Diffusion Models for Advances in Manufacturing Quality – Abrasive Grains

Srinika Sharma[1], Riddhi Aparajitha[1], Prafulla Kalapatapu[1], N. S. Karthiselva[2], Koshy Abraham[2], Shyam Rao[2], and Arya Bhattacharya[1]

[1] Mahindra University, Hyderabad, India
[2] Carborundum Universal Ltd., India

Abstract. Artificial Intelligence is playing a significant and growing role in manufacturing quality assessment and control. Some examples include methods for real-time condition monitoring and prediction of adverse digressions, online vision based systems that can obviate sample based testing, enabling optimal control that can run processes at near-optimal conditions most of the time, etc. For most of these systems and in particular for sophisticated methods like CNNs, model accuracy and performance depend on the number and representativeness of data samples, especially for samples from faulty conditions which are relatively rare and trigger training data imbalance. Generative Artificial Intelligence can potentially play a major role in creating new samples, especially those of the rarer kind, provided statistical distinction between synthetic and real samples can be minimized. Diffusion based techniques are the current frontline among generative models, and in this work we have developed mechanisms for using vision based Diffusion models to get real time quality indications in the manufacturing process of Abrasive Grains – which are critical from a functional viewpoint and quite challenging from a technical perspective. We have converted Stable Diffusion architectures that take text-based pre-conditions to generate images, into statistics-based quality demand vector conditions and generate new and diverse images of type and merit demanded in the pre-conditions. Generated images contribute significantly to training accurate CNN models that status quality of a running production process, allowing not only quality decisions but also potential in-process feedback control of relevant production parameters.

Keywords: Abrasive Grains, Generative Models, Stable Diffusion, Statistical Quality pre-conditioning, Online product quality monitoring.

1 Introduction

Abrasive Grains are small sized, hard and sharp-cornered particles that do the actual cutting, grinding, polishing and surface finishing of materials by removing tiny chips from their surface. These grains are typically embedded in bonded or coated abrasive cutting and finishing tools and grinding wheels, to remove material from workpieces in a controlled manner. Effectively, each grain acts as a tiny cutting tool. Abrasive grains are commonly composed of variants of Aluminum

oxides, with different manufacturing process and microstructure based on the specific functions that they are supposed to perform.

These abrasive grains, usually unobserved because of their small sizes, perform many critical roles in multiple manufacturing processes, with significant impact on quality of products and to an extent, on stability of processes. As a few non-exclusive examples, they are used for clean and controlled cutting of hard or brittle materials, removing scratches from and polishing aerospace and automotive parts, and generally improving surface appearance. In grinding operations, they contribute to achieving high dimensional accuracy and fine surface finish. They find use in honing engine cylinder bores, hydraulic and pneumatic cylinders, bearing surfaces, among many other applications.

An important process in the manufacture of these abrasive grains, comprises of the modification of their surfaces. A prevalent technique in this domain involves coating abrasive grains with ceramic oxides. This coating leads to a number of benefits, like providing a distinct color signature, and ensuring inter-grain uniformity in batch products. This uniformity is critical to ensure batch level consistency and quality, and attainment of performance and functional objectives.

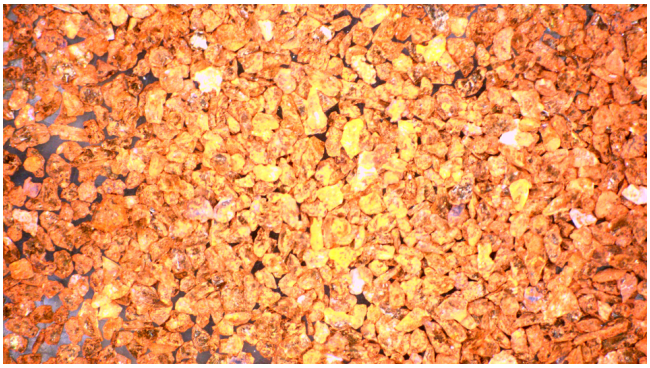

Fig. 1. Original Micrograph Sample, from Shop Floor in Production Phase

Quality assurance related evaluations are conducted at the end of the production line, for each batch (a few tonnes in a large bucket) of produced grains. Random samples from the batch are picked up at arbitrary spatial and temporal (i.e. even as the bucket is being populated) intervals, and micrographs (images under a microscope, here of size 300 x 300 pixels) are taken. Fig. 1 shows a typical micrograph. To get an indication of consistency and adherence to quality, first the Red, Green, Blue (RGB) values of each micrograph are measured in a Natural Color System Colorpin [1], and the deviation from a benchmark quality, i.e. colour distribution, noted. Next, six of these micrographs (in 2 x 3 format) accumulated from non-spatially-and-temporally adjacent samples are concatenated into a single large image (implying size of 600 x 900 pixels in the integral

image), and average RGB, and importantly the Standard Deviation (SD) of each of Red, Green, Blue are statistically computed. This SD provides a measure of the consistency of grains within a product batch. Fig. 2 shows a typical such concatenated image.

The question arises at this stage as to the relevance of any image processing and Vision-based algorithms? These immediately come into play when we translate the above manual sampling and measuring process into an algorithm and methodology for automated, real-time, quality and consistency extraction.

In such an online system, cameras are placed at a point in the flow of the produced grains just before they "stream" into the bucket. Note this has got good resemblance to characteristics of fluid flow, and images of same dimensions as in the original manual micrographs (300 x 300) are taken automatically at regular time intervals, and six such consecutive images are concatenated into an integral image in (2 x 3) format without manual intervention. Then this image is passed through an accurate Vision-based software to extract the average RGB as well the SD and get indications of quality and consistency. This Vision based algorithm is based on ResNet50 architecture and obviously has to be trained on sample images taken manually like described above, following principles of supervised learning and related transfer learning. This trained ResNet50 takes integral images as input and generates 6 values (RGB and their SDs) as output.

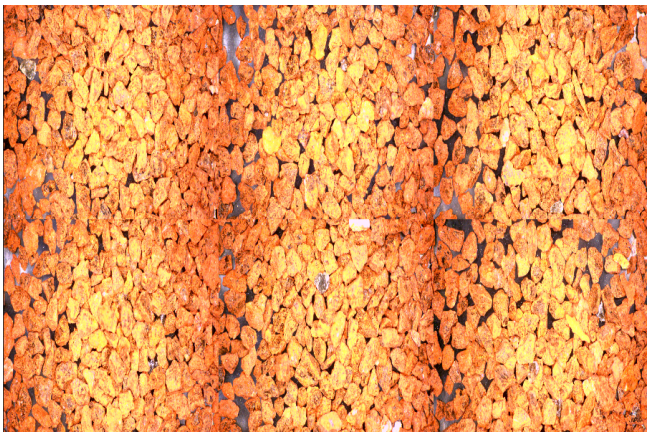

Fig. 2. Concatenated Integral Image - taken from 2 x 3 micrographs

Training a ResNet50 as mentioned even under transfer learning requires a very large set of manually extracted and then quality (RGB and SD) computed integral image samples. Taking samples in the production process, extracting micrograph images under microscope, followed by corresponding RGB values from the physical samples using NCSColorpin, and finally concatenating them into integral images with corresponding six statistical parameters, is an arduous process. In this work the authors have developed a procedure to start with a

small set of manual quality-adjudicated integral images, augment them using some novel techniques, and then use the expanded training set in a Diffusion model for generating a large number of new images. This is executed in a multi-stage process that is explained in subsequent sections of the paper. The new images are then used to train a ResNet50 model which is validated for new manually extracted original images and their 6-element statistical vector, and found to provide good performance, enabling use of the trained model for online streaming quality assessment as indicated above.

The rest of the paper consists of the following sections. Section 2 presents related work, while Sec. 3 explains the Diffusion model that is used and appropriately transformed for this application. Section 4 describes how our modified Diffusion model is implemented in the specific manufacturing application for generation of multiple new and diverse images for actual automated and online process improvement.

2 Related Work

Building on the industrial motivation and problem formulation outlined in the introduction section, we review prior work on modeling of microstructures, with an emphasis on diffusion-based generative frameworks. The synthesis and reconstruction of grain-structured and polycrystalline microstructures has long posed challenges due to the need to preserve complex spatial statistics, grain boundary connectivity, and physically meaningful morphology across multiple length scales. Diffusion-based generative models offer stable training dynamics and strong capacity for high-dimensional structured data, motivating their rapid adoption for microstructure synthesis, reconstruction, and inverse design in both two- and three-dimensional settings. While much of this literature emphasizes structural realism, our interest lies in adapting diffusion models for image-based industrial quality assessment, where statistics derived from micrographs are the primary signals.

Early diffusion-based studies largely focused on unconditional synthesis and reconstruction. Hoffman et al. [2] introduced GrainPaint, a multi-scale three-dimensional diffusion framework trained on SPPARKS grain-growth simulations, enabling large-scale volumetric polycrystalline synthesis via coarse-to-fine refinement. Despite strong scalability and morphological fidelity, validation was limited to simulated data, and the approach lacked conditioning mechanisms. Subsequent work expanded unconditional diffusion modeling to broader reconstruction tasks. Fernandez-Zelaia et al. [3] demonstrated that DDPMs can generate statistically consistent single-phase polycrystals and perform masked inpainting without handcrafted descriptors. Lee and Yun [4] and Lyu et al. [5] further showed improved reconstruction stability and agreement with correlation-based statistics across polycrystalline, composite, and random material systems, though explicit physical constraints and conditioning remained limited. Collectively, these studies established diffusion models as expressive generators for microstructure imagery, while highlighting gaps in physically informed validation.

Later work introduced weak forms of conditioning by linking diffusion models to processing parameters or predictive tasks. Azqadan et al. [6] applied DDPMs to predict microstructural images of cast–forged magnesium alloys under unseen processing conditions, capturing process–structure relationships with low error. However, this formulation remained image-centric, without explicit modeling of grain topology or interpretable grain-scale attributes. A more substantial shift occurred with the emergence of inverse and property-guided diffusion frameworks. Vlassis and Sun [7] coupled diffusion sampling with surrogate models of mechanical response, enabling property-driven microstructure generation. While conceptually significant, performance depended strongly on surrogate accuracy and training data coverage.

Conditional and multiphase diffusion models further expanded controllability. Baishnab et al. [8], Aouf et al. [9], and Nguyen et al. [10] introduced conditional and phase-fraction-aware diffusion models for multiphase and geological microstructures, improving compositional validity and phase consistency. In parallel, process-conditioned diffusion models sought to explicitly encode process–structure–property relationships. Ikeda et al. [11] and Choi et al. [12] demonstrated diffusion-based prediction of dendritic and grain morphologies from processing variables and target properties, though generalization remained sensitive to dataset diversity. Hybrid and physics-informed extensions followed, incorporating mechanical targets and physical priors into diffusion pipelines [13–15], while graph-based diffusion approaches explicitly modeled grain connectivity [16]. These methods advanced topological fidelity but largely overlooked appearance-level attributes critical for industrial inspection.

A complementary line of work treated microstructures directly as images or voxelized intensity fields, aligning diffusion models with experimental micrographs. Jung et al. [17] proposed multi-fidelity latent diffusion for three-dimensional microstructures, while Phan et al. [18] reconstructed full volumetric microstructures from single two-dimensional slices. Extensions to interactive, temporal, and physics-aware settings further broadened diffusion's scope, including command-based generation [19], temporal evolution modeling [20], physics-informed diffusion [21], and scalable latent diffusion for porous media [22]. These advances parallel diffusion successes in medical imaging and other scientific domains, where DDPMs outperform GANs under limited data [23–26], and are further reinforced by recent conditional and inverse-design diffusion frameworks for microstructure reconstruction and property-driven synthesis [27–30], and where property-conditioned diffusion is now standard in chemistry, biology, fluid mechanics, signal processing, and imaging [31–39], grounded in classifier and classifier-free guidance mechanisms [40].

Despite this progress, diffusion-based microstructure models remain weakly connected to industrial inspection and quantitative quality assessment. Prior work has prioritized geometric realism, phase distribution, or property-driven inverse design, with limited attention to color uniformity, coating coverage, and intensity variation—that dominate real-world manufacturing quality control. Moreover, diffusion models have rarely been leveraged to support image-derived

continuous regression tasks, nor systematically explored as data augmentation tools for improving regression robustness under realistic imaging constraints. By explicitly targeting RGB-based coating uniformity and color consistency, this work addresses a gap not covered by existing diffusion-based microstructure literature.

To summarize, the main contributions of this work to the state of the art are:

- Transformation of a Stable Diffusion architecture text preconditioner to one that takes 3D or 6D statistical parameters as prior conditioners to demand specific quality adherence of new and diverse generated image samples.
- Usage of generated samples to train vision based CNN models for online and accurate prediction of quality of manufactured Abrasive Grains with potential for feedback-based modulation of running process control parameters.
- Methodology for extraction of square sized patches from original micrograph, and concatenation of either such patches or the micrographs to yield larger integral images that Diffusion models can aspire to generate.

3 Diffusion Model

3.1 Latent Diffusion Formulation

Diffusion-based generative models provide a robust framework for structured image synthesis by leveraging stable, likelihood-based training and iterative denoising to capture complex, multimodal data distributions without adversarial instability, enabling high-fidelity and diverse generation. However, classical DDPMs[41] operate directly in pixel space and become computationally impractical at high resolutions due to their dimensionality and memory demands. Latent diffusion models[42] overcome this limitation by performing diffusion in a compact, learned latent space that preserves semantic and structural information, substantially improving efficiency while retaining expressive power. Building on this principle, the proposed framework adopts the Stable Diffusion v1.5 architecture[42], which enables computationally tractable yet semantically faithful modeling of high-resolution micrograph images.

Let

$$x \in R^{600\times900\times3} \tag{1}$$

denote an RGB integral image of spatial size 600×900. A pretrained Variational Autoencoder (VAE) encoder E_{VAE} maps the image x to a compact latent representation z_0:

$$z_0 = E_{\mathrm{VAE}}(x), \qquad z_0 \in R^{4\times75\times112}. \tag{2}$$

The latent tensor z_0 is a feature map with four learned channels over a 75×112 spatial grid, corresponding to an effective downsampling factor of 8 and capturing the high-level structure of the input image. The VAE encoder is directly adopted from the pretrained Stable Diffusion v1.5 model and kept frozen

throughout diffusion training to ensure stability of the latent manifold and to prevent distortion of the pretrained representation space.

The corresponding VAE decoder,

$$D_{\mathrm{VAE}} : R^{4\times 75\times 112} \to R^{600\times 900\times 3}, \tag{3}$$

is likewise inherited from Stable Diffusion v1.5 and used only during inference to reconstruct images from denoised latents, remaining frozen to preserve reconstruction consistency.

3.2 Latent-Space Forward Noising and UNet-Based Denoising

The forward diffusion process applies a Markovian Gaussian noising scheme directly in latent space, progressively corrupting the clean latent z_0 into an approximately isotropic Gaussian through a fixed variance schedule, with noisy latents constructed in closed form at uniformly sampled timesteps during training. The reverse process is parameterized by a UNet that learns to predict the injected noise given the noisy latent, diffusion timestep, and conditioning signal. A multiscale encoder–decoder architecture with skip connections preserves spatial structure, while sinusoidal timestep embeddings enable noise-aware denoising across diffusion stages. The UNet architecture is identical to Stable Diffusion v1.5, enabling reuse of pretrained weights and restricting modifications to conditioning alone.

3.3 Conditioning via RGB Embeddings

In standard Stable Diffusion, conditioning is performed using text embeddings produced by a pretrained language–image model (e.g., CLIP). In contrast, the proposed framework entirely removes text-based conditioning and replaces it with a conditioning mechanism based on measured color attributes.

Let

$$c_{\mathrm{RGB}} = (R_m, G_m, B_m) \in R^3 \tag{4}$$

denote the calibrated mean RGB values obtained from physical color measurements(refer Sec 1). These values are mapped into the conditioning embedding space via a lightweight Condition MLP:

$$f_\phi : R^3 \to R^{512} \to R^{768}. \tag{5}$$

The intermediate 512-dimensional hidden layer, combined with nonlinear activation functions, enables the network to learn nonlinear interactions between color channels. The final 768-dimensional output is chosen to exactly match the dimensionality of CLIP text embeddings, thereby ensuring architectural compatibility with the U Net's cross-attention layers without structural modification.

3.4 Token Replication and Cross-Attention Conditioning

To conform to the fixed-length token interface of text-conditioned diffusion models, the condition embedding c_{RGB} is replicated across 77 tokens:

$$c = \mathbf{1}_{77} \otimes f_\phi(c_{\mathrm{RGB}}), \qquad c \in R^{77\times 768}. \tag{6}$$

This replication enforces a global conditioning constraint, ensuring that the color specification uniformly influences all spatial locations. Unlike text prompts where different tokens encode different semantic entities, this design deliberately removes token-wise variation.

The conditioning tensor c is injected into every cross-attention layer of the UNet, where latent feature maps act as Queries (Q) and the replicated embeddings serve as Keys (K) and Values (V). Conditioning at multiple resolutions allows the RGB signal to modulate both coarse structural organization and fine micro structural textures throughout the denoising process. Q (Query) represents what information is being sought, K (Key) represents what information is available to match against the query, and V (Value) contains the actual information that is weighted and aggregated based on this match.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right) V \tag{7}$$

3.5 Reverse Diffusion Sampling and Training Objective

At inference, sampling begins from a Gaussian noise latent and iteratively applies the learned reverse diffusion process, with the UNet denoising the latent at each timestep conditioned on the RGB embedding. After the final denoising step, the clean latent is decoded by the frozen VAE decoder to generate a synthetic 600×900 integral image. Training optimizes the standard diffusion noise prediction objective, where the UNet learns to predict the injected Gaussian noise at randomly sampled timesteps. Only the UNet and conditioning MLP are trained, while the VAE encoder and decoder remain frozen to ensure latent stability and prevent catastrophic forgetting.

4 Diffusion based approach for Generation of new and diverse images for training

We describe here how a very small set of manually extracted integral images (600 x 900 pixels) are used to generate a sufficiently large and accurate set of new, diverse, and statistically traceable images through (a) augmentation and (b) our Diffusion model. The large image sets are used to train alternative ResNet50 architectures that take these images as inputs and generate 6-element vectors (RGB and corresponding SDs) as outputs. The trained ResNet50 models are first validated and ranked on a different set of manually extracted integral images,

and the best among these models implemented for online automated quality assessment of abrasive grains as explained in Sec. 1.

The first obvious question is - how accurate is the RESNET50 trained on the initial small set of manually generated data itself. We are not training the RESNET50 ab-initio but implementing transfer learning on an initial architecture trained on ImageNet, by retraining only the last layer with new data. We used 24 initial manually extracted samples for the retraining, and then validated it on 6 other manually extracted images. After convergence, the training MAPE values were 5.50%, 7.86% and 6.59% for R, G and B, 5.09%, 3.90% and 5.64% for the corresponding SDs, while the validation MAPE values were 99.71%, 99.61% and 99.55% for R, G and B, and 77.64%, 73.21% and 81.62% for the corresponding SDs. Obviously, the validation MAPEs are too poor for generalization, and that makes the entire exercise of synthetically generating many more data samples as imperative.

Our starting basis is a set of 30 manually extracted integral images. As discussed in Sec. 1, these are obtained by concatenating 6 parent images each extracted as micrographs from grain samples. These original micrographs are of size 300 x 300 pixels, an example of which may be seen in fig. 1. Each micrograph has a measured (by NCS Colourpin) set of RGB values, and on concatenation into an integral image, the latter will have an average RGB set and a SD set, forming a vector of 6 elements. Our first action is to extract "patches" of size 128 x 128 from within each of the 6 component micrographs, randomly from different locations but keeping orientation unchanged (as grain alignments may not be isotropic). 16 such patches are extracted from each micrograph, resulting in 96 such patches from the 6 parents of an integral image.

An integral image which is created by concatenating 2 x 3 parent micrographs, can now be alternatively synthesized by concatenating 5 x 8, i.e. 40 such patch images. Since that would result in a size of 640 x 1024, hence the last component patch on either axis is clipped to result in an exact size of 600 x 900. The 40 patches that contribute to an integral image are selected randomly out of 96 (only that near-equal distribution across the 6 parents is taken), and moreover, they are placed (i.e. order of concatenation) randomly to create the integral image. Hence, in principle a very large number of such integral images can be combinatorially synthesized from these patches. Moreover, these patches are originally extracted from 6 micrographs, hence the resultant average of any of the potential integral images will have the same RGB values (which also matches with that of the original micrograph-concatenated integral image) and the same internal variation, i.e. SD values. Implying the vector of 6 statistical parameters will be same across all potential new integral images and coincide with that of the original integral image.

Following the above procedure, we synthesize 4 (only) new integral images from each of the original 30 images, resulting in 150 integral images. Note that number of 6-element statistical vectors remain 30, and 5 integral images are aligned with one such vector. Next, by traditional augmentation process of modulation of brightness and contrast but avoiding rotation, we create 9 new integral

Fig. 3. Integral Image formed by concatenating 40 patches

images from each one of a set of 5. Since we started with a total of 30 such sets of 5, by this two-step augmentation process we create 1500 such integral images. We use this for training and validation of the first ResNet50, which we call `ResNet50_6D_A`. We are able to attain validation MAPE (Mean Absolute Percentage Error, with $\epsilon = 0.0001$) of 0.543% for Red, 0.536% for Green, 0.644% for Blue. The validation error is computed on 20% of the entire dataset, with the remaining 80% used for training.

Next, we use our customized Diffusion model as explained in Sec. 3 to generate 1500 new images, each pre-conditioned with one of 30 conditioning vectors of size 6 (we actually use only the RGB components of this vector for pre-conditioning as SDs are derived parameters). It is to be noted here that our Diffusion models are an adapted and customized version of the original Stable Diffusion v1.5 architecture, where conditioning with text-embeddings is replaced with a 3-element numerical vector representing RGB, without changing the fixed-length text tokens interface. The manner of attainment of this transformation of pre-conditioning, keeping Diffusion-architecture interfaces intact, is described in above Secs. 3.3 and 3.4. Fig. 4 shows two typical new synthesized Integral 600 x 900 images generated using this mechanism.

Combining the augmented dataset with the generated dataset yields a net training dataset of 1500 x 2 images, and we train another ResNet50 model on this combined dataset, which we call `ResNet50_6D_B`. The corresponding validation MAPEs for this new ResNet50 are 1.60% for R, 1.84% for G, 2.13% for B.

Finally, we use the 1500 generated new images from our Diffusion model, alone (i.e. not using the augmented-by-traditional-means dataset at all), to train the last `ResNet50_6D_C`. Note this is trained from Diffusion-synthesized images alone. The resultant validation errors are 2.61% for R, 3.68% for G, 3.80% for B.

For a rigorous comparison of the three ResNet50 models we have trained, we cannot take the corresponding MAPEs from the respective validation data alone,

Fig. 4. Newly generated images using our diffusion mechanism.

as these would not be statistically independent of their corresponding training data. Instead, we now use 89 alternative manually extracted integral images (see last line of first para of this section), which we have not engaged for training in any form, as an independent and authentic comparison set.

Table 1. Comparison of all 6 MAPEs from the 3 ResNet50s, on the alternative manually extracted integral images. All MAPEs are in percentages.

ResNet50 Version	MAPE_R	MAPE_G	MAPE_B	MAPE_SDR	MAPE_SDG	MAPE_SDB
ResNet50_6D_A	12.04	15.93	15.73	42.16	37.89	45.58
ResNet50_6D_B	9.66	20.63	22.41	37.20	35.74	42.49
ResNet50_6D_C	8.76	18.22	21.02	28.56	31.64	38.24

From Table 1, it is seen that the Computer Vision model, i.e. ResNet50 version, is performing best when trained with the data created from our Diffusion model alone. The augmented model results show better MAPE for G and B, and worse for all the others. It may be noted that the 89 manual integral images extracted as test data, are taken from a time window separation of two months from the production of the original 30 images and as such reflect some drift in production, but still the error levels can be considered as good-to-go for online in-process evaluation as described in Sec. 1.

Considering that SD is a derived parameter, the sample-to-sample variation as manifested in MAPE values may be considered large. Hence, in Table 2 we have noted the SDs (the SDs themselves, and not their MAPE values) from the 3 `ResNet50`s for the 89 new manual sets versus the SDs of the actual sets (appearing in first row). In each table cell we have entered (SD_max, SD_min, SD_median; across all 89 samples). It can be seen that the actual manually extracted data has substantial outliers, which is damped both for our augmented as well as Diffusion data sets, and it is possible that these outliers act to spoil the actual MAPE values, on any of the trained RESNET50s.

Table 2. Statistical comparison of standard deviation components (SD_R, SD_G, SD_B) across different ResNet models.

Model	SD_R	SD_G	SD_B
Original ResNet6D	6.609, 0.838, 3.007	5.152, 0.987, 2.330	4.822, 0.604, 2.237
ResNet50_6D_A	3.131, 1.713, 2.068	2.126, 1.370, 1.600	1.900, 1.203, 1.474
ResNet50_6D_B	4.700, 1.392, 1.855	2.066, 1.206, 1.639	2.269, 0.975, 1.342
ResNet50_6D_C	3.886, 1.253, 2.529	2.444, 1.044, 1.815	2.161, 0.283, 1.602

It may be mentioned here that, apart from the main line of investigations, we also experimented with different forms of new images to be generated. For example, instead of generating integral images of size 600 x 900, we tried generating patches of size 128 x 128 through the Diffusion model, and then concatenating them as in the image augmentation trials into Integral images. However, this did not lead to any improvements in image quality, so we preferred the above-described mainstream approach in comparison to this alternative.

5 Conclusion and Future Scope

This work demonstrates a practical and scalable approach for online quality assessment of abrasive grains using vision-based diffusion models.Starting with a small set of manually collected micrographs, statistically consistent and diverse images are generated through structured augmentation and diffusion-based generation. Replacing text-based conditions with measurable RGB parameters allows the model to generate images that directly reflect quality requirements. These generated samples effectively address data scarcity and imbalance, and enable reliable training of ResNet50 models for automated, near real-time monitoring of coating consistency on the shop floor.

The trained models show good accuracy in predicting color statistics, reducing dependence on extensive manual inspection and supporting the potential of feedback-driven process control during production. Although this paper focuses on red coated grains, the methodology has been extended to other colors, specifically purple, black and white coated grains, where accuracies attained are of similar levels. Work is under progress to further enhance accuracy levels of generated images using metrics that are alternative to the downstream Resnet50 based assessment.

As future work, this framework can be extended to other coating types and visual characteristics, including other colored grains, with suitable adaptation of statistical conditioning. Further validation across different abrasive materials, lighting conditions, and production lines will help establish the broader applicability of diffusion-based methods for manufacturing quality assessment and near-real-time feedback control.

References

1. Mittal, A., et al.: Diacetylene-based colorimetric sensors for γ-radiation detection in blood irradiation.
2. Hoffman, R., et al.: GrainPaint: A multi-scale diffusion-based generative model for microstructure reconstruction of large-scale objects. Acta Materialia (2025), arXiv preprint.
3. Fernandez-Zelaia, A., et al.: Digital polycrystalline microstructure generation using diffusion probabilistic models. Materialia **33** (2024).
4. Lyu, Y., et al.: Microstructure reconstruction of 2D/3D random materials using denoising diffusion probabilistic models. Scientific Reports (2024).
5. Lee, S., Yun, H.: Microstructure reconstruction using diffusion-based generative models. Materials and Manufacturing Processes (2024), arXiv preprint.
6. Azqadan, A., et al.: Predictive microstructure image generation using denoising diffusion probabilistic models. Acta Materialia **261** (2023).
7. Vlassis, N., Sun, W.: Denoising diffusion algorithm for inverse design of microstructures with fine-tuned nonlinear material properties. Computer Methods in Applied Mechanics and Engineering (2023).
8. Baishnab, R., et al.: 3D multiphase heterogeneous microstructure generation using conditional latent diffusion models. Digital Discovery (2025).
9. Aouf, M., et al.: 3D clay microstructure synthesis using denoising diffusion probabilistic models. Journal of Petroleum Science and Engineering (2025).
10. Nguyen, T., et al.: Phase-fraction guided denoising diffusion model for augmenting multiphase steel microstructure segmentation via micrograph image–mask pair synthesis (PF-DiffSeg). arXiv preprint (2025).
11. Ikeda, Y., et al.: Development of a conditional diffusion model to predict process parameters and microstructures of dendrite crystals of matrix resin based on mechanical properties. arXiv preprint (2024–2025).
12. Choi, J., et al.: Conditional diffusion modeling of grain morphology evolution in Li–Mn-rich cathode precursor synthesis. arXiv preprint (2025).
13. Mirzaee, M., et al.: Inverse design of microstructures using conditional diffusion models. Materials & Design (2025).

14. Zhang, H., et al.: HyperDiff: An inverse design framework for hyperelastic porous microstructures. Preprints (2025).
15. Author(s) unknown: Microstructure inverse design with generative hybrid models. Proceedings of the ACM (2025).
16. European Space Agency (ESA): Metallic microstructure design via conditional diffusion on graph representations. ESA Technical Reports (2023–2025).
17. Jung, S., et al.: Multi-fidelity learning-based latent diffusion model for three-dimensional microstructures. Materials & Design (2025).
18. Phan, T., et al.: Generating 3D images of material microstructures from a single 2D image using denoising diffusion probabilistic models. npj Computational Materials (2024).
19. Kartashov, D., Vlassis, N.: A large language model and denoising diffusion framework for targeted design of microstructures with commands in natural language. arXiv preprint (2024).
20. Zhang, Y., et al.: Diffusion-based text-to-image generative model for microstructure evolution. SSRN Electronic Journal (2024).
21. Bastek, J., et al.: Physics-informed diffusion models. arXiv preprint (2024).
22. Tahmasebi, P., et al.: Latent diffusion modeling of porous media informed by spatial statistics. Physical Review E (2025).
23. Khader, F., et al.: Three-dimensional medical image synthesis with denoising diffusion probabilistic models. Nature Machine Intelligence (2023).
24. Huijben, I., et al.: Denoising diffusion probabilistic models for addressing data limitations in chest X-ray classification. Artificial Intelligence in Medicine (2024).
25. Deshpande, A., et al.: Assessing the capacity of a denoising diffusion probabilistic model for medical imaging. Frontiers in Artificial Intelligence (2024).
26. Streiffer, C., et al.: Denoising diffusion models for improved performance in medical imaging tasks. Medical Image Analysis (2024).
27. Düreth, S., et al.: Conditional diffusion-based microstructure reconstruction. Computational Materials Science (2023).
28. Ikeda, Y., et al.: Conditional diffusion modeling of dendritic microstructures conditioned on elastic constants. Nature Computational Science (2024–2025).
29. Mirzaee, M., et al.: Inverse design of 3D plate microstructures using conditional diffusion models. Materials & Design (2025).
30. Zhuo, Y., et al.: An inverse design method of 3D plate microstructures using conditional diffusion models. Materials & Design (2025).
31. Vost, M., et al.: Improving structural plausibility in diffusion-based 3D molecule generation via property-conditioned training with distorted molecules. Digital Discovery (2025).
32. Villegas-Morcillo, A., et al.: Guiding diffusion models for antibody sequence and structure design. Proceedings of the Machine Learning Research (2023).
33. Yonekura, K., et al.: Preliminary study of airfoil design synthesis using a conditional diffusion model. Aerospace (2024).
34. Shu, Y., et al.: A physics-informed diffusion model for high-fidelity flow field reconstruction. Proceedings of the ACM (2023).
35. Sardar, M., et al.: Spectrally decomposed diffusion models for generative turbulence recovery. arXiv preprint (2023).
36. Li, Z., et al.: Multi-scale reconstruction of turbulent rotating flows with generative diffusion models. Fluids (2024).
37. Lu, Y., et al.: Conditional diffusion probabilistic model for speech enhancement. In: Proc. IEEE ICASSP, pp. 1–5 (2022).

38. Zhu, H., et al.: Cycle-conditional diffusion model for noise correction of diffusion-weighted imaging (Cycle-CDM). Medical Physics (2025).
39. Wang, X., et al.: Multi-decoder conditional denoising diffusion probabilistic model for dual-energy CT synthesis from cone-beam CT (2025).
40. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS) (2021).
41. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS) (2020).
42. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022).