

LPIPS-Based Visual Quality Inspection for Irregular-Shaped Products in Additive Manufacturing Post-Processing

Hyun-Chul Kang, Chang-Beom Kim, Min-Gi Kim, Eun Seo Lee, Ji-Yeon Son
ICT-enabled Intelligent Manufacturing Research Section
Electronics and Telecommunications Research Institute
Daejeon, Korea
{kauni, cbkim, kmk82, eslee, jyson}@etri.re.kr

Abstract—Additive Manufacturing (AM) enables the production of parts with irregular shapes, where variations in geometry, size, material, and color make it difficult to perform consistent quality inspection using conventional rule-based methods. In particular, each product often requires a different baseline, and repeated data collection, labeling, and training are needed posing significant limitations for real-time inspection systems. To address this, we apply the Learned Perceptual Image Patch Similarity (LPIPS), which models human visual perception and can identify perceptual differences without repeated learning. We compare its performance with traditional similarity metrics such as Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). Experimental results show that the proposed LPIPS-based method can effectively detect perceptual differences across the entire image of irregular products without requiring predefined Regions of Interest (ROI), enabling flexible and reliable quality evaluation across various product types without the need for additional retraining.

Keywords— Visual Quality Inspection, Additive Manufacturing, Post-Processing, LPIPS

I. INTRODUCTION

Additive Manufacturing (AM) enables the fabrication of highly complex and customized geometries that are impractical with traditional manufacturing methods. However, to achieve the required surface quality and mechanical integrity, precise post-processing and the associated quality inspection are indispensable [1–3]. In particular, the enhanced design freedom of AM parts amplifies the need for accurate, non-destructive evaluation during this stage. Moreover, mass customization has introduced a vast array of part geometries, sizes, materials, and colors. Conventional inspection systems relying on predefined “normal” baselines struggle with this diversity. Each new variant requires fresh data collection, manual labeling, and model retraining, resulting in significant time and resource overhead [4].

This challenge calls for a paradigm shift toward novel inspection technologies capable of emulating human-like cognitive perception to identify and rapidly quantify visually salient discrepancies. In the context of AM post-processing, such a flexible, AI-based inspection method has emerged as a crucial factor for enhancing production quality and advancing automation.

Previous studies have predominantly relied on traditional pixel-based similarity metrics like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) to detect shape anomalies or visual discrepancies [5]. However, these techniques have a limitation: their similarity values fluctuate significantly with even minor changes in a product's position or angle, making them unsuitable for unaligned, irregularly shaped products.

Consequently, these methods have been restricted to products with uniform shapes or required pre-defining a Region of Interest (ROI) or mask for comparison. In real-world AM environments, where a continuous stream of irregularly shaped products is introduced, setting an ROI is inherently difficult, and repeatable shape-based comparisons are often impossible. Therefore, there is a clear need for an approach that can quickly and automatically quantify similarity at a global image level without requiring pre-defined areas.

II. VISUAL QUALITY INSPECTION ALGORITHM

In this paper, to automatically evaluate the quality of irregularly shaped additively manufactured products, we measure perceptual feature differences over entire images using the Learned Perceptual Image Patch Similarity (LPIPS) score and compare its behavior with conventional similarity metrics such as SSIM and PSNR [5,6].

Existing quality evaluation methods often depend on predefined reference datasets or standardized shape criteria, which limits their consistency and generalization for diverse, non-aligned shapes. To address these shortcomings, we employ LPIPS, which is known to correlate well with human visual perception, to capture global feature discrepancies and detect quality anomalies without requiring manual region definitions. Specifically, we compute LPIPS between an original image and a target image by extracting deep feature maps from a pre-trained network (e.g., VGG16) [6,7].

In this study, we employ the LPIPS (Learned Perceptual Image Patch Similarity) metric proposed by Zhang et al. [6] to quantitatively evaluate perceptual similarity. Equation (1) represents the perceptual distance in LPIPS, quantifying the visual similarity between two images in a human perception [6].

Let $\widehat{y}_{hw}^l$ denote the unit-normalized deep feature vector at spatial location (h, w) in the l -th layer for image x , and $\widehat{y}_{0hw}^l$ for the corresponding location in reference image x_0 [6]. H_l and W_l represent the height and width of the feature map at layer l , and w_l is a learned per-channel weight vector that modulates the relative contribution of each channel. The perceptual discrepancy between x and x_0 is then computed as the spatial average of the weighted squared l_2 distance between feature vectors, aggregated over all spatial positions and summed across selected layers [6].

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} |w_l \odot (\widehat{y}_{hw}^l - \widehat{y}_{0hw}^l)|_2^2 \quad (1)$$

This formulation yields a single scalar distance reflecting perceptual similarity, where lower values indicate closer visual resemblance. By aggregating differences across

multiple feature layers with learned channel weighting, LPIPS captures human-like judgments of image quality and shape variation more robustly than pixel- or structure-based metrics.

III. PERFORMANCE EVALUATION AND RESULT ANALYSIS

To validate the performance of our product's visual quality evaluation, we conducted a performance assessment using a subset of the MVTEC AD (Anomaly Detection) dataset, a public dataset for industrial defect detection [8,9]. The MVTEC AD dataset consists of high-resolution images that include various shape and surface defects that can occur in real manufacturing environments [8,9]. For each product, both normal images and various types of abnormal (defective) images are provided.

This study compares the performance of three representative image similarity metrics used in automated visual quality inspection: LPIPS, SSIM, and PSNR [5,6]. These algorithms produce scores in different units LPIPS and SSIM yield unitless values in [0,1], whereas PSNR is expressed in decibels (dB).

LPIPS represents a learned perceptual distance, with lower values indicating greater visual similarity; SSIM measures structural similarity, where values closer to one denote higher likeness; and PSNR is based on pixel-wise error, with larger values corresponding to better fidelity. Although the raw scales differ, we normalize their outputs by computing, for each algorithm, the proportions of Normal, Warning, and Defect classifications along with the sensitivity, thereby enabling a consistent, comparative assessment of their anomaly detection performance.

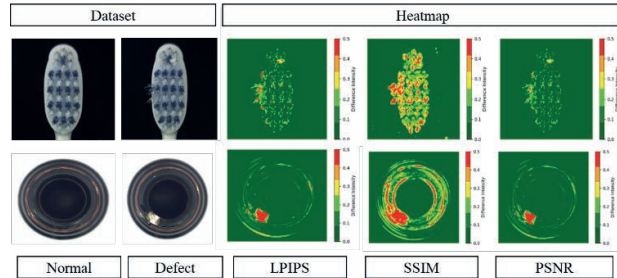


Fig. 1. Heatmaps of visual similarity metrics on the Toothbrush and Bottle test samples: comparison of Normal, Warning, and Defect classification ratios for LPIPS, SSIM, and PSNR.

As shown in Figure 1, we provide example images for the Toothbrush and Bottle classes, including both defect-free (Normal) and defective (Defect) samples, alongside the corresponding heatmaps generated by LPIPS, SSIM, and PSNR. For each class, the left side presents a clean and a defective image pair, while the right side displays the heatmaps from the three metrics. These heatmaps visualize how each algorithm distributes its outputs into Normal, Warning, and Defect categories for the given samples.

Table I summarizes the average results measured for each algorithm on the Bottle and Toothbrush datasets, comparing them based on three classification ratios: Normal, Warning, and Defect. For each class, the analysis was performed using one reference normal image and three test images (including defective examples) to enable comparative evaluation.

Experimental results show that while SSIM demonstrated the highest anomaly detection rate with a sensitivity of 48.7% on the Bottle dataset, this high rate implies a potential for over-

sensitivity and a high error rate on images that are actually Normal. Indeed, SSIM's Normal ratio was 53.6%, which is significantly lower than LPIPS (80.4%) and PSNR (79.8%), indicating a higher likelihood of misclassifying normal images as defective. This is because SSIM is highly sensitive to structural similarity, and minor surface or lighting changes can cause it to be classified as Defect. Although PSNR produces overall Normal/Defect patterns comparable to LPIPS, its low sensitivity can delay the detection of early anomalies, and its pixel-wise error focus limits its ability to reflect subtle, perceptually meaningful variations.

TABLE I. PERFORMANCE EVALUATION OF SIMILARITY MEASURES ON BOTTLE AND TOOTHBRUSH DATASET

Algorithms	Dataset	Score	Normal (%)	Warning (%)	Defect (%)	Sensitivity (%)
LPIPS	bottle	0.15	80.4	11.3	8.3	30.1
	brush tooth	0.15	89.4	5.7	4.9	16.3
PSNR	bottle	25.7	79.8	11.3	8.9	17
	brush tooth	28.0	89.1	5.8	5	9.2
SSIM	bottle	0.85	53.6	10.7	35.7	48.7
	brush tooth	0.8	76.1	3.8	20.1	25.1

In contrast, LPIPS maintained a relatively low sensitivity (30.1%) while recording a high Normal classification accuracy. This confirms that in practical applications, LPIPS can provide stable detection performance while reducing over-detection. LPIPS is evaluated as a more reliable algorithm for practical defect detection because it reflects perceptual quality differences rather than a single numerical error value.

IV. CONCLUSION

This study evaluated three image similarity metrics LPIPS, SSIM, and PSNR for post-processing quality inspection of irregular-shaped additively manufactured products. By normalizing their outputs into comparable classification ratios and sensitivity, we identified distinct trade-offs: SSIM tended to be over-sensitive, leading to frequent false positives, while PSNR's low sensitivity risked missing subtle early defects.

In contrast, LPIPS achieved a favorable balance, delivering high normal classification accuracy with moderate sensitivity and capturing perceptually meaningful differences without relying on predefined ROIs or retraining. These characteristics lead to LPIPS being evaluated as the best candidate among the tested metrics for real-world automated visual defect inspection in AM post-processing environments.

FUTURE WORK

To facilitate practical on-site deployment of the LPIPS inspection algorithm, future work will evaluate its performance on real-world datasets encompassing irregularly shaped AM parts under diverse process conditions. Building on these evaluations, we will pursue a hybrid approach that integrates LPIPS with traditional similarity metrics such as SSIM and PSNR. Under the integrated approach, SSIM and PSNR will first perform coarse alignment and global contrast assessment, after which LPIPS will provide fine-grained perceptual sensitivity to surface defects. We anticipate that this multi-metric strategy will enhance algorithmic consistency and reliability, ultimately yielding a robust and practical solution for real-world quality inspection.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No.2022-0-00969]. Additionally, this work was partly supported by the Technology Innovation Program (The Alchemist Project) (RS-2024-00410810, Development of Generative Manufacturing Technologies to Turn Ideas into Reality) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea).

REFERENCES

- [1] A. Diniță, A. Neacșa, A. I. Portoacă, M. Tănase, C. N. Ilinca, and I. N. Ramadan, "Additive manufacturing post-processing treatments, a review with emphasis on mechanical characteristics," *Materials*, vol. 16, no. 13, p. 4610, 2023.
- [2] S. A. Tofail, E. P. Koumoulos, A. Bandyopadhyay, S. Bose, L. O'Donoghue, and C. Charitidis, "Additive manufacturing: Scientific and technological challenges, market uptake and opportunities," *Materials Today*, vol. 21, no. 1, pp. 22–37, 2018.
- [3] H.-C. Kang, H.-N. Han, H.-C. Bae, M.-G. Kim, J.-Y. Son, and Y.-K. Kim, "HSV Color-Space-Based Automated Object Localization for Robot Grasping without Prior Knowledge," *Applied Sciences*, 2021.
- [4] S. E. Whang, Y. Roh, H. Song, and J. G. Lee, "Data collection and quality challenges in deep learning: A data-centric AI perspective," *The VLDB Journal*, vol. 32, no. 4, pp. 791–813, 2023.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [6] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.
- [7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.
- [8] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.
- [9] "MVTec Datasets" MVTec Website. [Online]. Available: <https://www.mvtec.com/company/research/datasets/mvtec-ad/downloads/>. [Accessed: Apr. 03, 2025].