17th Hypervelocity Impact Symposium
HVIS2024
September 9-13, 2024, Tsukuba, Japan

HVIS2024-045

THE APPLICATION OF ARTIFICIAL INTELLIGENCE AND DEEP LEARNING TO VISUALLY IDENTIFY MICROMETEOROID AND ORBITAL DEBRIS IMPACTS

Cameron M. Collins[1,*] Dana M. Lear[2], Kenton R. Fisher[3]

[1]Jacobs-JETS II, NASA Johnson Space Center, Mail Code XI5, Houston, TX 77058, USA
[2] NASA, NASA Johnson Space Center, Mail Code XI5, Houston, TX 77058, USA
[3] NASA, NASA Johnson Space Center, Mail Code XI4, Houston, TX 77058, USA

ABSTRACT

Recent advancements in artificial intelligence (AI) have made machine learning (ML) techniques readily available for practical applications while using a fraction of time that was previously required. In particular, the use of deep learning (DL) models has the potential to assist in the identification of micrometeoroid and orbital debris (MMOD) damage. The ability to detect MMOD damage with AI models has the potential to improve the safety of human space flight by reducing the time required for damage identification via automated methods. To demonstrate this, a binary classification dataset of 226 images was utilized to train and validate a MMOD damage detection model using a residual network architecture. The benefits of data augmentation were consistent and improved performance as expected. For performance, the MMOD damage detection model achieved 95.6% accuracy on the validation set and generalized to an accuracy of 87.5% on a test dataset of spacecraft images with MMOD damage. With the use of deep learning algorithms, MMOD damage can be identified with high accuracy and provides a new possibility to identify of hypervelocity impacts.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Hypervelocity Impacts, Micrometeoroid, Orbital Debris, Pattern Recognition, Convolution Neural Network.

NOMENCLATURE

Complexity	The size of the parameter space for a model; larger parameter spaces lead to more complex models.
Dimensionality	The number of features in the dataset.
Hyperparameters	Parameters set before training such as learning rate.
Overfitting	Models that perform poorly on unseen data but well on seen data.
Parameter Space	Range of possible values that define a particular model.

1. INTRODUCTION

The use of artificial intelligence (AI) and other machine learning (ML) algorithms to address complex problems has seen significant growth and garnered widespread public attention. Language models like generative pre-trained transformers (GPT) have made interactive AI more accessible than ever. These once niche topics are now prevalent in both the scientific community and the general public. The rapid advancements in these technologies continue to open new avenues for research and practical applications across various fields.

The AI revolution can be contributed to breakthroughs in deep learning (DL) techniques that occurred as early as 2018, combined with the availability of powerful computing hardware such as graphics processing units (GPUs) and tensor processing units (TPUs), along with advancements in data storage and collection. These factors made training and deploying AI models more feasible by eliminating previous barriers such as data accessibility and training efficiency.

These leaps in technological advancement have presented unique opportunities to those with niche data sources to train models with capabilities once thought not feasible. For example, the application of AI has improved the detection of breast cancer by 20% compared to traditional detection methods, according to Norver et al [1]. Similarly, AI has the capability to integrate into and improve current micrometeoroid and orbital debris (MMOD) analyses. The need to locate and characterize spacecraft MMOD impact damage is critical not only for categorizing the MMOD environments but also ensuring spacecraft integrity through visual inspection. This process is vital to ensure no damage to the spacecraft has occurred which could lead to the loss of crew, vehicle, and/or mission [2], [3].

Traditionally, MMOD damage identification has been completed by having a crew member manually take photographs of the spacecraft through a spacecraft's window using hand-held

* Corresponding author: cameron.m.collins@nasa.gov

camera or ground personnel directing externally mounted cameras to examine the vehicle. The photographs are then transmitted back to Earth for visual analysis. This method of MMOD damage inspection works well and has been used on various spacecraft, including the Space Shuttle and the International Space Station (ISS) [3].

One of the drawbacks of the current method is the speed and the accuracy in identifying MMOD impacts, which could be improved. It is essential to note that the process of detecting MMOD impacts in images can be both difficult and time consuming. The visual appearance of an MMOD impact can change drastically with lighting conditions, size of impact, depth of penetrations, material types, surface waviness, fabric coverings, camera type and lens, image distance, spacecraft orientation, and multiple other factors. Currently, this process involves a team of highly experienced specialists in both the fields of image analysis and MMOD impacts. The use of machine learning models is not new to the field of MMOD impact protection with prior research being conducted on use of machine learning models for ballistic limit equation (BLE) development which are utilized in risk assessments for spacecraft [4].

Thus, it was deemed imperative to test and observe whether AI could enhance critical capabilities, such as visually identifying micrometeoroid and orbital debris impact damage on spacecraft from images of the spacecraft exterior. This paper documents the initial research, training and performance evaluation of an AI model designed to identify both actual and laboratory tested MMOD impacts and perforations on exposed flat surfaces. While this initial goal seems modest, training such a model would have taken years of development with teams of individuals just ten years ago. The long-term goal is to continually increase the complexity and use-cases of the DL model being trained, with the aim of expanding its capabilities to identify MMOD impacts on all types of spacecraft surfaces.

2. THEORY, METHODS, AND APPROACH

2.1 Deep Learning Models

Artificial neural networks (ANNs) are a cornerstone of deep learning, a subset of machine learning. The foundational concepts of ANNs gained significant prominence in the 1980s with the introduction of backpropagation techniques [5]. These techniques were novel because it made the training process more efficient and systematic by enabling the creation of fully differentiable ANNs through innovative network architectures. The advent of backpropagation allowed for the effective and systematic adjustment of weights to minimize error.

Over time, the artificial neural network was improved by machine learners, mathematicians, and statisticians. This allowed the methodology of the approach to stabilized [6]. However, ANNs shortly fell out of favor with the adoption of other ML methods such as support vector machines (SVMs), boosting, and random forests. Unlike artificial neural networks, methods like SVMs, boosting, and random forests, had smaller parameter spaces, resulting in reduced model complexities and easier model generalization. Benefits were further demonstrated in cases of reduced data. Later, in the 2010s, artificial neural networks resurfaced with major advancements in image and video classification, in addition to speech and text modeling [7].

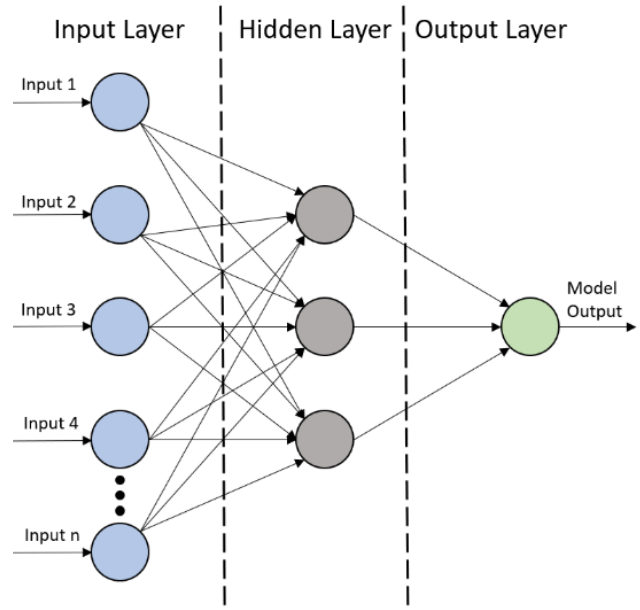

FIGURE 1: Simplified visual representation of a forward feeding artificial neural network model with three total layers, receiving an input vector of size n.

A simple artificial neural network, such as the network in Figure 1, will take a sample of data vectors, such as a flattened image array, as an input in the input layer. Subsequently, the hidden layer of the network then applies nonlinear transformations to the linear combinations of the input layer's components. In this process, the hidden layer updates weights by minimizing the backpropagated loss from the output layer to the hidden layer with an optimization algorithm such as gradient descent. After an artificial neural network has been trained, the output layer is a model using the hidden layer to predict results based on a given input. This complexity can be scaled and tuned by changing the number of hidden layers, the size of the dataset, the non-linear transformations utilized, and the number of hidden nodes per layer.

While there are several types of artificial neural networks, for the purposes of MMOD damage identification, a convolution neural network (CNN) was the learner utilized due to its established accuracy for image classification [7]. Specifically, a residual network (ResNet) was chosen since it efficiently addresses issues with gradient descent that occur in deep forward-feeding models [8]. In particular, the ResNet18 learner architecture was used for the model with the weights partially trained on the ImageNet dataset [9]. A simpler CNN model was considered but was not pursued due to computational constraints required to search for optimal hyperparameters and model architectures.

2.2 Dataset Description

The dataset used for all subsequent experiments had a distribution of both spacecraft surfaces and images from hypervelocity impact testing. All images in the dataset were classified beforehand into two categories: “MMOD Damage” and “No MMOD Damage”. All images were manually classified with the “MMOD Damage” classification including characteristics such as cratering, perforations, cracking, etc. These categories, which are commonly referred to as labels, are used to train and test the accuracy of the AI model. 80% of the dataset was used for model training while the remaining 20% was utilized as the validation set.

The dataset involved 130 images with MMOD damage and 96 images without MMOD damage, leading to a 58% and 42% split, respectively. Each image was initially 1280 pixels by 720 pixels before resizing. As for camera angle, 74% of the images were categorized as “Normal” to the surface, with the remaining 26% being at some other off normal angle. Camera distance relative to image location was another identifiable feature present in the dataset. Explicit camera distance measurements were not available for all images located in the dataset. To substitute the absence of quantitative measurements, qualitative descriptions of “Close”, “Medium Distance”, and “Far” were utilized for insight which resulted in a distribution of 32%, 44%, and 23%, respectively after manual classification.

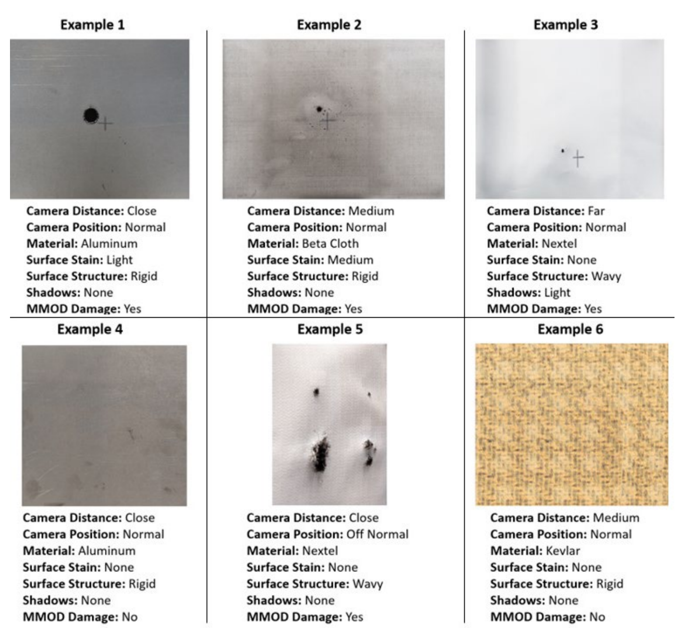

FIGURE 2: Example images in the dataset are show in the figure with characteristics described.

Examples of describable characteristics can be seen in Figure 2. One noticeable qualitative measurement is surface stain which is observable in example 2 in Figure 2. The more discolored the surface appeared, the higher the category bin was assigned to the image with “None” being labeled for images with no surface stain and “High” being assigned to images with significant amounts of surface discoloration. Additionally, the dataset can be broken up further by material representation of the inspected surface, as observed in Figure 3.

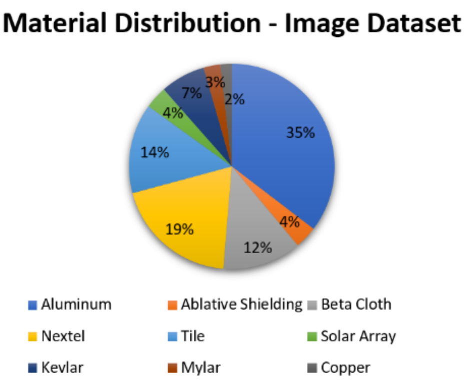

FIGURE 3: Primary material in each image of the dataset.

A plurality of images in the dataset were aluminum surfaces with other flight hardware materials present such as various ceramic fabrics and thermal protective shielding (TPS). Most images in the dataset (85%) appeared rigid with another 13% having a “wavy” surface appearance. Other notable visual features such as surface stain and shadowing had distributions present in the dataset as well, which is seen in Figure 4.

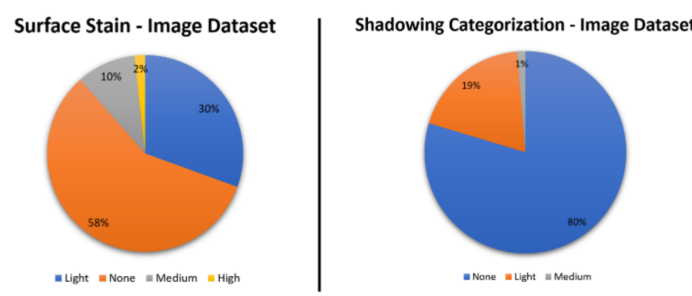

FIGURE 4: Shadowing and surface stain present in the dataset.

The dataset was representative of common spacecraft materials and future datasets will include other complexities. Due to computational limitations, this smaller subset of data was utilized. A much larger dataset will be utilized for future analysis and model training once computational resources are available.

2.3 Training and Testing Data

Since the ResNet18 model is accessible in PyTorch, equipped with pretrained weights from ImageNet, meaningful results were produced with 181 training images and 45 validation images. Previously, the limited data set would not be feasible to train a convolution neural network model, but recent advancements in DL algorithms, computing hardware, and data storage have allowed image models with preexisting weights to be finely tuned to domain-specific dataset.

Additionally, a second dataset was utilized for testing the model and was separate from the training and validation sets. The dataset consisted of 24 images of spacecrafts with 20 of them including MMOD damage and the other 4 not including MMOD damage. Images from this dataset are currently not available in the public domain and will not be featured in this paper due to export control regulation. However, model accuracy against the test dataset will be included to evaluate performance.

2.4 Data Augmentation and Preprocessing

For preprocessing, all images were resized to 400 pixels on the longest side. The resizing step ensured consistency in image dimensions across the dataset, which is crucial for uniform input into the model. This image sizing was found to be optimal after extensive tuning. Additionally, reducing the dimensionality of the images decreased the number of features the model needed to process, resulting in improved efficiency. This not only reduced the training time but also minimized memory usage, facilitating faster and more resource-efficient model training. Additionally, all whitespace present in the initial dataset was removed as well.

To increase the size and variability of the dataset, data augmentation was utilized to enhance generalization and reduce both underfitting and overfitting. Augmentations were randomly applied to each image during model training. This includes geometric transformations such as rotations, scaling, flipping, and cropping. Color augmentations such as brightness adjustments were utilized, in addition to other techniques such as shearing, perspective transformations, and noise injection.

The motivation behind the random application of these techniques on the dataset is to be representative of an ideal natural distribution of sampled data. This is due to the fact that not all images are taken at the same angle, aspect ratio, or size. Thus, data augmentation involves distorting the image in natural ways where the image is still recognizable so that human recognition remains unaffected [6]. An example of this can be seen in Figure 5.

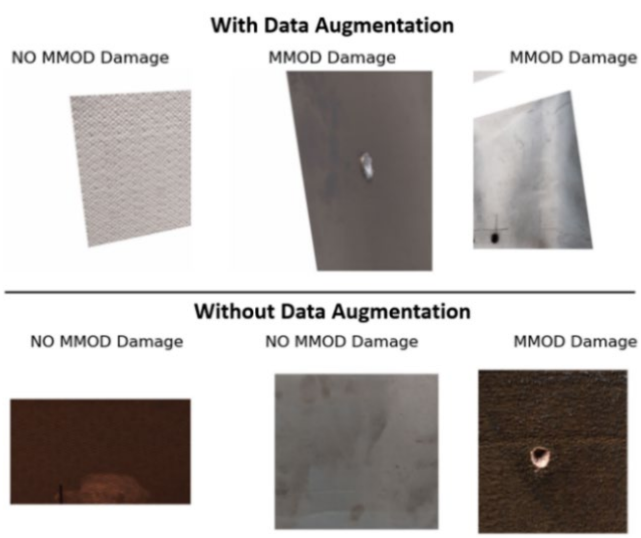

FIGURE 5: The visual difference data augmentation causes on the dataset.

For the model developed in this paper, data augmentation techniques had a 100% chance of being applied to each piece of data in the dataset, but the combination of techniques applied was randomly chosen. This approach expands the training set while reducing the likelihood of overfitting by utilizing data transformations. Lastly, the original dataset remains intact with augmentation only being performed when the data is sampled from the dataset.

2.5 Software and Coding

The MMOD impact identification model was built in Python version 3.9.19 using both the open source PyTorch library and the FastAI library [9], [10]. The FastAI library acts as a wrapper for the PyTorch library by increasing the ease of use through the application of data augmentation techniques and the visualization of results using built-in functions. An example batch of training images can be seen in Figure 6, after data augmentation have been applied to the sampled data.

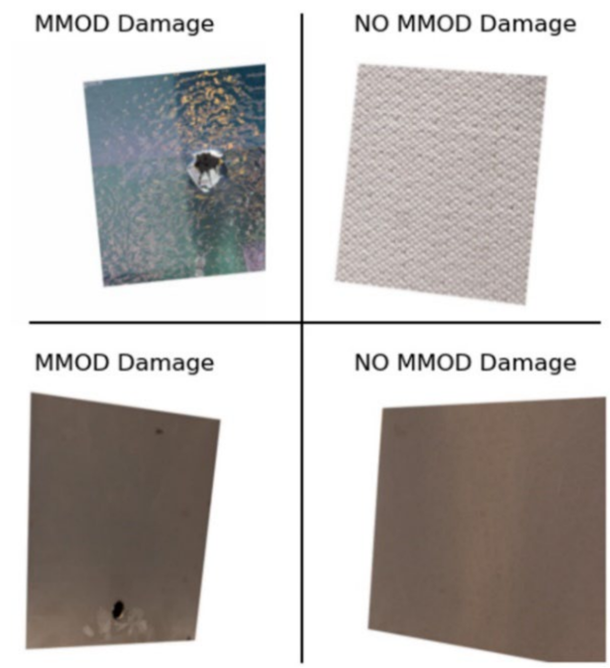

FIGURE 6: Four training images with random data augmentation applied and the proper classification of each image located above the image.

Afterward, the data set is loaded in a data loaders object which is used for training and hyperparameter optimization. The main hyperparameter of significance to this study is the maximum learning rate. The learning rate determines how much the weights can update between epochs. An optimal learning rate will train the model quickly and effectively. To determine this, the model and dataset are used to interpret the best learning rate. This is done by utilizing a very small learning rate and gradually increasing the learning rate. While this happens, the model is trained for increasing learning rate values. During this process, it records the loss at each learning rate which can then be interpreted.

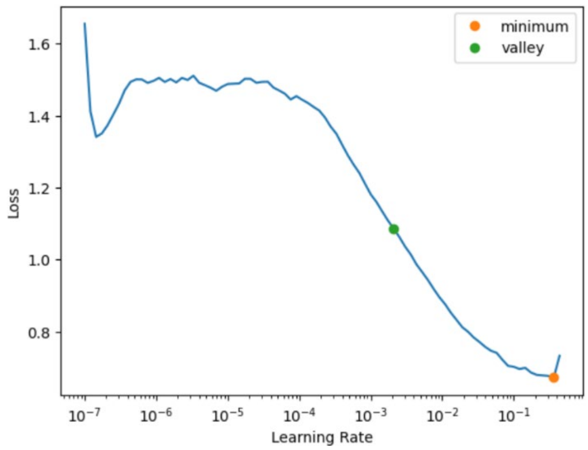

FIGURE 7: Loss vs Learning Rate for the MMOD damage detection model.

The plot in Figure 7, shows a loss that remains high in the beginning, indicating the model is updating weights minimally between epochs. As the learning rate increases, loss begins to decrease slightly before decreasing rapidly which signals that the learning rate value is enabling the model to meaningfully update weights between epochs. The orange marker signals where the loss is minimum with respect to learning rate. The loss then begins to increase at the largest learning rate values indicating that the model weights are updated aggressively between epochs, leading to model instability.

Each marker in Figure 7 represents different strategies for selecting a learning rate with each strategy having its advantages and tradeoffs. One marker to note on the plot is the valley marker which takes the steepest slope roughly 2/3 through the longest valley [10]. The minimum marker is denoted at the smallest calculated loss. Selecting the minimum point as the maximum learning rate would be the most stable, but it is also where the gradient of the loss function is not changing which leads to weights not being updated effectively. Instead, choosing the valley learning rate was a more reasonable choice for this problem as it was found to reduce the convergence time without affecting model stability. Valley was chosen to train all models. By implementing the following methods and knowledge listed above, an efficient AI model is trainable.

3. RESULTS AND DISCUSSION

3.1 Training and Validation Results

To illustrate the benefits of the data augmentation techniques applied in the previous section, two identical ResNet18 models were trained on the same dataset with the same random seed. The only difference being is that one model had randomly applied data augmentation techniques when sampling the training data while the other did not. Additionally, both learning rates were chosen with the same methodology. All parameters presented in the following are unitless.

For context, an epoch refers to a complete cycle through the training set, updating the weights of the model after each batch. The training loss measures how well the model is performing on the training dataset by comparing the classification predictions the model makes against the true labels of the dataset. The validation loss is the performance against the validation set. Additionally, datasets can be either balanced or unbalanced. The difference between a balanced and unbalanced dataset is class representation. Balanced datasets will have close to equal class representation of each label while unbalanced datasets will have disproportionate representation of one label over an another.

$$Error\ Rate = \frac{Number\ of\ Incorrect\ Predictions}{Total\ Number\ of\ Predictions} \quad (1)$$

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (2)$$

Common classification metrics to evaluate model performance include values such as error rate (1), accuracy (2), and F1 score. The F1 score is analogous to accuracy for unbalanced datasets. Acceptable thresholds for both F1 score and accuracy are typically above 0.7 with models above 0.9 typically deemed exceptional.

TABLE 1: Results and information regarding the trained MMOD impact identification model with no data augmentation.

Epoch	Training Loss	Validation Loss	Error Rate	Accuracy	F1 Score
0	1.195	0.909	0.356	0.644	0.704
1	1.060	0.555	0.244	0.756	0.766
2	0.939	0.349	0.178	0.822	0.789
3	0.844	0.322	0.156	0.844	0.800
…	…	…	…	…	…
8	0.351	0.229	0.067	0.933	0.923

Table 1 showcases effective training of the model, but the model is still underfitted since the training loss is much higher than the validation loss after eight epochs.

TABLE 2: Results and information regarding the trained MMOD impact identification model with data augmentation.

Epoch	Training Loss	Validation Loss	Error Rate	Accuracy	F1 Score
0	1.379	1.020	0.467	0.533	0.588
1	1.039	0.688	0.356	0.644	0.579
2	0.790	0.360	0.133	0.867	0.824
3	0.603	0.267	0.156	0.844	0.788
…	…	…	…	…	…
8	0.162	0.090	0.044	0.956	0.947

Table 2 also exhibits some underfitting behavior as observed in Table 1, but it is also showcasing the benefits previously mentioned about data augmentation. Additionally, the accuracy and F1 score was higher for the model trained with data augmentation. The learning curves for both models can be seen in Figure 8 and 9.

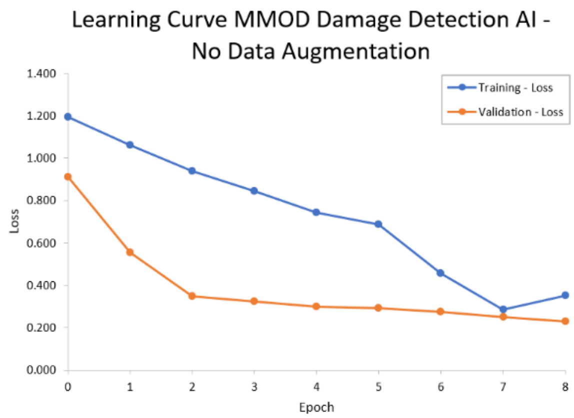

FIGURE 8: Learning curve for MMOD damage detection AI with no data augmentation.

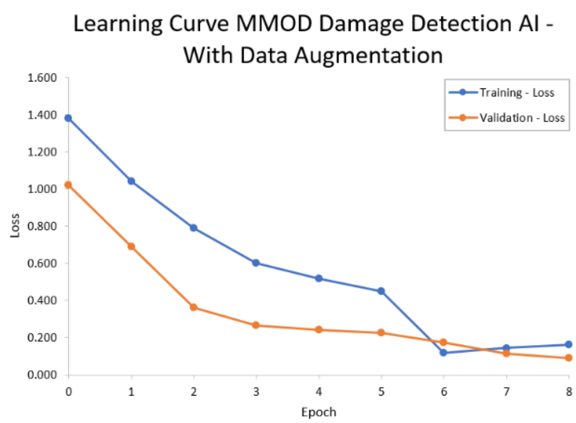

FIGURE 9: Learning curve for MMOD damage detection AI with data augmentation.

The results in Figure 8 and 9, graphically show the change in model performance as the number of epochs increases. Noticeably, both the training and validation loss for the model with data augmentation techniques applied converged to a lower value than the model without data augmentation techniques. The dataset utilized to train the following models was orders of magnitude smaller than most traditional datasets, and yet, a error rate of 4.4% was achieved when data augmentation was utilized. This lends credibility to the original point that DL models can be trained off very limited datasets using partial trained weights to initialize the learner.

When examining a batch of learner classification prediction examples from the data augmentation model against the validation data in Figure 10, it can be observed that the model accurately classified MMOD damage at various distances while handling different MMOD damage characteristics and materials. Figure 10 also shows one of the validation images that was labeled incorrectly by the model.

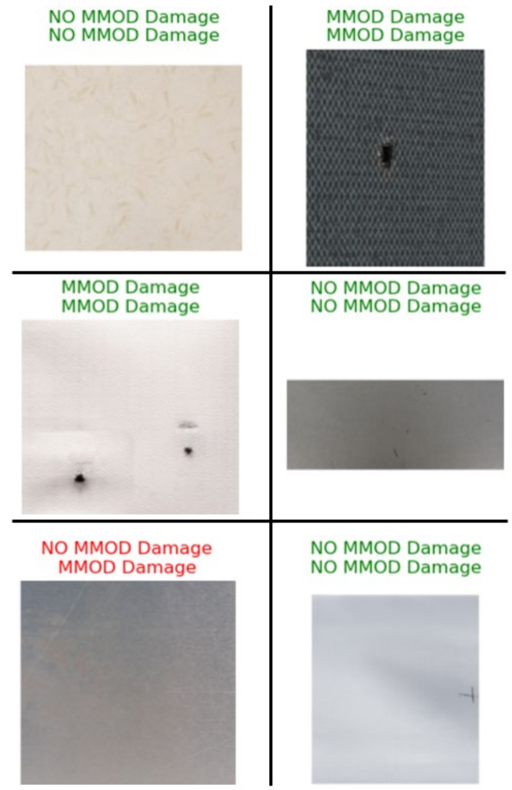

FIGURE 10: The topmost label above each image is the actual label while the bottom most label is the model prediction.

Additionally, another performance insight can be obtained by observing images where the trained model displayed the least confidence or confidently postured incorrect predictions on the validation dataset. This is useful since low probability correct predictions are viewed as areas of improvement for the model.

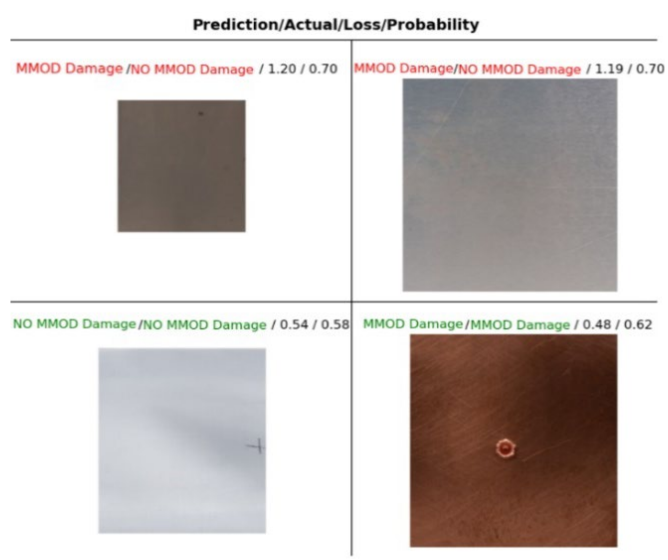

FIGURE 11: The worst predictions by performance made by the model with data augmentation against the validation set.

Figure 11 exhibits a couple of patterns the trained model struggled with. For instance, the MMOD impact identification model misclassified the two top aluminum plates with non-clean surfaces or other discolorations. The model classified these as MMOD damage, yet it is discernable via human recognition that the image does not display MMOD damage. One prediction on why this occurred is that the model lacks sufficient training data with images that have surface imperfections but no MMOD damage.

3.2 Spacecraft Test Data Results

The dataset with 24 images of spacecraft surfaces was utilized as a standalone set of data for testing. The MMOD damage detection model that was trained with the data augmentation techniques was tested on this dataset to see its performance. Once a model has been trained, the architecture of the model is frozen, and no further epochs are required to test the model.

TABLE 3: Test dataset results for trained MMOD impact identification model with data augmentation.

Validation Loss	Error Rate	Accuracy	F1 Score
0.273	0.125	0.875	0.666

Table 3 showcases that the model still has room for improvement, given the accuracy and F1 score. Particularly, the F1 score is below 70% which is a common model threshold that differentiates acceptable performance from unacceptable performance. The accuracy being above 70% is less significant since the test dataset is unbalanced. Model capability could be supplemented once a larger model is trained or more flight hardware imagery is included in the dataset.

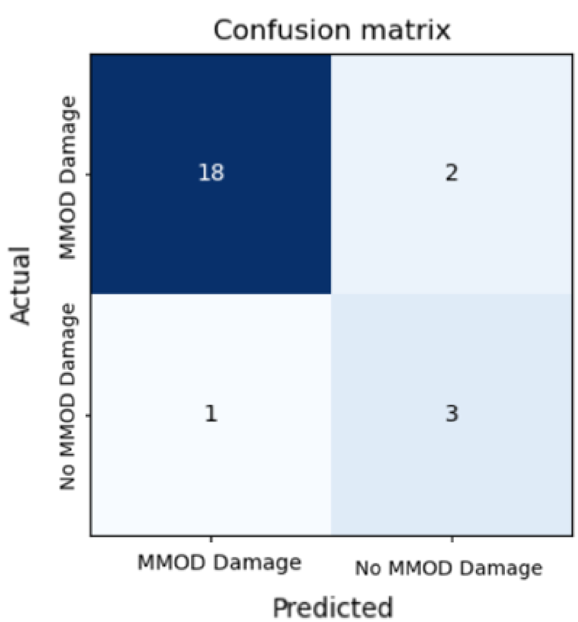

FIGURE 12: Confusion matrix for MMOD damage detection AI with data augmentation.

A confusion matrix is a common way to visualize the performance of a model against a test dataset. The main diagonal of the confusion matrix are all the correct classifications with all adjacent cells showing incorrect predictions. Insights such as model bias and precision can viewed as well. Figure 12 shows how the model classified much of the test dataset correctly but did misclassify images with MMOD damage and no MMOD damage. Expanding the training dataset to include more spacecraft data is expected to enhance performance. The test dataset provided to model could be manually classified to a higher accuracy, but with future improvements, the model is expected to reach acceptable classification rates at the benefit of reduced labor hours. Specific metrics such as time metrics will be included in future research.

3.3 Discussion and Future Research

Given a limited dataset of images showing MMOD damage and non-damage, the model achieved a high prediction accuracy against the validation set. Future efforts will focus on training with a larger dataset, which is expected to further improve the model generalization on unseen data. Additionally, there are other residual network models with additional layers to the neural network that can further increase the accuracy of predictions at the expense of other parameters such as training time.

Future research should explore the utilization of more advance hardware to minimize the need for reducing the size of images. The utilization of advanced hardware has both reduced training times and increased model accuracy as possible outcomes. Moreover, this becomes more paramount as the training dataset size continues to increase.

Furthermore, increasing the complexity of the images captured to assess MMOD damage is an important area of exploration as well. Spacecraft surfaces can include rivets, solar panels, guidance and navigation equipment, among other components, all of which must be considered when identifying MMOD damage. For the purposes of data collection, there is also different types of MMOD damage that can be categorized into additional distinct labels, instead of the initial binary that was set forth in this paper.

Finally, the use of DL models to investigate the process of predictive debris cloud formation due to hypervelocity impacts may lead to data mining of new relationships in MMOD shield performance, parameters, and types [11]. These new relationships have yet to be discovered, similar to how the DL model presented in this paper was able to learn features that allowed it to distinguish images between the two categorizations provided.

4. CONCLUSION

The renaissance of ANNs and the broader study of DL have unlocked new possibilities of model creation, image analysis and classification in addition to generative AI. Limited data can be utilized to train advance models with high levels of accuracy with minimal tuning while unlocking new possibilities of predictive modeling and data analysis. The weaker model performance on the spacecraft test dataset which was expected and gives opportunity for improvement for possible future research. The use of machine learning techniques increasing areas of interest that may expand to other areas of research such as predictive debris cloud generation and MMOD shield design improvements.

ACKNOWLEDGMENTS

The authors would like to express sincere gratitude to the NASA Johnson Space Center (JSC) Image Analysis team lead for help with this project. Additionally, the authors would like to express sincere thanks to Dr. Eric Christiansen for help with organizing and conducting this research. Lastly, the authors would like to thank the NASA JSC Hypervelocity Impact Technology (HVIT) team for providing essential feedback and support to make this a reality.

REFERENCES

[1] Nover, A. B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W. Y., Shih, W.-H., & Brooks, A. D. (2010). Modern Breast Cancer Detection: A Technological Review. *International Journal of Bio Imaging.* Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2804038/

[2] Hyde, J. L., Christiansen, E. L., Lear, D. M., Cline II, C. J., & Lozano, M. A. (2023). SpaceX Dragon Spacecraft Post Flight MMOD Inspection Campaign. In *2nd Orbital Debris Conf. Papers 2023.* Available: https://www.hou.usra.edu/meetings/orbitaldebris2023/pdf/6019.pdf

[3] Hyde, J. L., Christiansen, E. L., & Lear, D. M. (2019). Observations of MMOD Impact Damage to the ISS. In *First Int'l. Orbital Debris Conf. (2019).* Available: https://www.hou.usra.edu/meetings/orbitaldebris2019/orbital2019paper/pdf/6001.pdf

[4] S. Ryan, N. M. Sushma, H. Le, A. V. A. Kumar, J. Berk, T. M. Nguyen, S. Rana, S. Kandanaarachchi, and S. Venkatesh, "The application of machine learning in micrometeoroid and orbital debris impact protection and risk assessment for spacecraft," International Journal of Impact Engineering, vol. 181, 2023, Art. no. 104727. [Online]. Available: https://doi.org/10.1016/j.ijimpeng.2023.104727

[5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533–536, 1986. Available: Learning representations by back-propagating errors | Nature

[6] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning with Applications in Python* (pp. 399-413). Springer Texts in Statistics (STS).

[7] O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. *ArXiv e-prints.* Available: https://arxiv.org/pdf/1511.08458.pdf

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *ArXiv e-prints.* Available: https://arxiv.org/pdf/1512.03385.pdf

[9] PyTorch. (2024). *PyTorch: An open source machine learning library.* Available: https://pytorch.org/

[10] fastai. (2024). *fastai: A deep learning library for training fast and accurate neural networks using modern best practices.* Available: https://www.fast.ai/

[11] Christiansen, E.L et al., Handbook for Designing MMOD Protection, NASA Technical Memorandum, NASA/TM-2009-214785, 2009