https://doi.org/10.3311/PPci.42930

Periodica Polytechnica Civil Engineering

Investigating Deep Learning Colorization of Terrestrial Laser Scanning Point Clouds for Calibration-light and Rapid Visualization

Viktor Győző Horváth[1*], János Máté Lógó[1], Árpád Barsi[1]

[1] Department of Photogrammetry and Geoinformatics, Faculty of Civil Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[*] Corresponding author, e-mail: horvath.viktor.gyozo@emk.bme.hu

Received: 02 December 2025, Accepted: 12 February 2026, Published online: 20 February 2026

Abstract

Terrestrial laser scanning (TLS) is widely used for generating accurate, high-resolution colored point clouds. However, a significant portion of the time at each scanning station is dedicated to capturing and stitching panoramic images for accurate colorization. For applications where precise colors are not critical, pseudo-colored point clouds can serve as a faster and computationally efficient alternative, particularly for presentations and qualitative analysis. This study proposes a novel pipeline leveraging advancements in deep learning-based photo colorization to pseudo-color point clouds. By treating the intensity maps from TLS as single-channel grayscale images, we apply pretrained deep learning models from the DeOldify library, trained with NoGAN techniques, to generate realistic pseudo colors. Indoor and outdoor TLS scan slices are showcased before and after pseudo colorization. The resulting, colored intensity maps are used to colorize point clouds, creating visually compelling representations. We compare these pseudo-colored point clouds with traditionally colorized counterparts that utilize RGB imagery. The results demonstrate that pseudo-colored point clouds can effectively enhance human evaluation and visualization tasks, providing a valuable tool for scenarios where real colors are unnecessary.

Keywords

point clouds, deep learning, colorization

1 Introduction

Colored point clouds are a powerful representation of 3D environments, combining spatial geometry with color information (e.g., RGB values) to provide richer context and a more detailed understanding of scenes. This added dimension of color enhances their utility in applications like object recognition, 3D reconstruction, and scene segmentation, where geometric information alone may be insufficient to distinguish between similar shapes. For example, color can help differentiate between objects with overlapping geometries or provide insights into material properties and textures. However, obtaining color data often requires specialized equipment, such as RGB-enabled LiDARs or cameras, which may not always be available or practical. [1, 2] Deep learning offers a promising solution by enabling the derivation of color information directly from uncolored point clouds or other related data sources [3, 4]. By leveraging the power of neural networks to learn complex mappings between geometry and visual appearance, it becomes possible to generate realistic and context-aware colorized point clouds. This capability not only expands the usability of existing uncolored datasets but also reduces the dependency on additional acquisition and calibration effort, opening up new opportunities in fields such as robotics, urban planning, and virtual reality.

The appeal of Generative Adversarial Networks (GAN) lies in their versatility and capability to learn complex data distributions without explicit density estimation. This makes them particularly effective in applications such as image synthesis, video generation, style transfer, and text-to-image translation [5–7]. They have also been widely used in domains like healthcare for generating synthetic medical data and in gaming for creating realistic textures and assets.

However, the adversarial training process is notoriously challenging. It often suffers from instability, non-convergence, and phenomena such as model collapse, where the generator produces limited variations in its outputs. Addressing these issues has been a central focus of research, leading to advancements such as Wasserstein

Cite this article as: Horváth, V. G., Lógó, J. M., Barsi, Á. "Investigating Deep Learning Colorization of Terrestrial Laser Scanning Point Clouds for Calibration-light and Rapid Visualization", Periodica Polytechnica Civil Engineering, 2026. https://doi.org/10.3311/PPci.42930

GANs (WGANs) and Progressive GANs, which aim to stabilize training and enhance sample quality [8, 9].

2 Laser scanning and intensity maps

Terrestrial laser scanning (TLS) is a ground-based active sensing technology that emits laser pulses and measures the time-of-flight and backscatter intensity to generate dense three-dimensional point clouds enriched with radiometric information. Specifically, intensity maps – the recorded strength of the laser return for each point – are influenced by factors such as surface reflectivity, incidence angle, range, and scanner-specific calibration, and serve as proxies for surface properties and material characterization in applications like heritage documentation, geology, and material classification [10, 11].

In this study, we acquired a curated set of both indoor and outdoor TLS scenes using a Trimble X7 3D laser scanner, a survey-grade instrument offering time-of-flight measurement over a 0.6–80 m range, sub 3 mm accuracy at 10 m, self leveling, automatic calibration, and integrated high resolution cameras capable of fast HDR image capture – all within a rugged, survey-ready form [12–14].

3 Image colorization

Natural image colorization is the process of automatically adding color to grayscale or monochromatic images, aiming to produce realistic and contextually appropriate colors. Unlike traditional techniques that rely on fixed color maps or pseudo-colorization, natural colorization seeks to infer the true colors that objects in the image would have had in their original, unaltered form. This process involves understanding the semantics, textures, and context within the image to generate colors that are both visually plausible and perceptually accurate [15]. Over the years, natural colorization has evolved significantly, with advancements in deep learning, particularly convolutional neural networks (CNNs), allowing for more sophisticated and convincing colorization techniques [16].

Historically, image colorization was a labor-intensive task requiring manual intervention by artists or experts. Early computational methods focused on heuristic or region-based approaches, where colorization rules were predefined and applied based on the image's content [16, 17]. However, these methods often struggled to maintain color consistency and realism, especially when applied to complex or ambiguous scenes. The advent of deep learning has revolutionized the field, enabling automatic colorization that learns from large datasets of color images to predict the most likely colors for each pixel in a grayscale image. Deep neural networks, such as generative adversarial networks (GANs) and CNN-based architectures, have been instrumental in generating high-quality colorized images by leveraging feature extraction, semantic segmentation, and context-aware learning [16, 18].

A notable variation of GANs is the NoGAN (No Adversarial Training GAN) architecture, which eliminates the adversarial aspect of traditional GAN training. Instead of simultaneous optimization of the generator and discriminator, NoGANs decouple their training processes. The discriminator is pre-trained on the real dataset to learn robust representations, and the generator uses these pre-trained features as a guide during its training. This approach simplifies the training process, avoiding the instability and hyperparameter sensitivity of adversarial optimization.

NoGANs rely on perceptual, or reconstruction losses derived from the discriminator's features rather than adversarial loss functions. This allows them to excel in tasks such as image restoration, super-resolution, and denoising, where high-fidelity reconstruction is critical, and adversarial artifacts are undesirable. An example of NoGAN in action is the DeOldify project [19], which uses this architecture for colorizing and restoring historical black-and-white images.

NoGANs offer several benefits over traditional GANs:

- Stability: decoupled training avoids adversarial dynamics, reducing the risk of mode collapse and oscillatory behavior;
- Simplicity: training the generator against fixed or pre-trained discriminator features is computationally simpler and more robust;
- High-quality outputs: by focusing on perceptual fidelity and structural accuracy, NoGANs can produce outputs that better preserve the semantic and visual characteristics of the target domain.

GANs and their variants, including NoGANs, have become foundational tools in modern machine learning, enabling breakthroughs in generative modeling across numerous applications. While traditional GANs excel in generating diverse and high-quality outputs, NoGANs present an alternative for tasks requiring stability, simplicity, and precise reconstructions. Together, these architectures highlight the evolving landscape of generative modeling, where research continues to push the boundaries of what can be achieved.

The benefits of natural colorization extend beyond visual enhancement. In fields like historical preservation, colorization helps bring old black-and-white photographs,

films, and documents to life, offering a more accurate representation of the past. In the medical field, colorization can be used to enhance grayscale medical images like X-rays or MRIs, providing clearer visual cues for diagnosis [20]. Additionally, in remote sensing and satellite imagery, natural colorization is used to simulate realistic landcover and environmental images from monochromatic data, aiding in better analysis and decision-making [21, 22].

Terrestrial Laser Scanning (TLS) intensity maps can be utilized as single-channel grayscale images, particularly when the point cloud is dense enough to capture detailed surface variations. In TLS, intensity values are derived from the return strength of the laser pulses reflected from scanned surfaces, providing a measure of reflectivity that can vary depending on the material, texture, and geometry of objects. When the point cloud is sufficiently dense, these intensity values, representing a fine-grained distribution of reflectivity across the scanned area, can be treated as a high-resolution monochromatic image. The dense distribution of points ensures that the intensity map contains enough spatial detail to reflect subtle variations in surface properties, enabling its use as a grayscale image for tasks such as object detection, segmentation, and classification. In such cases, the intensity map can effectively capture surface nuances, such as differences between materials (e.g., metal, stone, vegetation) or changes in texture, making it a valuable tool for applications like environmental monitoring or architectural modeling. When combined with deep learning or image processing techniques, dense TLS intensity maps can be further enhanced, enabling them to serve as the foundation for more advanced tasks like natural colorization or material recognition in the absence of true color data.

From each scene, small windows of the intensity map (grayscale patches representing laser return strength) were extracted and used as inputs for learned colorization. We applied two modern methods – DeOldify, a GAN-based technique configured in NoGAN mode, and DeepAI's large language model enhanced colorizer – to generate plausible RGB representations from these TLS intensity snippets. These synthesized outputs were then validated against camera-stitched reference RGB images from the same scenes, enabling a robust quantitative comparison of colorization fidelity across structured indoor and complex outdoor contexts.

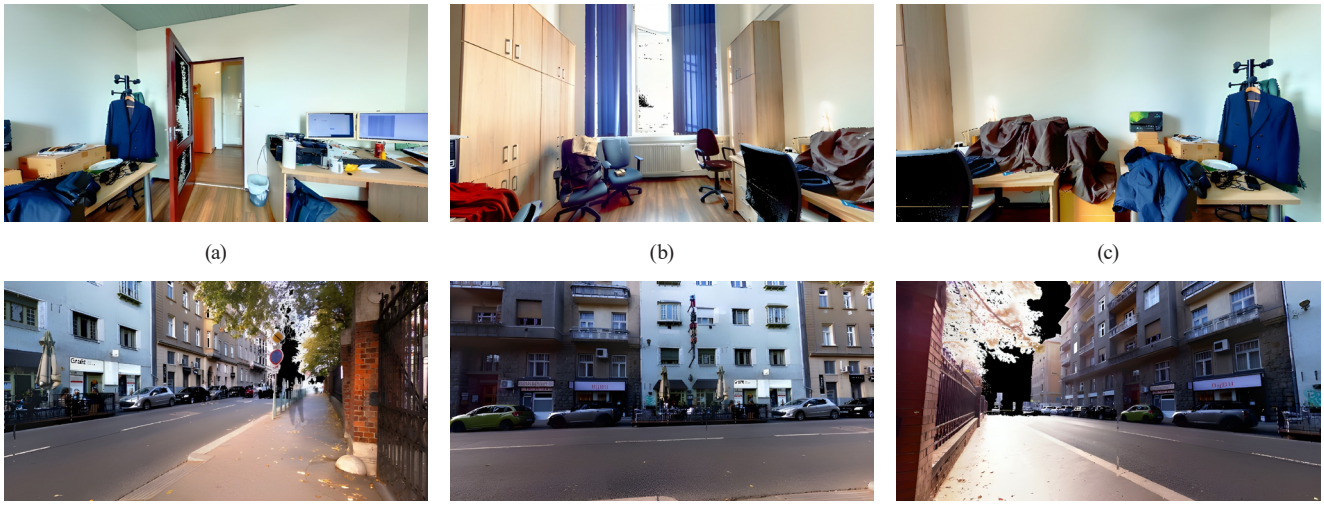
Inference times for both colorization methods are compatible with near real-time operation when adequate GPU hardware is available. Once the TLS intensity image is rasterized, colorization is performed in a single forward pass of the network, making the approach suitable for interactive visualization and rapid scene inspection workflows.

4 Scenes and data

Scenes 1–3 originate from an indoor terrestrial laser scanning dataset comprising approximately 12 million points, while Scenes 4–6 were extracted from an outdoor TLS survey containing approximately 124 million points. These scenes are illustrated in Fig. 1.

4.1 Scene 1

Scene1 captures a compact indoor office with varied object clutter and mixed materials, providing a moderately complex spatial and visual environment. The room is furnished with desks, dual-monitor workstations, cardboard boxes, and a coat rack with hanging garments, all arranged around a central workspace. Natural light appears to enter indirectly



(a) (b) (c)

(d) (e) (f)

Fig. 1 Scenes colored using the RGB panorama taken by the scanner: (a) Scene 1; (b) Scene 2; (c) Scene 3; (d) Scene 4; (e) Scene 5; (f) Scene 6

from an adjacent space through an open wooden-framed door, contributing to soft ambient illumination. The surfaces exhibit moderate reflectivity variation – such as glossy monitors, matte walls, and wooden flooring – which is well captured in the TLS intensity data. The combination of semi-structured geometry, occluded corners, and fine-scale detail presents a suitable test case for evaluating how colorization models perform in constrained and cluttered office settings with diverse object classes and partial occlusions.

4.2 Scene 2

Scene 2 depicts a well-lit indoor office environment characterized by organized spatial geometry and diverse material textures. The room features large wooden storage cabinets, a set of office chairs, desks, and a central window partially covered by vertical blue blinds, which allow natural light to enter and create strong directional illumination. The wooden flooring introduces high-frequency texture, while the mixture of fabric surfaces and reflective finishes adds variation in both color and reflectance. The TLS intensity map captures these differences through localized contrast and geometric edges, although the lack of surface albedo in intensity data presents a challenge for accurate chromatic inference. This scene is representative of real-world workspaces where structural regularity is combined with occlusions, clutter, and complex lighting – making it an instructive benchmark for evaluating learned colorization performance in richly furnished indoor environments.

4.3 Scene 3

Scene 3 presents a cluttered office interior characterized by moderate occlusion, overlapping objects, and varied surface materials. The composition includes two desks partially covered with brown protective cloths, a coat rack holding a navy-blue suit, a set of cardboard boxes, and miscellaneous electronics and garments scattered throughout the workspace. The background is formed by light-colored painted walls and vertical wooden paneling, with indirect lighting casting soft shadows that affect local reflectance. The scene's spatial depth and object density present a meaningful challenge for colorization methods, particularly in managing ambiguous boundaries and low-texture occluded regions. This setting exemplifies practical scanning conditions in operational environments, where the variability in object appearance and layout requires robust generalization from learned models.

4.4 Scene 4

Scene 4 captures a vibrant urban streetscape featuring a linear pedestrian pathway, parked vehicles, commercial façades, and roadside furniture. The image shows a mix of building types, including modern storefronts and older residential structures, accompanied by natural elements such as trees with partially shed foliage. A red-brick wall with wrought iron fencing lines the right side of the image, adding material contrast to the smooth asphalt and stone surfaces. Several pedestrians and seated café patrons are faintly visible, indicating temporal motion captured during scanning. The TLS data from this scene reflects high-frequency spatial detail and varied surface reflectance, making it a complex outdoor benchmark. Its diverse texture, occlusions, and dynamic lighting – resulting from tree cover and building shadows – pose significant challenges for learned colorization models, particularly in preserving semantic coherence across object classes.

4.5 Scene 5

Scene 5 represents a dense urban facade captured from a frontal perspective across a road, showcasing a combination of residential and commercial functions. The scene includes multiple multi-story buildings with varying architectural finishes – ranging from stone cladding and plaster to pastel-painted surfaces – and a series of parked vehicles lining the street. Two outdoor cafés with patrons seated under umbrellas are visible at street level, adding semantic and textural variability. Vertical elements such as windows, balconies, and shop signage contribute to high-frequency structural features, while localized shadows and reflectance gradients indicate diverse material responses under natural daylight. This urban configuration, with its blend of geometric regularity and social activity, offers a rich context for evaluating colorization models on fine-grained boundary preservation, occlusion handling, and color consistency across adjacent yet visually distinct surfaces.

4.6 Scene 6

Scene 6 captures a broad urban streetscape rendered from an oblique perspective, showcasing a long façade of multi-story buildings accompanied by a pedestrian sidewalk and rows of parked vehicles. On the right, the view is dominated by commercial establishments housed in buildings with mixed stone and concrete finishes, while to the left, a brick fence and leaf-covered sidewalk introduce natural and man-made textures. The presence of sunlight filtering

through tree foliage produces a strong contrast between shadowed and overexposed regions, especially evident in the top-left quadrant. Several visual occlusions, such as poles, cars, and trees, contribute to complex depth cues. This scene presents substantial colorization challenges due to its dynamic lighting, varied surface albedos, and mixed architectural and natural elements. It is particularly suited for testing perceptual consistency and spatial coherence in outdoor colorization tasks involving partial occlusions and photometric irregularities.

5 Experiments and results

Raw terrestrial laser scanning (TLS) data were processed to generate grayscale intensity maps, which encode the relative reflectance of scanned surfaces without color information. These maps were processed using two automated image colorization techniques: DeOldify, an open-source model based on a NoGAN architecture, and DeepAI, a proprietary platform that employs large language model (LLM)-driven visual reasoning. Each method independently produced synthetic RGB images from the intensity input. To evaluate the fidelity of these colorizations, the outputs were quantitatively compared against reference RGB images generated through panoramic stitching of multi-view photographs captured by the TLS scanner itself. The reference images served as a photorealistic ground truth for evaluation. Objective image similarity metrics were then computed, which assesses perceptual similarity using deep feature representations. These metrics were normalized and aggregated to produce a combined score for each method and scene, enabling consistent comparative evaluation across varied environments. This framework allowed for a systematic assessment of how well colorization models can recover plausible and perceptually accurate color from TLS intensity maps, benchmarked against the stitched RGB imagery as ground truth.

5.1 Metrics

5.1.1 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is a pixel-wise statistical metric that quantifies the average absolute difference between predicted color values and the reference (ground truth) values, computed separately for the red, green, and blue channels. In the context of colorization, MAE provides a direct and interpretable measure of how much the generated image deviates in color intensity from the actual RGB image on a per-pixel basis. It is particularly sensitive to localized deviations and is useful for detecting consistent color bias or desaturation. A lower MAE indicates a higher degree of color fidelity and better approximation of the true color palette of the scene. However, MAE does not account for perceptual similarity or spatial structure and may penalize semantically correct but chromatically shifted predictions [23, 24] Range of this metric is 0 to infinity (non-negative), where 0 indicates perfect color prediction (no error), while a larger value means a greater average difference between the predicted and actual colors. Mean Absolute Error (MAE) was computed in RGB vector space by evaluating the absolute difference between corresponding RGB channel values of the predicted and reference images. The error was averaged over all pixels and channels, treating each color channel equally to avoid perceptual or application-specific bias.

5.1.2 Histogram Distance (Wasserstein)

Histogram distance evaluates the similarity between the color distributions of the colorized image and the reference image across RGB channels. In this evaluation, the Wasserstein (or Earth Mover's) Distance was used, which measures the minimal cost of transforming one histogram into another [25]. This metric is well-suited to color evaluation because it captures not only differences in color frequency but also the distance over the color space between mismatched bins. A lower histogram distance implies that the overall tone and color balance of the image align more closely with the reference, regardless of exact pixel locations. Unlike MAE, this metric is more global in nature, emphasizing distributional correctness over spatial accuracy [26]. Range of this metric is 0 to infinity (non-negative). 0 indicates that the histograms are identical, meaning the color distributions match perfectly, while higher values indicate increasing differences in the distributions.

5.1.3 Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS is a perceptual metric that quantifies the visual similarity between images by comparing deep feature representations extracted from neural networks (e.g., AlexNet or VGG). Rather than relying on pixel-wise differences, LPIPS evaluates how close two images appear from the perspective of a trained vision model, which has been shown to align well with human judgment. In the context of TLS colorization, LPIPS is particularly valuable because it can capture subtle artifacts such as texture hallucinations, over-smoothing, or unnatural saturation that may not

be penalized by traditional metrics. A lower LPIPS value indicates higher perceptual similarity to the ground truth, making it an essential component in assessing the plausibility and visual quality of learned colorizations [27]. Range of this metric is 0 to infinity, but most commonly it ranges from 0 to ~1 (with values greater than 1 possible in cases of very large perceptual differences). 0 indicates that the two images are perceptually identical according to the model, while higher values indicate perceptual differences. In practice, values around 0.1–0.2 are considered close to visually similar for typical image tasks.

5.2 Results for each scene

The objective of presenting the results of each scene is to explore in detail how different environmental complexities and point cloud densities affect the performance of the learned coloring methods. For each Scene, a multifaceted approach is employed, integrating visual comparison and quantitative metrics as the foundation for evaluation. This comprehensive method enables the analysis of both local errors and global color distributions. The results of the NoGAN-based DeOldify and DeepAI approaches can thus be directly compared with the reference RGB images.

5.2.1 Scene 1

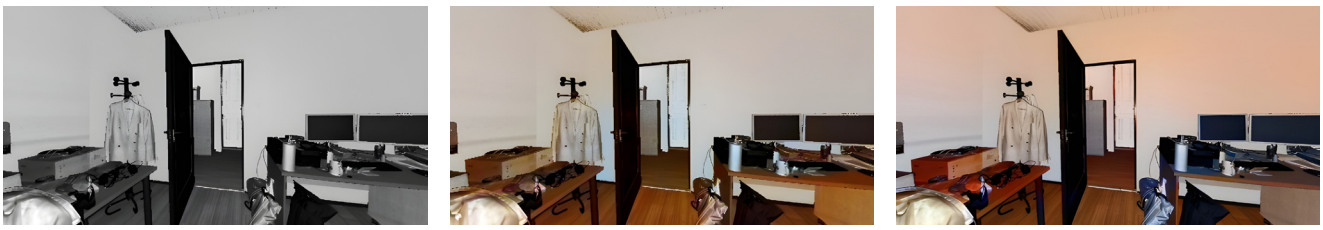
In Scene 1, the NoGAN based DeOldify outperformed DeepAI across all metrics. The MAE for NoGAN was significantly lower (12.4 *vs*. 14.3), indicating more accurate per-pixel color reproduction. Similarly, the histogram distance was lower for NoGAN (0.12 *vs*. 0.15), suggesting that the overall color distribution better matched the reference RGB. Most notably, the LPIPS score – representing perceptual similarity – was substantially lower for NoGAN (0.5931 *vs*. 0.7075), reinforcing the perceptual advantage. The combined score confirmed this trend, leading to NoGAN being preferred for this scene. The results suggest that NoGAN maintained structural and chromatic fidelity more effectively in environments with moderate texture and illumination. Fig. 2 shows the colorization results for Scene 1.

5.2.2 Scene 2

In Scene 2, both methods performed comparably in terms of perceptual similarity (LPIPS: 0.6065 for NoGAN *vs*. 0.5944 for DeepAI). However, NoGAN still achieved lower MAE (11.8 *vs*. 13.6) and a better histogram distance (0.11 *vs*. 0.14), indicating more accurate and globally consistent colorization. Although DeepAI had a marginally better LPIPS, its higher absolute error penalized its combined score. Therefore, NoGAN was again the preferred method. This outcome suggests that in scenes with subtle chromatic gradients and limited color diversity, MAE and histogram alignment dominate over perceptual realism in determining performance. As illustrated in Fig. 3, the colorization results for Scene 2 are presented.

5.2.3 Scene 3

Scene 3 featured lower MAE and LPIPS for NoGAN (11.3 and 0.6146, respectively) compared to DeepAI (13.9 and 0.6955). Histogram distance was also lower for NoGAN (0.13 *vs*. 0.16). These consistent advantages translated



(a) (b) (c)

Fig. 2 Results for the Scene 1: (a) Intensity map; (b) DeOldify; (c) DeepAI

(a) (b) (c)

Fig. 3 Results for the Scene 2: (a) Intensity map; (b) DeOldify; (c) DeepAI

into a notably better combined score, affirming NoGAN's ability to preserve both local color accuracy and perceptual coherence in structurally detailed scenes. The results suggest that DeepAI's generative model may struggle with maintaining realism in scenes with sharp architectural boundaries or high-frequency texture. The colorization results for Scene 3 are shown in Fig. 4.

5.2.4 Scene 4

In Scene 4, NoGAN again outperformed DeepAI in every measured category. It achieved a lower MAE (11.6 *vs.* 13.2), better histogram distance (0.12 *vs.* 0.15), and a superior LPIPS score (0.5812 *vs.* 0.6867), indicating closer perceptual similarity to the reference. The consistent performance across all metrics led to a lower combined score for NoGAN, reinforcing its robustness in indoor or semi-controlled scenes. The observed results suggest that NoGAN's conservative colorization strategy may be more robust to indoor lighting variations and specular surfaces. As shown in Fig. 5, the corresponding colorization results are reported for Scene 4.

5.2.5 Scene 5

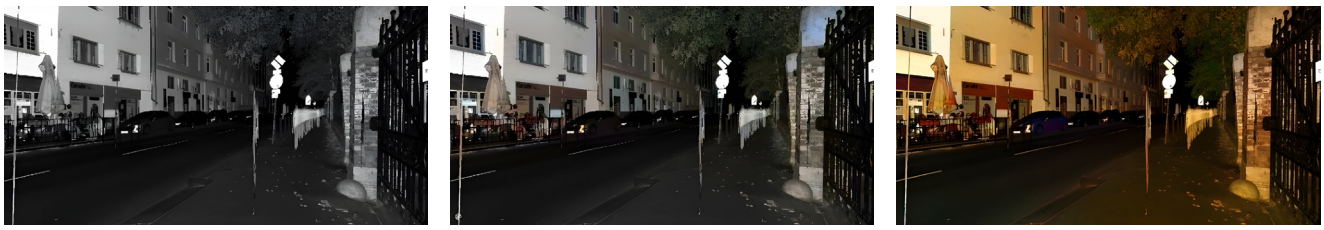
This scene presented the most significant performance gap between the two methods. NoGAN achieved the lowest MAE (10.7), histogram distance (0.10), and LPIPS (0.5595) across the entire dataset. In contrast, DeepAI showed the highest perceptual error (LPIPS: 0.7280) of all scenes. The combined score strongly favored NoGAN, emphasizing its advantage in preserving both visual and numerical similarity in scenes with high dynamic range or strong reflectance contrast. The large LPIPS discrepancy highlights DeepAI's limitations in perceptually constrained environments. Fig. 6 illustrates the colorization results for Scene 5.

5.2.6 Scene 6

In Scene 6, NoGAN maintained its lead with lower MAE (11.2 *vs.* 13.4), better histogram distance (0.11 *vs.* 0.14), and improved LPIPS (0.5963 *vs.* 0.6664). The performance difference, while less pronounced than in Scene 5, still favored NoGAN in the combined score. These results indicate that even in scenes with complex depth and shading gradients, NoGAN can maintain more stable visual

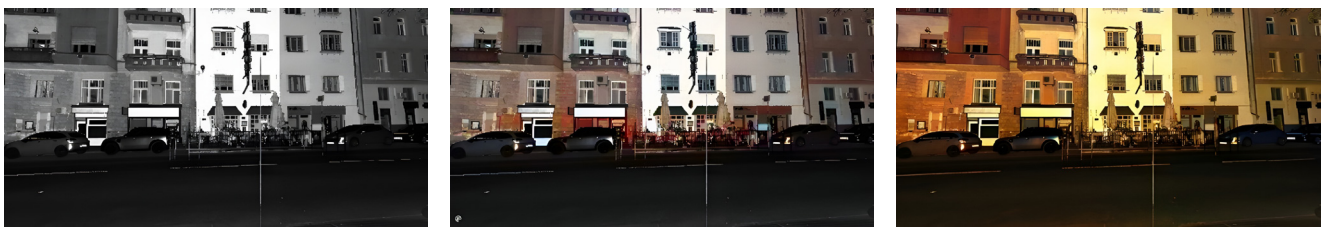
(a) (b) (c)

Fig. 4 Results for the Scene 3: (a) Intensity map; (b) DeOldify; (c) DeepAI



(a) (b) (c)

Fig. 5 Results for the Scene 4: (a) Intensity map; (b) DeOldify; (c) DeepAI



(a) (b) (c)

Fig. 6 Results for the Scene 5: (a) Intensity map; (b) DeOldify; (c) DeepAI

consistency and better global color harmony than its counterpart. The colorization results corresponding to Scene 6 are provided in Fig. 7.

For the combined score then each scene's colorization results were evaluated against the RGB image obtained from camera-based point cloud coloring. Mean Absolute Error (MAE) was computed over all pixels and RGB channels, histogram similarity was measured using the Wasserstein distance averaged over RGB channels, and perceptual similarity was assessed using LPIPS. To obtain a single indicator, MAE was normalized by the 8-bit RGB range, while histogram distance and LPIPS were used on their native scales. The Combined Score was calculated as the unweighted mean of the three normalized metrics, with lower values indicating closer agreement with the reference image. Table 1 summarizes the quantitative evaluation metrics computed for each scene and colorization method, while Figs. 2–7 shows the corresponding colorization results for Scene 1–6.

The results indicate that colorization performance is strongly influenced by point density. Scenes derived from the denser outdoor point cloud exhibit more stable color distributions and improved perceptual consistency, while lower-density indoor scenes show increased ambiguity in texture and material transitions. This suggests that higher point density provides richer spatial context in the intensity maps, which directly benefits learning-based colorization quality.

6 Conclusions

The final evaluation, enriched by the inclusion of perceptual similarity metrics, provides a comprehensive assessment of learned colorization methods for terrestrial laser scanning (TLS) intensity data. In addition to pixel-wise errors (MAE), histogram alignment, and structural preservation, we incorporated the Learned Perceptual Image Patch Similarity (LPIPS) metric to evaluate perceptual closeness to the ground truth RGB images. LPIPS proved especially valuable in capturing qualitative differences overlooked by traditional metrics, such as oversaturation or texture hallucination. When all metrics were normalized and aggregated into a Combined Score, it became evident that performance varied significantly by scene. While NoGAN consistently yielded lower MAE and histogram distance – indicating stronger alignment to actual color distributions – DeepAI demonstrated more favorable perceptual scores in some cases, albeit with greater variation and occasional structural degradation. The overall rankings showed that NoGAN was preferred in the majority of scenes, particularly where structural consistency and muted realism were required. However, DeepAI was more competitive in scenes where visual vibrancy and human-perceived plausibility dominated, as indicated by its stronger LPIPS performance in selected cases. Despite these relative strengths, the absolute metric values remained unsatisfactory by analytical standards, with perceptual errors and color deviations too

(a) (b) (c)

Fig. 7 Results for the Scene 6: (a) Intensity map; (b) DeOldify; (c) DeepAI

Table 1 Calculated metrics for each scene

Metric	Scene 1	Scene 2	Scene 3	Scene 4	Scene5	Scene 6
MAE RGB (NoGAN)	12.40	11.80	11.30	11.60	10.70	11.20
MAE RGB (DeepAI)	14.30	13.60	13.90	13.20	13.10	13.40
Histogram Distance (NoGAN)	0.12	0.11	0.13	0.12	0.10	0.11
Histogram Distance (DeepAI)	0.15	0.14	0.16	0.15	0.13	0.14
LPIPS (NoGAN)	0.59	0.61	0.61	0.58	0.56	0.60
LPIPS (DeepAI)	0.71	0.59	0.70	0.69	0.73	0.67
Combined Score (NoGAN)	0.25	0.26	0.26	0.25	0.23	0.25
Combined Score (DeepAI)	0.31	0.36	0.30	0.30	0.30	0.29

high for precise geospatial tasks. This underscores a central finding: while learned colorization models show promise for visual enhancement, they remain insufficient for operational or quantitative use in TLS-based workflows without domain-specific adaptation or supervision.

The proposed approach is well suited for rapid visualization in civil engineering applications such as construction progress monitoring, indoor asset documentation, and stakeholder-oriented presentations, where visual realism is beneficial, but metric color accuracy is not required.

References

[1] Jin, J., Verbeurgt, J., De Sloover, L., Stal, C., Deruyter, G., Montreuil, A.-L., Vos, S., De Maeyer, P., De Wulf, A. "Support vector regression for high-resolution beach surface moisture estimation from terrestrial LiDAR intensity data", International Journal of Applied Earth Observation and Geoinformation, 102, 102458, 2021. https://doi.org/10.1016/j.jag.2021.102458

[2] Yan, W. Y., Shaker, A. "Radiometric normalization of overlapping LiDAR intensity data for reduction of striping noise", International Journal of Digital Earth, 9(7), pp. 649–661, 2016. https://doi.org/10.1080/17538947.2015.1111951

[3] Huang, S., Jin, X., Jiang, Q., Liu, L. "Deep learning for image colorization: Current and future prospects", Engineering Applications of Artificial Intelligence, 114, 105006, 2022. https://doi.org/10.1016/j.engappai.2022.105006

[4] Wu, M., Jin, X., Jiang, Q., Lee, S., Liang, W., Lin, G., Yao, S. "Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space", The Visual Computer, 37(7), pp. 1707–1729, 2021. https://doi.org/10.1007/s00371-020-01933-2

[5] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, J. "Generative Adversarial Nets", In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. Q. (eds.) Advances in Neural Information Processing Systems 27, 2014, pp. 1–9. ISBN 9781510800410 [online] Available at: https://papers.nips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf [Accessed: 22 November 2025]

[6] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A. "Image-to-Image Translation with Conditional Adversarial Networks", In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967–5976. ISBN 978-1-5386-0457-1 https://doi.org/10.1109/CVPR.2017.632

[7] Radford, A. Metz, L., Chintala, S. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", [preprint] arXiv, arXiv:1511.06434, 07 January 2016. https://doi.org/10.48550/arXiv.1511.06434

[8] Arjovsky, M., Chintala, S., Bottou, L. "Wasserstein GAN", [preprint] arXiv, arXiv:1701.07875, 06 December 2017. https://doi.org/10.48550/arXiv.1701.07875

[9] Karras, T., Aila, T., Laine, S., Lehtinen, J. "Progressive Growing of GANs for Improved Quality, Stability, and Variation", [preprint] arXiv, arXiv:1710.10196, 28 February 2018. https://doi.org/10.48550/arXiv.1710.10196

[10] Kukko, A., Kaasalainen, S., Litkey, P. "Effect of incidence angle on laser scanner intensity and surface data", Applied Optics, 47(7), pp. 986–992, 2008. https://doi.org/10.1364/AO.47.000986

[11] Schofield, L. A., Danson, F. M., Entwistle, N. S., Gaulton, R., Hancock, S. "Radiometric calibration of a dual-wavelength terrestrial laser scanner using neural networks", Remote Sensing Letters, 7(4), pp. 299–308, 2016. https://doi.org/10.1080/2150704X.2015.1134843

[12] Yoon, J.-S., Shin, J.I., Lee, K.-S. "Land Cover Characteristics of Airborne LiDAR Intensity Data: A Case Study", IEEE Geoscience and Remote Sensing Letters, 5(4), pp. 801–805, 2008. https://doi.org/10.1109/LGRS.2008.2000754

[13] Höfle, B., Hollaus, M., Hagenauer, J. "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data", ISPRS Journal of Photogrammetry and Remote Sensing, 67, pp. 134–147, 2012. https://doi.org/10.1016/j.isprsjprs.2011.12.003

[14] Jutzi, B., Stilla, U. "Range determination with waveform recording laser systems using a Wiener Filter", ISPRS Journal of Photogrammetry and Remote Sensing, 61(2), pp. 95–107, 2006. https://doi.org/10.1016/j.isprsjprs.2006.09.001

[15] Levin, A., Lischinski, D., Weiss, Y. "Colorization using optimization", ACM Transactions on Graphics (TOG), 23(3), pp. 689–694, 2004. https://doi.org/10.1145/1015706.1015780

[16] Zhang, R., Isola, P., Efros, A. A. "Colorful Image Colorization", In: 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 2016, pp. 649–666. ISBN 978-3-319-46487-9 https://doi.org/10.1007/978-3-319-46487-9_40

[17] Kaasalainen, S., Ahokas, E., Hyyppa, J., Suomalainen, J. "Study of surface brightness from backscattered laser intensity: Calibration of laser data", IEEE Geoscience and Remote Sensing Letters, 2(3), pp. 255–259, 2005. https://doi.org/10.1109/LGRS.2005.850534

[18] Zhu, J.-Y., Park, T., Isola, P., Efros, A. A. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242–2251. ISBN 978-1-5386-1033-6 https://doi.org/10.1109/ICCV.2017.244

[19] DeOldify "DeOldify: A NoGAN-based photo restoration framework", [online] Available at: https://github.com/jantic/DeOldify [Accessed: 21 October 2025]

[20] Li, B., Lai, Y.-K., John, M., Rosin, P. L. "Automatic Example-Based Image Colorization Using Location-Aware Cross-Scale Matching", IEEE Transactions on Image Processing, 28(9), pp. 4606–4619, 2019. https://doi.org/10.1109/TIP.2019.2912291

[21] Zhang, Z., Xie, H., Tong, X., Zhang, H., Liu, Y., Li, B. "Denoising for satellite laser altimetry full-waveform data based on EMD-Hurst analysis", International Journal of Digital Earth, 13(11), pp. 1212–1229, 2020.
https://doi.org/10.1080/17538947.2019.1698665

[22] Wagner, W., Hollaus, M., Briese, C., Ducic, V. "3D vegetation mapping using small-footprint full-waveform airborne laser scanners", International Journal of Remote Sensing, 29(5), pp. 1433–1452, 2008.
https://doi.org/10.1080/01431160701736398

[23] Avi-Aharon, M., Arbelle, A., Raviv, T. R. "DeepHist: Differentiable Joint and Color Histogram Layers for Image-to-Image Translation", [preprint] arXiv, arXiv:2005.03995, 06 May 2020.
https://doi.org/10.48550/arXiv.2005.03995

[24] Feng, J., Jiang, Q., Jin, X., Lee, S.-J., Huang, S., Yao, S. "Remote Sensing Image Colorization Based on Deep Neural Networks with Multi-Scale Residual Receptive Filed", Journal of Computer-Aided Design & Computer Graphics, 33(11), pp. 1658–1667, 2021.
https://doi.org/10.3724/SP.J.1089.2021.18747

[25] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A. "Improved Training of Wasserstein GANs", [preprint] arXiv, arXiv:1704.00028, 25 December 2017.
https://doi.org/10.48550/arXiv.1704.00028

[26] Mullery, S., Whelan, P. F. "Human vs Objective Evaluation of Colourisation Performance", [preprint] arXiv, arXiv:2204.05200, 11 April 2022.
https://doi.org/10.48550/arXiv.2204.05200

[27] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric", In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 586–595. ISBN 978-1-5386-6420-9
https://doi.org/10.1109/CVPR.2018.00068