Int. J. Appl. Math. Comput. Sci., 2026, Vol. 36, No. 3, 419–432
DOI: 10.61822/amcs-2026-0027

COMBINED NO–REFERENCE METRICS FOR QUALITY ASSESSMENT OF RECONSTRUCTED SATELLITE LiDAR DATA

KRZYSZTOF OKARMA[a,*], MATEUSZ KOPYTEK[a], PIOTR LECH[a], OLEG IEREMEIEV[b], VLADIMIR V. LUKIN[b], ANDRES RAMIREZ-JAIME[c], NESTOR PORRAS-DIAZ[c], GONZALO R. ARCE[c]

[a] Department of Signal Processing and Multimedia Engineering
West Pomeranian University of Technology in Szczecin
Sikorskiego 37, 70-313 Szczecin, Poland
e-mail: {okarma,mateusz-kopytek,piotr.lech}@zut.edu.pl

[b] Department of Information-Communication Technologies
National Aerospace University "Kharkiv Aviation Institute"
Vadyma Manka 17, 61070 Kharkiv, Ukraine
e-mail: {o.ieremeiev,v.lukin}@khai.edu

[c] Department of Electrical and Computer Engineering
University of Delaware
Evans Hall, The Green, Newark, DE 19716, USA
e-mail: {aramjai,nestorfe,arce}@udel.edu

In this paper, experimental results concerning the application of no-reference image quality metrics for automatic quality assessment of reconstructed satellite light detection and ranging (LiDAR) data are discussed. Since satellite LiDARs provide much sparser data compared to airborne systems and their sending to Earth requires additional compression, the received data should be reconstructed. For this purpose, generative neural networks can be applied, including generative diffusion models; however, they require the choice of appropriate parameters, and the reconstruction results should be reliably assessed. Although the physical representation of data is related to the number of received photons, considering the novel data representation based on hyper-height data cubes, it is possible to analyze individual percentiles, representing canopy height models or digital terrain models, as grayscale images. Therefore, their quality may be evaluated using image quality assessment metrics. Due to the lack of reference data, in the application considered, only no-reference metrics may be efficiently applied, even though they have lower correlation with subjective quality perception compared to full-reference metrics. Therefore, the idea of designing combined no-reference metrics optimized for the highest correlation with subjective quality scores is discussed, along with the presentation of the obtained correlation results.

Keywords: satellite LiDARs, no-reference image quality metrics, LiDAR data reconstruction, compressive sensing, hyper-height data cubes.

1. Introduction

1.1. Motivation. Spaceborne light detection and ranging (LiDAR) instruments emit laser pulses to probe the Earth's surface and infer scene structure from the time-of-flight of the backscattered signal. By converting these returns into height profiles of terrain and surface features—including forest canopies—spaceborne LiDAR has become a foundational modality for quantitative remote sensing (Dubayah *et al.*, 2020). Despite its usefulness, LiDAR sensing from orbit entails constraints that motivate complementary strategies. Satellite platforms furnish global coverage but with coarser spatial footprints and relatively sparse sampling; additionally, the long atmospheric propagation path reduces the signal-to-noise ratio (SNR). Airborne LiDAR, by contrast, operates at lower altitudes and yields finer spatial detail, yet it is intrinsically limited in areal extent—typically only a few square kilometers per

*Corresponding author

acquisition—thereby constraining efficiency and resource utilization (Popescu *et al*., 2011).

To mitigate these limitations, artificial intelligence (AI) techniques have emerged as promising tools for enhancing and reconstructing compressed satellite LiDAR products, with the potential to improve overall data quality (Ramirez-Jaime *et al*., 2025b). Realizing this potential, however, demands rigorous quality control and systematic verification of results (Talebi and Milanfar, 2018). A central challenge is the frequent absence of pristine reference data. Recorded LiDAR signals arise from laser returns that traverse the atmosphere and are therefore subject to scattering, attenuation, and other environment-induced distortions (Weitkamp, 2005). Detector nonidealities introduce additional noise and artifacts, including dark counts, afterpulsing, shot noise, and read noise (Yang *et al*., 2023). Subsequent digital processing—encompassing recovery and reconstruction—can inadvertently amplify these degradations.

1.2. Hyper-height data cubes. In this work, we emphasize rasterized products derived from LiDAR waveforms or point clouds, such as percentile-based slices. These are commonly organized as hyper-height data cubes (HHDCs) (Ramirez-Jaime *et al*., 2024), i.e., three-dimensional tensors with two spatial dimensions (x, y) and a discretized height (bin) dimension, which serve as a canonical data structure for downstream analysis. Each percentile slice within an HHDC can be interpreted as a grayscale image that captures a distinct aspect of the underlying terrain structure (Fig. 1). For example, the highest percentile (typically 98%, to mitigate noise-related outliers without the need of using more sophisticated methods (Kulczycki *et al*., 2024)) in forested environments corresponds to the canopy height model (CHM), which exhibits seasonal variability due to changes in leaf coverage. The median percentile (50%) is frequently employed for biomass estimation, whereas the lowest percentile (often 2%, again to reduce measurement noise) represents the digital terrain model (DTM), delineating the ground topography. Image quality metrics offer a principled route to evaluate these LiDAR-derived representations; however, most established metrics are designed for optical imagery, whose statistical, radiometric, and geometric characteristics differ substantially from those of LiDAR products. While such metrics are promising, their suitability for LiDAR remains insufficiently validated, complicating reliable adoption. For restoration and reconstruction algorithms, fidelity assessment relative to an original is often performed using full-reference image quality assessment (FR-IQA) methods that quantify perceptual changes (Wang *et al*., 2004). In practical LiDAR scenarios, however, true references are often unavailable, motivating the use of no-reference (or "blind") image quality assessment (NR-IQA) metrics that operate without a benchmark (Mittal *et al*., 2012).

1.3. Verification of quality metrics. In optical imaging—a comparatively mature field—standardized validation relies on curated test databases in which images are corrupted by controlled artifact types and severities. These databases provide mean opinion scores (MOSs) derived from extensive subjective experiments, where aggregated expert and participant ratings constitute ground truth for perceptual quality. Because human vision is the primary consumer of image information, this paradigm has strong practical value. Subsequent analyses quantify the correlation between MOS and objective metrics to assess alignment with visual perception across distortion families. Numerous resources support this ecosystem for optical systems, including KADID-10k (Lin *et al*., 2019), PIPAL (Jinjin *et al*., 2020), TID2013 (Ponomarenko *et al*., 2015), CSIQ (Larson and Chandler, 2010), LIVE (Sheikh *et al*., 2006), and SPAQ (Fang *et al*., 2020). To our knowledge, publicly available counterparts tailored to LiDAR-derived products remain limited, reflecting the field's specialization and narrower scope.

Advancing LiDAR processing and reconstruction therefore depends on dedicated verification methodologies. Specifically, this entails constructing a bespoke test set of images with associated MOS estimates, followed by the evaluation and selection of appropriate objective quality metrics. Validated metrics, in turn, enable objective and controlled development of digital processing algorithms for remote sensing data, particularly LiDAR. While FR-IQA metrics are valuable when references exist, NR-IQA metrics offer distinct advantages in LiDAR contexts: they can act as loss functions during neural network-based reconstruction and can support quality evaluation when reference HHDCs are unavailable, which motivates the present focus on NR-IQA verification.

1.4. Objective of the study. The primary objective of this study is to assess the efficacy of NR-IQA metrics for evaluating the reconstruction quality of compressed satellite LiDAR data represented as HHDCs. In support of this objective, we introduce a new database of test images comprising selected percentile slices from HHDC tensors, each accompanied by subjective quality assessments. We then compute correlation coefficients between these MOS values and a suite of objective NR-IQA metrics, and we propose combined metrics to enhance overall performance.

Operationally, satellite LiDAR data must be downlinked to ground stations within limited visibility

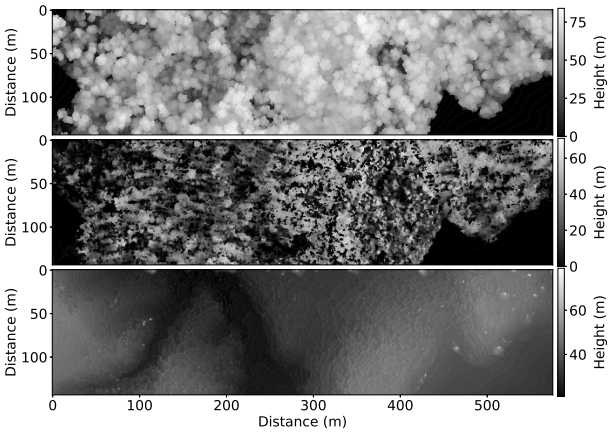


Fig. 1. HHDC-derived 2D products via percentile slicing. Top: canopy height model computed per footprint as $\mathrm{CHM}(i,j) = h_{98}(i,j) - h_2(i,j)$, where h_p is the p-th percentile of the within-footprint vertical return distribution from the hyper-height data cube. Middle: median elevation map $h_{50}(i,j)$, summarizing the central tendency of the canopy/understory structure and often used as a biomass proxy. Bottom: digital terrain model estimated from the 2nd percentile $h_2(i,j)$, approximating the bare-Earth surface.

windows. The intrinsic sparsity of HHDCs, characterized by many zero-valued entries, makes them amenable to compressive sensing for efficient transmission. Accurate post-downlink reconstruction, however, requires carefully tuned algorithms and parameters. Assessing these reconstructions necessitates appropriate quality metrics; given the lack of reference data in real-time satellite operations, such metrics must operate solely on the reconstructed outputs. Consequently, for individual percentile layers of the HHDC, no-reference image quality assessment metrics are the most suitable choice.

1.5. Organization of the paper. The remainder of the paper is organized as follows. Section 2 discusses the super-resolution reconstruction of satellite LiDAR data, with Section 3 providing a detailed treatment of diffusion models for inverse problems. Section 4 presents the constructed dataset of collected images and their subjective quality scores. Section 5 reviews the NR-IQA metrics considered in this work, and Section 6 introduces combined NR metrics and analyzes the resulting experimental findings. Finally, Section 7 concludes the paper.

2. Super-resolution of satellite LiDARs via diffusion models

Satellite-based LiDAR sensors are inherently constrained in spatial resolution due to factors such as limited photon flux and large beam footprints, which hinder detailed environmental monitoring. To overcome these limitations, super-resolution techniques aim to recover a high-resolution histogram of height-distributed counts tensor $\mathcal{X}_h \in \mathbb{R}^{c_h \times n_h \times m_h}$ from coarser low-resolution observations $\mathcal{Y}_l \in \mathbb{R}^{c_l \times n_l \times m_l}$. This process addresses an inherently ill-posed inverse problem, where numerous high-resolution tensors could potentially produce the same observed data.

We adopt a Bayesian approach to estimate $\mathcal{X}_h$ by maximizing the posterior probability $p(\mathcal{X}_h \mid \mathcal{Y}_l)$ (Ramirez-Jaime *et al.*, 2025a). Using Bayes' rule, this posterior is expressed as

$$p(\mathcal{X}_h \mid \mathcal{Y}_l) = \frac{p(\mathcal{Y}_l \mid \mathcal{X}_h)\, p(\mathcal{X}_h)}{p(\mathcal{Y}_l)}, \tag{1}$$

where $p(\mathcal{Y}_l \mid \mathcal{X}_h)$ represents the likelihood of observing the data given the high-resolution tensor and $p(\mathcal{X}_h)$ encodes prior assumptions about the underlying structure. Since the evidence $p(\mathcal{Y}_l)$ is constant with respect to $\mathcal{X}_h$, the maximization simplifies to minimizing the negative log-posterior:

$$\mathcal{X}_h^* = \arg\min_{\mathcal{X}_h} \left[-\log p(\mathcal{Y}_l \mid \mathcal{X}_h) - \log p(\mathcal{X}_h) \right]. \tag{2}$$

In this formulation, the data fidelity term $-\log p(\mathcal{Y}_l \mid \mathcal{X}_h)$ measures the consistency between the reconstructed $\mathcal{X}_h$ and the measurements $\mathcal{Y}_l$, mediated by the forward operator $\mathcal{A}(\cdot)$. This operator models the compressive sensing process in LiDAR, encompassing spatial downsampling through aggregation, subsampling of photon events, and application of randomized illumination patterns. Conversely, the regularization term $-\log p(\mathcal{X}_h)$ incorporates domain-specific priors to promote plausible reconstructions, such as sparsity, smoothness, or more sophisticated generative models that capture complex statistical distributions in natural scenes.

This Bayesian setup provides a principled balance between data adherence and prior-guided regularization, enabling robust super-resolution even under noisy or undersampled conditions. In subsequent sections, we explore advanced generative priors, including diffusion-based models, to further enhance the expressiveness of $p(\mathcal{X}_h)$.

3. Diffusion models for inverse problems

Building on the Bayesian framework for super-resolution introduced in Section 2, we now detail how diffusion models address the ill-posed inverse problem of recovering high-resolution LiDAR tensors $\mathcal{X}_h$ from low-resolution measurements $\mathcal{Y}_l$. This approach integrates learned data priors with physics-based measurement likelihoods, bridging the gap between the forward model (Section 2) and posterior estimation.

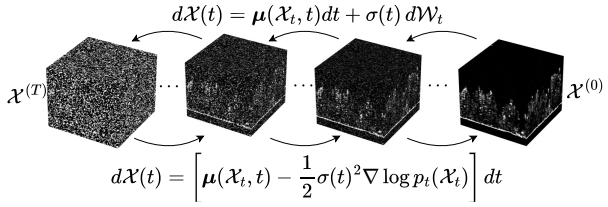


Fig. 2. Diffusion generative model for estimating $p(\mathcal{X})$. The forward process (top) uses an SDE to add noise to clean data $\mathcal{X}(0)$, shown as cubes transitioning from structured to noisy states. The reverse process (bottom), guided by a probability flow ODE and a neural network-estimated score, denoises $\mathcal{X}(T)$ back to $\mathcal{X}(0)$.

3.1. Diffusion generative models. Diffusion models learn to generate data by reversing a forward process that incrementally perturbs clean signals (Fig. 2). Let $\mathcal{X} \in \mathbb{R}^{c_2 \times n_2 \times m_2}$ represent high-resolution HHDC tensors. The forward diffusion process is governed by a stochastic differential equation (SDE) $\mathrm{d}\mathcal{X}_t = \boldsymbol{\mu}(\mathcal{X}_t, t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathcal{W}_t$, where $t \in [0, T]$ is the time, $\boldsymbol{\mu}(\cdot)$ and $\sigma(\cdot)$ are drift and diffusion coefficients, and $\mathcal{W}_t$ is tensor-valued Brownian motion (Song *et al*., 2020). The initial distribution $p_0(\mathcal{X})$ corresponds to the clean data distribution $p_{\text{data}}(\mathcal{X})$. As t increases, $\mathcal{X}_t$ becomes progressively perturbed. A deterministic counterpart to this process is described by the probability flow (PF) ordinary differential equation (ODE):

$$\mathrm{d}\mathcal{X}_t = \left[\boldsymbol{\mu}(\mathcal{X}_t, t) - \tfrac{1}{2}\sigma(t)^2 \nabla_{\mathcal{X}_t} \log p_t(\mathcal{X}_t)\right] \mathrm{d}t, \qquad (3)$$

which shares the same marginal distributions $p_t(\mathcal{X})$ as the SDE. Here, $\nabla_{\mathcal{X}_t} \log p_t(\mathcal{X}_t)$ is the score function—the gradient of the log-probability density of the diffused data. For complex LiDAR data distributions, this score is approximated via neural networks trained using score matching. Following Ho *et al*. (2020), we simplify the forward process by setting $\boldsymbol{\mu}(\mathcal{X}, t) = -\frac{1}{2}\beta(t)\mathcal{X}$ and $\sigma(t) = \sqrt{\beta(t)}$. This yields a variance-preserving diffusion where $p_t(\mathcal{X}) = \int p_{\text{data}}(\mathcal{X}_0)\mathcal{N}(\mathcal{X}; \sqrt{\alpha(t)}\mathcal{X}_0, (1-\alpha(t))\mathbf{I})\,\mathrm{d}\mathcal{X}_0$, with $\alpha(t) = \exp\left(-\int_0^t \frac{1}{2}\beta(s)\,\mathrm{d}s\right)$, and the PF ODE reduces to $\mathrm{d}\mathcal{X}_t = -\frac{1}{2}\beta(t)\left[\mathcal{X}_t + \nabla_{\mathcal{X}_t} \log p_t(\mathcal{X}_t)\right] \mathrm{d}t$. The score function $\nabla_{\mathcal{X}_t} \log p_t(\mathcal{X}_t)$ is estimated by a neural network $\mathbf{s}_\phi^{(h)}(\mathcal{X}_{h,t}, t)$ for high-resolution tensors.

3.2. Integrating measurement likelihoods. To solve the super-resolution inverse problem, we extend the PF ODE to incorporate the measurement likelihoods derived in Section 2. From Eqn. (1), the posterior estimation requires balancing the learned prior $p(\mathcal{X}_h)$ with the likelihood $p(\mathcal{Y}_l|\mathcal{X}_h)$. Given the forward model $\mathcal{A}(\cdot)$, which includes spatial aggregation, photon subsampling, and application of the random illumination pattern Ψ, the likelihood $p(\mathcal{Y}_l|\mathcal{X}_h)$ enforces consistency between the high-resolution tensor and the low-resolution measurements. To integrate this into the diffusion framework, we modify the PF ODE by adding a term that accounts for the likelihood score $\nabla_{\mathcal{X}_{h,t}} \log p(\mathcal{Y}_l|\mathcal{X}_{h,t})$. However, directly computing this score is challenging due to the stochastic nature of the forward model. Instead, we approximate it using the Kullback–Leibler divergence between the measurements $\mathcal{Y}_l$ and the expected low-resolution tensor obtained from $\mathcal{X}_{h,t}$ through the forward model (Ramirez-Jaime *et al*., 2025b). Specifically, we use $\nabla_{\mathcal{X}_{h,t}} D_{\text{KL}}\left(\mathcal{Y}_l \,\|\, \mathcal{G} * \hat{\mathcal{X}}_{h,t}\right)$, where $\hat{\mathcal{X}}_{h,t}$ is the estimated high-resolution tensor at time t, $\mathcal{G} * \hat{\mathcal{X}}_{h,t}$ represents the spatially aggregated tensor, and $\Psi \odot (\cdot)$ applies the illumination pattern. Thus, the augmented PF ODE for high-resolution reconstruction becomes

$$\begin{aligned}\frac{\mathrm{d}\mathcal{X}_{h,t}}{\mathrm{d}t} = -\frac{1}{2}\beta(t)\Big[&\mathcal{X}_{h,t} + \mathbf{s}_\phi^{(h)}(\mathcal{X}_{h,t}, t) \\ &+ \nabla_{\mathcal{X}_{h,t}} D_{\text{KL}}\left(\mathcal{Y}_l \,\|\, \mathcal{G} * \hat{\mathcal{X}}_{h,t}\right)\Big].\end{aligned} \qquad (4)$$

Solving this ODE from $t = T$ to $t = 0$ generates a high-resolution tensor $\mathcal{X}_h$ that is consistent with both the learned prior and the low-resolution measurements $\mathcal{Y}_l$.

4. SERC Subjective Quality LiDAR Image Dataset

A reliable quality assessment of the reconstructed HHDCs should be conducted using metrics that are as highly correlated with subjective quality perception as possible. Therefore, verification of available IQA approaches requires the development of a dataset containing the visualization of selected HHDC percentages in the form of digital images, along with their corresponding subjective quality scores. Such a database was created using 57 data cubes, reconstructed with the algorithm discussed in Section 2, intentionally selected to have varying quality. These data cubes originated from the raw data representing the Smithsonian Environmental Research Center (SERC) located in Maryland, USA (latitude 38.88° N, longitude 76.56° W), available on the National Ecological Observatory Network (NEON) website at `https://data.neonscience.org/data-products/DP1.30003.001`.

Each of the 57 HHDCs was used as the source of three grayscale and three additional pseudo-colored 192×192 pixel images, representing canopy height models (98th percentile), middle percentages (50%), and digital terrain models (2nd percentile). Pseudo-colored images are typically used for terrain and forest visualization purposes; therefore, they were included in the dataset. Sample reconstructed grayscale images, along

with their corresponding MOS values, are presented in Fig. 3. As can be seen, the MOS value is between 1 and 2 for the image pair shown in Fig. 3(a), where both images are extremely dissimilar, whilst the MOS is between 4 and 5 (close to maximal potential value) for the image pair presented in Fig. 3(d), for which images are very similar and only minor distortions can be noticed.

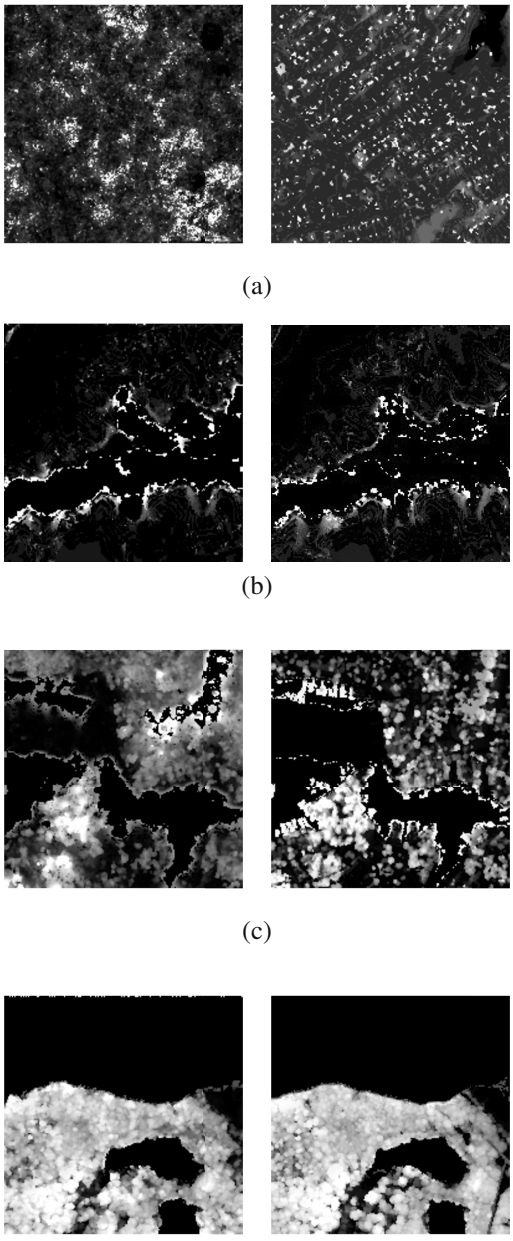
All images were subject to perceptual quality assessment by several independent volunteer observers from three countries (USA, Ukraine, and Poland), who assessed them by observing the pairs of reconstructed images and their reference "ground-truth" equivalents, using a 5-point scale (from 1: terrible to 5: the excellent). Considering the geographical distribution of volunteers, as well as various lighting and hardware conditions, all subjective experiments were preceded by the demonstration part, where some sample images along with their proposed quality scores were shown. All pairs of images were presented in a random order after a survey form, which contained relevant data related to lighting and display parameters (brightness, resolution, screen type, etc.), filled out by observers. Finally, at least 40 results were obtained for each image.

The volunteers assessing the quality of images were mainly ICT and IT students familiar with image processing. They were trained during the above-mentioned demonstration by being shown some examples of various quality images along with the corresponding subjective quality scores, which were agreed upon by a few experts who also participated in the subjective experiments.

After collecting the responses, using the designed software, their analysis was made to identify potential anomalous results, and those suggesting significantly differing visual differences compared to the others for the same image, or obtained after too long a response time (more than 10 seconds), were removed. Finally, between 40 and 50 scores were obtained for each pseudo-colored image from 54 original results submitted by volunteer observers, whereas 34 results were obtained for each of grayscale images.

The most relevant properties of the developed SERC Subjective Quality LiDAR Image Dataset are summarized in Table 1.

The assessment procedure was conducted independently for grayscale and pseudo-colored images. Pearson's correlations between the obtained mean opinion scores for grayscale and pseudo-colored images for individual percentiles are presented in Fig. 4. The overall correlation calculated for all 171 pairs of images representing all three percentiles is 0.9816. As may be observed, for two representations of the percentiles, correlations exceed 0.97, whereas the correlations of MOS values between different percentiles are significantly lower (shown as darker bars).



(a) (b) (c) (d)

Fig. 3. Sample pairs of reconstructed grayscale images (left) and their reference equivalents (right) from SERC Subjective Quality LiDAR Image Dataset: sample no. 24 (DTM), MOS = 1.677 (a), sample no. 37 (DTM), MOS = 3.135 (b), sample no. 41 (50th percentile), MOS = 2.710 (c), sample no. 53 (CHM), MOS = 4.333 (d).

The whole dataset is publicly available online at `https://doi.org/10.34808/zsv5-hn48` as well as on the website of the corresponding author at `https://okarma.zut.edu.pl/?id=dataset`.

Table 1. Selected properties of SERC Subjective Quality LiDAR Image Dataset.

HHDC percentiles	2nd, 50th, 98th
Image size in pixels	192 × 192
Number of image pairs	342
- pseudo-colored	171
- grayscale	171
Number of subsets (percentiles)	6
- pseudo-colored	3
- grayscale	3
Number of images in each subset	57
Number of scores per image	
- pseudo-colored	40–50
- grayscale	34
MOS range	1–5

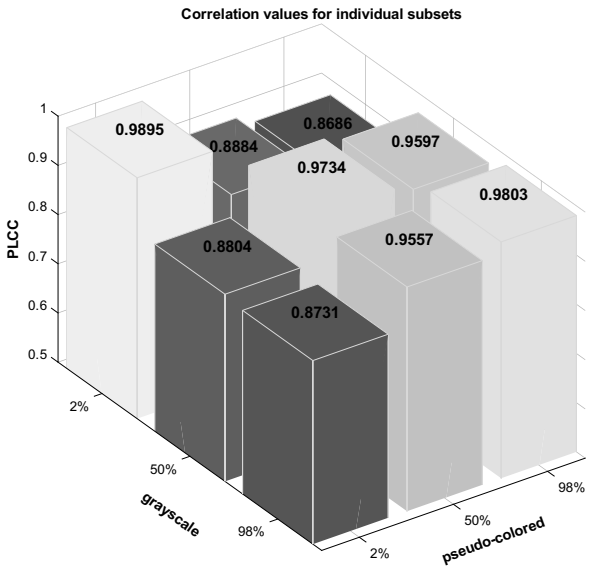


Fig. 4. PLCC values between the grayscale and pseudo-colored images representing individual percentiles.

5. No-reference image quality metrics for HHDC profiles

5.1. General assumptions. In contrast to general-purpose full-reference IQA metrics, "blind" methods are usually less universal and their correlation with subjective quality scores is significantly lower, and it is strongly dependent on the dataset. A similar situation occurs for more specialized datasets, including remote sensing images (Ieremeiev *et al.*, 2020; Rubel *et al.*, 2022), even though deep neural networks may be used for this purpose, assuming that the prior training of NR metrics should still be considered a "blind" IQA approach. Since NR-IQA metrics may be used without access to reference images, such models should predict the presence and amount of various types of image distortions. Therefore, their application to other image types may not always lead to satisfactory results.

To verify the usefulness of no-reference metrics for quality assessment of reconstructed HHDCs, several dozen such metrics have been calculated for the whole SERC Subjective Quality LiDAR Image Dataset, presented in Section 4. Many mainstream metrics, both FR and NR, are available in the PyTorch Toolbox for Image Quality Assessment (PyIQA) (Chen and Mo, 2022), although for some of the no-reference metrics their MATLAB® implementations have also been used. Since more detailed descriptions of most metrics, along with appropriate references to the original papers, are available, e.g., in the PyIQA documentation, further discussion will focus on the most relevant NR-IQA metrics, especially those demonstrated to be effective in the context of the application considered.

5.2. Overview of selected NR metrics. One of the most widely known "blind" IQA metrics is Natural Image Quality Evaluator (NIQE), based on the analysis of deviations from statistical regularities typical for natural images (Mittal *et al.*, 2013). The use of spatial-domain natural scene statistics (NSS) eliminates the need for model training; however, this metric is primarily applicable to quality assessment of photographs and other natural images. The NIQE index utilizes 36 NSS features from square patches, which are fitted to the multivariate Gaussian model (MVG) and further compared to the natural MVG one. The final NIQE value is computed as the distance between both models. Since various models may be applied, along with different representations of image data, available Python implementations of the metric slightly differ from the original MATLAB® code and produce different results.

Other popular no-reference IQA metrics include Perception-based Image Quality Evaluator (PIQE) (Venkatanath *et al.*, 2015) and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) (Mittal *et al.*, 2012). Both metrics are designed for natural images and are characterized by low computational complexity. While the PIQE metric relies on analyzing distortions within image blocks, BRISQUE utilizes hand-crafted natural scene statistics (NSS) data, similarly to the NIQE metric. On the other hand, NRQM (Ma *et al.*, 2017) utilizes the pyramid representation for global and local feature extraction; however, this metric was developed primarily for quality assessment of super-resolution images. The features considered include global and local frequency variations as well as spatial discontinuity, which are the inputs for the final two-stage regression model.

Recently proposed novel NR-IQA metrics employ more sophisticated approaches, including deep neural

networks trained using available databases, such as PIPAL (Jinjin *et al*., 2020), LIVE-FB Large-Scale Social Picture Quality Database (FLIVE) (Ying *et al*., 2020), containing nearly 40,000 real-world images, or LIVE Challenge (CLIVE) (Ghadiyaram and Bovik, 2016). One example of such metrics, known as MANIQA—Multi-dimension Attention Network for No-Reference Image Quality Assessment (Yang *et al*., 2022), was developed mainly for the assessment of images with distortions typical for generative adversarial networks (GANs). This approach is based on the use of four blocks, namely, Vision Transformer (ViT) for feature extraction, two blocks used as attention mechanisms (transposed attention block and scale swin transformer block), and the patch-weighting block utilizing a dual branch structure, employed for the calculation of the final quality score. It is worth noting that the MANIQA metric was ranked first during the NTIRE 2022 NR-IQA Challenge.

An earlier multi-scale NR metric based on neural networks is MUSIQ (Multi-scale Image Quality Transformer) (Ke *et al*., 2021), which was developed to eliminate the need for resizing and cropping input images to meet neural network requirements. Since these operations may affect image quality, the use of a multi-scale image representation is justified.

Another example of AI-based "blind" metrics is ARNIQA (Learning Distortion Manifold for Image Quality Assessment) Agnolucci *et al*. (2024), which belongs to the group of self-supervised approaches. The assessed images are initially degraded with the use of a large variety of randomly selected degradation patterns, and then the model is trained to maximize the similarity between different patches of equally distorted images, independently of the image contents, followed by a linear regression. The novelty of this approach lies in the analysis of the distortion manifold, where representations of images containing similar distortions are located close to each other in the manifold space even if these images are visually significantly different. Finally, their representations in the distortion manifold space are mapped to quality scores using a simple linear regression.

A simple network for quality prediction from both aesthetic and technical points of view was proposed by Talebi and Milanfar (2018), who developed the Neural Image Assessment (NIMA) metric utilizing the image classifier network with its last layer replaced by a fully-connected layer. Although this model, based on the VGG16 network pre-trained on ImageNet, was further trained using the AVA (Aesthetic Visual Analysis) dataset (Murray *et al*., 2012), containing more than 255,000 images, its applicability for IQA purposes, considering technical quality, was also verified. In the original paper, not only the VGG16 architecture was employed, but additional experiments were also provided using Inception-v2 and MobileNet.

Table 2. Summary of the discussed NR-IQA metrics in chronological order.

BLUR	Crete *et al.*, 2007
BRISQUE	Mittal *et al.*, 2012
NIQE	Mittal *et al.*, 2013
PIQE	Venkatanath *et al.*, 2015
NRQM	Ma *et al.*, 2017
NIMA	Talebi and Milanfar, 2018
MUSIQ	Ke *et al.*, 2021
MANIQA	Yang *et al.*, 2022
CLIP-IQA	Wang *et al.*, 2023
ARNIQA	Agnolucci *et al.*, 2024

A different approach, known as Contrastive Language–Image Pre-training (CLIP), was proposed as the result of the combination of text and image analysis for model training (Wang *et al*., 2023). The CLIP-IQA metric utilizes the cosine similarity between the provided images and pre-defined prompts describing the image quality. The additional extension of the metric based on the use of context optimization (Zhou *et al*., 2022), along with the application of standard backpropagation for updating the prompts, led to the CLIP-IQA+ metric, characterized by a better ability to generalize across various distortions and datasets.

Although certain modern NR-IQA metrics are not well-suited for specific types of images due to their intended application domains, the availability of Python reimplementations of many FR-IQA and NR-IQA metrics, along with their modifications, in the PyTorch Toolbox for Image Quality Assessment (PyIQA) (Chen and Mo, 2022) enabled the testing of all available "blind" metrics in the conducted experiments. Only those that consistently produced incorrect results were excluded from further analysis. An example of a metric available in the toolbox considered that cannot be applied to our task is DSL-FIQA (Chen *et al*., 2024), which is intended for no-reference facial image quality assessment.

5.3. Experimental results. The dynamic progress in the development of new metrics and increasing availability of modern graphics processing units (GPUs) have led to a growing interest in multi-scale IQA approaches. Unfortunately, these methods are often troublesome or even impossible to apply in the domain considered. Some other metrics could not be used due to the required minimal image size, GPU acceleration, or time constraints, whereas some others produced incorrect results. Therefore, some of the available metrics that could not be used were omitted. Finally, a set of 62 metrics was used in the experiments. For each metric, the Pearson linear correlation coefficient (PLCC) between

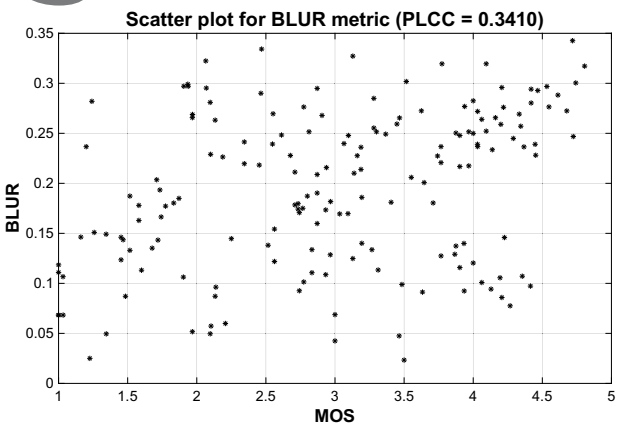


Fig. 5. Scatter plot obtained for the elementary metric demonstrating the highest correlation with MOS values.

Table 3. PLCC and SROCC values calculated between the ten "best" elementary NR metrics and subjective MOS values for 171 grayscale images representing three HHDC percentiles.

Metric	PLCC	SROCC
BLUR	0.3410	0.3239
ARNIQA (CSIQ)	0.3351	0.2618
ARNIQA (LIVE)	0.3238	0.2830
ARNIQA (FLIVE)	0.2928	0.2266
ARNIQA (KADID)	0.2923	0.2805
MUSIQ (SPAQ)	−0.2708	−0.2632
NIQE (PyIQA-M)	−0.2539	−0.2496
NRQM	0.2493	0.2286
ARNIQA (CLIVE)	0.2399	0.1808
NIMA (SPAQ)	−0.2313	−0.1867

its values and subjective scores (MOS) was calculated, representing the accuracy of the quality prediction using a given metric. Similarly, Spearman rank order correlation coefficient (SROCC) values, which reflect the prediction monotonicity, were calculated. We aim to employ NR metrics that exhibit high absolute values of the PLCC or, at the very least, large absolute SROCC values.

The values obtained for the "best" NR metrics, further referred to as elementary, are presented in Table 3. Since the highest correlation was obtained for the BLUR metric (Crete *et al*., 2007), a scatter plot, illustrating the weak correlation between its values and the MOS, is presented in Fig. 5 for all 171 grayscale images (representing three percentages) used in the experiments.

Analyzing the results obtained for the elementary NR metrics, it can be noted that some of them have positive correlation values, whereas the others have negative PLCC and SROCC values, which depend on the interpretation of the metric's values (the higher the better versus the lower the better). Nevertheless, the absolute values of correlations lower than 0.35 indicate no direct relation between the objective metrics and MOS values. Therefore, the highest PLCC value was achieved by a relatively simple BLUR metric, designed exclusively for blurry images, based on the comparison of variations and differences of the image and its blurred version in horizontal and vertical directions.

6. Combined metrics and discussion of experimental results

6.1. Methodology of the experiments. Since the correlation of elementary NR metrics with MOS values is very low, and the metrics considered are specialized, their application for the objective evaluation of the reconstructed HHDC profiles requires the use of more than a single elementary metric. An efficient method for the combination of elementary metrics is the employment of metrics with differing properties, also for remote sensing images, as shown by Ieremeiev *et al*. (2020) and Rubel *et al*. (2022), which allows complementing each other's strengths and weaknesses.

The combined metrics, investigated in the paper, are defined as

$$\mathrm{CM}_N = \sum_{n=1}^{N} \left(a_n \cdot Q_n^{b_n}\right) + a_{N+1} \cdot \prod_{n=1}^{N} Q_n^{c_n}, \qquad (5)$$

where CM_N stands for the combined metric with optimized weights a_n, b_n, c_n, and a_{N+1}. The metric is constructed from N elementary metrics, and values of the n-th elementary metric are denoted as Q_n. A similar approach has also been successfully applied, e.g., for binary images (Okarma and Kopytek, 2023) as well as full-reference metrics (Ieremeiev *et al*., 2025). The optimization is made by maximizing the absolute value of the PLCC between MOS values and the combined metric CM_N. For this purpose, the part of the dataset containing 171 grayscale images (57 images for each of three percentiles considered) was used, applying the MATLAB® *fminsearch* function, along with popular gradient-based optimization methods.

6.2. Results and discussion. The combination of two elementary metrics, obtained after the optimization of weighting parameters in Eqn. (5), denoted as CM_2, contains two various implementations of the NIQE metric. One of them is delivered in MATLAB® (version R2025a), whereas the other is available in the PyIQA toolbox. It is worth noting that the "best" combination of two metrics is not that of two elementary ones with the highest PLCC values, as their individual correlations with MOS values are only −0.1591 and −0.2539, respectively. The optimized combination of three metrics (CM_3) is built by adding the MANIQA metric trained on the PIPAL dataset.

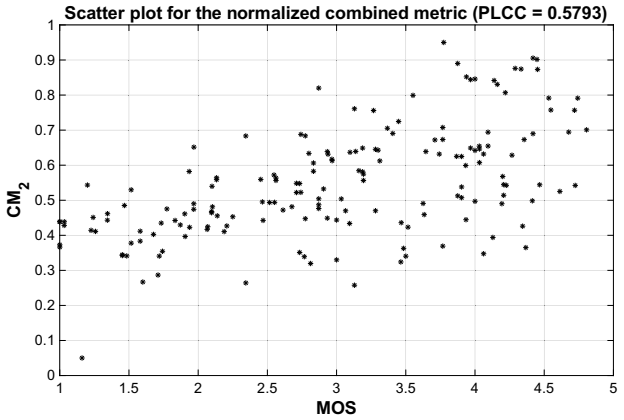


Fig. 6. Scatter plot obtained for the normalized combined metric constructed from two elementary metrics.

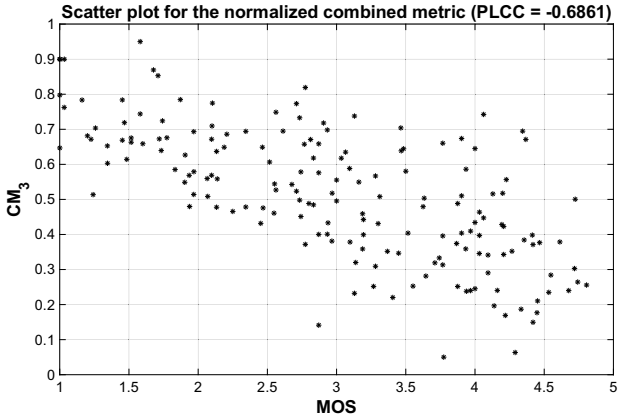


Fig. 7. Scatter plot obtained for the normalized combined metric constructed from three elementary metrics.

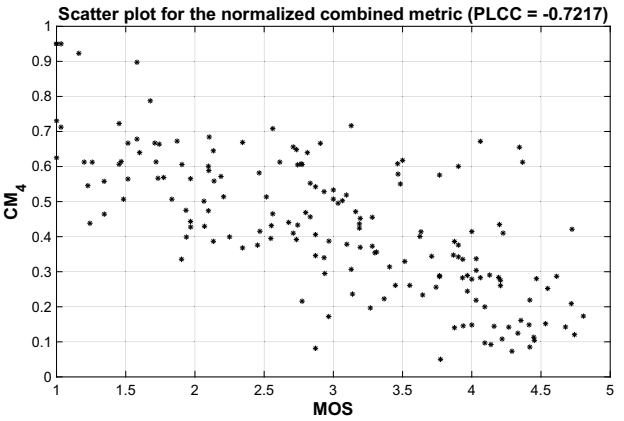


Fig. 8. Scatter plot obtained for the normalized combined metric constructed from four elementary metrics.

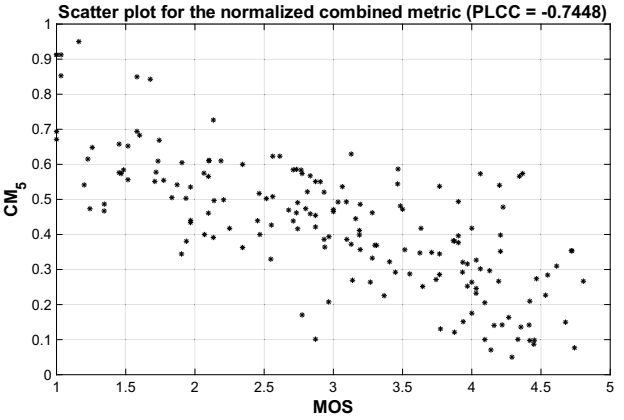


Fig. 9. Scatter plot obtained for the normalized combined metric constructed from five elementary metrics.

The combined metrics obtained for the combinations of more elementary metrics ($N > 3$) were also obtained for the metrics used in the CM_{N-1} metric with an additional one, since the other combinations led to lower correlations with MOS values. The additional metrics include ARNIQA (trained on CLIVE) for CM_4, ARNIQA (trained on FLIVE) for CM_5, and CLIP-IQA+ (trained on ResNet model RN50_512) for CM_6. The obtained correlation values for the "best" combined metrics are presented in Table 4, whereas the respective scatter plots are illustrated in Figs. 6–10.

Additionally, to illustrate the generalization abilities of the proposed approach, the average results obtained using a random train-test split are presented in Table 5. These PLCC and SROCC values were calculated assuming typical random splits of the dataset into 80% training and 20% testing parts. Then, the optimization of parameters of the combined metrics was performed to maximize the absolute PLCC for the training subset, and the additional correlation values were calculated for the "unseen" testing images. All presented results show average values obtained from 100 independent executions of the random splitting and optimization procedure.

These results additionally confirmed the advantages of using combined metrics, as they are much better than those possible to obtain using elementary metrics alone. The average results achieved for 100 independent executions, obtained for a smaller number of elementary metrics used in the combination, are even slightly higher than those for the entire dataset. This effect may be observed particularly using simpler models for relatively small testing subsets. When more elementary metrics are added, an increase in the PLCC and SROCC values may be observed for the training subset with a "flattening effect" for the testing subset (no further improvement is observed when adding more elementary metrics). This phenomenon is caused by the fact that our task involves maximizing the correlation with subjective scores for a dataset of a relatively small size.

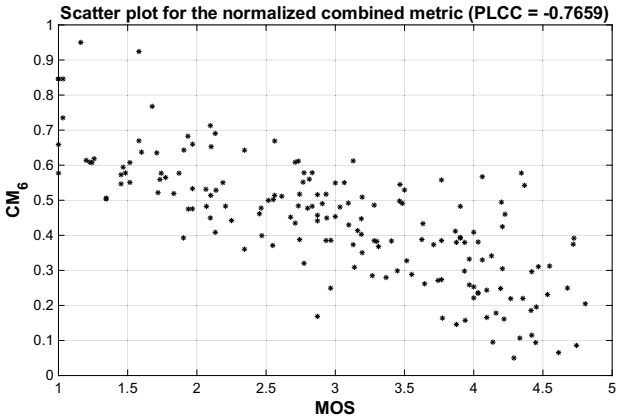


Fig. 10. Scatter plot obtained for the normalized combined metric constructed from six elementary metrics.

6.3. Verification for pseudo-colored images. The additional verification of the obtained results was performed using pseudo-colored images representing the same reconstructed HHDCs which volunteer observers independently assessed. As presented in Fig. 4, the obtained MOS results are highly correlated with those received for grayscale images; however, the additional analysis was conducted for each of the three percentiles. All PLCC and SROCC values were calculated using the same weighting coefficients in Eqn. (5), obtained as the results of their optimization for 171 grayscale images. The resulting correlations with MOS values are presented in Tables 6 and 7. In general, all respective correlations obtained for individual percentages and consecutive combined metrics for grayscale and pseudo-colored images are similar. An interesting observation is related to the relatively poor correlation of the CM_2 metric with MOS values for DTMs (2nd percentage) and significantly higher values for CHMs (98th percentage). For the metrics combined from more elementary ones, this difference is smaller, and the PLCC and SROCC values observed for DTMs are even higher than for the 50th percentile. Nevertheless, their highest values are always observed for the CHMs; therefore, their influence on the perceived quality seems to be essential. This observation may be particularly relevant for further research related to the use of various IQA metrics within loss functions during the reconstruction process.

6.4. Ablation study. Although the combination of elementary metrics according to Eqn. (5) yields significantly better results than those obtained for elementary metrics, an additional verification of the necessity of all parameters may be conducted. Therefore, two simplified models were analyzed that utilized exclusively the weighted sum or only the weighted product (without the unnecessary parameter a_{N+1}). As can be observed, the obtained correlation coefficients for the weighted sum are significantly better than those achieved for the weighted product, partly due to the higher number of weighting coefficients that are subject to optimization. The optimization procedure was conducted again for each of the simplified models, applying the same sets of elementary metrics. Nevertheless, the obtained correlations are noticeably worse than those presented in Table 4, hence the usefulness of the full model (5) is considered confirmed. In particular, it is possible to achieve higher correlations with MOS values using

Table 4. PLCC and SROCC values calculated between the combined NR metrics and subjective MOS values for 171 grayscale images representing three HHDC percentiles.

Number of elementary metrics	PLCC	SROCC
2 (CM_2)	0.5793	0.5807
3 (CM_3)	−0.6861	−0.6832
4 (CM_4)	−0.7217	−0.7158
5 (CM_5)	−0.7448	−0.7389
6 (CM_6)	−0.7659	−0.7663

Table 5. Average PLCC and SROCC values calculated for training (80%) and testing (20%) subsets randomly selected from 171 grayscale images: mean values for 100 independent executions.

Combined metric	Training subsets		Testing subsets	
	PLCC	SROCC	PLCC	SROCC
CM_2	0.6179	0.6273	0.6664	0.6875
CM_3	−0.7116	−0.7095	−0.7333	−0.7365
CM_4	−0.7362	−0.7350	−0.7355	−0.7427
CM_5	−0.7595	−0.7574	−0.7332	−0.7340
CM_6	−0.7736	−0.7750	−0.7253	−0.7422

Table 6. PLCC values calculated between the combined NR metrics and MOS values for individual HHDC percentiles.

Grayscale images				
Percentile	2nd	50th	98th	All
CM_2	0.3531	0.5568	0.7126	0.5793
CM_3	−0.6211	−0.6214	−0.7812	−0.6861
CM_4	−0.7057	−0.6264	−0.8032	−0.7217
CM_5	−0.7602	−0.6695	−0.7785	−0.7448
CM_6	−0.7531	−0.7063	−0.8300	−0.7659
Pseudo-colored images				
Percentile	2nd	50th	98th	All
CM_2	0.3727	0.5811	0.7147	0.5967
CM_3	−0.6178	−0.6106	−0.7790	−0.6840
CM_4	−0.7063	−0.6217	−0.8105	−0.7241
CM_5	−0.7540	−0.6554	−0.7959	−0.7463
CM_6	−0.7451	−0.6843	−0.8438	−0.7617

Table 7. SROCC values calculated between the combined NR metrics and MOS values for individual HHDC percentiles.

	Grayscale images			
Percentile	2nd	50th	98th	All
CM_2	0.3828	0.5406	0.7403	0.5807
CM_3	−0.6370	−0.6010	−0.8058	−0.6832
CM_4	−0.6779	−0.6050	−0.8148	−0.7158
CM_5	−0.7807	−0.6536	−0.7839	−0.7389
CM_6	−0.7679	−0.6812	−0.8452	−0.7663
	Pseudo-colored images			
Percentile	2nd	50th	98th	All
CM_2	0.3855	0.5567	0.7514	0.5958
CM_3	−0.6222	−0.5883	−0.8050	−0.6862
CM_4	−0.6826	−0.6001	−0.8172	−0.7188
CM_5	−0.7707	−0.6347	−0.7999	−0.7427
CM_6	−0.7578	−0.6534	−0.8559	−0.7648

Table 8. PLCC and SROCC values calculated between the simplified combined NR metrics and subjective MOS values for 171 grayscale images from SERC Subjective Quality LiDAR Image Dataset.

Metric	PLCC	SROCC
CM_{2p} (weighted product)	−0.4436	−0.4533
CM_{2s} (weighted sum)	0.5653	0.5732
CM_{3p} (weighted product)	−0.4437	−0.4521
CM_{3s} (weighted sum)	0.5660	0.5744
CM_{4p} (weighted product)	0.4901	0.5130
CM_{4s} (weighted sum)	0.6234	0.6389
CM_{5p} (weighted product)	0.5322	0.5223
CM_{5s} (weighted sum)	0.6374	0.6453
CM_{6p} (weighted product)	0.5543	0.5555
CM_{6s} (weighted sum)	0.6550	0.6725

the full model based on only three elementary metrics (CM_3) than with any of the simplified models (only the weighted sum or only the weighted product), even when employing six elementary metrics (CM_{6p} and CM_{6s}). It is also worth noting that the computational complexity of the full model is not significantly higher compared to the simplified models, as the longest calculation time is related to the calculations of elementary metrics, and their combination requires only several multiplications and additions; therefore, it is much faster. To decrease the calculation time, it is also possible to compute elementary metrics in parallel.

7. Concluding remarks

The proposed application of combined no-reference image quality metrics for assessing the reconstructed satellite LiDAR data, which represent forest environments, ensures a significantly increased correlation with perceived quality of the reconstructed images, particularly illustrating canopy height models. Due to the optimization of the parameters of the weighted sum and weighted product, even the use of three metrics makes it possible to achieve correlations with MOS values exceeding 0.7, and a combination of six metrics leads to the PLCC and SROCC values over 0.76.

The verification of the suitability of the NR-IQA metrics for the application considered and presented in the paper was possible thanks to the development of a novel dataset containing reconstructed HHDCs with reference data, as well as their mean opinion scores, collected from volunteer observers from three countries.

Since the application of the NR metrics is possible even without access to the reference data, they may be particularly useful for the assessment of the data reconstruction results for compressive sensing methods. Such techniques are used due to the necessity of fast data transmission during satellite flight over a specified area, when the transmission time is limited.

Another important application of such NR-IQA metrics may be related to the prediction of classification accuracy of satellite images (Wang *et al*., 2020) as well as in similar tasks when reconstructed LiDAR data may be used as the input.

Regardless of these applications, combined NR-IQA metrics may also be useful within the loss function employed in the data reconstruction process with the diffusion-based approach. Another interesting possibility for further extension of the whole reconstruction process may be related to the use of epistemic fuzzy data analysis methods (Grzegorzewski and Romaniuk, 2022). Nevertheless, these directions of research may be explored as part of our further work.

Acknowledgment

This research was funded in part by the US National Science Foundation (NSF) under the grant no. 2404740, the Science & Technology Center in Ukraine (STCU) within the agreement no. 7116, and the National Science Centre (NCN) in Poland under the grant no. 2023/05/Y/ST6/00197 within the joint IMPRESS-U project entitled *EAGER IMPRESS-U: Exploratory Research on Generative Compression for Compressive LiDAR*.

References

Agnolucci, L., Galteri, L., Bertini, M. and Del Bimbo, A. (2024). ARNIQA: Learning distortion manifold for image quality assessment, *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, USA*, pp. 188–197, DOI: 10.1109/WACV57701.2024.00026.

Chen, C. and Mo, J. (2022). IQA-PyTorch: Pytorch toolbox for image quality assessment, Available online at `https://github.com/chaofengc/IQA-PyTorch`.

Chen, W.-T., Krishnan, G., Gao, Q., Kuo, S.-Y., Ma, S. and Wang, J. (2024). DSL-FIQA: Assessing facial image quality via dual-set degradation learning and landmark-guided transformer, *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA*, pp. 2931–2941, DOI: 10.1109/CVPR52733.2024.00283.

Crete, F., Dolmiere, T., Ladret, P. and Nicolas, M. (2007). The blur effect: Perception and estimation with a new no-reference perceptual blur metric, *in* B.E. Rogowitz *et al.* (Eds), *Human Vision and Electronic Imaging XII*, Vol. 6492, International Society for Optics and Photonics, Bellingham, p. 64920I, DOI: 10.1117/12.702790.

Dubayah, R., Blair, J., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S. Jantz, P., Marselis, S., Patterson, P.L., Qi, W. and Silva, C. (2020). The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography, *Science of Remote Sensing* **1**: 100002, DOI: 10.1016/j.srs.2020.100002.

Fang, Y., Zhu, H., Zeng, Y., Ma, K. and Wang, Z. (2020). Perceptual quality assessment of smartphone photography, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA*, pp. 3674–3683, DOI: 10.1109/CVPR42600.2020.00373.

Ghadiyaram, D. and Bovik, A.C. (2016). Massive online crowdsourced study of subjective and objective picture quality, *IEEE Transactions on Image Processing* **25**(1): 372–387, DOI: 10.1109/TIP.2015.2500021.

Grzegorzewski, P. and Romaniuk, M. (2022). Bootstrap methods for epistemic fuzzy data, *International Journal of Applied Mathematics and Computer Science* **32**(2): 285–297, DOI: 10.34768/amcs-2022-0021.

Ho, J., Jain, A. and Abbeel, P. (2020). Denoising diffusion probabilistic models, *in* H. Larochelle *et al.* (Eds), *Advances in Neural Information Processing Systems (NeurIPS 2020)*, Vol. 33, Curran Associates, Red Hook, pp. 6840–6851, DOI: 10.5555/3495724.3496298.

Ieremeiev, O., Lukin, V., Okarma, K. and Egiazarian, K. (2020). Full-reference quality metric based on neural network to assess the visual quality of remote sensing images, *Remote Sensing* **12**(15): 2349, DOI: 10.3390/rs12152349.

Ieremeiev, O., Lukin, V., Ramirez-Jaime, A., Arce, G.R., Kopytek, M., Lech, P. and Okarma, K. (2025). Objective and subjective quality assessment of forest landscapes reconstructed using generative diffusion models from compressed satellite LiDAR data, *IEEE Access* **13**: 213107–213123, DOI: 10.1109/ACCESS.2025.3644825.

Jinjin, G., Haoming, C., Haoyu, C., Xiaoxing, Y., Ren, J.S. and Chao, D. (2020). PIPAL: A large-scale image quality assessment dataset for perceptual image restoration, *in* A. Vedaldi*et al.* (Eds), *Computer Vision—ECCV 2020*, Springer International Publishing, Cham, pp. 633–651, DOI: 10.1007/978-3-030-58621-8_37.

Ke, J., Wang, Q., Wang, Y., Milanfar, P. and Yang, F. (2021). MUSIQ: Multi-scale image quality transformer, *2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada*, pp. 5128–5137, DOI: 10.1109/ICCV48922.2021.00510.

Kulczycki, P., Franus, K. and Charytanowicz, M. (2024). A quality index for detection of atypical elements (outliers), *International Journal of Applied Mathematics and Computer Science* **34**(3): 439–451, DOI: 10.61822/amcs-2024-0031.

Larson, E. and Chandler, D. (2010). Most apparent distortion: Full-reference image quality assessment and the role of strategy, *Journal of Electronic Imaging* **19**(1): 011006, DOI: 10.1117/1.3267105.

Lin, H., Hosu, V. and Saupe, D. (2019). KADID-10k: A large-scale artificially distorted IQA database, *2019 11th International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany*, pp. 1–3, DOI:10.1109/QoMEX.2019.8743252.

Ma, C., Yang, C.-Y., Yang, X. and Yang, M.-H. (2017). Learning a no-reference quality metric for single-image super-resolution, *Computer Vision and Image Understanding* **158**: 1–16, DOI: 10.1016/j.cviu.2016.12.009.

Mittal, A., Moorthy, A.K. and Bovik, A.C. (2012). No-reference image quality assessment in the spatial domain, *IEEE Transactions on Image Processing* **21**(12): 4695–4708, DOI: 10.1109/TIP.2012.2214050.

Mittal, A., Soundararajan, R. and Bovik, A.C. (2013). Making a "completely blind" image quality analyzer, *IEEE Signal Processing Letters* **20**(3): 209–212, DOI: 10.1109/LSP.2012.2227726.

Murray, N., Marchesotti, L. and Perronnin, F. (2012). AVA: A large-scale database for aesthetic visual analysis, *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA*, pp. 2408–2415, DOI: 10.1109/CVPR.2012.6247954.

Okarma, K. and Kopytek, M. (2023). A hybrid method for objective quality assessment of binary images, *IEEE Access* **11**: 63388–63397, DOI: 10.1109/ACCESS.2023.3289168.

Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F. and Jay Kuo, C.-C. (2015). Image database TID2013: Peculiarities, results and perspectives, *Signal Processing: Image Communications* **30**: 57–77, DOI: 10.1016/j.image.2014.10.009.

Popescu, S.C., Zhao, K., Neuenschwander, A. and Lin, C. (2011). Satellite LiDAR vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level, *Remote Sensing of Environment* **115**(11): 2786–2797, DOI: 10.1016/j.rse.2011.01.026.

Ramirez-Jaime, A., Arce, G.R., Porras-Diaz, N., Ieremeiev, O., Rubel, A., Lukin, V., Kopytek, M., Lech, P., Fastowicz, J. and Okarma, K. (2025a). Generative diffusion models for compressed sensing of satellite LiDAR data: Evaluating

image quality metrics in forest landscape reconstruction, *Remote Sensing* **17**(7): 1215, DOI: 10.3390/rs17071215.

Ramirez-Jaime, A., Pena-Pena, K., Arce, G.R., Harding, D., Stephen, M. and MacKinnon, J. (2024). Hyperheight LiDAR compressive sampling and machine learning reconstruction of forested landscapes, *IEEE Transactions on Geoscience and Remote Sensing* **62**: 1–16, DOI: 10.1109/TGRS.2024.3356389.

Ramirez-Jaime, A., Porras-Diaz, N., Arce, G.R. and Stephen, M. (2025b). Super-resolved 3D satellite LiDAR imaging of Earth via generative diffusion models, *IEEE Transactions on Geoscience and Remote Sensing* **63**: 1–19, DOI: 10.1109/TGRS.2025.3543670.

Rubel, A., Ieremeiev, O., Lukin, V., Fastowicz, J. and Okarma, K. (2022). Combined no-reference image quality metrics for visual quality assessment optimized for remote sensing images, *Applied Sciences* **12**(4): 1986, DOI: 10.3390/app12041986.

Sheikh, H., Sabir, M. and Bovik, A. (2006). A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on Image Processing* **15**(11): 3440–3451, DOI: 10.1109/TIP.2006.881959.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S. and Poole, B. (2020). Score-based generative modeling through stochastic differential equations, *arXiv* 2011.13456, DOI: 10.48550/arXiv.2011.13456.

Talebi, H. and Milanfar, P. (2018). NIMA: Neural image assessment, *IEEE Transactions on Image Processing* **27**(8): 3998–4011, DOI: 10.1109/TIP.2018.2831899.

Venkatanath, N., Praneeth, D., Maruthi, C.B., Channappayya, S.S. and Medasani, S.S. (2015). Blind image quality evaluation using perception based features, *2015 21st National Conference on Communications (NCC), Mumbai, India*, pp. 1–6, DOI: 10.1109/NCC.2015.7084843.

Wang, J., Chan, K.C. and Loy, C.C. (2023). Exploring CLIP for assessing the look and feel of images, *Proceedings of the AAAI Conference on Artificial Intelligence* **37**(2): 2555–2563, DOI: 10.1609/aaai.v37i2.25353.

Wang, Y., Zhang, D. and Dai, G. (2020). Classification of high resolution satellite images using improved U–Net, *International Journal of Applied Mathematics and Computer Science* **30**(3): 399–413, DOI: 10.34768/amcs-2020-0030.

Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**(4): 600–612, DOI: 10.1109/TIP.2003.819861.

Weitkamp, C. (2005). *LiDAR: Range-Resolved Optical Remote Sensing of the Atmosphere*, Springer Science & Business Media, New York, DOI: 10.1007/b106786.

Yang, J., Zheng, H., Ma, Y., Zhao, P., Zhou, H., Li, S. and Wang, X.H. (2023). Background noise model of spaceborne photon-counting LiDARs over oceans and aerosol optical depth retrieval from ICESat-2 noise data, *Remote Sensing of Environment* **299**: 113858, DOI: 10.1016/j.rse.2023.113858.

Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J. and Yang, Y. (2022). MANIQA: Multi-dimension attention network for no-reference image quality assessment, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, USA*, pp. 1190–1199, DOI: 10.1109/CVPRW56347.2022.00126.

Ying, Z., Niu, H., Gupta, P., Mahajan, D., Ghadiyaram, D. and Bovik, A. (2020). From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA*, pp. 3572–3582, DOI: 10.1109/CVPR42600.2020.00363.

Zhou, K., Yang, J., Loy, C.C. and Liu, Z. (2022). Learning to prompt for vision-language models, *International Journal of Computer Vision* **130**(9): 2337–2348, DOI: 10.1007/s11263-022-01653-1.



Krzysztof Okarma is a graduate of the Szczecin University of Technology, Poland, in electronics and telecommunication (1999) and computer science (2001), and holds a PhD in electrical engineering (2003) and a habilitation in automation and robotics (2013), as well as the professorial title (2024). He has been with the Department of Signal Processing and Multimedia Engineering of the West Pomeranian University of Technology in Szczecin: from 2016 to 2024 as the head of department, and also as the dean of the Faculty of Electrical Engineering. He is the co-author of over 250 papers and holds two patents. His scientific interests include computer and machine vision, image processing and analysis, as well as signal processing.



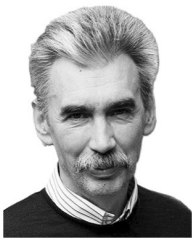
Mateusz Kopytek graduated in automatic control and robotics from the Faculty of Electrical Engineering of the West Pomeranian University of Technology in Szczecin, Poland, in 2024. He is a beneficiary of the ZUT Highfliers School (Szkoła Orłów ZUT) project within the framework of a program of the Polish Minister of Education and Science, co-financed by the European Social Fund. His engineering thesis is related to quality assessment of stitched panoramic images while his master's thesis to binary image quality assessment, both defended under the supervision of Prof. Krzysztof Okarma. Currently, he is a PhD student at ZUT Doctoral School.



Piotr Lech has been with the Department of Signal Processing and Multimedia Engineering of the Faculty of Electrical Engineering at the West Pomeranian University of Technology in Szczecin, Poland. He obtained an engineering degree in electrical engineering from the Faculty of Electrical Engineering of the Szczecin University of Technology in 1993 and a master's degree in 1995. In 2004, he was granted a PhD in technical sciences (electrical engineering). His scientific interests include industrial Internet of things applications, with particular emphasis on machine vision and artificial intelligence.



Oleg Ieremeiev received his master's degree in electronics and telecommunication from the National Aerospace University "Kharkiv Aviation Institute" (KhAI), Ukraine, in 2009, and his Candidate of Technical Science degree (PhD equivalent) in radioengineering devices and means of telecommunications from the same institution in 2015. He has been an associate professor at the Department of Information and Communication Technologies of the National Aerospace University since 2017. He is an author or a co-author of more than 50 journal articles and conference papers in the fields of image processing and analysis, as well as image visual quality assessment, including machine learning methods.



Vladimir V. Lukin graduated from the National Aerospace University "Kharkiv Aviation Institute" (Ukraine) in 1983. He also holds Candidate (1988) and Doctor of Technical Science (2002) degrees. In 2007, he received a professorial position. Since 1980, he has been with the Department of Transmitters, Receivers, and Signal Processing (now Information-Communication Technologies), since 2016 as the department head. He is a co-author of more than 800 journal and conference papers and book chapters, over 500 of which are indexed in Scopus. His research interests include quality assessment, image denoising and lossy compression, as well as remote sensing data processing.



Andres Ramirez-Jaime is a PhD candidate in the Electrical and Computer Engineering Department at the University of Delaware (UD). He received his BS and MSc degrees in electronic engineering from Universidad de Los Andes in Colombia in 2013 and 2016, respectively. He was also a visiting scientist at the Institute of Robotics and Industrial Computing at the Technical University of Catalonia in Barcelona in the years 2015–2016. Before joining the UD, he was a professor in the Faculty of Engineering at the University of La Sabana in Colombia, being the leader of the UNISABANA Herons SPL robotics team. His research focuses on generative models, computer vision, and deep learning.



Nestor Porras-Diaz received his BSc degree in computer engineering from Universidad de La Sabana, Bogotá, Colombia, in 2022. He is currently pursuing a PhD degree with the Department of Electrical and Computer Engineering, University of Delaware, USA. Nestor was the student-lead member of the UNISABANA Herons SPL Robotics Team and the UNISABANA Herons Ventilator Team in the years 2018–2021. His research interests include deep learning, computer vision, and natural language processing.



Gonzalo R. Arce is currently a Charles Black Evans Distinguished Professor of Electrical and Computer Engineering and a J.P. Morgan-Chase Senior Faculty Fellow with the Institute of Financial Services Analytics, University of Delaware. He is the editor-in-chief of *IEEE Transactions on Computational Imaging*. He holds 25 US patents and is the co-author of four books. His research interests include computational imaging, data science, and machine learning. He held the 2010 and 2017 Fulbright-Nokia Distinguished Chair of ICT with Aalto University, Espoo, Finland. He has received an NSF Research Initiation Award and has been elected Fellow of the IEEE, OPTICA, SPIE, AAIA, and the National Academy of Inventors.

Received: 31 October 2025
Revised: 28 January 2026
Accepted: 24 April 2026