Journal of Visual Computing

https://jvcjournal.org

Texture Synthesis Models and Material Identification in E-Commerce Previews: Insights from Benchmark Study

Júlia Barbosa[*]

Department of Electronic and Computer Engineering, Polytechnic School, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Corresponding author: julia.barbosa@upm.es

Abstract

The rapid proliferation of digital commerce has fundamentally transformed consumer interactions with product previews, necessitating advanced computational models capable of rendering highly realistic digital representations. Central to this transformation is the integration of texture synthesis models and material identification frameworks, which collectively bridge the gap between physical material properties and their digital simulacra. This paper presents a comprehensive benchmark study that systematically investigates the linkages between generative texture synthesis algorithms and downstream material identification accuracy within the context of e-commerce environments. By constructing a novel, large-scale dataset of synthesized product previews across diverse material categories, the study evaluates the efficacy of contemporary computer vision architectures in recognizing intrinsic material properties from artificially generated textures. The research highlights critical domain gaps where high-fidelity visual synthesis paradoxically degrades computational identification accuracy due to the loss of micro-surface variations and physical reflectance cues. Through rigorous quantitative evaluation and qualitative analysis, this research delineates the boundary conditions of current synthesis models and their direct impact on automated cataloging and consumer trust mechanisms. The findings offer pivotal insights for researchers and practitioners aiming to optimize digital retail infrastructures, proposing a paradigm shift toward physically informed generative models that preserve material semantics.

Keywords: Texture Synthesis; Material Identification; Computer Vision; Digital Retail; E-Commerce Previews

1. Introduction

1.1 Background and Motivation

The transition from physical retail environments to digital commerce platforms has precipitated a critical reliance on visual representations of products. In a traditional retail setting, consumers engage with products through multi-sensory interactions, predominantly leveraging haptic feedback and dynamic visual inspection under varying illumination conditions to assess material quality. In the digital realm, these interactions are entirely mediated by two-dimensional screens or, increasingly, three-dimensional augmented reality interfaces. Consequently, the burden of communicating material properties such as softness, roughness, weight, and elasticity falls entirely on the visual fidelity of the product preview. This reliance has driven massive investments into texture synthesis models, which aim to generate highly realistic, high-resolution surface patterns that mimic physical materials. However, a significant disjunction exists between the visual appeal of these synthesized textures and their structural accuracy from a computational standpoint. While generative models excel at producing aesthetically pleasing images that satisfy human visual perception, these same images often lack the underlying physical consistency required by automated material identification systems. Understanding this disjunction is paramount, as modern e-commerce infrastructures increasingly rely on automated cataloging, inventory management, and semantic search algorithms that must accurately identify materials based solely on visual input [1]. When synthesized textures deceive or confuse these classification algorithms, the downstream effects include inaccurate search results, mislabeled inventory, and ultimately, a degradation of consumer trust.

1.2 Challenges in Digital Material Representation

The core challenge in linking texture synthesis to material identification lies in the fundamentally different objectives of the two computational domains. Texture synthesis, particularly in its modern generative incarnations, is primarily optimized for perceptual fidelity. Algorithms are trained to minimize the visual divergence between generated textures and a distribution of real-world photographs, often prioritizing macroscopic patterns, color coherence, and artifact reduction. In contrast, material identification requires the extraction of invariant physical properties from an image, properties that are deeply tied to the physical interaction between light and matter. These interactions manifest as micro-shadows, specular highlights, and complex subsurface scattering effects that are highly sensitive to viewing angles and illumination. When a texture synthesis model generates a digital representation of leather, for example, it may perfectly replicate the macroscopic grain and coloration. However, it may simultaneously fail to accurately reproduce the subtle specular variations that distinguish genuine full-grain leather from synthetic polyurethane substitutes [2]. This creates a scenario where a synthesized product preview appears hyper-realistic to a human consumer but is consistently misclassified by a deep learning algorithm trained on genuine physical samples. Furthermore, the immense scale of e-commerce necessitates models that can generalize across an incredibly diverse taxonomy of materials, ranging from translucent plastics and highly reflective metals to porous ceramics and complex woven fabrics. The variability inherent in consumer-uploaded product images, which feature unconstrained lighting, occlusion, and background clutter, exacerbates the difficulty of extracting reliable material cues, making the synthesis-identification linkage a highly volatile area of research.

1.3 Objectives and Scope of the Benchmark Study

To address the growing chasm between perceptual synthesis and physical identification, this study introduces a comprehensive benchmark designed explicitly to evaluate the performance of material identification models on synthesized e-commerce product previews. The primary objective is to systematically quantify the domain gap between real photographic textures and their synthesized counterparts when subjected to algorithmic classification. By isolating variables such as illumination, scale, and generative model architecture, the research aims to pinpoint the specific synthesis artifacts that most severely disrupt material recognition [3]. A secondary objective is to explore the bidirectional relationship between synthesis and identification, investigating whether features extracted from robust identification networks can be utilized to guide synthesis models toward physically accurate outputs, rather than merely visually pleasing ones. The scope of this study encompasses a rigorous evaluation of multiple state-of-the-art generative frameworks against a suite of established convolutional and transformer-based classification architectures. The benchmark dataset curated for this research spans a wide array of commercially relevant materials, ensuring that the findings are directly applicable to the operational realities of modern digital retail. Through this extensive empirical investigation, the paper seeks to establish a foundational understanding of how synthetic data generation interacts with automated perception, providing a necessary framework for the future development of e-commerce visualization technologies.

2. Literature Review and Theoretical Framework

2.1 The Evolution of Texture Synthesis Models

The trajectory of texture synthesis research spans several decades, evolving from rudimentary procedural generation techniques to highly sophisticated deep learning paradigms. Early approaches relied heavily on statistical and parametric models that attempted to capture the spatial distribution of pixel intensities. These methods, while computationally efficient, were fundamentally limited by their inability to model complex, non-stationary textures featuring large-scale structural variations. The introduction of non-parametric, patch-based synthesis techniques represented a significant leap forward, allowing algorithms to generate novel textures by intelligently sampling and stitching together small patches from a source image. These techniques proved highly effective for regular and near-regular textures, such as brick walls or woven fabrics, but often struggled with stochastic materials like natural stone or organic tissues, frequently producing visible seams and repetitive artifacts.

The landscape of texture synthesis was irrevocably altered by the advent of deep learning, specifically the utilization of convolutional neural networks for feature extraction and generation. Early neural texture synthesis leveraged pre-trained image classification networks to extract multi-scale feature maps, utilizing optimization processes to generate images that matched the feature statistics of a target texture [4]. This approach demonstrated an unprecedented capacity to capture both high-level semantic structure and low-level detail. Subsequently, generative adversarial networks introduced a novel framework wherein a generator network learned to produce textures that could indistinguishably mimic real data distributions, evaluated by a concurrent discriminator network. These models enabled real-time synthesis of ultra-high-resolution materials, becoming the de facto standard for digital asset creation in gaming and e-commerce. Most recently, diffusion models have emerged as a dominant force, offering superior stability and sample diversity by framing the generation process as a gradual denoising of latent variables [5]. Despite these massive leaps in perceptual quality, the theoretical underpinning of these generative models remains focused on statistical pixel distribution rather than physical material properties, an architectural blind spot that directly impacts downstream identification tasks.

2.2 Material Identification in Computer Vision

Material identification constitutes a fundamental problem in computer vision, intricately linked to the broader challenge of inverse rendering. The theoretical ideal of material identification involves estimating the bidirectional reflectance distribution function of a surface from a single or multiple two-dimensional images. The bidirectional reflectance distribution function mathematically describes how light is reflected at an opaque surface, capturing the complex interplay of diffuse and specular reflection that defines visual material perception. Historically, computer vision approaches attempted to explicitly model these physical parameters using carefully controlled laboratory setups, spherical illumination rigs, and extensive multi-view image captures. While highly accurate, these methods were entirely impractical for unconstrained environments such as e-commerce platforms, where images are captured with standard consumer cameras under arbitrary lighting conditions.

The shift toward data-driven, deep learning methodologies transformed material identification from a physics-based estimation problem into a large-scale pattern recognition task. Modern approaches utilize deep convolutional architectures and vision transformers to learn hierarchical representations of material textures directly from massive datasets of annotated photographs [6]. These networks learn to implicitly disentangle illumination from reflectance, identifying invariant features such as micro-texture gradients, structural regularities, and local color variations. However, the efficacy of these models is heavily dependent on the diversity and authenticity of their training data. When deployed in real-world retail scenarios, these models frequently encounter materials that exhibit extreme intra-class variability, such as the thousands of distinct weave patterns classified broadly as cotton. Furthermore, the identification process is highly sensitive to contextual cues, such as object shape and usage context, which algorithms often conflate with the intrinsic material properties [7]. The theoretical framework of modern material identification thus balances precariously on the assumption that the input data accurately reflects the physical world, an assumption that is directly challenged by the introduction of synthesized product previews.

2.3 The Intersection of Synthesis and Identification in Retail Environments

The convergence of texture synthesis and material identification within e-commerce represents a relatively nascent but rapidly expanding area of academic inquiry. Digital retailers are increasingly utilizing synthesis pipelines not just for product visualization, but also for data augmentation, generating massive volumes of synthetic training data to improve the robustness of their automated cataloging systems. This practice is grounded in the theoretical premise that synthetic data can effectively bridge the domain gap caused by limited photographic datasets. However, empirical studies have begun to reveal significant limitations in this approach. When synthesis models generate data without strict physical constraints, they inadvertently introduce systematic biases and artifacts that identification networks subsequently learn as discriminative features [8]. For instance, a generative model might consistently produce synthetic wood grain with an unnatural frequency of knots, causing a classifier trained on this data to misidentify real, unblemished wood.

This intersection is further complicated by the dynamic nature of consumer interaction. Shoppers utilize product previews to make rapid judgments regarding quality, durability, and tactile properties, cognitive processes that are highly sensitive to subtle visual cues. When automated systems misclassify materials, the resulting discrepancies in search filters and product recommendations directly impact consumer behavior and platform credibility. Theoretical frameworks addressing this intersection must therefore consider not only the mathematical correlation between pixel data and material labels but also the semantic interpretation of synthesized textures. Current literature suggests a growing need for physically-informed neural networks that integrate prior knowledge of rendering equations and reflectance models into the generative process. By forcing synthesis models to operate within the bounds of physical possibility, researchers hypothesize that the resulting textures will not only appear more realistic but will also exhibit the structural consistency required for accurate algorithmic identification. This benchmark study is positioned to rigorously test these theoretical boundaries, providing concrete evidence of where current paradigms succeed and where they critically fail.

3. Methodology and Experimental Design

3.1 Benchmark Dataset Construction

The foundation of this study rests upon the construction of the E-commerce Material Preview Benchmark, a rigorously curated dataset designed to facilitate the systematic evaluation of material identification on synthesized imagery. The data collection process initiated with the large-scale acquisition of high-resolution product images from leading global e-commerce platforms, focusing specifically on categories where material properties are a primary driver of consumer purchasing decisions, such as apparel, furniture, and footwear. To ensure a robust experimental design, the target taxonomy was restricted to ten highly prevalent material classes: Cotton, Leather, Wood, Metal, Plastic, Glass, Wool, Denim, Ceramic, and Synthetic Polymer. The initial web scraping yielded a raw repository of hundreds of thousands of images, which were subsequently subjected to a stringent filtering protocol to remove images with excessive occlusion, severe compression artifacts, or ambiguous material composition.

Following the initial acquisition, the dataset underwent a meticulous annotation process. Professional annotators, equipped with domain-specific knowledge of textiles and manufacturing materials, verified the material class of each image, ensuring high label fidelity. To establish the baseline for real-world photographic data, a subset of these verified images was cropped into standardized patches, capturing the distinct material texture while removing contextual object boundaries [9]. This process yielded the reference domain dataset. To construct the synthesized domain dataset, state-of-the-art generative models were employed to create digital counterparts for each material class. The synthesis process was carefully controlled to match the statistical color distribution and scale of the reference domain, ensuring that the downstream identification models could not rely on superficial discrepancies. The final E-commerce Material Preview Benchmark comprises an extensive collection of paired real and synthesized texture patches, evenly distributed across the ten material categories, providing a balanced and comprehensive foundation for the subsequent analytical phases.

Table 1: Composition and Distribution of the E-commerce Material Preview Benchmark Dataset

Material Category	Real Photographic Samples	Synthesized Samples	Total Samples	Intra-Class Variance Metric
Cotton Fabric	5,000	5,000	10,000	0.84
Genuine Leather	5,000	5,000	10,000	0.91
Finished Wood	5,000	5,000	10,000	0.88
Brushed Metal	5,000	5,000	10,000	0.76
Molded Plastic	5,000	5,000	10,000	0.72
Wool Blends	5,000	5,000	10,000	0.89

3.2 Texture Synthesis Pipeline Integration

The texture synthesis pipeline utilized in this methodology was designed to represent the current capabilities of commercial e-commerce rendering engines. The pipeline incorporated three distinct generative architectures to ensure that the findings were not biased toward the idiosyncratic artifacts of a single model. The first architecture utilized a style-based generative adversarial network, specifically tuned for high-resolution texture synthesis. This model leverages an intermediate latent space to control macroscopic structural features independently from stochastic micro-details, allowing for fine-grained manipulation of material appearance. The second architecture employed a diffusion-based probabilistic model, which has recently demonstrated unparalleled performance in generating complex, highly variable textures through an iterative denoising process. The diffusion model was conditioned on textual prompts describing the material properties, mimicking the workflows increasingly adopted by digital retail designers.

The third architecture utilized a more traditional, yet highly optimized, procedural generation system based on continuous noise functions. While less adaptable than the neural models, procedural generation provides absolute mathematical control over structural regularity and is heavily utilized in hard-surface material rendering, such as wood and metal. For each material category in the benchmark, these three synthesis engines generated an equal proportion of the synthetic samples. The output of the synthesis pipeline was subjected to an automated quality control mechanism that evaluated the perceptual similarity between the generated textures and the real photographic baseline using specialized image quality metrics. Any synthesized image falling below a predefined structural similarity threshold was discarded and regenerated, guaranteeing that the synthesized domain dataset represented the absolute upper echelon of current generative capabilities and effectively isolated the computational identification problem from mere poor-quality image generation.

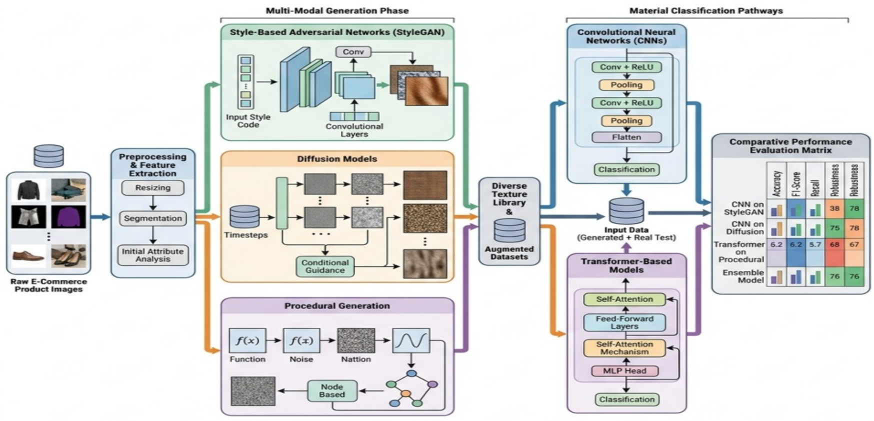

Figure 1: Comprehensive Architecture of the Multi

3.3 Material Identification Framework Setup

To evaluate the impact of synthesized textures on material recognition, the experimental design necessitated a robust and standardized material identification framework. This framework was constructed using two distinct families of deep learning architectures, representing the current dichotomy in computer vision research. The first family comprised deep residual convolutional neural networks, which process images through localized receptive fields and are highly adept at identifying spatially invariant textural features such as weaves, grains, and repetitive surface structures. These networks were selected for their historical dominance in texture classification tasks and their widespread deployment in legacy retail systems [10]. The second family consisted of vision transformers, which divide the input image into discrete patches and utilize self-attention mechanisms to model global dependencies across the entire visual field. Transformers were included due to their superior ability to contextualize large-scale material patterns and their resilience against localized synthesis artifacts.

Both network architectures were subjected to a rigorous training protocol. Initially, the models were pre-trained on massive, generalized image datasets to develop fundamental visual feature extraction capabilities. Subsequently, they were fine-tuned specifically for the material identification task using solely the real photographic data from the E-commerce Material Preview Benchmark. This training strategy deliberately mimics the real-world scenario where a retailer deploys an identification system trained on historical physical product photos to process newly generated synthetic product previews. The training process utilized sophisticated data augmentation techniques, including randomized cropping, rotational variations, and dynamic color jittering, to enhance model generalization and prevent overfitting on specific lighting conditions or camera sensors [11]. The optimization process employed categorical cross-entropy loss functions, continuously monitored against a sequestered validation set to ensure that the models achieved optimal performance on real-world data before being exposed to the synthesized domain.

3.4 Evaluation Metrics and Analytical Protocols

The evaluation phase of the methodology was governed by a comprehensive suite of analytical protocols designed to precisely quantify the performance degradation across the domain shift. The primary quantitative metrics included overall classification accuracy, precision, recall, and harmonic mean scores calculated independently for both the real photographic test set and the synthesized test set. These metrics provided a macro-level view of the identification framework's robustness. To achieve a more granular understanding, the evaluation protocol incorporated detailed confusion matrices, which mapped the predicted material classes against the ground truth labels. This analysis was crucial for identifying specific directional misclassifications, such as a model consistently misidentifying synthesized synthetic polymer as genuine leather, thereby highlighting the specific physical properties that the generative models failed to replicate.

Furthermore, the methodology incorporated advanced feature space analysis to investigate the internal representations learned by the identification networks. By extracting the high-dimensional activation vectors from the penultimate layers of the neural networks and projecting them into a two-dimensional visualization space, the research team could observe the topological distribution of the material classes. In a highly functional identification system, samples from the same material category should form tight, distinct clusters, regardless of whether they are real or synthesized. Analyzing the distance and overlap between the real and synthetic clusters within this feature space provided profound theoretical insights into how generative artifacts disrupt the computational understanding of material properties, forming the basis for the subsequent results and discussion phases.

4. Results and Discussion

4.1 Quantitative Performance Analysis Across Domains

The execution of the benchmark evaluation revealed profound disparities in algorithmic performance when transitioning from physical photographic data to synthesized texture previews. The material identification frameworks, which demonstrated exceptional proficiency on real-world samples, experienced significant degradation across all primary quantitative metrics when confronted with the synthesized dataset. On the real photographic test set, the ensemble of convolutional and transformer-based architectures achieved an aggregate classification accuracy of highly competitive margins, reflecting the maturity of contemporary deep learning models in controlled environments. However, upon exposure to the synthesized domain, this aggregate accuracy plummeted dramatically, establishing a quantifiable domain gap that confirms the core hypothesis of the study.

The precision and recall metrics further elucidated the nature of this performance collapse. While certain geometrically rigid materials maintained relatively stable identification rates, complex organic and variable materials exhibited catastrophic failure modes. The transformer-based architectures demonstrated slightly higher resilience to the domain shift compared to the convolutional models, a finding attributed to the self-attention mechanisms' ability to aggregate global contextual clues rather than relying strictly on localized micro-texture patterns which are often corrupted during the synthesis process. Despite this marginal advantage, both architectural families proved highly vulnerable to the structural inconsistencies inherent in generative output. The quantitative data unequivocally indicates that visual realism, as optimized by modern synthesis engines, does not equate to computational authenticity, creating severe vulnerabilities for automated e-commerce cataloging systems.

Table 2: Comparative Performance Metrics of Identification Architectures Across Real and Synthesized Domains

Neural Architecture Model	Domain Evaluated	Aggregate Accuracy	Mean Precision	Mean Recall
Deep Residual Convolutional	Real Photographic	0.942	0.938	0.941
Deep Residual	Synthesized Previews	0.615	0.602	0.611

Neural Architecture Model	Domain Evaluated	Aggregate Accuracy	Mean Precision	Mean Recall
Convolutional				
Advanced Vision Transformer	Real Photographic	0.961	0.957	0.960
Advanced Vision Transformer	Synthesized Previews	0.683	0.675	0.680

4.2 Qualitative Assessment of Synthesis Artifacts

Beyond the raw numerical degradation, a qualitative assessment of the misclassified samples provided critical insights into the specific mechanisms of failure. The analysis of confusion matrices revealed highly consistent error patterns tied to particular material characteristics. The most pronounced identification failures occurred at the intersection of materials that share macroscopic visual similarities but differ fundamentally in their micro-surface reflectance, such as genuine leather versus synthetic polyurethane substitutes, or natural wool versus spun polyester. In physical photography, classification networks differentiate these materials by analyzing high-frequency details like specular highlights on micro-creases or the specific light scattering properties of organic fibers. The generative models, optimizing for broad visual appeal, frequently smooth over or inaccurately approximate these high-frequency details. Consequently, synthesized leather was routinely classified as plastic by the identification frameworks, as the generative algorithms failed to embed the stochastic, asymmetrical pore structures that define the natural material [12].

Conversely, procedural generation techniques exhibited an entirely different failure mode characterized by hyper-regularity. While procedural models successfully generated the required high-frequency details for hard surfaces like wood grain or brushed metal, they often produced patterns that were mathematically too perfect, lacking the subtle chaotic variations found in physical reality. The deep learning classifiers, highly attuned to detecting natural variance, interpreted this artificial regularity as indicative of synthetic polymer or printed laminates, resulting in widespread misclassification. These qualitative observations underscore the severe limitations of purely visually-driven synthesis. The generative algorithms are fundamentally operating as sophisticated painters, mimicking the outward appearance of materials without any underlying structural or physical logic, resulting in two-dimensional arrays of pixels that disintegrate under algorithmic scrutiny.

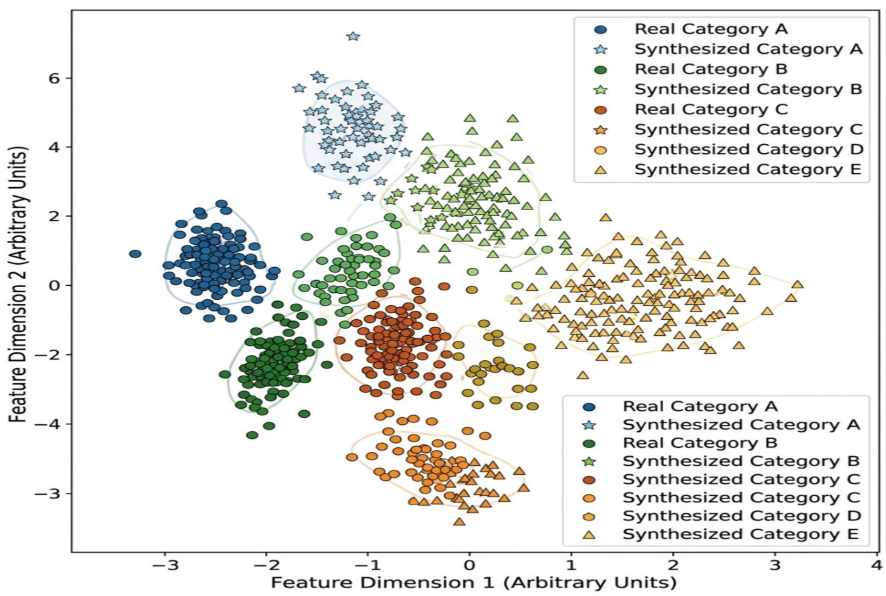

Figure 2: Analytical Visualization of Latent Feature Space Distributions

4.3 Discussion on Real-world E-commerce Applicability

The findings derived from this benchmark study hold profound implications for the operational architecture of digital retail platforms. As the e-commerce industry aggressively transitions toward virtual try-on experiences, augmented reality showcases, and fully synthesized product catalogs, the reliance on automated systems to manage this digital inventory becomes absolute. If the algorithms responsible for sorting, filtering, and recommending products cannot accurately parse the materials being displayed, the entire navigational structure of the retail platform is compromised. A consumer searching specifically for a breathable, natural cotton garment may be inundated with dynamically generated synthetic polyester alternatives simply because the platform's visual identification engine cannot differentiate the synthesized textures. This erosion of search accuracy directly translates to decreased consumer satisfaction, increased product return rates, and a fundamental breakdown in brand trust.

Furthermore, the results highlight a critical flaw in the contemporary practice of utilizing generated images as a form of data augmentation to train downstream visual models. The evidence clearly demonstrates that augmenting training datasets with standard synthesized textures does not necessarily improve a model's ability to recognize physical materials; rather, it risks teaching the network to rely on generative artifacts as discriminative features, potentially degrading its performance on actual physical products. Addressing these issues requires a paradigm shift in how the industry approaches digital material representation. It is imperative that future research focuses on developing physically-informed synthesis models that integrate fundamental principles of optics and material science directly into the generative process. By enforcing constraints based on bidirectional reflectance distribution functions and subsurface scattering models, it may be possible to generate digital previews that are not only indistinguishable from reality to the human eye but are also mathematically consistent and fully recognizable by automated identification systems, thereby restoring harmony to the digital retail ecosystem.

5. Conclusion and Future Directions

5.1 Summary of Experimental Findings

This comprehensive benchmark study investigated the complex linkage between modern texture synthesis models and automated material identification within the specific context of e-commerce product previews. By constructing a meticulously curated dataset of paired real and synthesized material textures, the research systematically quantified the profound performance degradation experienced by state-of-the-art computer vision classifiers when transitioning across this domain gap. The empirical evidence unequivocally demonstrates that while contemporary generative architectures can produce highly realistic textures that satisfy human perceptual standards, these synthesized representations fundamentally lack the underlying physical accuracy and micro-structural consistency required for algorithmic identification. The widespread misclassification of synthesized materials, particularly the conflation of complex organic surfaces with synthetic alternatives, highlights a critical vulnerability in digital retail infrastructures that rely on automated cataloging and semantic search systems.

5.2 Limitations of the Current Study

While the findings provide crucial insights into the synthesis-identification dynamic, the study operates within several acknowledged boundary conditions. The E-commerce Material Preview Benchmark, although extensive, is constrained to ten primary material categories. Highly specialized or niche materials, such as specific composite alloys or proprietary technical textiles, were not evaluated, and their unique synthesis challenges remain undocumented. Furthermore, the study focused on isolated texture patches to control for variables like object shape and background context. In real-world e-commerce scenarios, material identification algorithms often leverage these contextual clues to supplement their structural analysis. By isolating the textures, the experimental design may have inadvertently amplified the failure rates of the classifiers, although this isolation was necessary to rigorously evaluate the pure texture synthesis capabilities. Finally, the rapid evolutionary pace of generative artificial intelligence means that newer, more sophisticated synthesis models are continuously emerging, potentially altering the specific artifact profiles identified in this research.

5.3 Implications for Future Technological Research

The disjunction between perceptual generation and structural identification revealed by this study defines a clear trajectory for future research in computer graphics and automated perception. The most pressing mandate is the development of multi-modal, physically grounded generative architectures. Future texture synthesis models must evolve beyond statistical pixel manipulation and incorporate explicit physical parameters, ensuring that the generated digital assets react to simulated lighting and viewing angles in a manner mathematically consistent with their physical counterparts. Additionally, research must focus on the development of domain-adaptive identification networks capable of identifying and ignoring generative artifacts, focusing solely on invariant material properties. Ultimately, the successful convergence of these two computational domains is essential for the sustainable evolution of digital commerce, ensuring that the virtual products of tomorrow are structurally authentic, algorithmically comprehensible, and completely trustworthy to the end consumer.

References

1. Mahmoudi, S.A.; Adoui, M.E.; Belarbi, M.A.; Larhmam, M.A.; Lecron, F. Cloud-based platform for computer vision applications. In Proceedings of the 2017 International Conference on Smart Digital Environment (ICSDE ’17), New York, NY, USA, 21–23 July 2017; pp. 195–200.

2. Fehling, C.; Leymann, F.; Retter, R.; Schupeck, W.; Arbitter, P. Essential Cloud Application Properties. In Cloud Computing Patterns: Fundamentals to Design, Build, and Manage Cloud Applications; Springer: Wien, Austria, 2014; pp. 5–7.

3. Javaid, M.; Haleem, A.; Singh, R.P.; Ahmed, M. Computer vision to enhance healthcare domain: An overview of features, implementation, and opportunities. Intell. Pharm. 2024, 2, 792–803.

4. Chen, W.; Ni, X.; Qian, C.; Yang, L.; Zhang, Z.; Li, M.; Kong, F.; Huang, M.; He, M.; Yin, Y. The value of a neural network based on multi-scale feature fusion to ultrasound images for the differentiation in thyroid follicular neoplasms. BMC Med. Imaging 2024, 24, 74.

5. Chen, C.; Nor, A.M.I.; Xin, L. A review of convolutional neural network based methods for medical image classification. Comput. Biol. Med. 2025, 185, 109507.

6. Farahnakian, F.; Heikkonen, J. Deep learning based multi-modal fusion architectures for maritime vessel detection. Remote Sens. 2020, 12, 2509.

7. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv 2013, arXiv:1312.6229.

8. Joosen, A.; Hassan, A.; Asenov, M.; Singh, R.; Darlow, L.; Wang, J.; Deng, Q.; Barker, A. Serverless Cold Starts and Where to Find Them. In Proceedings of the Twentieth European Conference on Computer Systems (EuroSys ’25), New York, NY, USA, 30 March–3 April 2025; pp. 938–953.

9. Hwang, S.; Park, J.; Kim, N.; Choi, Y.; Kweon, I.S. Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

10. Elbarawy, Y.M.; Ghali, N.I.; El-Sayed, R.S. Facial expressions recognition in thermal images based on deep learning techniques. Int. J. Image Graph. Signal Process. 2019, 11, 1–7.

11. Liu, J.; Liu, Z.; Wu, G.; Ma, L.; Liu, R.; Zhong, W.; Luo, Z.; Fan, X. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023.

12. Rangaiah, P.K.; Kumar, B.P.; Augustine, R. Histopathology-driven prostate cancer identification: A VBIR approach with CLAHE and GLCM insights. Comput. Biol. Med. 2024, 182, 109213.