

GUSL-Dehaze: A Green U-Shaped Learning Approach to Image Dehazing

Mahtab Movaheddrad
University of Southern California
Los Angeles, California, USA
movahhed@usc.edu

Laurence Palmer
University of Southern California
Los Angeles, California, USA
palmerla@usc.edu

C.-C. Jay Kuo
University of Southern California
Los Angeles, California, USA
jckuo@usc.edu

Abstract—Image dehazing is a restoration task that aims to recover a clear image from a single hazy input. Traditional approaches rely on statistical priors and the physics-based atmospheric scattering model to reconstruct the haze-free image. While recent state-of-the-art methods are predominantly based on deep learning architectures, these models often involve high computational costs and large parameter sizes, making them unsuitable for resource-constrained devices. In this work, we propose GUSL-Dehaze, a Green U-Shaped Learning approach to image dehazing. Our method integrates a physics-based model with a green learning (GL) framework, offering a lightweight, transparent alternative to conventional deep learning techniques. Unlike neural network-based solutions, GUSL-Dehaze completely avoids deep learning. Instead, we begin with an initial dehazing step using a modified Dark Channel Prior (DCP), which is followed by a green learning pipeline implemented through a U-shaped architecture. This architecture employs unsupervised representation learning for effective feature extraction, together with feature-engineering techniques such as Relevant Feature Test (RFT) [1] and Least-Squares Normal Transform (LNT) [2] to maintain a compact model size. Finally, the dehazed image is obtained via a transparent supervised learning strategy. GUSL-Dehaze significantly reduces parameter count while ensuring mathematical interpretability and achieving performance on par with state-of-the-art deep learning models.

Index Terms—Image Dehazing, Green Learning, Machine Learning, Image Restoration

I. INTRODUCTION

Haze, caused by light scattering from small particles, reduces visibility and distorts colors, affecting image quality and degrading the performance of higher-level vision tasks such as image recognition and scene understanding. Consequently, dehazing is vital for recovering occluded details in hazy images.

Traditional dehazing methods rely on a physics-based atmospheric scattering model to describe how the haze forms in an image [3]. The observed haze is viewed as a mixture of scene radiance and scattered ambient light. Therefore, recovering a clear image requires estimating quantities such as the amount of haze at each pixel, the overall airlight color, or the scene depth. Because this inverse problem is underdetermined, most techniques introduce statistical priors to guide these estimates, the dark channel prior (DCP) being a prime example [4]. Despite their practical success, these approaches rest on simplifications that limit their ability to capture more complex, nonhomogeneous scattering phenomena.

Recently, deep learning architectures have come to dominate predictive tasks [5]–[7]. Current state-of-the-art methods rely on deep learning architectures, which are black-box methods that require large amounts of computational power [8], [9]. Due to these limitations, they are generally not feasible for deployment on resource-constrained devices. In addition, they demand substantial amounts of clear-hazy pairs to learn the ideal mapping. When provided with synthetic hazy data, this is not an issue but becomes a greater issue in real-world applications. A new method, termed Green Learning (GL), has been introduced to overcome several limitations of deep learning. Unlike conventional approaches, GL operates without neural networks. This paper presents GUSL-Dehaze, a novel dehazing framework incorporating a Green U-Shaped dehazing pipeline. Our method uses a physics-based model as its foundation and then employs a U-shaped architecture to predict the dehazed images. First, our method adopts a modified DCP as an initial dehazing step, and then it uses a U-shaped architecture to predict the final dehazed images. This approach consists of three modules:

- **Modified DCP:** We improve the traditional dark channel prior by incorporating a learning-based approach. This allows the model to adaptively estimate the haze density and improve the transmission map accuracy in challenging scenes.
- **Representation Learning:** This module extracts rich, multiscale feature representations from the hazy input using PixelHop units. The extracted features are further refined to select the most relevant components and generate strong secondary features, enhancing the overall representational capacity.
- **Decision Learning:** Guided by a U-shaped architecture, this module integrates an XGBoost regressor to predict the final dehazed image. The combination of structured learning and boosting enables accurate and interpretable decision making with low computational overhead.

II. RELATED WORKS

A. Dehazing

Traditional dehazing methods are generally based on the atmospheric scattering model (ASM) [3], a physics-based model of the formation of hazy images. It assumes that a

hazy image I is a convex combination of direct attenuation and airlight. Most prior-based methods make an informed estimate of parameters within the ASM supported by statistical observations in the data, allowing them to invert the ASM formulation and recover the clear image. Methods such as the DCP [4] and the color attenuation prior [10] represent prior-based approaches. However, the ASM approximates the complex interactions between particles and light with some problematic simplifications. For example, it assumes uniform haze, which makes it unsuitable for modeling nonhomogeneous haze.

Deep learning methods comprise the current SOTA in the dehazing domain. Within this category, we can partition the approaches into physics-aware and physics-unaware. Physics-aware methods consider some aspects of the ASM to guide feature creation or the overall dehazing process. For example, DehazeNet [11] and All-in-One-Dehazing Net [12] regress the medium transmission maps or rely on a reformulation of the ASM, respectively. Although these models improve over pure prior-based methods, they regress intermediate results and use ASM to recover the scene radiance. Thus, they suffer similar issues to those with prior-based methods. To overcome this limitation, other deep learning methods attempt to regress the clear image directly. These models, including FFA-Net [13], were the next generation of SOTA dehazing.

Recently, vision transformer architectures, such as DehazeFormer [14], have become the most successful dehazing schemes, with variants comprising most of the best-performing models. Other interesting approaches include physics-aware deep learning models, which use the ASM as a guide for discriminant feature creation. For example, C^2 PNet [15] uses the ASM in a dual-branch unit, approximating features associated with the atmospheric light and transmission map for more precise dehazing.

B. Green Learning

Green Learning (GL) [16] represents a paradigm shift away from the resource-heavy deep learning frameworks that dominate AI. By forgoing the iterative gradient-based updates of backpropagation, GL attains substantial savings in computation. Instead, it relies on unsupervised feature extraction techniques, most notably the Saab transform [17] and its channel-wise variant [18] to distill salient representations without the overhead of end-to-end weight tuning.

To further refine its set of characteristics, GL incorporates discriminative selection mechanisms such as the Discriminant Feature Test (DFT) and the Relevant Feature Test (RFT) [1]. These tests isolate the most informative components, streamline the model's inputs, and boost predictive accuracy. Once optimal features are identified, GL leverages powerful learners such as XGBoost [19] and Subspace Learning Machines (SLM) [20] to build robust classifiers that adapt gracefully across varied datasets and tasks.

A key advantage of GL lies in its lightweight nature. GL delivers scalable solutions suitable for real-world deployment by

removing backpropagation and eschewing monolithic end-to-end training. For example, a Green Learning-based demosaicing approach (GID) has been proposed in [21], demonstrating lower computational demands while maintaining high-quality image reconstruction.

III. PROPOSED METHOD

In this section, we present the GUSL-Dehaze method for image dehazing. Our model skips backpropagation altogether. Instead, it iteratively drives down the loss with boosted trees through the XGBoost regressors. The U-shaped layout in our proposed pipeline calls to mind the U-Net architecture in many neural networks. However, all parameters within each module are set in a feed-forward manner. Figure 1 illustrates the GUSL-Dehaze block diagram, which is composed of three distinct processing modules. Raw images are first projected into a spatial-spectral feature domain during the representation-learning stage. This projection is accomplished through a cascade of PixelHop units that extract and refine latent representations. At each hop, the Saab transform exploits the input patches' intrinsic spatial and spectral attributes, revealing feature representations of increasing abstraction. At each layer, the kernel size is tailored to the image size. Consequently, the filters become progressively smaller in the deeper layers.

A. Modified DCP

The DCP is a prior-based method that relies on the following formulation of the ASM:

$$I(x) = J(x)t(x) + A(1 - t(x))$$

$$t(x) = e^{-\beta d(x)}$$

where x is a given pixel, J is the scene radiance, t is the transmission map, A is the airlight, β is the atmospheric scattering coefficient, and d is the depth map. The DCP objective is to estimate $t(x)$ and A to recover J .

The first step in our pipeline is to perform preliminary dehazing with a modified DCP algorithm. The DCP for a given image J is defined as

$$J^{dark}(x) = \min_{c \in \{r,g,b\}} (\min_{y \in \Omega(x)} (J^c(y)))$$

where $\Omega(x)$ is a patch centered a pixel x . Statistically, J^{dark} tends to 0 for clear (outdoor) images. Using this property in the context of the ASM, an estimated transmission map can be recovered using

$$\tilde{t}(x) = 1 - \omega \min_{c \in \{r,g,b\}} \left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right)$$

The term on the right is the DCP of the normalized haze image $I^c(y)/A^c$ multiplied by $\omega \in (0, 1]$, which is a term to preserve aerial perspective and control the amount of haze removal. A^c is the atmospheric light estimate for the channel c . We refine $\tilde{t}(x)$ via a guided filtering algorithm [22] similar

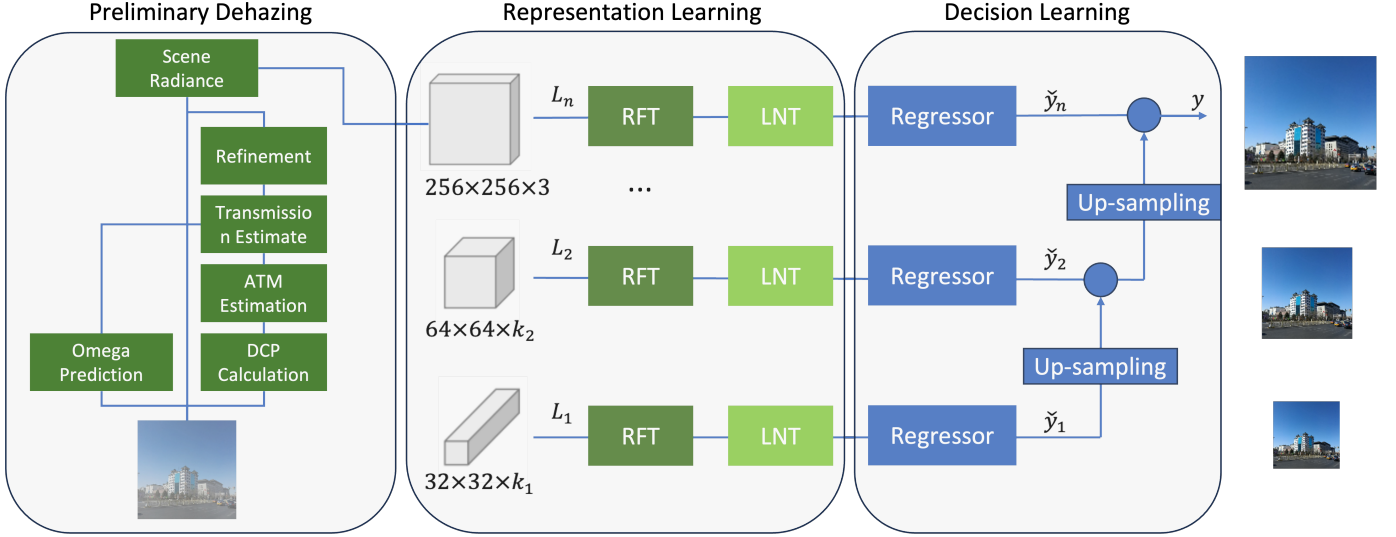


Fig. 1. An overview of the proposed pipeline in three modules. Module 1 represents the preliminary dehazing using a modified DCP method. Module 2 learns robust representations via the Saab transform, selects the most relevant raw features, and then generates secondary features through the LNT. Module 3 employs two XGBoost regressors at each level to estimate the residuals based on the initial predictions of the preceding (coarser) level.

to the soft-matting procedure in the original DCP. Then, the scene radiance can be recovered with

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A$$

where t_0 is a lower bound on the transmission map. The A and the A^c are estimated using the DCP. A percentage of the brightest pixels, $\frac{1}{k}$, in the DCP are identified. Among these pixels, we find the pixels with the highest intensity in the input image and use those as our A estimates.

The DCP has four main parameters of size ω , t_0 , Ω , and k , excluding the guided filter parameters. The ω has the most significant effect on the recovered image, depending on the intensity of the haze. For example, if ω is high and the intensity of the haze is low, the recovered image will suffer significant artifacts. Similarly, if ω is low and the intensity of the haze is high, the recovered image will contain high levels of residual haze.

To overcome this, we fix all other parameters and regress the ω value using a random forest model [23]. The RESIDE dataset provides β parameters associated with the OTS set, which were used to synthesize the hazy images with the ASM equations. These *beta*'s serve as a GT of haze intensity. For each β , we tested for the ideal ω and trained the random forest model to regress the optimal ω given the global mean, min, max, and variance within RGB and YUV color spaces. This model generalized well to other datasets while training only on 7K images.

B. Representation Learning

1) *PixelHop feature extraction*: PixelHop, introduced in [18], applies the Saab filters to extract features from raw image data. Every PixelHop unit takes a central pixel and its neighborhood $n \times n$, then flattens them into a vector \mathbf{v} .

A Saab filter of dimensions $m \times m$ (where $m < n$) is applied to this patch. Because the Saab filter is equivalent to a PCA rotation followed by a bias in a successive manner, the patch is projected onto an orthonormal set of principal vectors

$$\mathbf{U} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{K-1}\},$$

, each accompanied by a bias c_k . The k -th projection coefficient is therefore

$$z_k = \mathbf{u}_k^T \mathbf{v} + c_k, \quad k = 0, 1, \dots, K-1.$$

The leading vector \mathbf{u}_0 is linked to the largest eigenvalue and captures the DC (Direct Current) energy of the patch. All remaining vectors $\mathbf{u}_1, \dots, \mathbf{u}_{K-1}$ encode the AC (Alternating Current) components. Hence, the feature space splits orthogonally into

$$\mathcal{Z} = \mathcal{Z}_{\text{DC}} \oplus \mathcal{Z}_{\text{AC}},$$

with $\mathcal{Z}_{\text{DC}} = \text{span}\{\mathbf{u}_0\}$ and $\mathcal{Z}_{\text{AC}} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{K-1}\}$.

Let $\mathbf{v}_{\text{DC}} = \mathbf{u}_0^T \mathbf{v} + c_0$ denote the DC projection of the patch. Subtracting this dominant component yields the residual AC part,

$$\mathbf{v}_{\text{AC}} = \sum_{k=1}^{K-1} (\mathbf{u}_k^T (\mathbf{v} - \mathbf{v}_{\text{DC}}) + c_k).$$

Once the dominant DC energy has been removed, the residual \mathbf{v}_{AC} is subjected to a further PCA so that the remaining variance is concentrated along a compact set of axes. Let $\Phi = \{\phi_0, \phi_1, \dots, \phi_{L-1}\}$ be the orthonormal eigenvectors obtained from this PCA and let d_l be their associated biases. The refined AC representation is

$$\tilde{\mathbf{v}}_{\text{AC}} = \sum_{l=0}^{L-1} (\phi_l^T \mathbf{v}_{\text{AC}} + d_l) \phi_l,$$

providing an economical encoding that preserves delicate variations in the data.

At hop i the algorithm receives an $n \times n$ neighbourhood. Convolving this patch with an $m \times m$ Saab kernel ($m < n$) produces a feature tensor

$$I^{(i)} \in \mathbb{R}^{N \times N \times K^{i-1}}, \quad N = n - m + 1.$$

Because each hop appends new channels to those of the previous stage, the channel count can grow rapidly. A max-pool layer is applied immediately after the convolution to maintain tractability, reducing the spatial dimensions while retaining the strongest responses. The pooled tensor serves as the input to hop $i + 1$, ensuring that the hierarchical representation remains both informative and computationally manageable. Successive multi-hop transformations distill a compact and expressive latent embedding that preserves high-level global context and fine-grained local structure, providing a stronger foundation for accurate and robust downstream tasks.

2) *Relevant Feature Selection*: After the PixelHop feature extraction phase, we have three high-dimensional feature matrices—one per hop. To keep only the most predictive dimensions, we use a supervised feature selection method based on the Relevant Feature Test (RFT) for regression [1]. RFT measures each feature’s effect on prediction error and iteratively removes those with little impact, yielding compact, informative representations for each hop. For each dimension j and each candidate cut value f_t^j , the RFT partitions the training data into

$$S_{L,t}^j = \{ \mathbf{x}_i : x_{ij} \leq f_t^j \}, \quad S_{R,t}^j = \{ \mathbf{x}_i : x_{ij} > f_t^j \}.$$

Then, we fit separate regression models on the left and right subsets and record their mean-squared errors $\mathcal{L}_{L,t}^j$ and $\mathcal{L}_{R,t}^j$. The loss associated with threshold t is the sample-size-weighted average:

$$\mathcal{L}_t^j = \frac{N_{L,t}^j \mathcal{L}_{L,t}^j + N_{R,t}^j \mathcal{L}_{R,t}^j}{N},$$

where $N_{L,t}^j$ and $N_{R,t}^j$ denote the subset sizes and N is the total number of observations.

The optimal threshold for feature j minimizes this loss:

$$\mathcal{L}_{\min}^j = \min_{t \in T} \mathcal{L}_t^j,$$

with T the set of all candidate thresholds. Because smaller values of \mathcal{L}_{\min}^j indicate a cleaner and more predictable split, they serve as a relevance score.

Finally, we rank all features by \mathcal{L}_{\min}^j in ascending order and retain the K features with the lowest scores. The supervised ranking yields a small subset of variables that improve the performance of the regression.

The RFT loss curves reveal which dimensions carry the greatest predictive power. In Figure 2, the features are sorted by ascending RFT loss. Thus, variables on the left (low loss) are judged most informative.

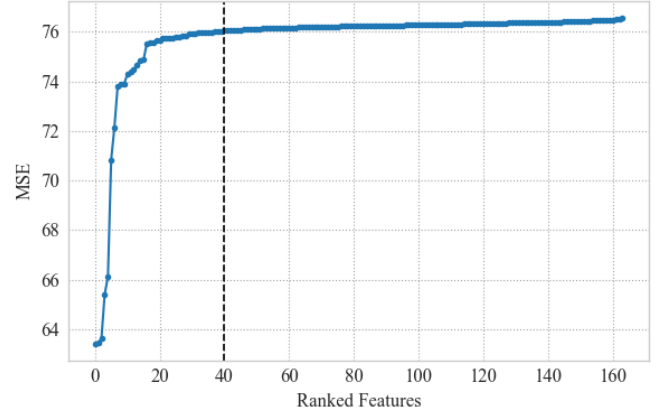


Fig. 2. Ranked features based on MSE loss using RFT feature selection.

TABLE I
MEAN SQUARED ERROR (MSE) FOR DIFFERENT FEATURE INCORPORATION METHODS.

Feature incorporated	MSE		
	R	G	B
Raw	11.35	11.04	12.72
L1 + L2	11.44	11.28	13.03
Raw + L1 + L2	10.17	9.80	11.85

3) *Feature Generation*: We selected a subset of the most discriminative features extracted from the images in the previous module. However, expanding the feature space with additional complementary features offers clear advantages. This is particularly important in complex tasks where a limited feature set may not capture all the subtle variations in the data. A diverse feature pool increases the likelihood of capturing relevant patterns.

We employ the least squares normal transform (LNT) [2] to obtain these complementary features. This method reformulates the original multi-class classification task into a linear regression framework, deriving discriminative transformations that enhance the features selected in earlier stages. Suppose we have ℓ training instances drawn from κ distinct classes. Introduce an indicator matrix $\mathcal{T} \in \mathbb{R}^{m \times \ell}$ whose entry $\tau_{m,\ell}$ is 1 when the sample ℓ belongs to superclass m and 0 otherwise. The task is to learn a weight matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ that linearly projects the feature matrix $\mathbf{X} \in \mathbb{R}^{n \times \ell}$ onto the target space, producing the least squares system

$$\mathbf{A}\mathbf{X} + \mathbf{B} = \mathcal{T},$$

where \mathbf{B} is a rank one bias term that shifts only the feature mean. Taking expectations, we can isolate the bias:

$$\mathbf{B} = \mathbb{E}[\mathcal{T}] - \mathbf{A} \mathbb{E}[\mathbf{X}].$$

Because this bias only translates the mean, it does not influence the choice of discriminative directions. The core problem

TABLE II
QUANTITATIVE COMPARISON ON VARIOUS DEHAZING BENCHMARKS.

Method	Venue & Year	SOTS-indoor		SOTS-outdoor		Dense-Haze		NH-Haze2		#Params
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
DCP [4]	TPAMI2010	16.62	0.8179	19.13	0.8148	11.01	0.4165	11.68	0.6475	–
AODNet [12]	ICCV2017	19.06	0.8504	20.29	0.8765	12.82	0.4683	12.33	0.6311	x0.001
GDN [24]	ICCV2019	32.16	0.9836	30.86	0.9819	14.96	0.5326	19.26	0.8046	x0.56
GCANet [25]	WACV2019	30.06	0.9596	22.76	0.8887	12.62	0.4208	18.79	0.7729	x0.41
MSBDN [26]	CVPR2020	32.77	0.9812	34.81	0.9857	15.13	0.5551	20.11	0.8004	x18.22
FFA-Net [13]	AAAI2020	36.39	0.9886	33.57	0.9840	12.22	0.4440	20.00	0.8225	x2.59
AECr-Net [27]	CVPR2021	37.17	0.9901	–	–	15.80	0.4660	20.68	0.8282	x1.52
DeHamer [8]	CVPR2022	36.63	0.9881	35.18	0.9860	16.62	0.5602	19.18	0.7939	x77.0
MAXIM-2S [28]	CVPR2022	38.11	0.9908	34.19	0.9846	–	–	–	–	x8.26
UDN [29]	AAAI2022	38.62	0.9909	34.92	0.9871	–	–	–	–	x2.47
MB-TaylorFormer-B [30]	ICCV-2023	40.71	0.9920	37.42	0.9890	–	–	25.05	0.7880	x1.56
C ² PNet [15]	CVPR-2023	42.56	0.9954	36.68	0.9900	16.88	0.5728	21.19	0.8334	x4.17
GUSL-Dehaze	–	39.14	0.9911	36.83	0.9894	16.21	0.5719	20.74	0.8154	x1

is estimating \mathbf{A} . Applying the normal equations of linear regression gives

$$\mathbf{A} = \mathcal{T} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^{-1}.$$

Once \mathbf{A} is known, any feature vector $\mathbf{x} \in \mathbb{R}^n$ can be mapped to its LNT representation $\mathbf{d} \in \mathbb{R}^m$ via

$$\mathbf{d} = \mathbf{A} \mathbf{x} = (d_1, \dots, d_m)^\top. \quad (1)$$

We group the features into n categories to improve classification accuracy and use XGBoost decision trees to select the most effective subsets.

Following this pipeline, we refer to the features taken directly from the raw input as Level 1 features. Additionally, we build Level 2 features, not directly from the raw data, but from Level 1 features. This layered approach lets Level 2 features capture more complex patterns and deeper connections, making the overall feature space more expressive and structured. To evaluate how effective this two-level feature generation is, we compare the MSE results of the validation set by models trained with the raw and combined raw generated features in Table 1.

C. Decision learning

The **Decision Learning** module utilizes two **XGBoost regressors** $f^{(\ell)}$ at each processing level ℓ . The input to this module is a feature vector $\mathbf{x}^{(\ell)} \in \mathbb{R}^{n \times n \times d}$, composed of both selected raw features and synthesized features generated at the corresponding level. The target output at level ℓ is the prediction value $\mathbf{y}^{(\ell)} \in \mathbb{R}^{n \times n \times 1}$, corresponding to the pixel value of the respective channel in the downsampled image. XGBoost performs regression by minimizing a regularized objective function:

$$\mathcal{L}^{(\ell)}(\phi) = \sum_{i=1}^N l(y_i^{(\ell)}, \hat{y}_i^{(\ell,t)}) + \sum_{k=1}^t \Omega(f_k^{(\ell)}),$$

where

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2,$$

with l a convex loss (e.g., squared error), $f_k^{(\ell)}$ the k -th tree, T its number of leaves, and w the leaf weights. Using a second-order Taylor expansion around the current prediction $\hat{y}^{(\ell,t-1)}$:

$$\mathcal{L}^{(\ell)} \approx \sum_{i=1}^N \left[g_i^{(\ell)} f_t^{(\ell)}(x_i^{(\ell)}) + \frac{1}{2} h_i^{(\ell)} f_t^{(\ell)}(x_i^{(\ell)})^2 \right] + \Omega(f_t^{(\ell)}),$$

where

$$g_i^{(\ell)} = \frac{\partial}{\partial \hat{y}_i^{(\ell,t-1)}} l(y_i^{(\ell)}, \hat{y}_i^{(\ell,t-1)}),$$

$$h_i^{(\ell)} = \frac{\partial^2}{\partial (\hat{y}_i^{(\ell,t-1)})^2} l(y_i^{(\ell)}, \hat{y}_i^{(\ell,t-1)}).$$

This formulation yields fast, accurate updates and explicit complexity control, making XGBoost particularly well-suited for pixel-wise regression in the Decision Learning module.

IV. EXPERIMENTS

We used the RESIDE dataset to train our modified DCP model, which provides synthetic indoor and outdoor images paired with ground truth scattering coefficients (β values) for the outdoor training set. The indoor training subset (ITS) comprises 13,990 pairs generated via ground truth depth maps. The Outdoor Training Subset (OTS) contains 72135 pairs synthesized from estimated depth maps, ensuring a diverse range of haze densities and lighting conditions. We initially applied the DCP algorithm to resize hazy images (256×256) to obtain preliminary dehazed outputs. Based on pre-processed images of size 256×256 , we employ a U-shaped network with four successive downsampling stages, generating feature maps at resolutions of 128×128 , 64×64 , and 32×32 . The model predicts residual components at each layer, which are subsequently resampled and integrated into the upper layer to refine the reconstruction. We train two variants of

this architecture, one trained in the ITS subset for indoor scenes and another trained in the OTS subset for outdoor environments, and evaluate both in several benchmark datasets. We report the total number of parameters for our proposed model to support our goal of green learning of competitive performance with reduced complexity. Table II presents our quantitative results alongside parameter counts. Our model achieves PSNR and SSIM scores comparable to state-of-the-art deep learning models, significantly outperforming traditional physics-based dehazing methods. Our model contains 1.72 million parameters in total—approximately 0.43 million per layer, including 0.36 million attributed to the XGBoost regressor. We report the smallest size among the leading deep learning approaches.

V. CONCLUSION

In this work, we proposed a physics-aware green learning framework for image dehazing, GUSL-Dehaze. It delivers performance comparable to state-of-the-art deep learning methods, with the smallest model size reported among them. GUSL-Dehaze is mathematically transparent and computationally efficient. They are well-suited for deployment on edge devices, offering minimal latency, low battery consumption, and efficient memory usage.

ACKNOWLEDGMENT

The computational work was supported by the University of Southern California Center for Advanced Research Computing (carc.usc.edu).

REFERENCES

- [1] Y. Yang, W. Wang, H. Fu, C.-C. J. Kuo *et al.*, “On supervised feature selection from high dimensional feature spaces,” *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.
- [2] X. Wang, V. K. Mishra, and C.-C. J. Kuo, “Enhancing edge intelligence with highly discriminant Int features,” in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 3880–3887.
- [3] S. G. Narasimhan and S. K. Nayar, “Contrast restoration of weather degraded images,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [4] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
- [5] A. Fayyazi, M. Kamal, and M. Pedram, “Fair-sight: Fairness assurance in image recognition via simultaneous conformal thresholding and dynamic output repair,” *arXiv preprint arXiv:2504.07395*, 2025.
- [6] M. A. Ranjbar, A. Ghaleh, H. B. Dogaheh, P. Razmara, and M. Baghani, “Beyond subjective measures: Systematic review of deep learning in chronic pain: Modalities, methods, and applications,” <https://doi.org/10.21203/rs.3.rs-6217737/v1>, 2025.
- [7] Y. Torabi, S. Shirani, and J. P. Reilly, “Large language model-based nonnegative matrix factorization for cardiorespiratory sound separation,” *arXiv preprint arXiv:2502.05757*, 2025.
- [8] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, “Image dehazing transformer with transmission-aware 3d position embedding,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5812–5820.
- [9] A. Abdollahi, N. Ashrafi, and M. Pishgar, “Advanced predictive modeling for enhanced mortality prediction in icu stroke patients using clinical data,” *arXiv preprint arXiv:2407.14211*, 2024.
- [10] Q. Zhu, J. Mai, and L. Shao, “A fast single image haze removal algorithm using color attenuation prior,” *IEEE transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end system for single image haze removal,” *IEEE transactions on image processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “Aod-net: All-in-one dehazing network,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.
- [13] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, “Ffa-net: Feature fusion attention network for single image dehazing,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.
- [14] Y. Song, Z. He, H. Qian, and X. Du, “Vision transformers for single image dehazing,” *IEEE Transactions on Image Processing*, vol. 32, pp. 1927–1941, 2023.
- [15] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du, “Curricular contrastive regularization for physics-aware single image dehazing,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 5785–5794.
- [16] C.-C. J. Kuo and A. M. Madni, “Green learning: Introduction, examples and outlook,” *Journal of Visual Communication and Image Representation*, vol. 90, p. 103685, 2023.
- [17] Y. Chen and C.-C. J. Kuo, “Pixelhop: A successive subspace learning (ssl) method for object recognition,” *Journal of Visual Communication and Image Representation*, vol. 70, p. 102749, 2020.
- [18] —, “Pixelhop: A successive subspace learning (ssl) method for object classification,” *arXiv preprint arXiv:1909.08190*, 2019.
- [19] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [20] H. Fu, Y. Yang, V. K. Mishra, and C.-C. J. Kuo, “Subspace learning machine (slm): Methodology and performance evaluation,” *Journal of Visual Communication and Image Representation*, vol. 98, p. 104058, 2024.
- [21] M. Movahhedrad, Z. Chen, and C.-C. J. Kuo, “A green learning approach to efficient image demosaicking,” in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 1067–1074.
- [22] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.
- [23] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
- [24] X. Liu, Y. Ma, Z. Shi, and J. Chen, “Griddehazenet: Attention-based multi-scale network for image dehazing,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7314–7323.
- [25] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, “Gated context aggregation network for image dehazing and deraining,” in *2019 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2019, pp. 1375–1383.
- [26] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, “Multi-scale boosted dehazing network with dense feature fusion,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2157–2167.
- [27] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, “Contrastive learning for compact single image dehazing,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10 551–10 560.
- [28] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, “Maxim: Multi-axis mlp for image processing,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5769–5780.
- [29] M. Hong, J. Liu, C. Li, and Y. Qu, “Uncertainty-driven dehazing network,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 1, 2022, pp. 906–913.
- [30] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, and Z. Jin, “Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 802–12 813.