DISC-GAN: Disentangling Style and Content for Cluster-Specific Synthetic Underwater Image Generation

Sneha Varur Anirudh R Hanchinamani Tarun S Bagewadi Uma Mudenagudi
Chaitra D Desai Sujata C Padmashree Desai Sumit Meharwade
KLE Technological University, Hubballi, Karnataka, India
{ sneha.varur, 01fe22bcs264, 01fe22bcs151, uma}@kletech.ac.in
{ chaitra.desai, sujata_c, padmashri, 01fe21bci032}@kletech.ac.in

Abstract

In this paper, we propose a novel framework, Disentangled Style-Content GAN (DISC-GAN), which integrates style-content disentanglement with a cluster-specific training strategy towards photorealistic underwater image synthesis. The quality of synthetic underwater images is challenged by optical distortions due to phenomena such as color attenuation and turbidity. These phenomena are represented by distinct stylistic variations across different waterbodies, such as changes in tint and haze. While generative models are well-suited to capture complex patterns, they often lack the ability to model the non-uniform stylistic conditions of diverse underwater environments. To address these challenges, we employ K-means clustering to partition a dataset into style-specific domains. We use separate encoders to get latent spaces for style and content; we further integrate these latent representations via Adaptive Instance Normalization (AdaIN) and decode the result to produce the final synthetic image. The model is trained independently on each style cluster to preserve domain-specific characteristics. Our framework demonstrates state-of-the-art performance, obtaining a Structural Similarity Index (SSIM) of 0.9012, an average Peak Signal-to-Noise Ratio (PSNR) of 32.5118 dB, and a Fréchet Inception Distance (FID) of 13.3728.

1. Introduction

In this paper, we propose a novel generative framework for underwater image synthesis, termed as Disentangled Style-Content GAN (DISC-GAN). The aim is to generate photorealistic underwater images by modeling the diverse optical conditions of different waterbodies. Synthesizing underwater images is challenging because of the complex optical interactions that take place during image formation. Unlike atmospheric environments, underwater scenes are affected by severe, depth-dependent light attenuation and absorption, which cause a rapid loss of longer wavelengths and result in a dominant blue-green color distortion [5, 6, 16]. Simultaneously, suspended particles scatter light in forward and backward directions, creating image blur and a veiling haze that reduces contrast and obscures distant objects [1, 2]. These degradations affect the visual quality of images and limit the performance of tasks such as marine monitoring, exploration, and robotics.

Researchers in this domain address underwater synthesis and restoration in two ways: physics-based and learning-based. Physics-based models, for instance, have proposed using monocular depth estimation as a prior to guide restoration [8, 28], fusion-based techniques using color and contrast priors [2], or models based on wavelength compensation and dehazing [1, 5, 6]. However, these methods often rely on hand-crafted features, which limits their generalizability across diverse underwater scenes.

Alternatively, learning-based methods train neural networks on real-world or synthetic datasets to directly map between domains. Methods such as WaterGAN [19] demonstrated how to generate synthetic data by simulating light attenuation on in-air RGB-D pairs, while other generative approaches have used variational models [7] or cycle-consistency loss for unpaired enhancement [10, 29]. An extensive body of work has explored GANs for this task, focusing on multi-domain translation [3], texture consistency [23], perceptual losses [25], and lightweight networks for real-time performance [14, 24]. However, many of these methods do not explicitly model the fine-grained stylistic differences between waterbody types, making them less controllable and limiting their realism for domain-specific applications.

Hybrid methods and advanced GAN architectures have attempted to merge physical priors with learning-based techniques to address these limitations. Style transfer using Adaptive Instance Normalization (AdaIN) [13] and percep-

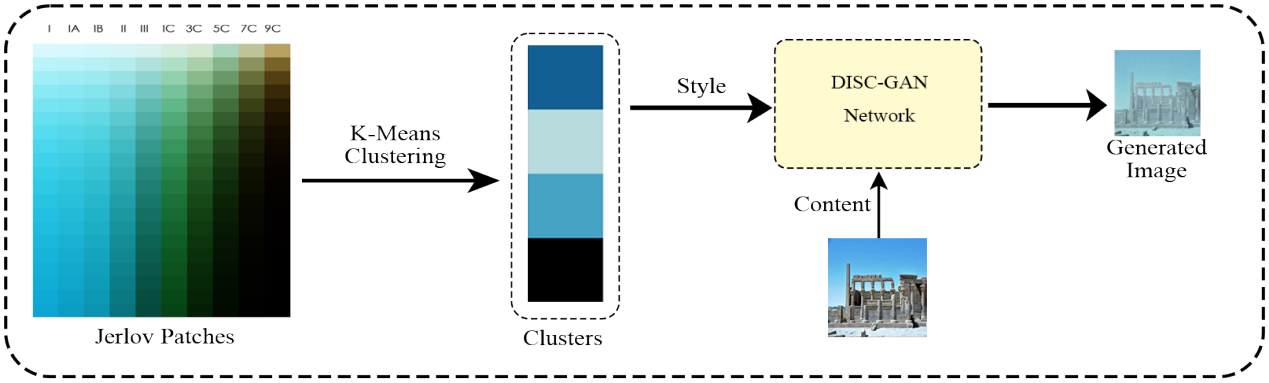


Figure 1. The high-level design of our proposed DISC-GAN framework. DISC-GAN is trained using a clusters obtained from the dataset composed of Jerlov water image patches, which were introduced by [9]

tual losses [17] have enabled greater structural preservation. Disentangled GANs, such as the architecture proposed by Kazemi et al. [18] and others [27], have shown how to learn latent representations of content and style independently. Recent works have also incorporated depth-guidance [4, 8], camera-awareness [22], or dual-domain learning [19, 21] to improve results. However, the explicit partitioning of the underwater domain into distinct stylistic clusters, inspired by physical water characteristics, remains underexplored within these learning frameworks.

Towards this, we propose DISC-GAN, a framework that includes a style clustering module to partition the underwater domain before training a style-content disentangled GAN. Our work adopts the Realistic Synthetic Underwater Image Generation with Image Formation Model (RSUIGM) dataset [9], which provides physically accurate stylistic variations across different Jerlov water types [16]. The usage of this dataset ensures that our learned style clusters are grounded in real-world optical properties. The proposed two-phase pipeline is shown in Figure 1. We combine physics-informed data partitioning with a learning-based synthesis model and make the following contributions:

- We propose a method to partition underwater images into distinct style domains via K-means clustering on a feature space combining color histograms and mean depth values, inspired by Jerlov water types.
- We introduce a novel, cluster-specific synthesis framework that generates realistic underwater images by training a separate GAN on each style domain to prevent style leakage.
- We develop an effective style-content disentangled GAN, which uses AdaIN to preserve content structure from terrestrial images while transferring style from the target underwater domain, validated by high SSIM, PSNR and low FID scores.

In Section II, we describe our synthesis framework, outlining each component from style clustering to image synthesis. Section III provides the implementation details of our model. In Section IV, we discuss the results of our experiments. Finally, Section V concludes the paper and suggests future work.

2. Style Clustering and Image Synthesis

In this section, we propose a novel framework, Disentangled Style-Content GAN (DISC-GAN), for cluster-specific underwater image generation that integrates domain partitioning with a deep learning model. The synthesis framework has three modules corresponding to three phases:

- dataset preprocessing and style clustering,
- style-content feature disentanglement and fusion, and
- learning-based synthesis,

as shown in Figure 6. In Section 2.1, we explain the datasets and our preprocessing pipeline. In Section 2.2, we detail the methodology for partitioning the dataset into distinct style clusters. In Section 2.3, we focus on the disentangled synthesis architecture and the composite loss function. Finally, in Section 2.4, we provide an overview of the training process.

2.1. Dataset and Preprocessing

The Parameter Estimator module is designed to predict the parameters that govern underwater image formation, enabling physics-informed underwater image restoration. For style modeling and training, we utilize the RSUIGM (Realistic Synthetic Underwater Image Generation Model)

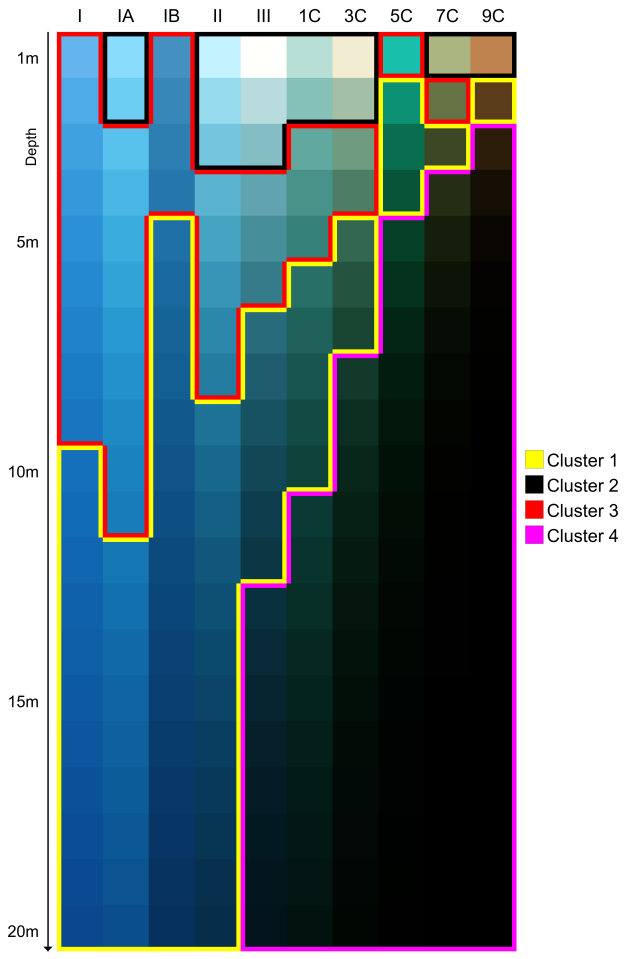


Figure 2. Example patches from the RSUIGM dataset used for style clustering.

dataset [9], a synthetic benchmark generated by simulating underwater image formation under various water conditions. Its consistency with a physically grounded image formation model makes it ideal for training our style-content disentanglement framework. Each synthetically degraded image is paired with a clean reference and a corresponding depth map, allowing for supervised learning. In our pipeline, clean images serve as content inputs, while the distorted underwater versions are used to model style vectors.

For training our proposed DISC-GAN framework, we specifically employ the RSUIGM dataset due to its physically inspired modeling of underwater light propagation. RSUIGM is generated using an image formation model that accounts for both downwelling depth (z) and line-of-sight distance (d), thereby incorporating realistic spatial variations in light attenuation and scattering. The dataset consists of 6000 synthetic underwater images rendered across diverse ocean bed and coastal bed scenes, spanning the full range of Jerlov water types. These images simulate characteristic degradations such as haze, blur, color attenuation, and tint, making RSUIGM particularly suitable for training data-driven generative models. In our work, RSUIGM serves as the ground-truth reference for guiding DISC-GAN synthesis and ensuring that generated images retain the statistical and perceptual characteristics of real underwater imagery.

In addition to RSUIGM, we use the SUID (Standard Underwater Image Dataset) as a source of clean terrestrial images for content extraction. The dataset provides high-resolution images with diverse structural features, which helps the content encoder generalize across complex scene geometries. To ensure consistency, all input images are resized to a 256x256 resolution and augmented via horizontal flipping. Corresponding depth maps are downsampled to match, and RGB histograms are extracted from the RSUIGM style images for the clustering phase. For training and validation, the dataset is split into a 80:20 ratio.

2.2. Style Domain Clustering

To model the diverse appearances of underwater scenes, we first partition the RSUIGM dataset into visually coherent style domains. This process is designed to group images based on both color and physical depth properties, inspired by the classification of oceanic water into Jerlov water types [16]. The physical basis for this clustering comes from the RSUIGM image formation model [9], which simulates degradation using the Beer-Lambert law:

$$I_c(x) = J_c(x) \cdot e^{-K_c \cdot d(x)} + B_c \cdot \left(1 - e^{-K_c \cdot d(x)}\right) \quad (1)$$

Here, $I_c(x)$ is the observed intensity, $J_c(x)$ is the clean radiance, K_c is the attenuation coefficient, $d(x)$ is depth, and B_c is background light. To capture these degradation patterns, each image is converted into a feature vector composed of its RGB histogram and normalized mean depth.

We then apply K-means clustering to partition the data by minimizing the intra-cluster variance according to the objective:

$$\arg\min_{C} \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

where C_i is the i^{th} cluster and μ_i is its centroid. The optimal number of clusters was determined to be $k = 4$ using the Elbow Method. This yields four dominant underwater styles: **blue**, **light-blue**, **dark-blue**, and **black**, as shown in Figure 3. The clear separation of these clusters in 3D RGB space, shown in Figure 4, validates that each group represents a distinct and visually coherent style domain.

To visually validate this choice, we compared the clustering results for $k \in \{3, 4, 5, 6\}$, as shown in Figure 4.

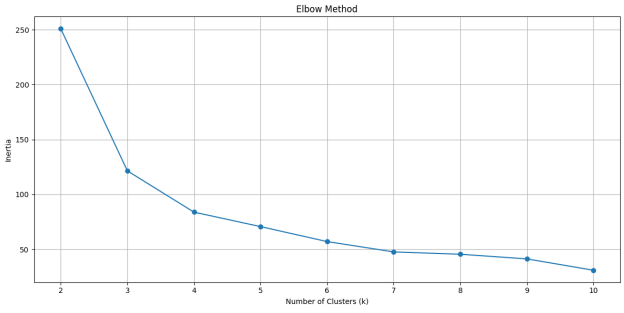


Figure 3. The Elbow Method plot for K-means clustering on the 200 classes of Jerlov considering the rbg values for each class. The "elbow" point at k=4 suggests it is the optimal number of clusters.

This comparison includes both the 3D RGB space visualization and the corresponding clustered image patches for each value of k. For $k = 3$, distinct style domains appear to be merged into single, less coherent groups. Conversely, for $k = 5$ and $k = 6$, visually similar styles are unnecessarily split, leading to over-segmentation and redundancy. The configuration for $k = 4$ (Figures 4(c) and 4(d)) provides the most meaningful separation, yielding four dominant and visually distinct underwater styles, which we label as **blue**, **light-blue**, **dark-blue**, and **black**. The clear separation of these clusters in 3D RGB space validates that each group represents a distinct and coherent style domain.

2.3. Disentangled Synthesis Framework

The final stage of the framework is the synthesis pipeline, which transforms latent style and content features into a photorealistic image, drawing inspiration from the original SC-GAN architecture [18]. Given a clean content image I_c and a style reference I_s from a target cluster, the framework first uses separate encoders. The content encoder $\mathcal{E}_{content}$ produces a latent tensor z_c:

$$z_c = \mathcal{E}_{content}(I_c) \tag{3}$$

Simultaneously, the style encoder $\mathcal{E}_{style}$ extracts a style vector z_s:

$$z_s = \mathcal{E}_{style}(I_s) \tag{4}$$

These features are fused within the generator G using Adaptive Instance Normalization (AdaIN) [13], which aligns the feature statistics to inject the style without altering content structure. The generator then synthesizes the final stylized output $\hat{I}$:

$$\hat{I} = G(z_c, z_s) \tag{5}$$

The operation yields a final stylized image $\hat{I}$ reconstructed from disentangled and physics-guided latent features. To ensure the generator produces high-fidelity outputs, we use a composite loss function that combines L1 reconstruction loss with an adversarial loss. The final generator loss $\mathcal{L}_G$ is given by:

$$\mathcal{L}_G = \mathcal{L}_{L1(\hat{y},y)} + \lambda \cdot \mathcal{L}_2 \tag{6}$$

Here, the adversarial term $\mathcal{L}_{G_{\text{GAN}}}$ enforces perceptual realism, while the L1 term $\mathcal{L}_{G_{\text{L1}}}$ encourages structural preservation. The hyperparameter λ balances these two objectives. This composite loss approach follows standard practice in conditional image synthesis, where an adversarial loss improves visual realism [12] and a pixel-wise loss like L1 maintains structural fidelity to the target [15].

2.4. Training Strategy

The training process of the model happens in two phases, as outlined in Algorithm 1.

1. **Style Domain Partitioning:** The RSUIGM dataset is first partitioned into four distinct style clusters using the K-means algorithm based on color and depth features. This ensures that the model can learn waterbody-specific styles in a controlled manner.
2. **Cluster-Specific GAN Training:** The generator and discriminator are then trained in an alternating fashion. For each training step, a content image from SUID and a style reference from one of the four RSUIGM clusters are used. The composite loss is back-propagated to update the generator and discriminator. The entire GAN is trained independently for each of the four style clusters to prevent style leakage and enhance control. This sequential procedure ensures that the synthesis model benefits from a stable and meaningful feature space shaped by the physics-informed clustering.

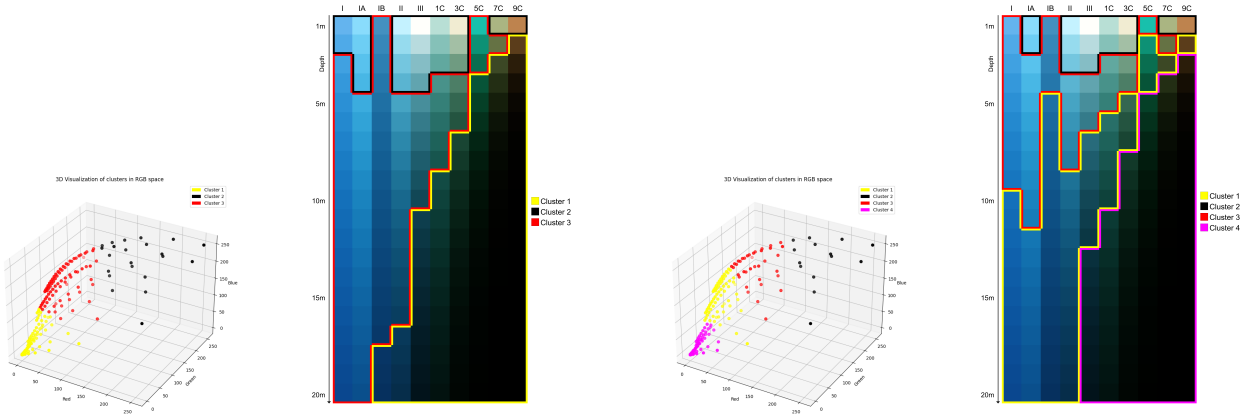
3. Implementation Details

The end-to-end architecture of the proposed Disentangled Style-Content GAN (DISC-GAN) is illustrated in Figure 6. The framework is composed of three primary modules: a content encoder, a style encoder, and a generator with a corresponding discriminator. These components are designed to learn disentangled mappings for controlled and realistic underwater image synthesis.

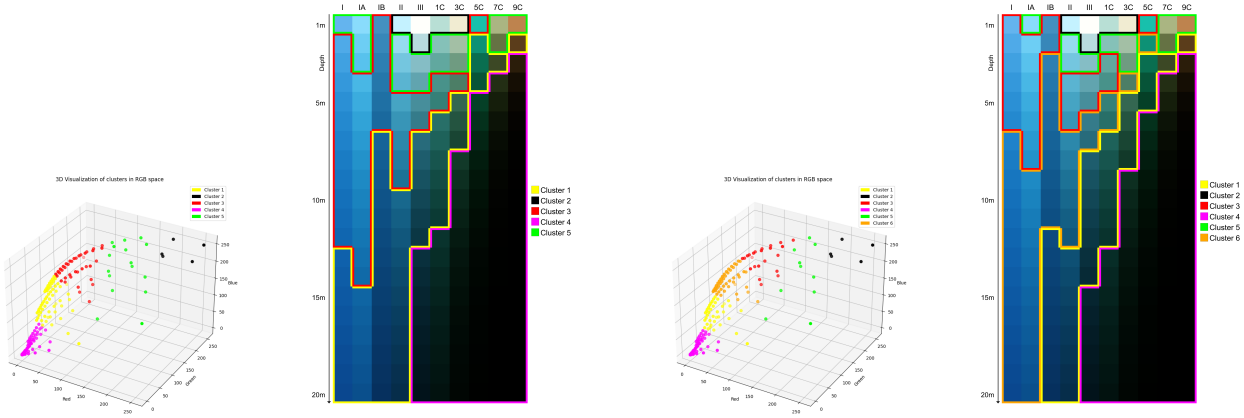
3.1. Network Architecture

Our framework utilizes encoders built upon the VGG19 network [20], pretrained on ImageNet, to extract high-quality content and style representations.

Content and Style Encoders. To achieve disentanglement, feature extraction is stratified. The content encoder uses the deeper `relu4_2` layer of the VGG19 network to capture high-level semantic and structural information from the clean input image [11, 17]. In contrast, the style encoder is designed to extract global appearance statistics. It computes Gram matrices from the feature maps of shallower layers (`relu1_1`, `relu2_1`, `relu3_1`) to effec-



(a) 3D, k=3 (b) Patches, k=3 (c) 3D, k=4 (Optimal) (d) Patches, k=4 (Optimal)



(e) 3D, k=5 (f) Patches, k=5 (g) 3D, k=6 (h) Patches, k=6

Figure 4. Visual comparison of K-means clustering results for k=3, 4, 5, and 6. The top row compares k=3 and k=4 (optimal), while the bottom row compares k=5 and k=6. Each pair shows the 3D RGB plot and its corresponding clustered jerlov classes. The results for k=4 show the most visually coherent and distinct style separation.

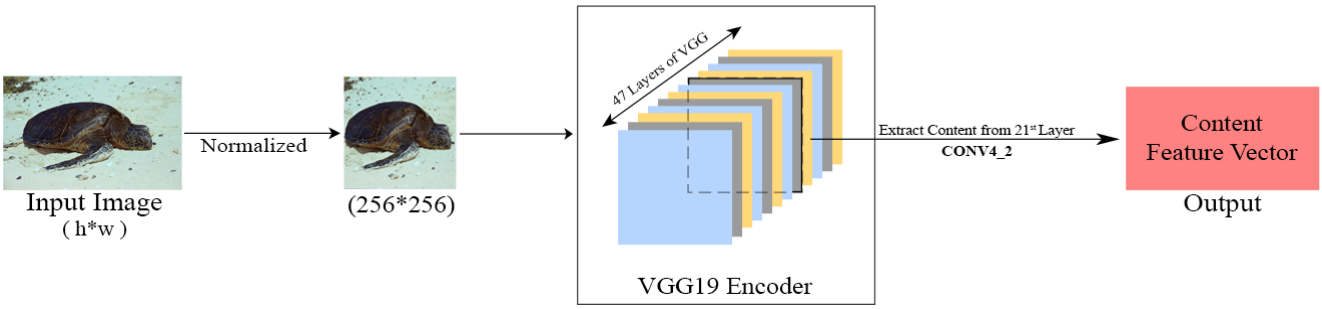


Figure 5. A content feature vector is extracted from the input image using the conv4_2 layer of a VGG19 encoder, preserving high-level structural information.

tively model the low-level texture and color tint from the underwater style reference image [11].

Generator and Discriminator. The generator receives the extracted content features and fuses them with the style representation using Adaptive Instance Normalization (AdaIN) layers [13]. This fused tensor is then processed through a series of residual blocks and transposed convolutions to decode the final, stylized output image. For adversarial training, we employ a PatchGAN discriminator [15], which evaluates realism by classifying 70x70 overlapping

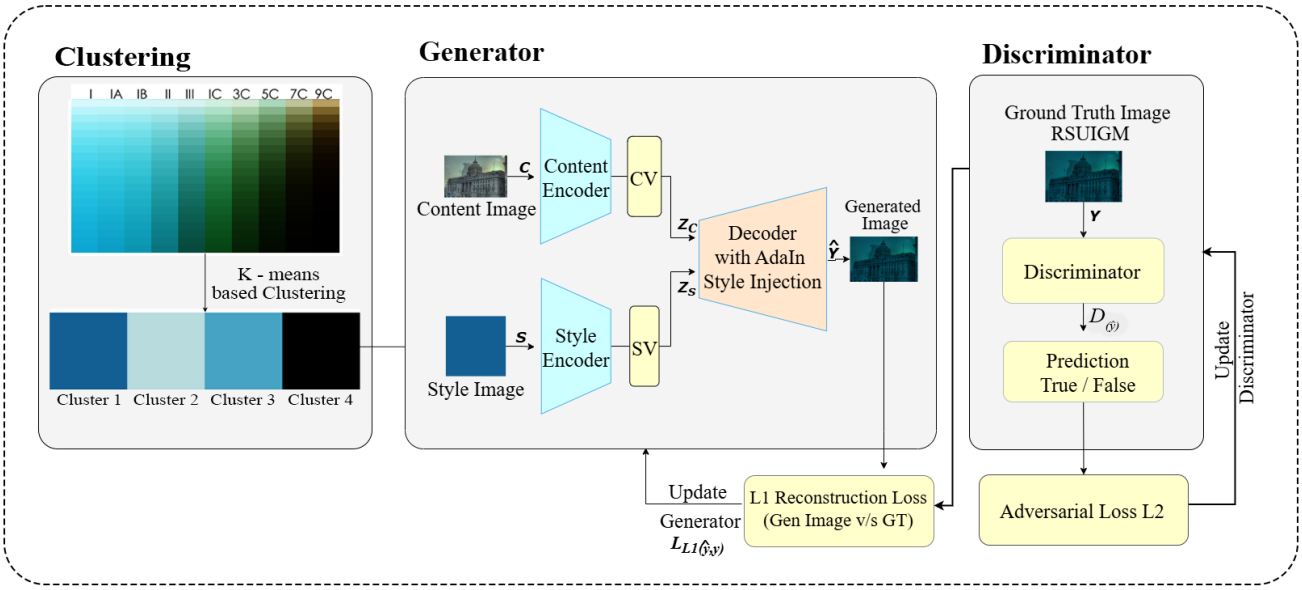


Figure 6. The complete architecture of DISC-GAN. The pipeline shows the parallel encoding of content and style images, feature fusion via Adaptive Instance Normalization (AdaIN), the generative decoder, and the PatchGAN discriminator that guides training using a composite loss.

patches of the image. This approach provides localized feedback, better preserving high-frequency details.

3.2. Training Parameters

The DISC-GAN framework is trained end-to-end by optimizing the composite loss function described in Equation 6, which combines an L1 reconstruction loss with an L2 adversarial loss. The final objective for the generator is to minimize:

$$\mathcal{L}_G = \lambda_{\text{rec}} \|I_s - \hat{I}\|_1 + \lambda_{\text{adv}} \|D(\hat{I}) - 1\|_2^2 \tag{7}$$

where I_s is the target style image, $\hat{I}$ is the generated image, and $D(\cdot)$ is the discriminator's output. The hyperparameters λ_{rec} and λ_{adv} balance structural fidelity with perceptual realism.

The training is optimized using the Adam optimizer with a learning rate of 2×10^{-4} and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The models were implemented using PyTorch and trained for 100 epochs for each of the four style clusters on an NVIDIA Tesla V100 GPU. The full training and inference procedure is formalized in Algorithm 1.

4. Results

In this section, we present the evaluation of our proposed DISC-GAN framework on the RSUIGM dataset [9]. We report results on the synthesis quality across the four predefined style clusters and analyze the visual fidelity of the generated images using full-reference image quality metrics. We also compare the performance of our data-driven method against the physics-informed principles used to create the ground-truth dataset through both quantitative and qualitative analysis.

4.1. Quantitative Analysis

We use a learning-based framework, DISC-GAN, for the synthesis of underwater images. To evaluate its performance, we use three standard quantitative metrics: the Structural Similarity Index Measure (SSIM), the Peak Signal-to-Noise Ratio (PSNR), and the Fréchet Inception Distance (FID) [26]. SSIM and PSNR assess structural and pixel-level similarity to the ground truth, while FID measures the distributional similarity between real and generated images. The Fréchet Inception Distance is given by:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}\right) \tag{8}$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean and covariance of features from real and generated images, respectively. A lower FID score indicates better perceptual quality.

The model was evaluated for its ability to generate high-fidelity images for each of the four style clusters. The quantitative scores are reported in Table 1. The model demon-

Algorithm 1 Cluster-Specific Synthetic Underwater Image Generation using DISC-GAN

Initialize: Generator θ_G, Discriminator θ_D, dataset $\mathcal{D}$, clusters $\mathcal{C}$, encoders $\mathcal{E}_{content}, \mathcal{E}_{style}$
Phase I: Clustering
Partition dataset $\mathcal{D}$ into $k = 4$ style clusters $\mathcal{C}$ using K-means.

Phase II: Training
for each style cluster $c \in \mathcal{C}$ **do**
  for each content image I_{cont} and style image $I_{style} \in c$ **do**
    $z_c \leftarrow \mathcal{E}_{content}(I_{cont})$
    $z_s \leftarrow \mathcal{E}_{style}(I_{style})$
    $\hat{I} \leftarrow G(z_c, z_s)$
    Calculate $\mathcal{L}_G$ using Eq. 6
    Update θ_G and θ_D via gradient descent
  end for
end for

Inference:
Given content image I_c and target cluster c
Select random style image I_s from cluster c
$\hat{I} \leftarrow G(\mathcal{E}_{content}(I_c), \mathcal{E}_{style}(I_s))$
return $\hat{I}$

strates robust performance, achieving high SSIM and PSNR values alongside low FID scores. Notably, Cluster 1 (blue) achieved an SSIM of 0.9012 and a PSNR of 32.5118, indicating strong structural preservation. The low FID scores across all clusters, such as 3.8576 for Cluster 4 (black), confirm that the generated images closely match the statistical distribution of the real underwater scenes in the RSUIGM dataset.

Table 1. Quantitative evaluation of DISC-GAN across the four style clusters using SSIM, PSNR, and FID metrics.

Metric	Blue	Light-Blue	Dark-Blue	Black
SSIM	0.9012	0.8107	0.7551	0.7212
PSNR	32.5118	27.8125	31.7876	40.8871
FID	8.3728	7.6894	7.0021	3.8576

4.2. Qualitative Analysis

We provide a qualitative comparison of the synthesized images for each style cluster in Figure 7. We observe that DISC-GAN successfully applies distinct underwater styles to a variety of clean content images. The model capably adapts tint, haze, and illumination to match the target cluster while preserving the structural integrity of the input content. The generated outputs are visually consistent with real underwater scenes and show minimal structural artifacts. This modular generation capability validates the effectiveness of our cluster-wise training approach and the successful disentanglement of style and content.

4.3. Analysis of Data-Driven Synthesis

A significant outcome of our work is the demonstration that a purely data-driven approach can achieve a level of realism comparable to that of physics-informed methodologies. The RSUIGM dataset, used here as ground truth, was generated with a model cued by the physical principles of underwater light attenuation. Our DISC-GAN framework, without being explicitly programmed with these physical constraints, has learned to synthesize images that are quantitatively and qualitatively similar. This result validates that deep learning models, when structured for effective disentanglement, can internalize complex physical phenomena from data alone, presenting a powerful and flexible alternative to traditional physics-based rendering. Although spatial variation is not explicitly modeled in our pipeline, the GAN leverages the RSUIGM dataset to implicitly capture and reproduce these variations, thereby contributing to the realism of the generated underwater images.

5. Conclusion

In this paper, we have proposed a novel framework for the synthesis of underwater images that integrates style-domain partitioning and deep learning. The system is structured into three modules: K-means clustering to define style domains, disentangled feature learning using VGG19 encoders, and image generation via a decoder with Adaptive Instance Normalization. Unlike traditional methods that treat the underwater environment as visually homogenous, our approach benefits from cluster-specific training, which provides fine-grained control over the generated appearance. We have demonstrated the qualitative and quantitative effectiveness of the model by synthesizing realistic underwater images and evaluating performance using metrics such as PSNR, SSIM, and FID.

To improve its performance and expand its applicability, future work may explore the integration of temporal consistency for video synthesis, which would enable real-time simulation for underwater video feeds. The framework could be further enhanced by incorporating attention mechanisms or depth-aware style modulation to achieve more precise blending. Expanding the clustering mechanism to consider additional water quality metrics such as turbidity or salinity could also increase realism for specialized marine domains. These strategies can enhance the model's adaptability and accuracy, strengthening its applications in fields such as autonomous navigation, data augmentation for object detection, and environmental modeling.

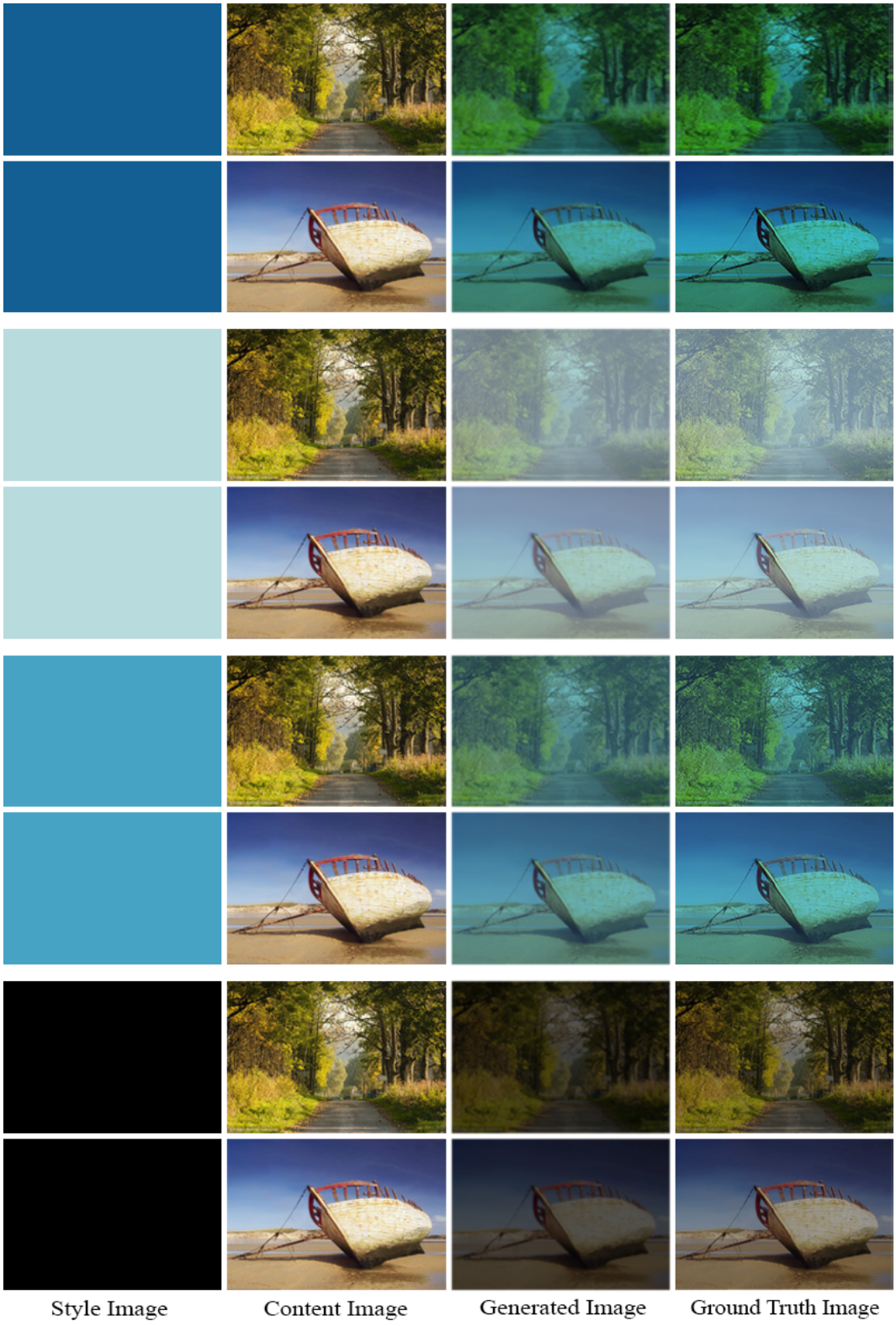


Figure 7. Qualitative results of the DISC-GAN framework. Each row shows a different content image, while each column represents a different target style cluster (Blue, Light-Blue, Dark-Blue, and Black). The content remains consistent while the underwater style changes across columns.

References

[1] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1682–1691, 2019. 1

[2] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *2012 IEEE conference on computer vision and pattern recognition*, pages 81–88. IEEE, 2012. 1

[3] Ahsan B Bakht, Zikai Jia, Muhayy Ud Din, Waseem Akram, Lyes Saad Saoud, Lakmal Seneviratne, Defu Lin, Shaoming He, and Irfan Hussain. Mula-gan: Multi-level attention gan for enhanced underwater visibility. *Ecological Informatics*, 81:102631, 2024. 1

[4] Dhiraj Chaurasia and Prateek Chhikara. Sea-pix-gan: Underwater image enhancement using adversarial neural network. *Journal of Visual Communication and Image Representation*, 98:104021, 2024. 2

[5] John Y Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4):1756–1769, 2011. 1

[6] Xiaofeng Cong, Yu Zhao, Jie Gui, Junming Hou, and Dacheng Tao. A comprehensive survey on underwater image enhancement based on deep learning. *arXiv preprint arXiv:2405.19684*, 2024. 1

[7] Chenggang Dai and Mingxing Lin. Single underwater image restoration using variational framework guided by imaging model with noise. *IEEE Access*, 12:82427–82442, 2024. 1

[8] Chaitra Desai, Sujay Benur, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi. Depthcue: Restoration of underwater images using monocular depth as a clue. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 196–205, 2023. 1, 2

[9] Chaitra Desai, Sujay Benur, Ujwala Patil, and Uma Mudenagudi. Rsuigm: Realistic synthetic underwater image generation with image formation model. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(1):1–22, 2024. 2, 3, 6

[10] Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 7159–7165. IEEE, 2018. 1

[11] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015. 4, 5

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 4

[13] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 1, 4, 5

[14] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE robotics and automation letters*, 5(2):3227–3234, 2020. 1

[15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei 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, 5

[16] Nils Gunnar Jerlov. *Marine optics*. Elsevier, 1976. 1, 2, 3

[17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2, 4

[18] Hadi Kazemi, Seyed Mehdi Iranmanesh, and Nasser Nasrabadi. Style and content disentanglement in generative adversarial networks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 848–856. IEEE, 2019. 2, 4

[19] Jie Li, Katherine A Skinner, Ryan M Eustice, and Matthew Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation letters*, 3 (1):387–394, 2017. 1, 2

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[21] Anparasy Sivaanpu, Rasika Priyadarshani, Thanikasalam Kokul, and Amirthalingam Ramanan. Underwater image enhancement using dual convolutional neural network with skip connections. In *2022 2nd International Conference on Advanced Research in Computing (ICARC)*, pages 224–229. IEEE, 2022. 2

[22] Hamidreza Farhadi Tolie, Jinchang Ren, Md Junayed Hasan, and Somasundar Kannan. Enhancing underwater situational awareness: Realsense camera integration with deep learning for improved depth perception and distance measurement. In *Artificial Intelligence for Security and Defence Applications II*, pages 34–42. SPIE, 2024. 2

[23] Junjun Wu, Xilin Liu, Qinghua Lu, Zeqin Lin, Ningwei Qin, and Qingwu Shi. Fw-gan: Underwater image enhancement using generative adversarial network with multi-scale fusion. *Signal Processing: Image Communication*, 109: 116855, 2022. 1

[24] Haodong Yang, Jisheng Xu, Zhiliang Lin, and Jianping He. Lu2net: a lightweight network for real-time underwater image enhancement. *arXiv preprint arXiv:2406.14973*, 2024. 1

[25] Yang Yu and Chenfeng Qin. An end-to-end underwater-image-enhancement framework based on fractional integral retinex and unsupervised autoencoder. *Fractal and Fractional*, 7(1):70, 2023. 1

[26] Yu Yu, Weibin Zhang, and Yun Deng. Frechet inception distance (fid) for evaluating gans. *China University of Mining Technology Beijing Graduate School*, 3(11), 2021. 6

[27] Yexun Zhang, Ya Zhang, and Wenbin Cai. Separating style and content for generalized style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8447–8455, 2018. 2

[28] Qi Zhao, Zhichao Xin, Zhibin Yu, and Bing Zheng. Unpaired underwater image synthesis with a disentangled representation for underwater depth map prediction. *Sensors*, 21(9): 3268, 2021. 1

[29] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 1