a pretrained model. The generator in the conditional GAN consists of an 8-layer architecture using 4×4 convolutional filters with a stride of 2. The number of kernels begins at 64 and doubles at each subsequent layer, up to a maximum of 512.

4.3. Evaluation Metrics

We evaluate our model using the metrics shown in Shen *et al*. [14] and also defined below; normalized root mean square (NRMSE), spectral angle mapper (SAM), and multi-dimensional extended version of the universal image quality index (Q4).

$$NRMSE = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{RMSE(n)}{\mu(n)}\right)$$

$$SAM(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{<\mathbf{x}, \mathbf{y}>}{||\mathbf{x}||_2 \cdot ||\mathbf{y}||_2}\right)$$

$$Q4 = \frac{4|\sigma_{z_1 z_2}| \cdot |\mu_{z_1}| \cdot |\mu_{z_2}|}{(\sigma_{z_1}^2 + \sigma_{z_2}^2)(\mu_{z_1}^2 + \mu_{z_2}^2)}$$

Each metric captures specific aspects of the colorization. Optimizing against a particular metric will depend on the application. For our experiments we evaluate and report results for each metric individually.

- **NRMSE** focuses on the individual pixel differences, looking at errors between pixel values

- **SAM** assesses the spectral angle between two bands, emphasizing spectral fidelity in images

- **Q4** evaluates structural and perceptual similarity between two images

5. Experiments

We conduct a series of experiments to demonstrate how our approach improves the recolorization process for SAR satellite images. We have two main steps in our approach, the train step and the test step. In the train step, the goal is to understand which model performs the best for each type of terrain. In our test set, we first classify each image into a type of terrain, and then use the best corresponding colorization model for that type of terrain.

The data for both steps is sourced from a subset of the SEN12MS-CR dataset, which focuses on the San Francisco Bay Area. We split the images into a training set and a test set with a 65:35 split, with roughly 470 images in the training set and 250 images in the test set.

5.1. Step One: Train

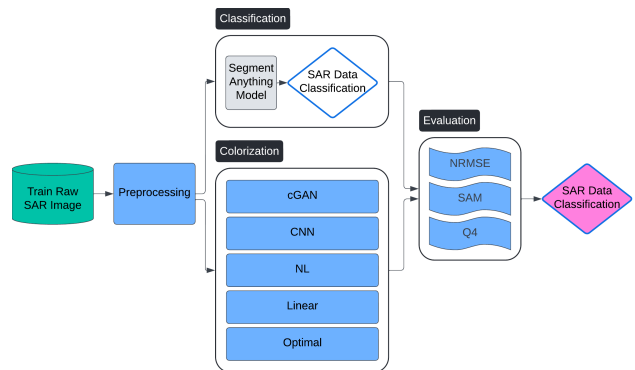Our train step can be summarized by the following pipeline.

Figure 5: Train Step Pipeline

We first begin the with raw SAR image data files, and preprocess each of the roughly 470 images to use in our experiments. The first step is classification, using the method we describe in 4.1 to assign categories to each SAR image.

Terrain Class	Count
Mountain	22
Urban	53
Water	248
Water-Urban	31
Water-Mountain	117

Table 2: Training Data Terrain Counts

This table summarizes the counts of each type of terrain in our training data. Because the data focuses on the Bay Area, the majority of the images are of the Water and Water-Mountain terrain types.

Independently of the classification process, we run the four colorization models (cGAN, CNN, NL and LR) described in 4.2 on each of the SAR images. Finally, we score all the outputs using the three evaluation metrics described in 4.3, paying special attention on how each model performs on the different terrain classes. The evaluation results are outlined in Table 6, where the best model performance for each terrain class and metric is bolded.

Looking at the table, we see that for the Mountain class, the cGAN model earns the best score using the NRMSE and Q4 metrics, while the CNN performs slightly better than the cGAN for the SAM metric. Overall, the cGAN colorization model performs the best for the Mountain class, so we assign that model to the class. The CNN model performs the best across all three metrics for the Urban class, so we assign that model to the class. The LR model performs the best using the SAM and Q4 metrics, while cGAN is slightly better using the NRMSE metric. Overall, the LR model performs the best and we assign it to the Water terrain. For Water-Urban and Water-Mountain, the cGAN performs the best for all three metrics, so we assign that model to the classes. The best overall performing models for each terrain class are summarized in the following table.

Terrain Class	Optimal Model
Mountain	cGAN
Urban	CNN
Water	LR
Water-Urban	cGAN
Water-Mountain	cGAN
All	cGAN

Table 3: Optimal Colorization Model for each Terrain Class

5.2. Step Two: Test

Now that we understand which models are optimal for each type of terrain class, we can test our methodology using the following pipeline.
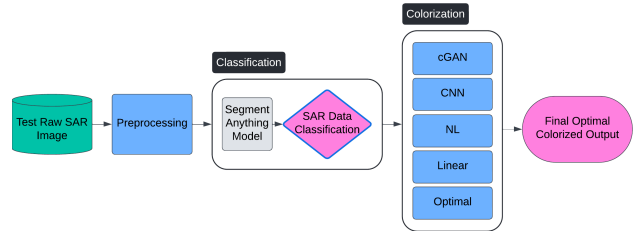

Figure 6: Test Step Pipeline

We begin with the raw SAR image data files within the test set, and preprocess the images like with the train images. Next, using the same terrain classification method described in 4.1, we classify each image, and the counts of each class are summarized in the following table.

Terrain Class	Count
Mountain	37
Urban	2
Water	154
Water-Urban	30
Water-Mountain	27

Table 4: Training Data Terrain Counts

Like with the training set, the most common terrain class is Water. However, compared to the training set, the test set has more Water-Mountain and Mountain cases, a similar amount of Water-Urban cases, and much fewer Urban cases.

Next, we color the SAR image using the optimal model that is assigned to that SAR image's terrain class. We then evaluate the resulting colored SAR image using the three metrics defined in 4.3. Finally, we calculate the average score for each evaluation metric, and compare to the cGAN baseline metrics for the test set. We compare against the cGAN model as a baseline because it was the optimal model identified by Shen et al. [14]. The evaluation results are summarized in the following table.

Method	RMSE	SAM	Q4
cGAN [14]	0.350	9.511	0.424
Terrain-based (ours)	**0.239**	**6.256**	**0.441**
Ideal value	0	0	1

Table 5: SAR image colorization metrics on the test set

Through our terrain-based approach, we achieve better performance than the cGAN method on all three metrics. Our RMSE and SAM scores are significantly lower than the baseline, and our Q4 metric also achieves greater performance than the baseline.

Qualitatively, the models assigned to each terrain class also visually look the most accurate to the ground truth. In Figure 4, the image is of a region that contains both water and urban areas, and is classified as Water-Urban. The model associated with this terrain class is the cGAN model, and visually the cGAN output does produce a closer result to the ground truth than the CNN output.
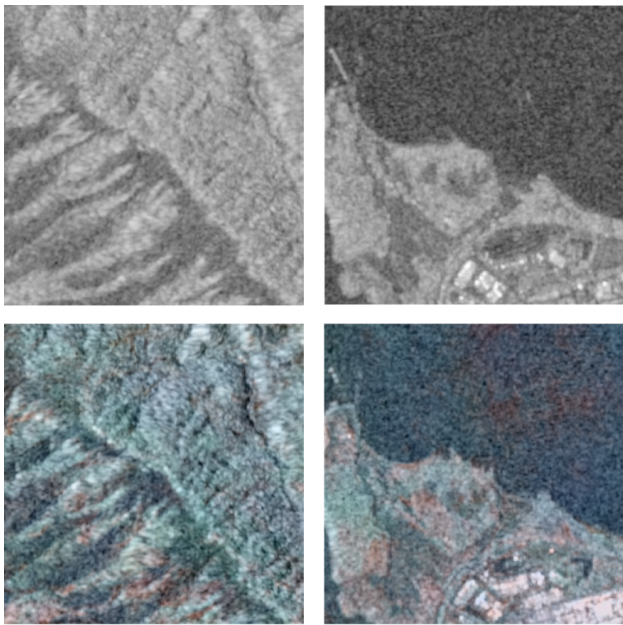

Figure 7: cGAN Colorized Outputs
(Top Left: Mountain Class SAR Image, Top Right: Water-Urban Class SAR Image, Bottom Left: Mountain Class cGAN Colorized Image, Bottom Right: Water-Urban Class cGAN Colorized Image)

Here are a few more cGAN colorized output examples for the Mountain class and Water-Urban class. For both of these classes, the cGAN model performs the best across almost all the metrics. And, both of the cGAN colorized outputs look very close to the baseline ground truth, and visually look very realistic to how optical satellite imagery might appear for these same regions.

6. Conclusion

The terrain-based approach presented in this paper demonstrates a significant improvement in recolorization accuracy for SAR satellite images comparing to the cGAN baseline model. By classifying images into distinct terrain classes (Mountain, Urban, Water, Water-Urban, and Water-Mountain) and assigning the optimal colorization model for each class, we achieve superior results across all three evaluation metrics: RMSE, SAM, Q4. This suggests that a one-size-fits-all model may not be the most effective approach for SAR image recolorization. The terrain-based approach allows for greater customization and flexibility, resulting in more accurate and visually appealing colorized images.

6.1. Future Work

We see several avenues to improve the results in this paper. First, roughly 94% of the images were classified correctly by the GPT-4o model, which means there is room for improvement for the classification process. Further, we chose the five classes based on visual inspection of the training set, and while these terrain classes work well for the San Francisco Bay Area subset, they will not work in other regions. To improve the classification process, we are interested in fine-tuning a model to augment the process.

Second, there are some classes with a mix of terrains, like the Water-Urban and Water-Mountain classes. We are interested in how segmenting these images into their specific terrains, and recolor based on the masks of the terrains, can improve the results.

Figure 8: Segment Anything Model output
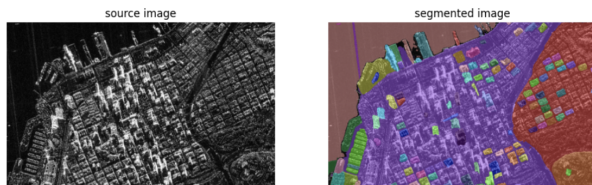(Left: cGAN output, Right: Segmentation output)

Figure 9: Visualization of Satellite Imagery Segmentation
(Left: Original SAR Image, Right: Segmentation output)

Here, we began experimenting with how we could most efficiently segment the SAR images into their respective terrains using Meta's Segment Anything model[6]. For example, both figures above would be classified as Water-Urban,

Terrain Class	Metric	cGAN	CNN	NL	LR	Ideal
	NRMSE	**0.247**	0.324	0.261	0.254	0
Mountain	SAM	7.764	**7.578**	8.810	7.616	0
	Q4	**0.763**	0.698	0.698	0.684	1
	NRMSE	0.241	**0.205**	0.402	0.423	0
Urban	SAM	6.103	**5.795**	16.401	17.529	0
	Q4	0.728	**0.720**	0.721	0.667	1
	NRMSE	**0.226**	0.231	0.285	0.230	0
Water	SAM	5.104	4.990	7.899	**4.864**	0
	Q4	0.306	0.272	0.304	**0.391**	1
	NRMSE	**0.206**	0.225	0.349	0.345	0
Water-Urban	SAM	**5.832**	5.922	11.203	11.250	0
	Q4	**0.778**	0.739	0.716	0.753	1
	NRMSE	**0.204**	0.226	0.258	0.236	0
Water-Mountain	SAM	**5.902**	6.041	7.602	6.199	0
	Q4	**0.752**	0.694	0.689	0.738	1
	NRMSE	**0.235**	0.277	0.286	0.269	0
All	SAM	6.689	**6.560**	9.297	7.928	0
	Q4	**0.649**	0.599	0.603	0.621	1

Table 6: Metrics for training conducted on the SEN12MS-CR Subset

so we could produce masks that distinguish between the Water and Urban regions of these images.

Additionally, evaluating the approach on a wider range of dataset will ensure its robustness and generalizability across different geographic locations and image acquisition conditions. Finally, applying this approach to real-world scenarios like environmental monitoring will demonstrate its practical value and potential impact.

7. Contributions & Acknowledgements

- **Sukrut Oak** worked on the terrain classification process and produced final graphics and visualizations, and experimented with Segment Anything.

- **Young Chol Song** worked on training and running the SAR recolorization methods described in this report and generated visualizations and final metrics.

- **Da Sun** worked on implementing and running the Spectral-based baselines for SAR colorization presented in the paper[14] for comparative analysis.

References

[1] C. M. Bishop. Mixture density networks. 1994.

[2] P. Ebel, A. Meraner, M. Schmitt, and X. X. Zhu. Multisensor Data Fusion for Cloud Removal in Global and All-Season Sentinel-2 Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[4] G. Ji, Z. Wang, L. Zhou, Y. Xia, S. Zhong, and S. Gong. Sar image colorization using multidomain cycle-consistency generative adversarial network. *IEEE Geoscience and Remote Sensing Letters*, 18(2):296–300, 2021.

[5] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[6] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[7] W. Ku and D. Chun. The method for colorizing sar images of kompsat-5 using cycle gan with multi-scale discriminators. *Korean Journal of Remote Sensing*, 34(6_3):1415–1425, 2018.

[8] J.-H. Lee, K. Kim, and J.-H. Kim. Design of cyclegan model for sar image colorization. In *2021 IEEE VTS 17th Asia Pacific Wireless Communications Symposium (APWCS)*, pages 1–5, 2021.

[9] S.-Y. Lee and D.-W. Chung. Labeling dataset based colorization of sar images using cycle gan. *The Journal of Korean Institute of Electromagnetic Engineering and Science*, 33:776–783, 10 2022.

[10] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone. Haze correction for contrast-based multispectral pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2255–2259, 2017.

[11] G. Macelloni, G. Nesti, P. Pampaloni, S. Sigismondi, D. Tarchi, and S. Lolli. Experimental validation of surface scattering and emission models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(1):459–469, 2000.

[12] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[13] M. Schmitt, L. Hughes, M. Körner, and X. X. Zhu. Colorizing sentinel-1 sar images using a variational autoencoder conditioned on sentinel-2 imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:1045–1051, 2018.

[14] K. Shen, G. Vivone, X. Yang, S. Lolli, and M. Schmitt. A benchmarking protocol for sar colorization: From regression to deep learning approaches, 2023.

[15] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang. A fast intensity-hue-saturation fusion technique with spectral adjustment for ikonos imagery. *IEEE Geoscience and Remote sensing letters*, 1(4):309–312, 2004.