

ENHANCING DESIGN GUIDELINES FOR METAL POWDER BED FUSION: ANALYZING GEOMETRIC FEATURES TO IMPROVE PART QUALITY

Jannatul Bushra¹, Hannah D. Budinoff¹, Pablo Luna Falcon¹, Marat Latypov¹

¹The University of Arizona, Tucson, AZ

ABSTRACT

Additive manufacturing (AM) part quality relies on many factors, including part geometry that impacts both the manufacturability and resulting dimensional accuracy of the part. To improve the dimensional accuracy of AM parts, data-driven approaches can be utilized to explore the effect of different process parameters on both simple and complex geometries. However, to provide general design guidelines, it is necessary to develop models and tools that accurately predict geometry-driven distortion across a broad range of geometries, while also being user-interpretable. Identifying and analyzing common part features that contribute to geometrical deviations and using them to design better parts could improve AM part quality. In this paper, a Gaussian process regression surrogate model was trained using 21 geometric features (selected from a set of 92 shape descriptors) from 324 different axisymmetric parts to predict maximum part distortion and identify the features that impact part distortion the most. Validated high-fidelity finite element analysis simulations were used to determine the maximum distortion corresponding to each part. Our results show the surrogate model approach can accurately predict part distortion, with a predictive error of approximately 0.07 mm for the testing set. The findings of this study can have implications for the exploration of new part designs by adjusting these identified features or incorporating them as design rules in AM product designs.

Keywords: AM part geometry, surrogate model, machine learning, geometric features, part distortion

1. INTRODUCTION

1.1 Prediction of AM outcomes

Additive manufacturing (AM) processes enable the manufacture of parts with complex geometry. In metal powder bed fusion processes (PBF), such as selective laser melting (SLM), the layer-by-layer fabrication and laser scanning process

results in a complex thermal history, with different regions of the part heating and cooling over different periods of time. The thermal history determines local microstructure formation and can lead to residual stresses that may result in poor part quality or premature failure [1].

Residual stresses can be predicted for SLM-fabricated parts using a variety of physics-based models. Commercial software, including ANSYS Additive, Netfabb Simulation, and Additive Works Amphyon, can predict residual stresses with reasonable accuracy [2,3]. Custom codes have also been developed (e.g., [4]). Most modeling approaches are computationally expensive. Some strategies for reducing the computation time include adaptive meshing and making simplifying assumptions regarding how thermal loads are applied as in the inherent strain method [3]. An alternative approach to decreasing the computation time is to develop surrogate models. Surrogate models have been developed to predict AM process outcomes with little decrease in accuracy [5].

1.2 Surrogate models

Data-driven surrogate models are used to approximate system behavior when simulations or experiments are expensive. They help reduce computational cost and time for analysis, exploring the relationship between the input and output of the system. Surrogate models are widely used in additive manufacturing to analyze the complex relationships between process parameters (i.e., print speed, layer thickness) and the resulting quality (i.e., melt pool geometry, surface roughness, dimensional accuracy, and mechanical properties) of manufactured parts. Researchers have utilized these models to explore relationships between parameters and quality, predict performance, and to optimize process parameters [6]. These models are built using experimental data [7] or simulation results [8–10]. Researchers also improve the prediction accuracy of physics based surrogate models by calibrating them using experimental data [11]. Considering different geometries as inputs and geometrical deviation/distortion as output,

researchers have developed surrogate models for AM applications based on machine learning (ML) techniques such as Neural Networks [6], Recurrent Neural Networks [12], Graph Neural Networks [9], Convolutional Neural Networks (CNN) [8], and Gaussian Processes [11]. Some surrogate models allow predictions of residual stresses for a rich variety of geometries [8,9], whereas other studies focus on simpler geometries (i.e., cubes, cylinders, thin walls) but additionally incorporate process parameters as inputs for prediction of thermal histories and their fields [11,12]. Surrogate models have the potential for good predictive accuracy for distortion: for a simple wall and strut geometry, studies report errors of 6.6% for geometries outside of the training set for models trained on simulation data [13].

Beyond surrogate models, many examples of ML techniques to predict/optimize geometrical accuracy have been reported to date. Many studies only consider process parameters as ML model inputs to predict volumetric error [6], dimensional accuracy [14,15], distortion [16], and shrinkage ratio [17]. Most of these studies train their models using experimental data obtained by printing a single geometric part [14–17]. Some researchers considered 3D part geometry/model as input of their ML model, using neural networks such as Convolutional Neural Networks, Graph Neural Networks, and Bayesian Neural Networks to perform deformation compensation [18], to predict residual stress field [8,9], or to predict geometric errors [19,20].

Many of these adopted ML techniques require a significant amount of training data, and lack interpretability, which means their ability to provide a clear understanding of the common geometric features that are responsible for distortions are limited. Designers may be able to input a computer-aided design (CAD) model and receive a numerical prediction while providing little guidance on which geometrical features lead to large distortions or high residual stresses. The lack of such insights hinders design improvements. Decision trees and other rule-based ML models provide a direct means of interpretation, although this interpretability can come at the cost of reduced prediction accuracy, and there are limits on the number of features [21]. Other options for interpreting ML models include evaluating model sensitivity to input perturbations or analyzing local or global surrogate approximations of the ML model [21]. In prior work, one study employed a random forest method and identified significance values of features to predict dimensional inaccuracies in complex 3D shapes, but the study focused on small deviations of fused filament fabrication parts [22].

1.3 AM design guidance

Existing design guidance for SLM is typically derived through trial-and-error experience. Literature has several examples of guidelines for geometric features associated with high stresses. The GAPS worksheet and sources cited indicated that parts with large solid volumes were associated with high warpage, suggested using fillets to decrease stresses at geometry changes, and recommended that the size of unsupported downward facing surfaces be minimized [23–25]. Adam and Zimmer provided thresholds for the length of overhanging features and the maximum allowable amount of material

accumulation for SLM processes [26]. Other features associated with high distortion are changes in the centroid of layer cross sections [3]. The standard ISO/ASTM 52910, which provides design guidelines and recommendations, advises that for PBF, abrupt changes in the thickness of printed geometry be minimized [27]. While helpful, this ad hoc guidance has not been rigorously validated to the authors’ knowledge and the relative severity of the various geometric features is not given.

Data-driven design exploration methods can enable timely and efficient product realization cycles [28]. Jiang et al. suggest an ML-integrated design for AM approach that uses a deep neural network as a surrogate model to design products with desired mechanical property distributions by tailoring part geometry [29]. Several studies have utilized ML clustering algorithms to create design feature recommendation systems for novice designers [30] and design modification recommendation systems based on the assessment of AM feasibility [31].

In this paper, we seek to identify geometric metrics that are correlated to part distortion in SLM, to determine the relative sensitivity of part distortion to the variables. We also develop design guidance based on systematic part exploration, rather than anecdotal trial-and-error experimentation. The contribution of this work is to evaluate prior hypotheses about features leading to distortion. Given the infinite design space of part geometries and features, we focused our study on axisymmetric parts with fixed process parameters.

The paper is organized as follows: Section 2 presents the methodology, Section 3 outlines the results of our study, Section 4 contains the discussion of our findings, and Section 5 is the conclusion of the paper.

2. MATERIALS AND METHODS

In this study, we focused on axisymmetric parts for simplicity. To this end, we generated CAD models for axisymmetric parts and created 2D images to represent each part, calculated shape descriptor features for each shape based on the image, used a commercial PBF process-simulation package to generate labels for each image, and trained our model (different stages are depicted in Fig. 1). In future, our approach can be expanded to extract similar shape descriptor features from multiple cross-section images of complex geometries other than axisymmetric parts.

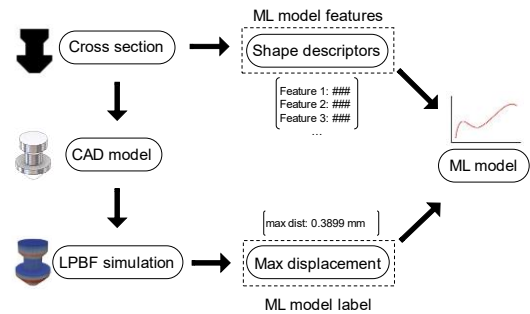


FIGURE 1: OVERVIEW OF STAGES INVOLVED IN OUR METHOD

2.1 CAD models

The 324 parts investigated in this study have 54 unique cross-section designs (examples shown in Fig. 2). For every basic cross-section design, we created variations with two different height-to-base ratios. For each of those 108 cross-sections variations, we made parts with three different sizes. The overall part sizes range from 20mm × 20mm × 25mm to 52mm × 52mm × 50mm. To make parts feasible and self-supporting, we considered the design guidelines, i.e., minimum wall thickness, overhang angles $\leq 45^\circ$ for AM parts. To be consistent with the build direction (z-axis) all sketches were done on the bottom view where positive z-axis is pointing up, and positive x-axis is pointing to the right.

Each variation contains a unique set of features within the guideline limitations. We varied the position and dimensions of overhang angles ensuring that a distinct variation of outer and inner features was present in each design. We verified that there were no enclosed geometries generated when revolving the cross-section by incorporating a hole with at least 1-mm radius to allow the removal of powder. Some variations were the result of either rotating existing cross sections by 180-degrees, changing the axis of rotation, or a combination of both. Other variations resulted from modifying geometries that would need support or cause errors (e.g., overlaps) when revolving.

Each cross-section image was generated as a drawing in CAD software, with the dimensions of the sheet set to 4 inches × 4 inches. We created a fixed point at the center of the drawings to use as reference for the sketch. Each cross-section was then copied from the 3D model sketches, centered, rotated 180-degrees, and mirrored. We then added a solid hatch. All files were saved as PDF files and were exported as PNG images.

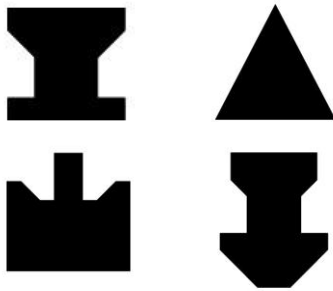


FIGURE 2: EXAMPLES OF PART CROSS-SECTIONS

2.2 LPBF process simulations

We used ANSYS Additive Print thermal strain (TS) simulations to generate the data labels for our study. We ran process simulations corresponding to the 324 axisymmetric parts

and obtained their on-plate part distortions at the end of the build. From those simulation outputs, we obtained maximum distortion (magnitude) for each part.

In ANSYS Additive Print, TS simulation considers part geometry, material configuration, and machine parameters. TS simulations utilize a rotating stripe scan pattern established by machine parameters such as layer thickness, starting layer angle, and layer rotation angle. Other machine parameters considered in the TS simulation are hatch spacing, baseplate temperature, laser beam diameter, slicing stripe width, scan speed, and laser power. TS simulation utilizes thermal and mechanics solvers to generate the output. For full model details, the reader is referred to the ANSYS Additive User Guide [32], but a brief overview is provided here. The thermal solver models the periodic heating and quick cooling observed in the LPBF process. The thermal solver generates a spatiotemporal temperature field by simulating scan tracks and thermal conduction through the component and then determines the strain magnitude for each location. The strain magnitude determined for each location by the thermal solver is transferred to the mechanics solver. The mesh size used by the mechanics solver is determined from user-specified parameters (i.e., voxel size and voxel sample rate). The mechanics solver assumes the entire powder layer is melted at once. Thus, it simulates thermal shrinkage during solidification by combining multiple micron-scale physical layers into one. This thermal shrinkage causes part deformation and drives the mechanics analysis.

As this study aims to analyze the relationship between part geometry and maximum on-plate part distortion, we used constant machine parameters in all the simulations. To conduct TS simulation, it is necessary to calibrate both the isotropic (Strain Scaling Factor) and anisotropic (Anisotropic Strain Coefficients) strain scaling coefficients. Calibration of SSF compensates for machine-related variations and enhances simulation accuracy. Anisotropic strain coefficients represent anisotropic strain behavior, enabling the computation of anisotropic strains along longitudinal, transverse, and build directions. For this study, we used Ansys pre-defined Ti-6Al-4V material and used previously validated experimental calibration factors (SSF, ASCs) for the EOS M290 machine and Ti-6Al-4V [33]. We also used the same voxel size (0.25 mm), material behavior (elastic-plastic) and process parameters set used for determining and validating the calibration factors [33]. We employed J2 (Von Mises) plasticity model for elastoplastic computations of generated stresses and associated deformations. In all our simulations, a voxel sample rate of 5 and a mesh resolution factor of 5 were utilized.

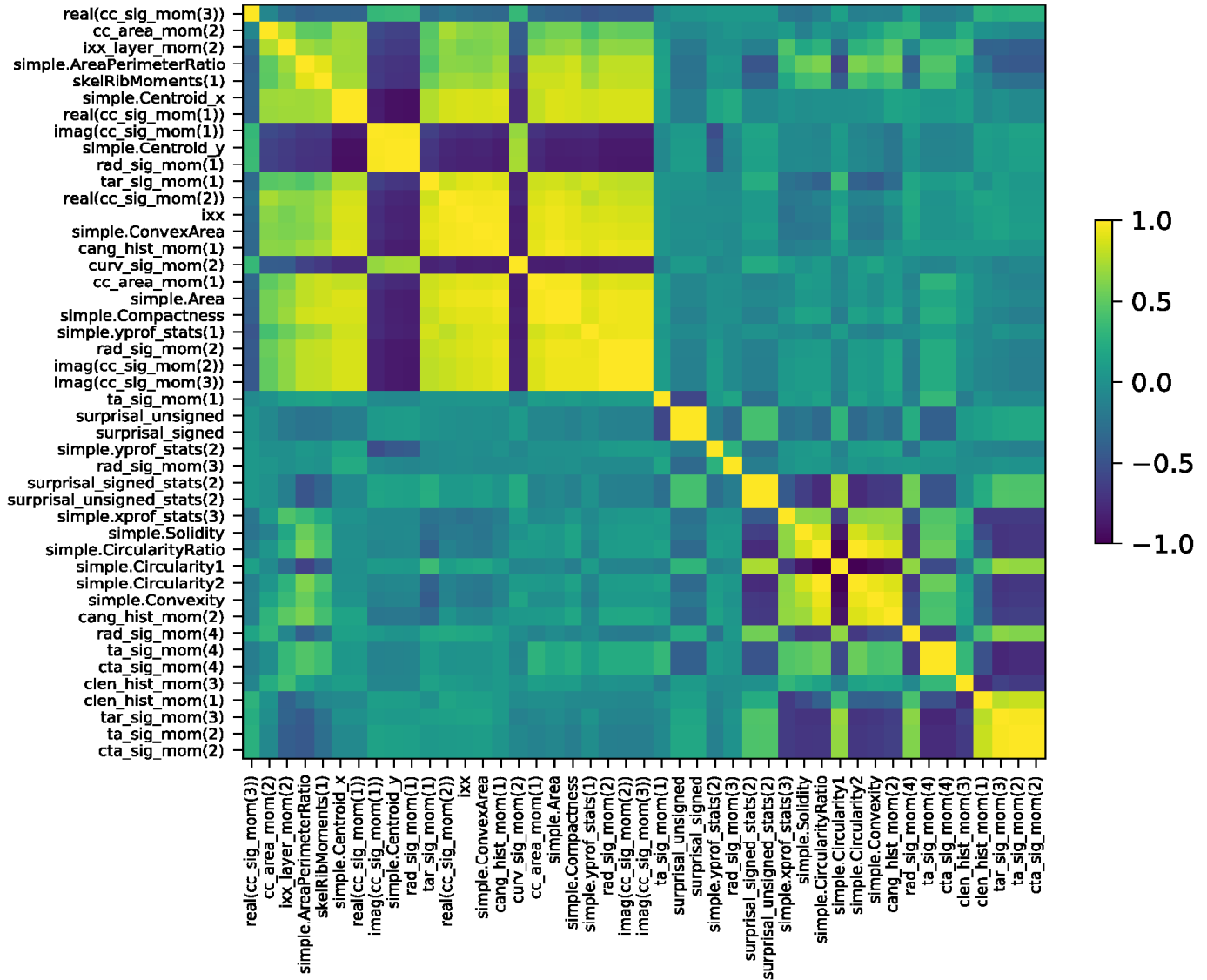


FIGURE 3: CORRELATION COEFFICIENTS OF THE 45 PRELIMINARY FEATURES

2.3 Shape descriptors

Morgenstern et al. developed a model that uses shape features extracted from 2D shapes - ranging from basic descriptors like area and perimeter to more intricate descriptors like shape skeleton - to predict human shape similarity judgments between pairs of shapes [34]. In our study, we adopted the same shape descriptors including simple shape descriptors (i.e., area, perimeter, eccentricity), shape context (i.e., chord lengths and angles), shape signatures (i.e., radius signatures, curvature signatures), and shape skeleton descriptors (i.e., the total number of skeletal branches, mean skeletal depth). We used only descriptors which resulted in a single value per shape (e.g.,

mean of radius signature, rather than the radius signature itself) and eliminated any features that could not be computed for all of our shapes. In addition to these shape descriptors from Morgenstern et al., we added some custom descriptors based on ad hoc guidance for avoiding distortion (as summarized in the introduction). We created histograms of the second moment of area of each row of pixels in the input images (denoted *ixx_layer*) and the number of pixels in each row of pixels (denoted *cc_area*). Features derived from each of these histograms were the mean, standard deviation, skewness, and kurtosis. We also included a feature (denoted *ixx*) for the second moment of area for the entire image.

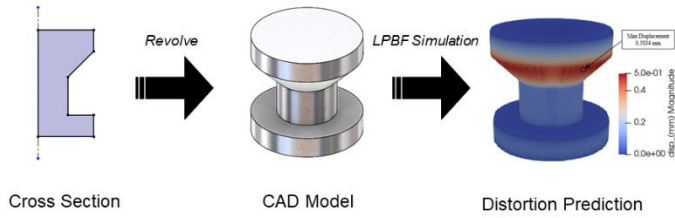


FIGURE 4: DATA GENERATION PROCESS TO GENERATE LABELS USED FOR TRAINING AND TESTING

2.4 Machine learning feature selection

To prevent overfitting, simplify resulting models, and eliminate irrelevant inputs, we conducted feature selection prior to training our model. Initially, we utilized a feature selection technique that evaluates the relationship of each feature (shape descriptors defined in Section 2.3) with the target variable (maximum part distortion) individually to select the most important features. Accordingly, among 92 total features, we kept only the top 45 features with the highest univariate scores (determined by the F-statistic and p-values). Then, to account for interactions or relationships between features, we generated a correlation map (Fig. 3) of those 45 features and excluded features that exhibited a large linear correlation (with Pearson correlation coefficients greater than 0.84 or less than -0.84). The 21 final features (shape descriptors) included are listed in Table 1 and Table 2.

TABLE 1: FEATURES INCLUDED IN FINAL MODEL AND THEIR DESCRIPTIONS

Feature name	Description
simple.Area	Area of cross section
simple.Solidity	Proportion of pixels in the convex hull that are also in the image
simple.Circularity1	$4\pi(\text{Area}/\text{Perimeter}^2)$
simple.Centroid_y	Y-coordinate of shape centroid in image
simple.AreaPerimeterRatio	Ratio of Area to perimeter
simple.xprof_stats(3)	Kurtosis for horizontal distribution of image pixels
simple.yprof_stats(2)	Skewness for vertical distribution of image pixels
surprisal_signed	Information along part contour, calculated at base
clen_hist_mom(1)	Mean for histogram of chord lengths
clen_hist_mom(3)	Skewness for histogram of chord lengths
rad_sig_mom(3)	Skewness for radius signature

rad_sig_mom(4)	Kurtosis for radius signature
curv_sig_mom(2)	Standard deviation for curvature signature
tar_sig_mom(1)	Mean for triangular area signature
tar_sig_mom(3)	Skewness for triangular area signature
ta_sig_mom(1)	Mean for tangent angle signature
cta_sig_mom(4)	Kurtosis for cumulative tangent angle signature
real(cc_sig_mom(3))	Skewness for distribution of x-values of image boundary
cc_area_mom(2)	Skewness for distribution of number of pixels in each row of image
ixx_layer_mom(2)	Standard deviation for histogram of second moment of area of each row of image
surprisal_signed_stats(2)	Standard deviation for surprisal (information along contour)

2.5 Machine learning pipeline

To train and test our ML model, we used the following pipeline: designing the cross-section and axis of the axisymmetric parts, creating 3D models from the cross-sections and axis information, running simulations with those 3D models (Fig. 4). By following these steps, we generated 324 data points. The necessary size of the dataset for ML tasks depends on how complex the problem and model are, and because we had a simple regression problem with a tabular dataset, we started with several hundred data points. We also reduced the number of features by about 75% through feature selection, thus increasing the ratio between dataset size and number of features. After this reduction, our dataset size follows a rule of thumb that recommends having ten times more samples than features [35].

Our goal was to use an ML regression model to predict the maximum on-plate part distortion for the part. The ML model's inputs are the shape descriptors extracted from the images of the cross-sections (Fig. 5).

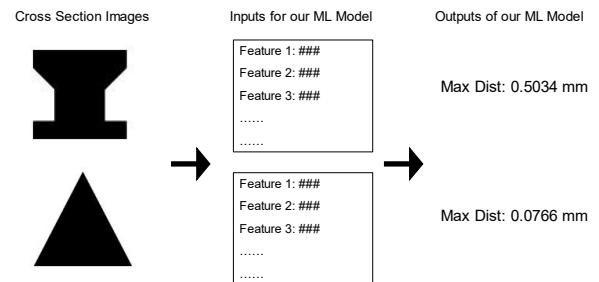


FIGURE 5: INPUTS AND OUTPUTS OF THE ML MODEL

TABLE 2: FEATURES INCLUDED IN FINAL MODEL AND HIGHLY CORRELATED FEATURES WITH THEM

Feature name	Highly correlated ($ r > 0.84$) features
simple.Area	simple.Compactness, simple.Centroid_x, simple.ConvexArea, simple.yprof_stats(1), cang_hist_mom(1), rad_sig_mom(2), real(cc_sig_mom(1)), real(cc_sig_mom(2)), imag(cc_sig_mom(2)), imag(cc_sig_mom(3)), ixx, cc_area_mom(1)
simple.Solidity	simple.Circularity2, simple.CircularityRatio
simple.Circularity1	simple.Circularity2, simple.Convexity, simple.CircularityRatio, cang_hist_mom(2)
simple.Centroid_y	simple.Centroid_x, rad_sig_mom(1), rad_sig_mom(2), real(cc_sig_mom(1)), imag(cc_sig_mom(1)), imag(cc_sig_mom(2)), imag(cc_sig_mom(3))
simple.AreaPerimeterRatio	simple.Compactness, skelRibMoments(1)
simple.xprof_stats(3)	N/A
simple.yprof_stats(2)	N/A
surprisal_signed	surprisal_unsigned
clen_hist_mom(1)	ta_sig_mom(2), cta_sig_mom(2)
clen_hist_mom(3)	N/A
rad_sig_mom(3)	N/A
rad_sig_mom(4)	N/A
curv_sig_mom(2)	N/A
tar_sig_mom(1)	simple.ConvexArea, simple.yprof_stats(1), cang_hist_mom(1)
tar_sig_mom(3)	ta_sig_mom(2), cta_sig_mom(2)
ta_sig_mom(1)	N/A
cta_sig_mom(4)	ta_sig_mom(4)
real(cc_sig_mom(3))	N/A
cc_area_mom(2)	N/A

ixx_layer_mom(2)	N/A
surprisal_signed_stats(2)	surprisal_unsigned_stats(2)

3. RESULTS AND DISCUSSION

3.1 Model training

Among 324 data points, we used 243 data points for training the model, and the remaining for testing the model. The 21 features (listed in Table 1) selected in the feature selection step along with their labels, were used to train a variety of regression models using Auto-Sklearn [36], limiting the ensemble size to one for simplicity. The considered regressors include adaboost, bayesian ARD (automatic relevance determination), decision tree, extra-trees, gaussian process, gradient boosting, k nearest neighbors, random forest, epsilon-support vector, and multi-layer perceptron [36]. The final selected model was a Gaussian process regressor. The Gaussian process regressor is a supervised regression method that models the relationship between input and output variables as a Gaussian Process [37]. A Gaussian process can be interpreted as defining a probability distribution over functions, and inferences are made within that function space [38]. This estimator considered hyperparameters, such as the kernel that determines the covariance function of the Gaussian process, the alpha value added to the diagonal of the kernel matrix during fitting, the number of times the optimizer is restarted to find the optimal kernel parameters, whether or not to normalize the target values, and the random state that determines random number generation used to initialize the centers [39]. After fitting to the training set, the parameters' values were set as follows: kernel was radial basis function (RBF) with length scale of 1 for each feature, alpha was 0.00217, the number of restarts of the optimizer was 10, the random state was 1, and target normalization was True which means target values were normalized by subtracting the mean and scaling to unit-variance.

3.2 Model performance

The performance (parity plot) of our trained model is shown in Fig. 6, with both training and testing data shown. The R^2 of the model, evaluated using the test samples, was 92.2%, and the root mean squared error (RMSE) was 0.068 mm. The parity plot demonstrates that while there are some predictions that exhibit noticeable deviations from this expected value, most predictions for both the training and testing sets agree well with the simulation-predicted maximum distortion values.

To identify the critical factors that impact max distortion prediction accuracy, we used the permutation importance technique, which is particularly helpful for non-linear predictors. The significance of each feature is determined by its importance value, which quantifies the degree to which the model's accuracy is affected by shuffling that feature randomly. We evaluated permutation importance using the testing dataset and generated Fig. 7 which displays the mean and standard deviation of feature importance scores over ten permutations. From Fig. 7, we can see the top five features include surprisal_signed, ixx_layer_mom(2), clen_hist_mom(1), tar_sig_mom(1),

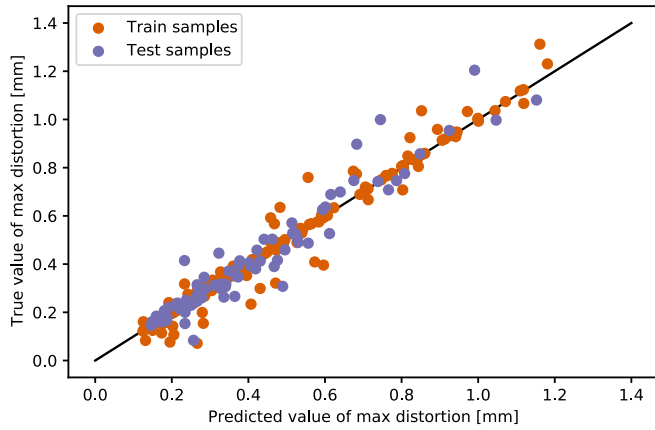


FIGURE 6: ANSYS ADDITIVE PRINT PREDICTED MAX DISTORTION VS ML PREDICTED MAX DISTORTION

cta_sig_mom(4). Here, considering the preliminary 45 features, surprisal_signed is highly correlated with surprisal_unsigned. ix_x_layer_mom(2) is not highly correlated ($|r| > 0.84$) with any other features from the set of the 45 features. clen_hist_mom(1) is highly correlated with ta_sig_mom(2) and cta_sig_mom(2). tar_sig_mom(1) is highly correlated with simple.ConvexArea, simple.yprof_stats(1), and cang_hist_mom(1). cta_sig_mom(4) is highly correlated with ta_sig_mom(4).

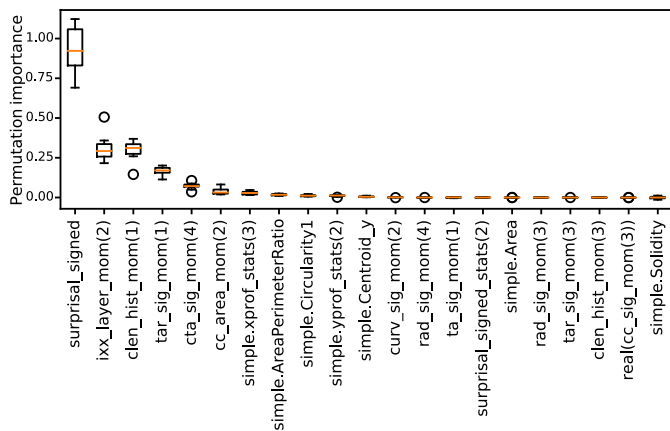


FIGURE 7: PERMUTATION IMPORTANCE SCORES OF GEOMETRY FEATURES USED IN FINAL MODEL

4. DISCUSSION

We found that for our relatively small dataset, the surrogate model performance was relatively accurate for both the test and training data. The trained Gaussian process regression provided good predictive performance, with error far less than the typical predicted distortion.

One advantage of our data-driven approach is computation time. Computation time is minimal for computing the shape descriptors of 2D images, with all 92 shape descriptors being computed in <1s per shape. In comparison, the physics-based simulations took between 1 to 8 hrs for smaller shapes and up to

90 hrs for larger parts. Designers could use trained models to generate quick design feedback inside CAD packages.

Using our trained model, we were able to identify important features. Surprisal, a measure of contour curvature [40], had the highest importance. Surprisal is largest in regions of high magnitude of curvature. Here, we calculate the surprisal at the bottom left point of each image, which provides information about the shape at its connection to the build plate. Our results indicate that distortion during printing is sensitive to the shape curvature at its connection to the build plate.

The next most important feature was ix_x_layer_mom(2), the standard deviation for histogram of second moment of area of each row of the shape image. The ix_x_layer features were created to capture the change in material distribution on a layer-by-layer basis. The other most important features, namely clen_hist_mom(1) (mean for histogram of chord lengths), tar_sig_mom(1) (mean for triangular area signature), and cta_sig_mom(4) (kurtosis for cumulative tangent angle signature) were also related to how material is distributed around the center axis. Our results agree with prior design guidance cautioning designers about drastic changes to the parts' cross-section. While shape area was selected as a top-21 important feature, it was not among the most important features based on permutation importance, indicating designers should focus more on minimizing abrupt changes in cross-section rather than minimizing overall part size. In addition to confirming design guidance, our model can be used to provide more specific thresholds (e.g., changes of no more than xx mm/layer should be allowed) or to quickly compare potential designs' distortion potential.

While we evaluated a large set of shape descriptors, our selection of 92 descriptors was not exhaustive. For example, additional features such as size of part at connection to build platform could be added. Many of our features tended to provide more global information about the shape, rather than local information. Based on the importance of surprisal at the base, local features computed just for the part base may be highly relevant and will be explored in future work.

Although we have focused on LPBF part distortion for a single material, our approach can be generalized for other metal AM processes and to predict other manufacturing outcomes. The geometric features we identified as important could be used in other data-driven or physics-informed distortion prediction models. By providing these models with information on how the variation in geometric features relates to the predicted manufacturing outcomes (e.g., residual stress), we can deliver more insights and achieve more informed decisions.

One important area for future development is to expand our dataset of parts. This distortion prediction approach can be tested for geometries other than axisymmetric parts. Because our parts were limited to axisymmetric parts, we could represent the input geometry as a single cross section. For non-axisymmetric parts, a single cross section is no longer sufficient. Many 3D shape descriptors exist (some which are extensions of the 2D shape descriptors we used here) and some application-specific shape descriptor features could be designed.

Sequential vertical cross-sections can be taken at various intervals to expand the applicability of the distortion prediction approach beyond axisymmetric parts. These cross-sections can be used to create images, and the important feature metrics for those part cross-sections can then be related to the maximum distortion. By examining the combinations of important feature values at different vertical cross-sections, we can gain insight into how they contribute to the maximum distortion.

In future work, this analysis can be expanded to predict distortions in specific locations, such as five key points on the CAD model, or to predict the distortion field across the entire part. We hope to also compare the prediction error of CNN models which directly take the image of the cross section as input to our traditional ML approach. With CNN, post-hoc interpretability methods would be needed, such as permuting the CAD models and evaluating the effect of the permutation on the predicted outcomes, but such methods may provide predictions with less error.

5. CONCLUSION

Data-driven distortion prediction models have the potential to accelerate the product development cycle by identifying interpretable geometric features that contribute to higher part distortion. In this study, a surrogate model was used to predict the maximum distortion of axisymmetric parts and identify important geometric features impacting the maximum distortion. With a relatively small training data set of 243 unique part shapes, model performance was relatively accurate with a RMSE of less than 0.10 mm. Specific geometric features, including surprisal and a feature related to the second moment of area, were found to have a significant effect on the maximum distortion, agreeing with existing ad-hoc design guidance. Our data-driven approach and the identified geometric features can be helpful to develop distortion models that can provide fast predictions for a wide range of parts.

ACKNOWLEDGEMENTS

We are grateful to Yaniv Morgenstern and the other authors of [34] for sharing their code to compute shape descriptors for our images.

REFERENCES

- [1] Bartlett, J. L., and Li, X., 2019, "An Overview of Residual Stresses in Metal Powder Bed Fusion," *Additive Manufacturing*, **27**(March), pp. 131–149.
- [2] Gouge, M., Denlinger, E., Irwin, J., Li, C., and Michaleris, P., 2019, "Experimental Validation of Thermo-Mechanical Part-Scale Modeling for Laser Powder Bed Fusion Processes," *Additive Manufacturing*, **29**, p. 100771.
- [3] Peter, N., Pitts, Z., Thompson, S., and Saharan, A., 2020, "Benchmarking Build Simulation Software for Laser Powder Bed Fusion of Metals," *Additive Manufacturing*, **36**.
- [4] Yavari, R., Smoqi, Z., Riensche, A., Bevans, B., Kobir, H., Mendoza, H., Song, H., Cole, K., and Rao, P., 2021, "Part-Scale Thermal Simulation of Laser Powder Bed Fusion Using Graph Theory: Effect of Thermal History on Porosity, Microstructure Evolution, and Recoater Crash," *Materials and Design*, **204**, p. 109685.
- [5] Kouraytem, N., Li, X., Tan, W., Kappes, B., and Spear, A. D., 2021, "Modeling Process–Structure–Property Relationships in Metal Additive Manufacturing: A Review on Physics-Driven versus Data-Driven Approaches," *J. Phys. Mater.*, **4**(3), p. 032002.
- [6] Vosniakos, G.-C., Maroulis, T., and Pantelis, D., 2007, "A Method for Optimizing Process Parameters in Layer-Based Rapid Prototyping," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, **221**(8), pp. 1329–1340.
- [7] Tapia, G., Khairallah, S., Matthews, M., King, W. E., and Elwany, A., 2018, "Gaussian Process-Based Surrogate Modeling Framework for Process Planning in Laser Powder-Bed Fusion Additive Manufacturing of 316L Stainless Steel," *Int J Adv Manuf Technol*, **94**(9), pp. 3591–3603.
- [8] Dong, G., Wong, J. C., Lestandi, L., Mikula, J., Vastola, G., Jhon, M. H., Dao, M. H., Kizhakkian, U., Ford, C. S., and Rosen, D. W., 2022, "A Part-Scale, Feature-Based Surrogate Model for Residual Stresses in the Laser Powder Bed Fusion Process," *Journal of Materials Processing Technology*, **304**, p. 117541.
- [9] Wong, J. C., Ooi, C. C., Chattoraj, J., Lestandi, L., Dong, G., Kizhakkian, U., Rosen, D. W., Jhon, M. H., and Dao, M. H., 2022, "Graph Neural Network Based Surrogate Model of Physics Simulations for Geometry Design," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1469–1475.
- [10] Ahmadi Dastjerdi, A., Movahhedy, M. R., and Akbari, J., 2017, "Optimization of Process Parameters for Reducing Warpage in Selected Laser Sintering of Polymer Parts," *Additive Manufacturing*, **18**, pp. 285–294.
- [11] Li, J., Jin, R., and Yu, H. Z., 2018, "Integration of Physically-Based and Data-Driven Approaches for Thermal Field Prediction in Additive Manufacturing," *Materials & Design*, **139**, pp. 473–485.
- [12] Mozaffar, M., Paul, A., Al-Bahrani, R., Wolff, S., Choudhary, A., Agrawal, A., Ehmann, K., and Cao, J., 2018, "Data-Driven Prediction of the High-Dimensional Thermal History in Directed Energy Deposition Processes via Recurrent Neural Networks," *Manufacturing Letters*, **18**, pp. 35–39.
- [13] Dong, G., Wong, J. C., Lestandi, L., Mikula, J., Vastola, G., Jhon, M. H., Dao, M. H., Kizhakkian, U., Ford, C. S., and Rosen, D. W., 2022, "A Part-Scale, Feature-Based Surrogate Model for Residual Stresses in the Laser Powder Bed Fusion Process," *Journal of Materials Processing Technology*, **304**, p. 117541.
- [14] Baturynska, I., and Martinsen, K., 2021, "Prediction of Geometry Deviations in Additive Manufactured Parts: Comparison of Linear Regression with Machine Learning Algorithms," *J Intell Manuf*, **32**(1), pp. 179–200.

- [15] Lee, S. H., Park, W. S., Cho, H. S., Zhang, W., and Leu, M. C., 2001, “A Neural Network Approach to the Modelling and Analysis of Stereolithography Processes,” *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, **215**(12), pp. 1719–1733.
- [16] Francis, J., and Bian, L., 2019, “Deep Learning for Distortion Prediction in Laser-Based Additive Manufacturing Using Big Data,” *Manufacturing Letters*, **20**, pp. 10–14.
- [17] Rong-Ji, W., Xin-hua, L., Qing-ding, W., and Lingling, W., 2009, “Optimizing Process Parameters for Selective Laser Sintering Based on Neural Network and Genetic Algorithm,” *Int J Adv Manuf Technol*, **42**(11), pp. 1035–1042.
- [18] Shen, Z., Shang, X., Zhao, M., Dong, X., Xiong, G., and Wang, F.-Y., 2019, “A Learning-Based Framework for Error Compensation in 3D Printing,” *IEEE Trans. Cybern.*, **49**(11), pp. 4042–4050.
- [19] de Souza Borges Ferreira, R., Sabbaghi, A., and Huang, Q., 2020, “Automated Geometric Shape Deviation Modeling for Additive Manufacturing Systems via Bayesian Neural Networks,” *IEEE Trans. Automat. Sci. Eng.*, **17**(2), pp. 584–598.
- [20] Samie Tootooni, M., Dsouza, A., Donovan, R., Rao, P. K., Kong, Z. (James), and Borgesen, P., 2017, “Classifying the Dimensional Variation in Additive Manufactured Parts From Laser-Scanned Three-Dimensional Point Cloud Data Using Machine Learning Approaches,” *Journal of Manufacturing Science and Engineering*, **139**(9), p. 091005.
- [21] Molnar, C., Casalicchio, G., and Bischl, B., 2020, “Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges,” *ECML PKDD 2020 Workshops*, I. Koprinska, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R.P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P.M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, E. Masciari, Z.W. Ras, P. Christen, E. Ntoutsi, E. Schubert, A. Zimek, A. Monreale, P. Biecek, S. Rinzivillo, B. Kille, A. Lommatzsch, and J.A. Gulla, eds., Springer International Publishing, Cham, pp. 417–431.
- [22] Decker, N., and Huang, Q., 2019, “Geometric Accuracy Prediction for Additive Manufacturing Through Machine Learning of Triangular Mesh Data,” *Volume 1: Additive Manufacturing; Manufacturing Equipment and Systems; Bio and Sustainable Manufacturing*, American Society of Mechanical Engineers, Erie, Pennsylvania, USA, p. V001T02A043.
- [23] Bracken, J., Pomorski, T., Armstrong, C., Prabhu, R., Simpson, T. W., Jablokow, K., Cleary, W., and Meisel, N. A., 2020, “Design for Metal Powder Bed Fusion: The Geometry for Additive Part Selection (GAPS) Worksheet,” *Additive Manufacturing*, **35**.
- [24] Utley, E., 2017, “An Introduction to Designing for Metal 3D Printing,” *The SOLIDWORKS Blog* [Online]. Available: <https://blogs.solidworks.com/solidworksblog/2017/06/introduction-designing-metal-3d-printing.html>. [Accessed: 22-Feb-2023].
- [25] “Xact Metal | Learning Resources,” Xact Metal [Online]. Available: <https://xactmetal.com/learn/>. [Accessed: 22-Feb-2023].
- [26] Adam, G. A. O., and Zimmer, D., 2014, “Design for Additive Manufacturing-Element Transitions and Aggregated Structures,” *CIRP Journal of Manufacturing Science and Technology*, **7**(1), pp. 20–28.
- [27] ISO/ASTM, 2018, *Additive Manufacturing — Design — Requirements, Guidelines and Recommendations*, ISO/ASTM 52910:2018.
- [28] Kang, S., Deng, X., and Jin, R., 2021, “A Cost-Efficient Data-Driven Approach to Design Space Exploration for Personalized Geometric Design in Additive Manufacturing,” *Journal of Computing and Information Science in Engineering*, **21**(6), p. 061008.
- [29] Jiang, J., Xiong, Y., Zhang, Z., and Rosen, D. W., 2022, “Machine Learning Integrated Design for Additive Manufacturing,” *J Intell Manuf*, **33**(4), pp. 1073–1086.
- [30] Yao, X., Moon, S. K., and Bi, G., 2017, “A Hybrid Machine Learning Approach for Additive Manufacturing Design Feature Recommendation,” *Rapid Prototyping Journal*, **23**(6), pp. 983–997.
- [31] Ghiasian, S. E., and Lewis, K., 2020, “A Machine Learning-Based Design Recommender System for Additive Manufacturing,” *Volume 11A: 46th Design Automation Conference (DAC)*, American Society of Mechanical Engineers, Virtual, Online, p. V11AT11A025.
- [32] ANSYS, Inc, “Additive User’s Guide (Print and Science),” *ANSYS®Additive Manufacturing 2021 R2*.
- [33] Mayer, T., Brändle, G., Schönenberger, A., and Eberlein, R., 2020, “Simulation and Validation of Residual Deformations in Additive Manufacturing of Metal Parts,” *Heliyon*, **6**(5), p. e03987.
- [34] Morgenstern, Y., Hartmann, F., Schmidt, F., Tiedemann, H., Prokott, E., Maiello, G., and Fleming, R. W., 2021, “An Image-Computable Model of Human Visual Shape Similarity,” *PLoS Comput Biol*, **17**(6), p. e1008981.
- [35] 2022, “How Much Data Is Needed For Machine Learning? | Graphite Note” [Online]. Available: <https://graphite-note.com/how-much-data-is-needed-for-machine-learning>. [Accessed: 11-May-2023].
- [36] Matthias, F., Eggenberger, K., Falkner, S., Lindauer, M., and Hutter, F., 2020, “Auto-Sklearn 2.0: Hands-Free AutoML via Meta-Learning,” *arXiv:2007.04074 [cs.LG]*.
- [37] “1.7. Gaussian Processes,” *scikit-learn* [Online]. Available: https://scikit-learn.org/stable/modules/gaussian_process.html. [Accessed: 09-May-2023].
- [38] Rasmussen, C. E., and Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass.
- [39] “Sklearn.Gaussian_process.GaussianProcessRegressor,” *scikit-learn* [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_proce

ss.GaussianProcessRegressor.html. [Accessed: 08-May-2023].

[40] Feldman, J., and Singh, M., 2005, "Information along Contours and Object Boundaries," *Psychol Rev*, **112**(1), pp. 243–252.