Structural Preservation Governs Data Augmentation in Deep Learning-Based Laser Speckle Material Classification

Mohamed Abdallah Salem
College of Engineering
North Dakota State University
Fargo, ND, USA, 58102
ORCID: 0009-0004-4673-420X

Nourhan Zein Diab
College of Computer Science and Engineering
New Mansoura University
New Masoura, Egypt
nourdiab736@gmail.com

arXiv:2607.22725v1 [cs.CV] 22 Jul 2026

***Abstract*—Data augmentation is routinely used to improve generalization in image classification, but the assumptions underlying standard policies are poorly matched to coherent imaging. Laser speckle patterns are not generic textures; they arise from coherent interference, and their discriminative content is carried by structured stochastic spatial and frequency statistics. This study examines how controlled augmentation perturbations influence speckle-based material classification on the SensiCut dataset. We train ResNet18 and EfficientNet-B0 under a parametric augmentation framework comprising rotation, Gaussian blur, independent Gaussian noise, spatially correlated speckle-aware noise, intensity jitter, and spatial masking, and evaluate test performance using macro F1-score averaged over three random seeds. Separate ordinary least squares models link augmentation parameters to performance for each architecture. Across both models, Gaussian blur exerts a strong negative effect ($p < 0.001$), indicating that low-pass filtering suppresses high-frequency structure that is informative for material discrimination. Independent pixel-wise noise is likewise harmful ($p = 0.003$ for EfficientNet-B0 and $p = 0.001$ for ResNet18), consistent with disruption of local spatial coherence. In contrast, spatially correlated perturbations yield significant positive coefficients ($p = 0.004$ for EfficientNet-B0 and $p = 0.001$ for ResNet18), showing that variability can improve robustness when it preserves speckle organization. The fitted models explain a substantial fraction of performance variation ($R^2 = 0.796$ for EfficientNet-B0 and $R^2 = 0.879$ for ResNet18). These results show that, in laser speckle imaging, augmentation effectiveness is determined primarily by structural preservation rather than perturbation magnitude. The findings motivate physics-aware augmentation design for coherent optical sensing.**

***Index Terms*—laser speckle imaging, data augmentation, coherent imaging, material classification, robustness, deep learning**

I. Introduction

LASER speckle sensing exploits the fact that coherent illumination interacting with a rough or heterogeneous surface produces an interference pattern whose spatial statistics depend on the underlying scattering structure. In material recognition settings, those patterns are informative precisely because they retain sensitivity to microstructure, surface finish, and optical response rather than to conventional object appearance alone. Prior work has shown that speckle images can support material-aware sensing and classification, including in the SensiCut system for laser cutting, but the signal is fundamentally different from the semantic content usually assumed in natural-image classification pipelines [1]–[3]. In speckle imagery, class-relevant information is embedded in local coherence, interference-driven texture, and frequency content, which makes the choice of augmentation substantially more delicate than in standard photographic datasets.

Data augmentation has become a central component of modern deep learning because it can enlarge the effective support of the training distribution and improve generalization under limited data [4], [5]. Most widely used policies, however, were developed for natural images and encode assumptions of label invariance to pose changes, photometric perturbations, occlusion, or synthetic corruption. The robustness literature has repeatedly shown that these gains are not uniform across corruption types and do not necessarily transfer to natural distribution shifts [6]–[8]. Frequency-domain analyses further indicate that augmentation can alter the spectral sensitivity of convolutional models in nontrivial ways [9]. These observations are especially relevant in coherent imaging, where a transformation that appears visually mild may nevertheless alter the physics-bearing statistics of the measurement.

Recent work in scientific imaging has reinforced the need for domain-specific robustness analysis. Deep learning is now widely used in holography and coherent imaging [10], and robustness studies in microscopy have shown that conclusions derived from natural-image benchmarks do not automatically transfer to scientific acquisition settings [11]. Yet, despite growing interest in speckle-based learning, there remains limited understanding of which augmentation families preserve the information required for accurate classification and which ones distort that information beyond usefulness. The central question in this paper is therefore not simply whether augmentation helps, but which physical properties of speckle must be preserved for augmentation to remain beneficial.

The present study addresses this question through a con-

979-8-3315-8488-7/26/$31.00 © 2026 IEEE

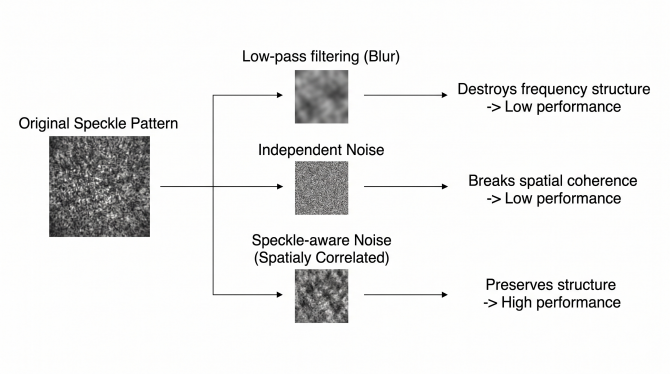


Fig. 1. Conceptual summary of the structural-preservation hypothesis examined in this work. Low-pass blur suppresses frequency content, independent pixel-wise noise breaks local coherence, and speckle-aware perturbations preserve spatial organization while introducing controlled variability.

trolled, parametric analysis of augmentation effects in laser speckle material classification. Using the SensiCut dataset and two representative convolutional backbones, ResNet18 and EfficientNet-B0, we vary rotation, blur, independent noise, spatially correlated speckle-aware noise, intensity jitter, and spatial masking over continuous ranges and fit architecture-specific regression models to relate augmentation parameters to macro F1-score. This design allows augmentation to be studied as a structured perturbation problem rather than as a set of isolated recipes. The resulting manuscript argues, and the experiments support, that augmentation effectiveness in speckle imaging is governed primarily by structural preservation of spatial coherence and frequency content. Figure 1 summarizes this working hypothesis and the three main transformation behaviors evaluated in the study.

II. Related Work

The literature on augmentation for deep learning is extensive, but much of it is grounded in natural-image benchmarks. Surveys such as that of Shorten and Khoshgoftaar [4] document the broad utility of geometric and photometric augmentation, while policy-search methods such as AutoAugment show that learned augmentation schedules can improve benchmark accuracy when the search space is well aligned with the task [5]. Occlusion-style perturbations, including random erasing, have likewise been effective as regularizers when partial spatial removal is a plausible nuisance factor [12]. The common thread across these methods is an implicit assumption that semantic identity is preserved under the augmentation family being applied.

Robustness studies complicate this picture by showing that augmentation gains are strongly corruption-dependent. Hendrycks and Dietterich demonstrated that many high-performing image classifiers remain fragile to blur, noise, and other common corruptions [6]. Rusak *et al.* showed that robustness can improve when training perturbations more closely match the corruption family encountered at test time [7]. Taori *et al.* further reported that robustness to synthetic corruptions and robustness to natural distribution shifts are related but distinct objectives [8]. From a spectral perspective, Yin *et al.* showed that corruption sensitivity can be understood in part through the Fourier characteristics of the perturbation and the model [9]. Taken together, these studies imply that augmentation should be interpreted as a distribution-design problem rather than a universally beneficial preprocessing step.

The coherent imaging and speckle literature provides additional motivation for task-specific analysis. Reviews of laser speckle contrast imaging emphasize that speckle statistics are tightly coupled to optical coherence, acquisition geometry, and sample dynamics [1], [2]. At the same time, deep learning is increasingly used in coherent imaging pipelines, where the interaction between network priors and wave-based measurements has become an active research topic [10]. Learning from speckle patterns has already proven useful in several applications, including material-aware laser cutting [3], [13], [14] and the interpretation of coherent X-ray speckle patterns [15]. Subsequent application-specific studies have extended this line of work to laser-cutting material classification with red-laser speckle acquisition, hazardous-material detection during laser cutting, robustness to laser-color changes, and lightweight edge-compatible CNN design for speckle-based recognition [13], [16], [17]. However, the augmentation question remains underdeveloped. Existing work demonstrates that speckle can be informative, but it does not systematically identify which augmentation mechanisms preserve the statistics that make speckle informative in the first place. This gap motivates the present study.

III. Methodology

The experiments were designed to isolate how different augmentation mechanisms alter performance in speckle-based material classification.

A. Dataset Preparation

We used the public SensiCut dataset introduced by Dogan *et al.* [3]. The dataset consists of laser speckle images acquired from multiple physical material surfaces using a controlled optical sensing system. The SensiCut taxonomy naturally supports two complementary label granularities. In the fine-grained formulation, each physical sample is treated as an independent class, resulting in 59 sample-level classes that capture intra-material variability associated with surface finish, manufacturing conditions, and acquisition instances. In the coarse formulation, the same samples are grouped according to their underlying material type, yielding 30 classes consistent with the original material taxonomy. For the augmentation analysis reported in this paper, results are presented for the 30-class material-level formulation, while the 59-sample hierarchy is retained during dataset preparation to clarify label granularity and split semantics.

After curation for the present study, the working dataset comprised 39,609 images. We created training, validation, and test subsets using a stratified image-level split of 70%, 15%, and 15%, respectively, so that every class remained represented

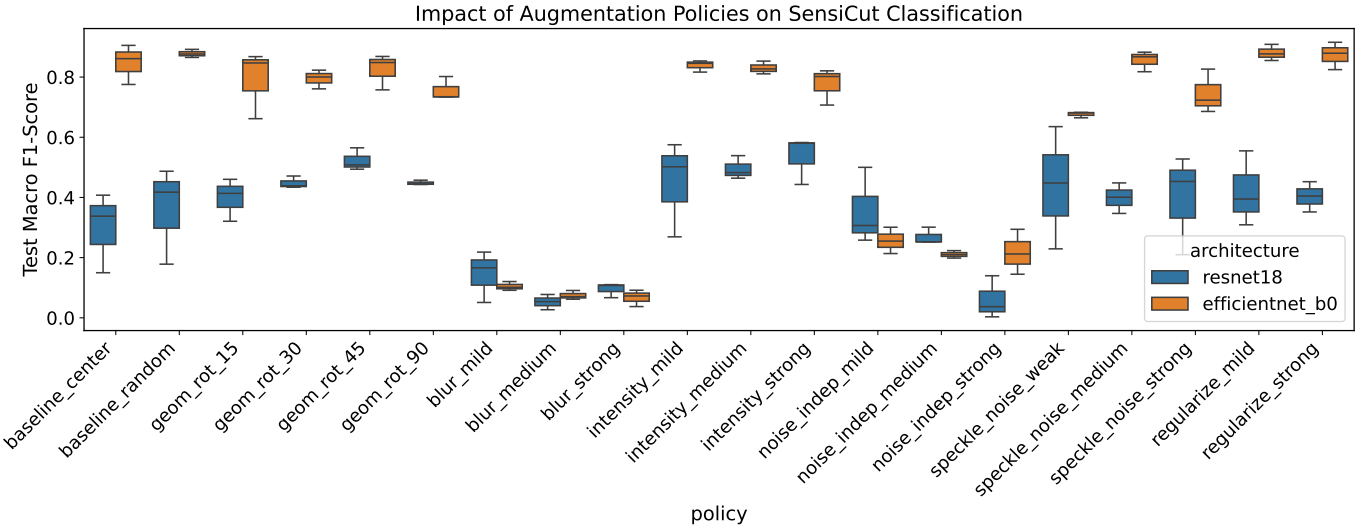


Fig. 2. Distribution of test macro F1 across augmentation policies and random seeds. The main qualitative pattern is the collapse caused by blur and independent noise, contrasted with the more favorable behavior of moderate structure-preserving policies.

while image instances stayed disjoint across subsets. The resulting evaluation therefore measures generalization to previously unseen observations of the same physical samples rather than to entirely unseen material instances. Class frequencies were approximately balanced in both label formulations, with only minor deviations attributable to acquisition constraints. Performance was therefore evaluated using macro-averaged F1-score.

Preprocessing was designed to preserve the structural characteristics of speckle patterns. Interpolative resizing was avoided because it can introduce smoothing and aliasing that alter the frequency content of the signal. Instead, fixed-size 224×224 patches were extracted from the native images by cropping. Within the experimental design, center and random crops were retained as explicit baselines, allowing positional variability to be introduced without disturbing the native spatial statistics of the speckle field. This formulation supports controlled evaluation of augmentation robustness while minimizing confounding effects arising from sample-level distribution shifts.

B. Model Architectures

To test whether the augmentation effects were architecture-specific or instead reflected broader properties of the data, we used two convolutional architectures. ResNet18 was chosen as a canonical residual network with comparatively modest depth and a well-established optimization profile [18]. EfficientNet-B0 was used as a stronger baseline that combines compound scaling with a more parameter-efficient architecture [19]. Evaluating both models allows the analysis to compare a conventional residual backbone with a more modern architecture that often exhibits stronger baseline accuracy.

C. Parametric Augmentation and Training

Augmentation was formulated parametrically so that each transformation could be interpreted as a controllable perturbation of the input distribution rather than as a binary on/off choice. The geometric variable was in-plane rotation, parameterized by angle θ. Frequency-domain distortion was introduced with Gaussian blur of standard deviation σ_b, which acts as a low-pass filter. Independent stochastic perturbation was implemented as additive Gaussian noise with standard deviation σ_n, sampled independently at each pixel. Speckle-aware perturbation was implemented as spatially correlated noise with correlation scale σ_s, constructed to introduce local variation while preserving neighborhood structure. Intensity jitter of magnitude j modulated global brightness and contrast, and spatial masking was controlled by erase probability p_e, following the general idea of random erasing [12]. In addition to these augmented policies, center-crop and random-crop baselines were retained for comparison. The full design comprised 20 policies, evaluated for both architectures across three random seeds, for a total of 120 trained models.

All models were trained under the same optimization protocol, with identical stopping criteria and seed averaging so that comparisons reflect augmentation effects rather than changes in optimization procedure. Performance was measured on the held-out test set using macro F1-score, which weights all classes equally and is therefore appropriate when per-class behavior is as important as aggregate accuracy. For each policy, the reported score corresponds to the mean across three seeds.

D. Regression Formulation

To quantify the contribution of each augmentation variable, we fit a separate ordinary least squares model for each architecture. The response variable was test macro F1-score, and the predictors were the continuous augmentation parameters:

$$y = \beta_0 + \beta_1\theta + \beta_2\sigma_b + \beta_3\sigma_n + \beta_4\sigma_s + \beta_5 p_e + \beta_6 j + \varepsilon, \quad (1)$$

where y denotes mean test macro F1-score and ε is the residual term. This regression formulation is useful because it estimates the direction and strength of each transformation while controlling for the others. It also makes it possible to distinguish whether a policy performs well because of one dominant perturbation or because several transformations act synergistically.

IV. Results

The augmentation landscape is summarized qualitatively by the performance distribution in Fig. 2 and the mean–variance view in Fig. 4. EfficientNet-B0 achieved higher absolute macro F1 than ResNet18 under most policies, but the ranking of harmful and beneficial perturbations was similar across both architectures. Gaussian blur consistently occupied the lowest-performing region, with mean macro F1 values below 0.11 for all three blur settings in both models. Independent Gaussian noise also substantially degraded performance, especially for EfficientNet-B0, for which all independent-noise policies remained near 0.23 macro F1 or below. By contrast, moderate structure-preserving policies remained competitive. For EfficientNet-B0, moderate spatial masking and medium speckle-aware noise were among the strongest configurations, alongside the random-crop baseline. For ResNet18, performance was more variable in absolute terms, but blur and independent noise again defined the low-performance regime, while moderate rotations and intensity modulation were comparatively more tolerant.

The regression analysis in Table I formalizes these observations. For EfficientNet-B0, the model explained 79.6% of the observed variance in macro F1 ($R^2 = 0.796$, model $p = 7.10 \times 10^{-4}$). Gaussian blur had a strongly negative coefficient ($\beta = -0.2693$, $p < 0.001$), indicating that low-pass filtering systematically reduced performance. Independent Gaussian noise was also significantly harmful ($\beta = -3.2102$, $p = 0.003$). In contrast, the speckle-aware perturbation term was positive and significant ($\beta = 0.2233$, $p = 0.004$), showing that spatially correlated variability improved robustness when it preserved local structure. Rotation, erase probability, and intensity jitter were not significant in the EfficientNet-B0 model.

For ResNet18, the fitted model explained 87.9% of the variance in test macro F1 ($R^2 = 0.879$, model $p = 2.82 \times 10^{-5}$). The signs of the key coefficients matched those of EfficientNet-B0. Blur again showed a strong negative effect ($\beta = -0.1217$, $p < 0.001$), and independent noise remained significantly harmful ($\beta = -1.4016$, $p = 0.001$). The coefficient of the speckle-aware term was positive and statistically significant ($\beta = 0.0984$, $p = 0.001$). Rotation was marginal but did not cross the 0.05 threshold, and spatial masking remained nonsignificant after controlling for the other variables. Intensity jitter differed from EfficientNet-B0 in that it showed a positive association with performance in ResNet18 ($\beta = 0.2210$, $p = 0.014$), suggesting a model-dependent tolerance to global photometric modulation. Nevertheless, the dominant pattern across architectures was the same: transformations that suppressed frequency structure or broke local coherence were detrimental, whereas structure-preserving perturbations were beneficial.

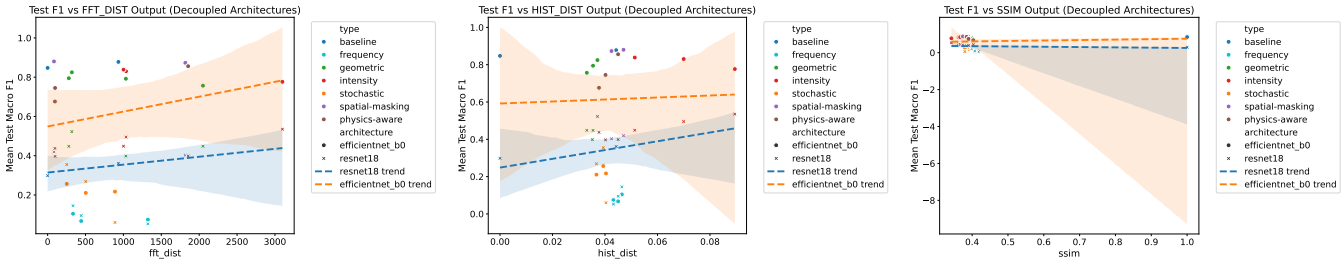
The distortion-oriented plots in Fig. 3 and the Pareto analysis in Fig. 4 provide a complementary view. Among the scalar distortion descriptors, the SSIM-based view showed the clearest monotonic association with macro F1 across both architectures, whereas histogram distance was comparatively diffuse, indicating that simple intensity-distribution changes alone do not explain the performance ordering. The Fourier-distance view is consistent with the regression result that frequency-disruptive transformations are harmful, but it also shows that spectral change by itself is not sufficient to predict performance without considering spatial organization. The Pareto plot further shows that the most useful policies are not merely those that maximize average F1, but those that do so without inflating variance across seeds. In this regard, moderate structure-preserving perturbations occupy a more favorable frontier than blur or independent noise, which are simultaneously low-performing and unstable.

V. Discussion

The results support a simple but consequential interpretation: in laser speckle classification, the utility of augmentation depends less on perturbation magnitude than on whether the perturbation respects the statistical structure generated by coherent scattering. Gaussian blur was consistently harmful because it removes high-frequency content and reduces local contrast fluctuations that are intrinsic to speckle formation. In natural images, blur may remove nuisance texture while leaving semantic shape relatively intact. In speckle imagery, by contrast, the fine-scale fluctuations are not nuisance details but part of the class-discriminative signal itself [1], [2]. The negative blur coefficients observed in both architectures are therefore physically interpretable rather than merely empirical.

Independent Gaussian noise failed for a related but distinct reason. Even when its variance is controlled, pixel-wise noise perturbs each location without regard to the correlation structure that coherent interference induces across neighboring pixels. This creates samples that are statistically implausible under the acquisition process. The fact that independent noise was significantly harmful in both regression models suggests that robustness in this domain cannot be improved by arbitrary stochasticity. What matters is not simply exposing the model to variation, but exposing it to variation that lies on a plausible manifold for speckle formation. This conclusion is consistent with broader robustness findings showing that mismatch between training perturbations and evaluation distortions can negate the intended regularization benefit [7], [8].

The positive effect of speckle-aware perturbations is therefore particularly informative. Spatially correlated noise introduces variability while preserving local organization, and in doing so it acts more like a physically plausible perturbation



(a) FFT distance (b) Histogram distance (c) SSIM

Fig. 3. Relationships between test macro F1 and scalar descriptors of augmentation-induced distortion. The plots show that structural descriptors are more informative than histogram change alone, while also indicating that no single scalar metric fully captures the interaction between augmentation and speckle statistics.

TABLE I
ORDINARY LEAST SQUARES COEFFICIENTS RELATING AUGMENTATION PARAMETERS TO TEST MACRO F1-SCORE. SEPARATE MODELS WERE FITTED FOR EFFICIENTNET-B0 AND RESNET18.

Predictor	EfficientNet-B0 coef.	EfficientNet-B0 p	ResNet18 coef.	ResNet18 p
Rotation θ	0.0013	0.536	0.0014	0.070
Blur σ_b	-0.2693	< 0.001	-0.1217	< 0.001
Independent noise σ_n	-3.2102	0.003	-1.4016	0.001
Speckle-aware noise σ_s	0.2233	0.004	0.0984	0.001
Erase probability p_e	0.6830	0.287	0.1351	0.558
Intensity jitter j	0.1589	0.468	0.2210	0.014
R^2	0.796	—	0.879	—
Model p-value	7.10×10^{-4}	—	2.82×10^{-5}	—

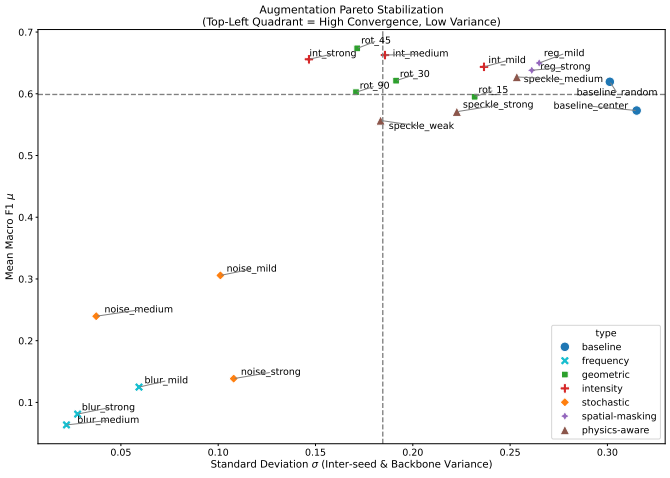


Fig. 4. Pareto view of mean test macro F1 versus variance across augmentation policies. Policies near the favorable frontier combine strong average performance with stable behavior across seeds.

than like an artificial corruption. The effect was significant in both EfficientNet-B0 and ResNet18, despite the substantial difference in their absolute performance levels. This sign consistency indicates that the phenomenon is data-centric rather than architecture-specific. It also suggests that physics-aware augmentation need not be highly elaborate to be effective; even relatively simple correlation-preserving perturbations can move the training distribution in a beneficial direction when they respect the structure of the sensing process.

The remaining augmentation families help refine this interpretation. Rotation produced small coefficients and no clear regression significance, which suggests that modest geometric variation is neither especially beneficial nor especially destructive once the stronger structural effects are accounted for. Spatial masking performed well in some settings, particularly for EfficientNet-B0, but its regression coefficient was not significant. This indicates that masking may function as a secondary regularizer rather than as a primary mechanism of robustness in this domain. The positive jitter coefficient for ResNet18, absent in EfficientNet-B0, points to a model-dependent tolerance to global photometric shifts. Importantly, however, these secondary effects do not overturn the central result. Across both architectures, the dominant determinants of augmentation success are preservation of spatial coherence and retention of informative frequency structure.

These findings have practical implications for coherent optical sensing. Standard augmentation search spaces imported from natural-image classification are unlikely to be optimal when the measurement itself is a stochastic interference pattern. Instead, augmentation design should be constrained by optical plausibility. For speckle-based recognition, this means favoring transformations that preserve correlation length, local continuity, and spectral envelope while avoiding operations that sever those relationships. More generally, the results align with emerging evidence from coherent and scientific imaging that model robustness depends on respecting the

generative physics of the measurement process [10], [11], [15]. Future work should therefore move beyond generic policy transfer toward adaptive augmentation schemes informed by wave-optics models, class-conditional distortion tolerances, or learned measures of speckle fidelity.

VI. Conclusion

This paper presented a structured analysis of data augmentation for laser speckle material classification using the SensiCut dataset and two convolutional architectures. By treating augmentation parameters as continuous variables and relating them to macro F1-score through ordinary least squares regression, the study moved beyond recipe-level comparison and identified the dominant factors associated with performance changes. The results were consistent across EfficientNet-B0 and ResNet18: Gaussian blur and independent pixel-wise noise degraded classification because they suppress frequency information or disrupt local coherence, whereas spatially correlated speckle-aware perturbations improved robustness by preserving the organization of the speckle field.

The main implication is that augmentation effectiveness in coherent imaging is governed by structural preservation. In speckle-based learning, useful variability is not arbitrary variability. It must remain compatible with the spatial and frequency statistics induced by coherent scattering. This principle provides a concrete foundation for future augmentation design in optical sensing, including physics-aware policies, adaptive strategies tied to measured distortion, and joint learning frameworks that incorporate forward models of image formation.

References

[1] D. A. Boas and A. K. Dunn, "Laser speckle contrast imaging in biomedical optics," *Journal of Biomedical Optics*, vol. 15, no. 1, p. 011109, 2010.

[2] D. Briers, D. D. Duncan, E. Hirst, S. J. Kirkpatrick, M. Larsson, W. Steenbergen, T. Stȓomberg, and O. B. Thompson, "Laser speckle contrast imaging: Theoretical and practical limitations," *Journal of Biomedical Optics*, vol. 18, no. 6, p. 066018, 2013.

[3] M. D. Dogan, S. V. Acevedo Colon, V. Sinha, K. Akşit, and S. Mueller, "Sensicut: Material-aware laser cutting using speckle sensing and deep learning," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 24–38.

[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[5] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113–123.

[6] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *International Conference on Learning Representations*, 2019.

[7] E. Rusak, L. Schott, R. S. Zimmermann, J. Bitterwolf, O. Bringmann, M. Bethge, and W. Brendel, "A simple way to make neural networks robust against diverse image corruptions," in *Computer Vision – ECCV 2020*, ser. Lecture Notes in Computer Science. Springer, 2020, vol. 12348, pp. 53–69.

[8] R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, and L. Schmidt, "Measuring robustness to natural distribution shifts in image classification," in *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[9] D. Yin, R. G. Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, "A fourier perspective on model robustness in computer vision," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 13 255–13 265.

[10] Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light: Science & Applications*, vol. 8, p. 85, 2019.

[11] L. Zhong, L. Li, and G. Yang, "Benchmarking robustness of deep neural networks in semantic segmentation of fluorescence microscopy images," *BMC Bioinformatics*, vol. 25, no. 1, p. 269, 2024.

[12] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, 2020, pp. 13 001–13 008.

[13] M. A. Salem, A. Elshenawy, and H. A. Ashour, "Material classification in laser cutting using deep learning," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*. IEEE, 2023, pp. 167–173.

[14] M. A. Salem, H. A. Ashur, and A. Elshinnawy, "Towards a safer and sustainable manufacturing process: Material classification in laser cutting using deep learning," 2025. [Online]. Available: https://arxiv.org/abs/2511.16026

[15] M. Shen, D. Sheyfer, T. D. Loeffler, S. K. R. S. Sankaranarayanan, G. B. Stephenson, M. K. Y. Chan, and D. Morgan, "Machine learning for interpreting coherent x-ray speckle patterns," *Computational Materials Science*, vol. 230, p. 112500, 2023.

[16] M. A. Salem, A. K. ElShenawy, and H. A. Ashour, "Detection of hazardous materials in laser cutting using deep learning and speckle sensing," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-1/W2-2023, pp. 497–503, 2023.

[17] M. A. Salem and N. Z. Diab, "Efficient edge-compatible cnn for speckle-based material recognition in laser cutting systems," in *2025 35th International Conference on Computer Theory and Applications (ICCTA)*, 2025, pp. 296–301.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[19] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 97. PMLR, 2019, pp. 6105–6114.