Facilitating Scientific Material Discovery via Deep Learning on Small Image Datasets

PhD Dissertation

Biao Yin
Worcester Polytechnic Institute
Data Science Program

April 18th, 2024

Committee Members:
Dr. Elke A. Rundensteiner, Professor, WPI. Advisor.
Dr. Ziming Zhang, Assistant Professor, WPI. Co-Advisor.
Dr. Jian (Frank) Zou, Associate Professor, WPI.
Dr. Robert Jensen, Team Lead, Materials Data Science, DEVCOM Army Research Laboratory. External Advisor

Abstract

Scientific material discovery, important for economic prosperity and well-being from transportation, construction, and security, to healthcare; has traditionally been tackled both by intensive mathematical modeling and extensive physical experimentation. Recently, popular deep learning models have emerged as a promising solution approach. However, challenges remaining include overfitting due to the typically small size of datasets in this domain, tiny but critical pixels, strong false positives, lack of domain knowledge, etc.

In this dissertation, I tackle four particular directions of research related to deep learning models on experimental complex data such as images derived from real-world projects for scientific material discovery.

In the first part, we designed and developed the first open-source corrosion image dataset, annotated for data-driven automation in scientific corrosion assessment using expert labeling. Using this dataset, we built an AI platform, incorporating our published deep learning model, for real-world anti-corrosive material discovery rating via automatic data collection, exchange, and visual analytics embedded with our published deep learning models.

In the second part, we focused on deep learning models in image-based scientific corrosion assessment for existing alloys. Techniques like augmentation, transfer learning, contrastive learning, as well as generative self-supervised learning were incorporated into the solution to improve its effectiveness.

In the third part, we innovated a science-informed deep learning model named DeepSC-Edge, enhanced by a novel edge guidance submodel. This submodel focuses

attention on high-level edge shapes while utilizing our unique loss function to prevent overfitting to edges. Additionally, our model incorporates a class-balanced loss, improving segmentation, particularly with challenging yet essential edges crucial for scientific corrosion assessment.

In the fourth part, we created a domain-promptable AlloyGAN model aimed at producing microstructure images for alloys that have not previously existed in the world, based on their chemical composition and manufacturing parameters. By integrating domain knowledge into the model, my research empowers material scientists to effectively handle hypothetical alloys through instant and scientifically validated material simulation and evaluation. This approach represents a quicker and equally precise alternative to conventional methods in material science for evaluating alloy microstructures, while also showcasing the potential of GAN-based models in advancing scientific exploration in the realm of materials discovery.

This work is based on a collaboration with material scientists at the DEVCOM Army Research Laboratory (ARL) - with the later testing and working with the resulting technology. In a collaboration between ARL, WPI, and ASM, the technology is being transitioned into practice for marketing and release by ASM. In general, the application of AI techniques to material science challenges promises to save time and effort for scientific material discovery.

Acknowledgments

Contents

1 Introduction

The field of material discovery plays a pivotal role in driving scientific advancements and expanding our technological horizons. However, traditional approaches heavily reliant on complex mathematical modeling often encounter limitations in terms of efficiency and practical applicability. To address this challenge, this dissertation focuses on harnessing the power of deep learning, a highly data-centric technique, and applying it to expedite the process of material discovery. Despite the promising potential of this approach, a significant obstacle persists - the scarcity of datasets in the realm of material science, leading to issues such as overfitting when employing deep learning models. To overcome these hurdles, this research explores innovative methods of embedding material science knowledge into deep neural networks.

1.1 Motivation

Motivated by the immense possibilities that arise from integrating AI with material science and the opportunity to make substantial strides in material discovery, our collaboration with the Army Research Laboratory (ARL) was initiated in 2019. The primary objective was to address the scarcity of data and the challenge of overfitting when applying deep learning models to material science research.

With the growing use of AI to advance understanding in diverse fields, such as engineering, medicine, healthcare, agriculture, and environmental science, we recognize the potential of a semi-automated model. In this model, human scientists work alongside AI systems, necessitating precise communication of existing knowledge to the AI and comprehensible presentation of the AI's findings. Given the remarkable successes of deep neural works in various tasks, there remains a critical gap in how corrosion science knowledge can be effectively integrated into deep learning models to automate material science research.

1.1.1 Motivating Example: WPI-ARL GQP Project
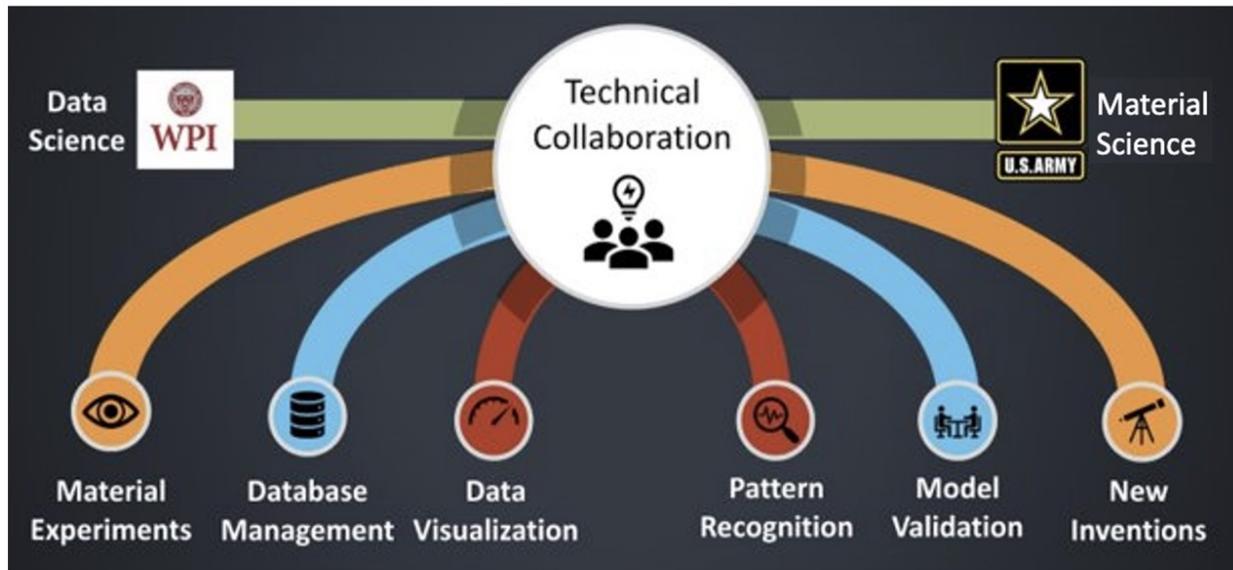
Figure 1: WPI-ARL Collaboration in Data Science GQP Projects

Collaborating closely with researchers from the Army Research Lab (ARL), we embarked on developing digital technologies to support the study and design of materials with enhanced corrosive properties. This endeavor included the collection of real-world data from corrosion experiments conducted in both indoor and outdoor testing environments, enabling detailed analytics. Additionally, we developed an automatic smartphone-based tool to collect and rate corrosion data in the field, along with the scalable and secure management of experimental data in a big data store. By optimizing the data collection tool, we significantly reduced the workload for effective data collection by ARL researchers in the field. Details can be viewed at our project home website: https://arl.wpi.edu/

During these projects, we have mentored over 100 WPI students majoring in Data Science, Computer Science, Mechanical Engineering, Material Science, IT, Statistics, etc.

Material science researchers and marketing specialists have been joined from ARL and ASM to collaborate, utilize, and commercialize the products.

1.1.2 Overall Objectives

The overarching aim of this work is to bridge the gap between corrosion science and AI research, accelerate the discovery of new materials, and facilitate efficient knowledge transfer between domain experts and AI models. We endeavor to usher in a paradigm shift in material science, transitioning from a labor-intensive, slow-paced process to a rapid, automated approach powered by advanced AI techniques.

In this dissertation, we aim to explore how corrosion science knowledge can be seamlessly incorporated into deep learning models, thus effectively automating material science research. We identify domain knowledge as expert annotations that measure anti-corrosive capability or material composition, such as chemical elements and manufacturing temperature. Through a series of machine learning tasks, in collaboration with our domain collaborator at ARL, we tackle the challenges of scientific corrosion classification, regression, segmentation, and material generation using deep learning techniques.

1.2 Part I, II, III – Automating Scientific Corrosion Assessment on Existing Materials

1.2.1 State-of-the-Arts

For corrosion science, machine learning solutions have been applied to automate engineering tasks such as defect detection [1–4] and corroded pipe detection [5]. However, there is no well-performing ML work that exists for scientific corrosion assessment [6]. Due to the difficulty and expertise required to annotate the corrosion, to the best of our knowledge, our work provides the first expert-level image, rating, and segmentation datasets for scientific corrosion assessment and corresponding high-performing deep learning methods for standardized corrosion assessment with innovations in both data

and material science fields. In the broader sense, deep learning has been widely used in image prediction tasks. UNet is a popular deep learning model that has been successfully applied in various image prediction tasks [7–10], such as image segmentation. It was first introduced for biomedical image segmentation and primarily introduced the utilization of a skip-connection structure to propagate information from the encoder to the decoder. Furthermore, there exist works that aim to refine predicted segmentation by integrating additional boundary and edge information [11–14]. Losses to address class imbalance during training in tasks like object detection by applying a modulating term to the cross entropy loss to focus learning on hard, misclassified examples has been shown to be effective in improving deep learning performance in real-world tasks [15]. In this work, we also incorporate edge information into UNet, however, critically demonstrate methods involving ground-truth edges and a class-balanced loss to benefit scientific corrosion segmentation, especially on difficult-to-segment images that pose both domain-related and general-purpose segmentation challenges.

1.2.2 Challenges

However, precisely, quickly, and safely assessing corrosion progression remains a challenging task, hampering the understanding of corrosion and material discovery as a whole. Several factors contribute to this challenge, including the costs of manufacturing processes and running tests, the danger of hazardous chemicals, the need for expert knowledge to identify corrosion, long observation periods for corrosion progression, and inherent biases associated with human analysis and measurements.

From the perspective of computer vision, applying deep learning to corrosion assessment faces specific challenges: 1) Limited High-Quality Data: Collecting high-quality corrosion data suitable for training data-driven models is challenging due to the limited observation of high-quality samples, which are essential for accurate scientific results. 2) Small Dataset Size: Corrosion datasets often have limited samples, making it difficult to

train deep learning models effectively. 3) Diverse and Complex Patterns: Corrosion images exhibit intricate patterns, tiny scales, corrosion underneath, and similar colors and textures with non-corrosion objects, posing challenges for accurate classification. 4) Transfer Learning and Pre-training: Assessing the suitability of transfer learning from pre-trained models on ImageNet to corrosion classification is essential, as corrosion images differ significantly from natural images. 5) Manual augmentation and Self-supervision: Verifying which ways to field semantic information well is necessary, manual augmentation, contrastive or generative self-supervision are conducted to compare the performance.

1.3   Part IV - Discovering Unknown from Known via Generative Models

1.3.1   State-of-the-Arts

Conventional approaches in alloy microstructure modeling involve extensive computational efforts and expertise. Techniques such as Phase Field modeling, Monte Carlo methods, and Cellular Automation have been used to simulate various aspects of recrystallization and solidification but face challenges in complex metallurgical processes and transition rules. On the other hand, deep learning methods, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Convolutional Neural Networks (CNNs), have shown promise in material generation and microstructure analysis. We propose to structure utilizes fundamental material compositions as conditions to prompt microstructure image generation, thereby facilitating the process of scientific alloy discovery.

1.3.2   Challenges

Traditional numerical methods struggle to simulate the final microstructure in alloy formation accurately due to the complex nonlinear chemical and physical interactions in-

volved. These methods, which often rely on intricate Partial Differential Equations (PDEs) and Finite Element Methods (FEM), can be computationally demanding and require specific expertise in material science, limiting their accessibility.

1.4 Overview Dissertation Tasks

To address the aforementioned challenges and achieve the objectives of this dissertation, we define a series of tasks:

1.4.1 Part I: Open Corrosion Image Dataset and MOSS AI Platform

The initial task involves creating the first-ever open corrosion image dataset with expert annotations. This dataset aims to provide a valuable resource for the computer vision community, enabling research in scientific corrosion analysis.

Additionally, we develop an AI platform to automate scientific corrosion assessment, facilitating seamless collaboration between AI models and domain experts.

1.4.2 Part II: Automatic Corrosion Rating

In this task, we explore the use of CNN and Transformer-based models for corrosion classification. We investigate the effectiveness of manual domain-specific data augmentation compared to self-supervised deep learning models. In self-supervised models, we compared PIRL and Masked Auto-Encoder (MAE) models to find if a contrastive or generative approach works with corrosion ratings. By leveraging transfer learning and self-supervised learning techniques, we aim to improve classification performance despite the limited dataset size. Additionally, we incorporate ordinal regression instead of classification, utilizing corresponding loss functions on the best-performing model to further suit the nature of corrosion rating.

1.4.3 Part III: Learning Segmentation via Ground-truth Edges

Thirdly, we delve into corrosion segmentation and introduce a novel ground-truth edge guidance technique to enhance the performance of the UNet architecture. By leveraging innovative loss functions that prevent overfitting on edges, we empower UNet to effectively learn intricate corrosion segmentation, utilizing high-level edge shapes guided by the Decoder.

1.4.4 Part IV: GAN for Unseen Alloy Discovery

Lastly, we develop promptable generative models that allow material scientists to view instantaneous corrosion simulations based on established parameters, such as chemical reactions. This facilitates rapid material performance assessment without the need for complex and inefficient mathematical equations.

1.5 Roadmap

We commence our journey by presenting the first-ever open corrosion image data set with expert annotations and an AI platform incorporated with our published deep learning model for (*Part I*). We then explore the use of Convolutional Neural Networks (CNN) and Transformers for corrosion classification, leveraging techniques like manual augmentation, transfer learning, and self-supervised learning to improve performance, including ordinal regression on corrosion data using generative self-supervised models (*Part II*). Further, we devise a ground-truth edge guidance methodology in the decoder with innovative loss functions, enabling UNet to learn complex corrosion segmentation via ground-truth edge shape in the Decoder (*Part III*).

Lastly, we invent promptable generative models for innovative Alloy simulations on microstructure images (Part IV). This model is now openly available, offering fast and scientifically valid results instead of the slow and equivalent outcomes produced by com-

plex conventional material science models.

By overcoming the challenges associated with small data sizes in deep learning and their applications to real-world scientific material discovery, we aspire to create a robust bridge between material science and AI. The resulting technology is set to significantly save time and labor in the pursuit of new material discovery, thereby bringing us closer to the next wave of technological advancements.

Our resulting methodologies and findings have been published in renowned conferences and are accessible through public platforms, enabling the broader research community to benefit from our work. Moreover, our software products, including an iPadOS app and web dashboard, embedded with our published deep learning models, are in use by ARL, under standardized by ASM, and going to apply with PPG and NASA, further affirming the practical impact of our research. We thank ARL, NSF, and DoE for sponsoring this research.

Part I

Open Corrosion Image Dataset and MOSS AI Platform

2 First-ever Open Corrosion Image Dataset and MOSS AI Platform used for Machine Learning to Accelerate Real-world Material Discovery (Task 1)

In this part, we introduce a novel corrosion image dataset, the first of its kind, featuring expert annotations and ratings obtained through standardized material science experiments. Comprising 600 real-world images of experimentally-tested material panels, the dataset provides a unique resource for advancing the application of AI in materials research, specifically in accelerating the discovery of new materials. Figure 1 illustrates examples of the dataset.

One of the challenges posed by our dataset is the inherent complexity of the images, which may exhibit single or double scribe lines, variable background colors, and noisy spots unrelated to the actual corrosion. Furthermore, the thinness of certain corrosion instances can make them virtually invisible to the untrained eye.

2.1 Motivation: Expert vs. Non-expert Ratings

The complexity of our dataset and the intricacies of corrosion assessment become apparent when comparing the performance of trained non-experts to that of experienced corrosion scientists. We carried out a study in which non-experts, who received training from a corrosion expert, attempted to rate corrosion on a subset of 60 image panels from our dataset.

Despite following the ASTM D1654 standards, the non-experts achieved a classification accuracy of only 0.38 when compared to the ground truth expert ratings. This underscores the importance of domain expertise and experience in accurately identifying and rating corrosion, and motivates the use of deep learning as a means to replicate this expertise.

2.2 Dataset Details and Contributions

Our open dataset includes 600 corrosion panel images of 512x512 pixels each, with each panel having a ground truth rating assigned by a corrosion scientist. To avoid imbalances in the data and to focus on the most domain-relevant corrosion ratings, we included only panels with ratings in the range of 5-9.

In addition to the images, the dataset includes expertly curated binary segmentation masks for each panel. These masks were obtained through an iterative process involving the use of the OpenCV GrabCut algorithm and consultation with domain experts to ensure accuracy.

This dataset represents a significant contribution to the field, not only due to the unique data it provides, but also due to the challenges overcome in its creation. These challenges include the domain knowledge required for panel assessment, the time and cost of experimental procedures, and the need for appropriate laboratory facilities.
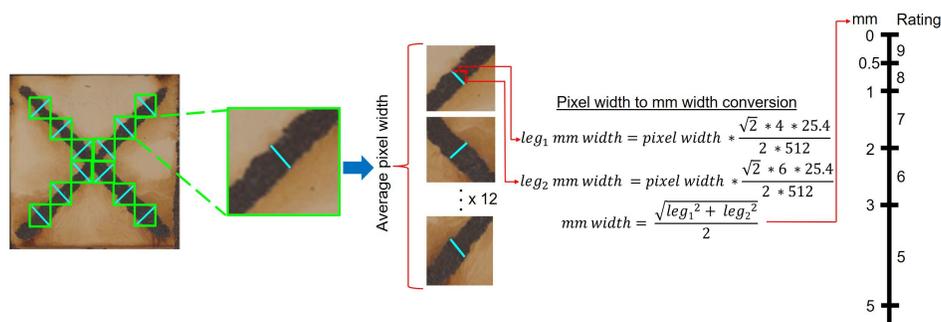
Figure 2: Non-expert rating procedure. Corrosion areas are identified in 12 locations on an image, measured on the computer in a pixel width, averaged across the 12 boxes, converted to a mm width, and then assigned the appropriate scribe corrosion rating 0-10.
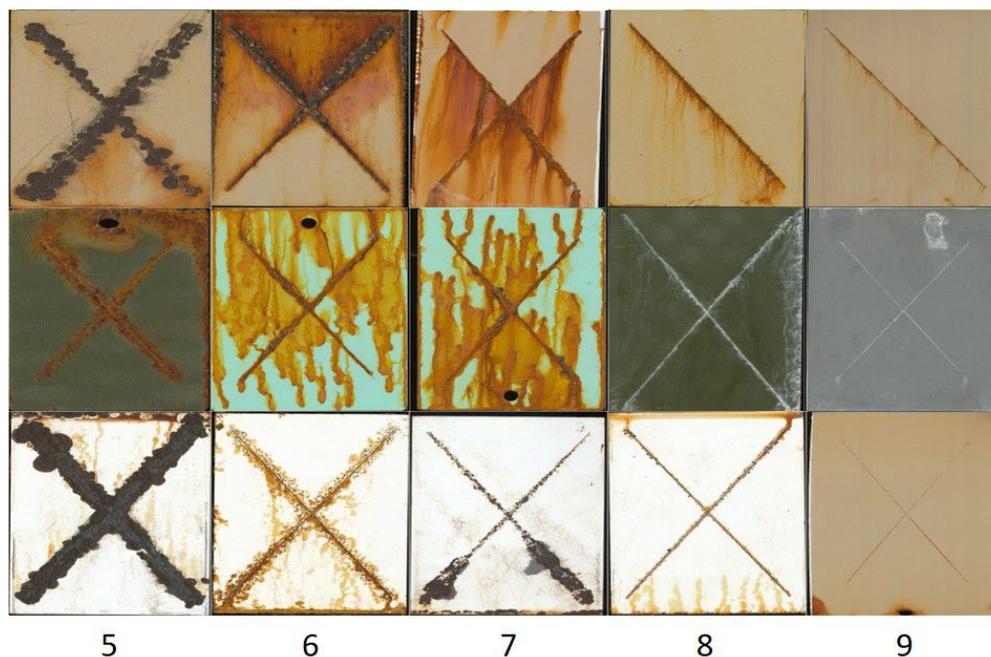
Figure 3: Three sample images per corrosion rating class 5 - 9 from our data set.

2.3 Related Publications and Software Products

The creation of this unique dataset and its implications for automating the scientific assessment of materials have been detailed in our paper published in the 2021 British Machine Vision Conference (BMVC).

In addition, we have developed an innovative AI-based digital platform, MOSS (Materials Open Science Software), to support materials science corrosion research. This platform, which was developed in collaboration with the Army Research Lab, includes a user-friendly iPadOS app for in-field corrosion progression data collection, a deep learning module for tasks like automatic corrosion assessment, a robust data repository for long-term experimental data archiving and modeling, and a visual analytics web portal for materials science research. The architecture of the MOSS platform is shown in Figure 2. The platform is transited to ARL and more details, including user cases, have been published to the 2023 Conference on Information and Knowledge Management (CIKM).

Figure 4: Capture, record, and save all experimental information about a new experiment.

The MOSS platform, in a real-world setting, is utilized in field studies to analyze corrosion progression in different materials. The accompanying app facilitates comprehensive real-time data collection, including photos, material condition notes, and environmental readings, even under limited internet connectivity. Once data is collected, it is uploaded for analysis on the server and web portal. The platform's deep-learning module assesses each material sample's corrosion state, providing clear and automatic insights into corrosion understanding.

iPadOS App to Serve Data Collection. Initial data collection is carried out using the user-friendly iPadOS app. A technician can collect, annotate, and save their experimental procedures and observations in an organized fashion. After logging in, the technician can begin to review or create new experiment sessions and tasks. Users can review past experiments and detailed records. Each material sample, complete with its details such as surface substrate, pretreatment, profile, primer, topcoat, timestamp, and a unique QR code, can be reviewed and shared among a team.

As seen in Figure 4, a new experiment can be created, wherein a user can photograph the material and input corrosion measurements. Corresponding rating scores are then
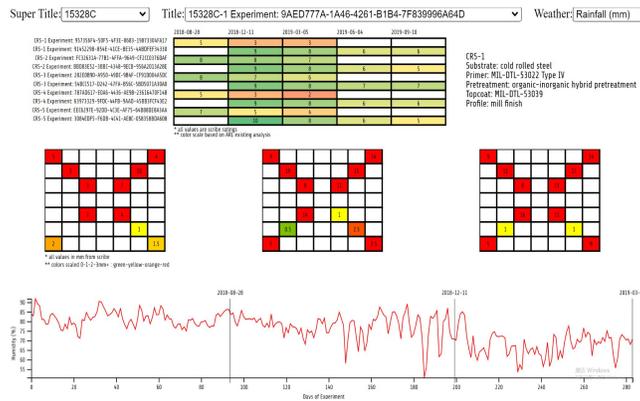
Figure 5: Select from a variety of plotting tools and material features to visualize trends and relationships across observations of corrosion progression ratings.
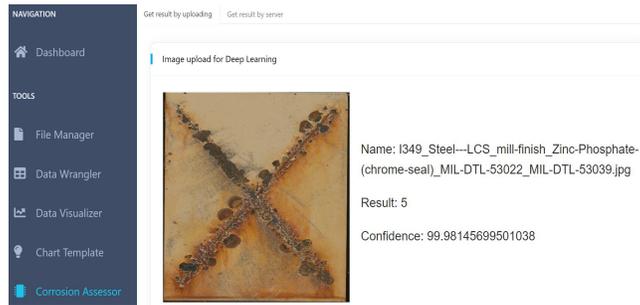
Figure 6: Automatically assess a corrosion panel image with embedded deep learning classification model. Confidence shows the softmax probability of the predicted class.

calculated and displayed in boxes, with our deep learning algorithm offering a predicted corrosion rating for optional inclusion.

Data Repository to enable long-term data integration. After collecting data for an experiment session, the user can select and back the data up to the data repository. This data repository allows for long-term and comparative studies, assisting in building a comprehensive understanding of corrosion progression. The system's advanced search capabilities, along with data versioning features, ensure efficient data management and uphold data integrity.

Web portal to enable Deep Learning and Visual Analysis. After collecting and saving data, further analysis and discoveries can be done through our web portal. As seen in Figure 5, the web portal allows for interactive data visualization from MOSS uni-

fied database to recognize patterns and interrelationships of detailed corrosion ratings on samples, weather conditions, and materials. Most importantly, corrosion and AI researchers can easily access the data and view manual and automatic assessments by interacting with the embedded deep-learning models. Sample image results from the database can be displayed on the web portal (See "Get results by server" tab in Figure 6), where researchers can also upload local images to view deep-learning results (See "Get results by uploading" tab and results shown in Figure 6)

In this demonstration, we showcase through real-world use cases how MOSS supports and expedites corrosion assessments over time and with it the development of new corrosion-resistant materials. This ultimately may contribute to the reduction of the economic burden and safety problems caused by the corrosion of our national infrastructure. Our vision is to inspire a broader adoption of AI technologies in corrosion research in particular and, in doing so, ignite a revolution in the field of materials sciences in general. Our MOSS platform has been deployed and attempted at the Army Research Lab. We provide a video demo at: https://www.youtube.com/watch?v=CzcxMMRsxkE

Part II

Automatic Corrosion Rating

3 Corrosion Classification from CNN and Transformer via Manual or Self Augmentation (Task 2)

Deep learning models, such as Convolutional Neural Networks (CNN) and Transformers, have shown significant potential in computer vision tasks. However, these models often require large amounts of data for training, which may not be readily available in all fields. In the domain of material science, particularly in corrosion classification, presents several challenges, primarily due to the scarcity of annotated data and the need for a comprehensive understanding of material degradation mechanisms. The small dataset size restricts the ability of traditional regression models to generalize effectively, leading to potential overfitting and inaccurate predictions. To overcome this hurdle, we investigate the use of manual and self-augmentation from the perspective of self-supervision to improve the model performance, including generative self-supervision, as the complex interplay between various corrosion factors demands a robust method capable of capturing nuanced correlations and generating meaningful representations.

Our task is crucial to advancing material discovery and establishing a more efficient connection between material science and AI research.

3.1 Manual Augmentation

Manual augmentation is a widely adopted technique in machine learning and computer vision tasks to increase the amount of training data. This process involves applying various transformations, such as rotation, scaling, and translation, to the original images to create new, 'augmented' images.

Loss	Focus to learn	Parameter name to tune	Decay type	Interact with	Motivation	ARL Application
Focal loss	hard samples	gamma	power of softmax learnt in process	softmax learnt in process	Increase test accuracy	Imbalanced in/outdoor class
Balanced cross entroy loss	minority samples	alpha	fixed multiplier	softmax learnt in process	Increase test accuracy	Imbalanced in/outdoor class
GHC loss	outliers (extreme hard samples)	beta	denominator of a function of softmax learnt in process	softmax learnt in process	Increase test accuracy	Failure coupons
[NIPS 20'] Debiased Contrastive Loss (involve energy-based GAN?)	true negative samples	tau	fixed denominator and multiplier	multiplication of logits learnt in process	A better embedding which is to increase downstream test accuracy	?
[ICLR 20'] Hardness-biased Contrastive loss	hard negative samples	beta	fixed denominator and multiplier	multiplication of logits learnt in process	A better embedding which is to increase downstream test accuracy	?
[ICML 20'] a loss to constrain alignment and uniformity on hypersphere	Embedding space	Alpha, t	power of l2 distance multiplier in RBF	functions of logits	A better embedding which is to increase downstream test accuracy	How well it can be applied to ARL data?
A part in any loss above	**Focus to learn**	**Parameter name to tune**	**Smooth type**	**Interact with**	**Motivation**	
Temperture scarling	Smoothed space 1. Softmax -- Euclidean 2. Softmax dot product -- Hypersphere	t	fixed denominator	logit learnt in process	Amplify loss to avoid local minimum while optimization	-

Figure 7: Literature Review: Self-supervised Learning

We used manual augmentation to increase our dataset size, which originally consisted of 600 images. Our augmentation methods included rotations and flips, resulting in an augmented dataset that was four times the size of the original one. By leveraging these manually augmented images, we were able to improve the performance of our corrosion classification model.

3.2 Self Augmentation

While manual augmentation is an effective method to increase the dataset size, it may not be sufficient for complex tasks like corrosion classification, which require an understanding of high-level features. To further boost the performance of our model, we explored self-augmentation techniques, specifically self-supervised learning.

Self-supervised learning is a learning paradigm where the model learns to predict a part of the data from other parts of the same data. In our case, we used the MixMatch algorithm, a semi-supervised learning method, which treats the predictions on augmented

unlabeled data as targets, and then refines these targets using the labeled data.

3.3 Results

- *Learning with manual data augmentation:* ResNet-18, ResNet-50, DenseNet, and HR-Net [16–18] are trained from scratch with extensive tuning with 9 different data augmentation approaches, achieving 0.81, 0.77, 0.80, and 0.77 best test accuracy using 10-fold cross-validation, respectively.

- *Self-supervised learning:* We investigate the potential of a recent self-supervised learning approach, PIRL [19], which takes advantage of data augmentation automatically for representation learning. We compare the pretrained PIRL on ImageNet and the PIRL trained from scratch on our data, leading to 0.75 and 0.72 test accuracy, respectively.
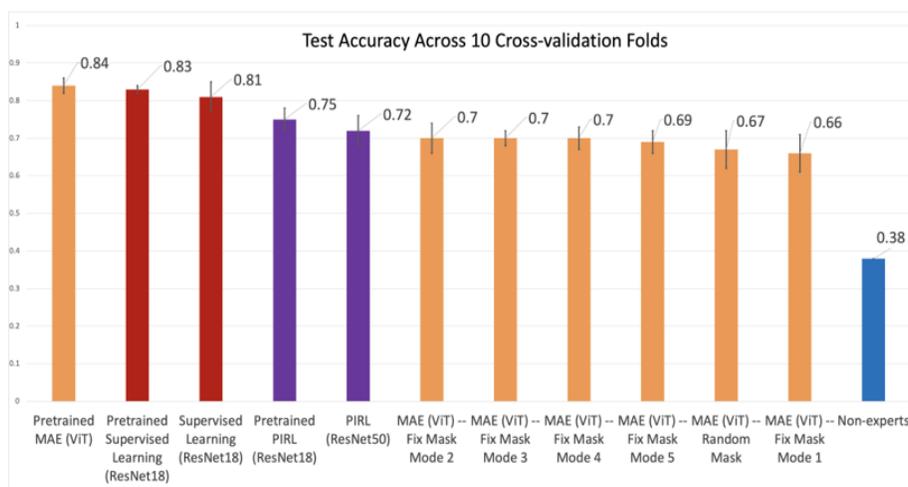
Figure 8: Corrosion Classification Results show MAE, a generative self-supervised learning model, outperforms other types of deep learning models if with transfer learning

By guiding generative self-supervised learning with material science knowledge and expert annotations, our methods aim to revolutionize corrosion regression, thereby enabling a deeper understanding of material performance under corrosion conditions and accelerating the discovery of novel corrosion-resistant materials. Our results using MAE with ordinal regression loss compared to cross-entropy loss can be seen in Table 2.

Augmentation	Parameters	ResNet-18	ResNet-50	DenseNet	HRNet
None	N/A	0.78 ± 0.03	0.72 ± 0.03	0.79 ± 0.01	0.76 ± 0.04
Color Jitter	Prob. 25%, Brightness (1.5, 2), Contrast (0.5, 1.5), Hue 0.5, Saturation (0.5, 1.5)	0.79 ± 0.03	0.74 ± 0.03	0.76 ± 0.02	0.68 ± 0.04
Gaussian Blur	Prob. 75%, Kernel 11, Sigma 5	0.75 ± 0.04	0.71 ± 0.05	0.74 ± 0.03	0.75 ± 0.02
Horiz. Flip	Prob. 25%	0.74 ± 0.03	0.70 ± 0.03	0.78 ± 0.03	$\mathbf{0.77 \pm 0.04}$
Rand. Erasing	Prob. 25%, Max Attempt 5, Area Ratio (0, 0.05)	0.78 ± 0.04	0.74 ± 0.04	$\mathbf{0.80 \pm 0.02}$	0.75 ± 0.02
Rand. Perspective	Prob. 75%, Distortion Scale 25%	0.76 ± 0.03	0.74 ± 0.04	0.74 ± 0.02	0.76 ± 0.05
Rand. Resized Crop	Prob. 25%, Scale (0.3, 0.7)	0.77 ± 0.03	0.76 ± 0.04	0.77 ± 0.03	$\mathbf{0.77 \pm 0.04}$
Rand. Rotation	Prob. 75%, Degrees (-25, 25)	0.77 ± 0.03	0.69 ± 0.03	0.74 ± 0.04	0.75 ± 0.03
Vert. Flip	Prob. 50%	0.75 ± 0.03	0.72 ± 0.04	$\mathbf{0.80 \pm 0.02}$	0.76 ± 0.04
Rand. Crop	Prob. 50%, Padding 4, Padding Mode Constant	$\mathbf{0.81 \pm 0.04}$	0.74 ± 0.02	0.79 ± 0.02	$\mathbf{0.77 \pm 0.03}$
Combination	Same settings	$\mathbf{0.81 \pm 0.04}$	$\mathbf{0.77 \pm 0.03}$	$\mathbf{0.80 \pm 0.03}$	0.76 ± 0.04
Pretrained + Combination	Same settings	$\mathbf{0.83 \pm 0.01}$	0.76 ± 0.02	0.79 ± 0.04	$\mathbf{0.83 \pm 0.03}$

Table 1: Test accuracy comparison using 10-fold cross-validation.

Method	Mean Absolute Error (MAE)
MAE with Ordinal Regression Loss	84.99 ± 2.64
MAE with Cross-Entropy Loss	83.662 ± 2.17

Table 2: Current results – Comparison of MAE using different loss functions.

3.4 Conclusion

With our data set in Part I, we demonstrate that we can leverage deep learning techniques to automate corrosion assessment. We demonstrate that image augmentation methods can be tuned to our data achieving 0.83 classification accuracy in corrosion assessment. Our longer-term goal is to build quality assessment models and integrate the assessment model with standardized experimental procedures to speed up experimental workflows. Our data set will drive innovation and development of deep learning techniques such as generative models for corrosion progression prediction and new representation learning techniques for small data sets – over time bridging computer vision and material innovation.

Part III

Learning Segmentation via Ground-truth Edges

4 Corrosion Ordinal Regression in Generative Self-Supervised Learning with Intrinsic Long-tail Data (Task 3)

4.1 Background

Corrosion is defined as the gradual degradation of a metal over time due to chemical interactions with its environment. It results in major safety risks worldwide and negatively impacts the environment, societal health, national infrastructure, manufacturing, and transportation. The associated economic burden is substantial, with global losses estimated to be around 4% of the gross domestic product (GDP), equivalent to approximately $2.5 trillion [20, 21]. Consequently, studying corrosion is an active research field in material science that aims to design new materials capable of preventing corrosion. This involves conducting corrosion tests by various industries, government agencies, and countries [22–28].

However, it is challenging to assess corrosion progression precisely, quickly, and safely. This impedes the understanding of corrosion and material discovery in general. Traditional manual segmentation methods to understand and assess corrosion are time-consuming, labor-intensive, and prone to errors. Therefore, an automated corrosion segmentation tool using Machine Learning (ML) algorithms is highly needed. However, this is hindered in the AI and computer vision communities due to the lack of high-quality, scientific, corrosion segmentation data. Obtaining such data is hard due to domain reasons including varied corrosion patterns on invented materials, complex expert-dependent annotations, water stains with similar colors or textures to corrosion, and millimeter-level observations, especially when defining the boundaries of corrosion. For example, shown
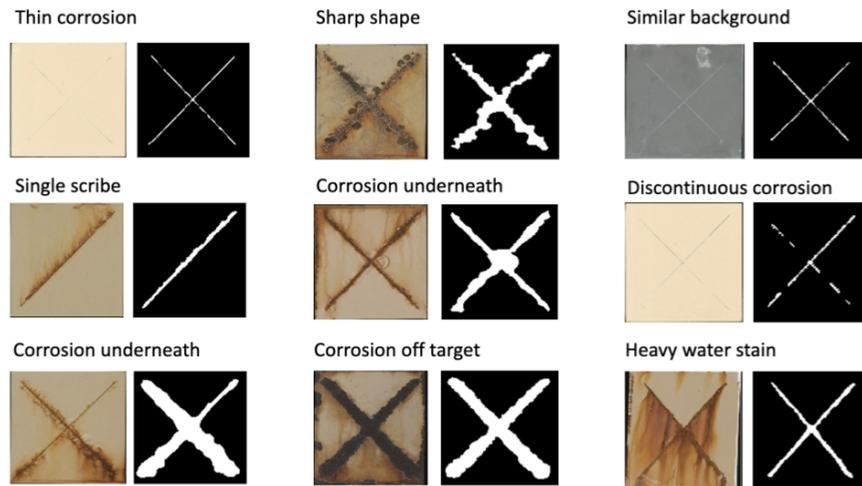
Figure 9: Sample corrosion images with the segmentation ground truth we provide with corrosion experts.

in Figure 9, corrosion underneath, i.e. corrosion under coating, has not broken through the panel's topcoat but the measured corrosion of interest should include the area. To assess this, the expert needs to scrape the corresponding coating out, which is labor-intensive and may result in a harmful environment. Other examples like thin or discontinuous corrosion show visually apparent corrosion to be segmented where the knowledge and effort of experts are highly needed to distinguish these small areas where corrosion exists. In addition, experts only assess corrosion along the scribe in a corrosion test while corrosion in other areas is not of interest although their textures or colors may be the same.

Data and Ground Truth In corrosion science, experimentalists develop and validate new corrosion-inhibiting materials following standard material science procedures. Each panel is scribed, exposed, scraped (if necessary), and assessed for its amount of corrosion present as defined in the ASTM standard [26]. Additional details on the procedure can be found in [29], including access to the 600 original images of corroded panels for scientific corrosion classification.

Ground truth, expertly curated, binary segmentations are produced for each of the 600 images in this work. To obtain the ground truth segmentations, we use the OpenCV Grab-

Cut algorithm to segment out scribe corrosion areas of interest from the background [30]. In consultation with domain experts, we develop criteria on how to segment scribed areas of corrosion such that we include all necessary areas of corrosion and exclude superfluous background pixels. The binary segmentations we obtained were reviewed by corrosion domain experts for their accuracy and refined based on an iterative feedback methodology.

4.2 Challenges

From examples shown in Figure 9, challenges that are both domain-specific and which extend to the general computer vision community [31], include i) a variety of textures and shapes in limited images, ii) tiny-scale and detailed areas to segment, iii) similarities in color with coating or image artifacts such as water staining, iv) areas of corrosion not visibly apparent on the images but identifiable to domain experts, and v) class imbalance between corrosion and background pixels, which arises naturally due to the relatively small size and irregular distribution of interest, such as any item above. These challenges lead us to think:

- To what extent can a deep learning architecture be trained to learn and recover these domain-specific and challenging corrosion areas?
- How to further alleviate these challenges in the deep learning architecture by exploiting strategies tied to these observations, such as edge detection or class imbalance methods?

4.3 Methodology

We establish baseline segmentation performance using UNet [7] – a deep learning architecture widely used and proven successful for machine segmentation tasks on small data. We observe that UNet is not fully learning sharp-edge information of small-scaled corrosion crucial for corrosion understanding and assessment. This motivates us to enhance

UNet with edge maps generated by deterministic computer vision techniques such as the Canny edge operator [1].

Thus, we develop an edge-guided baseline architecture, ET-UNet. This architecture inherits the backbone from UNet [7] and edge-guidance modules from ET-Net [32]. This represents a second baseline comparable with the UNet baseline. Although ET-UNet may guide UNet to learn low-level features of corrosion edges in early encoder layers, it may not provide direct guidance that could further verify the edge of segmentation prediction; we aim to address this with our proposed strategies.

Adding onto the ET-UNet baseline, we thus propose combinations of 3 different strategies to overcome challenges of difficult edges and class-imbalance for scientific corrosion segmentation. Accordingly, our three strategies are: 1) targeted edge guidance on ET-UNet from segmentation prediction, 2) soft involvement to further regularize the model to predict segmentation rather than the guidance itself, and 3) class-balanced loss, Focal Tversky Loss (FTL), to better distinguish challenging and minority corrosion of interest. Models reflecting these strategies and their combinations are listed in Table 3.

4.3.1 Innovative Loss to Achieve Ground-truth Edge Guidance via Shapes in Decoder

Our strategies are added to our DeepSC-Edge models listed in Table 3. We define the loss functions of these deep learning models below. Combining Binary Cross Entropy (BCE) and Dice loss ($Dice\_Loss$) [7], UNet loss function is calculated by ground truth segmentation Y and the segmentation prediction $\hat{Y}$:

$$\mathcal{L}_{UNet} = BCE(Y, \hat{Y}) + Dice\_Loss(Y, \hat{Y}) \tag{1}$$

---

[1] https://kornia-tutorials.readthedocs.io/en/latest/_nbs/filtering_edges.html#canny-edges

Other than UNet, ET-UNet also learns edges of ground truth segmentation in early encoder layers so that its loss function can be described as:

$$\mathcal{L}_{ET-UNet} = \mathcal{L}_{UNet} + \lambda \cdot \mathcal{L}_{edge} \tag{2}$$

$$where\ \mathcal{L}_{edge} = BCE(Yedge, \hat{Y}edge)$$
$$+ Dice\_Loss(Yedge, \hat{Y}edge), \tag{3}$$

In Equation (3), $Yedge$ is the edge map of ground truth segmentation, $\hat{Y}edge$ is the predicted edge map from its Edge Guidance Module illustrated in Figure 10, and λ controls the strength of this edge guidance. We then involve the additional edge guidance shown in Table 3:

$$\mathcal{L}_{edge^*} = BCE(Yedge, \hat{Y}edge^*)$$
$$+ Dice\_Loss(Yedge, \hat{Y}edge^*). \tag{4}$$

where $\hat{Y}edge^*$ denotes the edge map of segmentation prediction output from the Weighted Aggregation Module of ET-UNet – instead of $\hat{Y}edge$ that is output from the Edge Guidance Module as shown in Figure 10. We control this additional guidance using γ. So that:

$$\mathcal{L}_{Guided\ ET-UNet} = \mathcal{L}_{ET-UNet} + \gamma \cdot \mathcal{L}_{edge^*} \tag{5}$$

$$\mathcal{L}_{S-Guided\ ET-UNet} = \mathcal{L}_{ET-UNet}$$
$$+ \frac{1}{\max{(\mathcal{L}_{seg} - \theta,\ 0)} + 1} \cdot \mathcal{L}_{edge^*} \tag{6}$$

Model	Description (Equation)	Method
UNet	UNet baseline for image segmentation (1)	Baseline 1
ET-UNet	Edge guidance architecture based on UNet (2)	Baseline 2
Guided ET-UNet	ET-UNet (2) + additional edge guidance (5)	Proposed Strategy 1
S-Guided ET-UNet	ET-UNet (2) + additional edge guidance in a soft manner (6)	Proposed Strategy 1, 2
UNet + FTL	UNet (1) + class balanced loss (7)	Baseline 1 + Proposed Strategy 3
ET-UNet + FTL	ET-UNet (2) + class balanced loss (7)	Baseline 2 + Proposed Strategy 3
Guided ET-UNet + FTL	ET-UNet (2) + additional edge guidance (5) + class balanced loss (7)	Proposed Strategy 1, 3
S-Guided ET-UNet + FTL	ET-UNet (2) + additional edge guidance in a soft manner (6) + class balanced loss (7)	Proposed Strategy 1, 2, 3

Table 3: Model descriptions: UNet-based models that reflect our methods combining baselines and proposed strategies. Strategy 1 or 2 is proposed upon ET-UNet. DeepSC-Edge A is Guided ET-UNet. DeepSC-Edge B is S-Guided ET-UNet.

By assigning θ appropriately, we involve the additional guidance, $\mathcal{L}_{edge^*}$, in the following soft (S) manner: the prediction edge, $\hat{Y}edge^*$, will provide its guidance in inverse proportion to $\mathcal{L}_{seg}$, and the full guidance to the prediction segmentation only if $\mathcal{L}_{seg}$ is smaller than or equal to θ. For Strategy 3, class balanced loss FTL [33] combines Tversky Index [33] and Focal loss [15]:

$$FTL = (1 - TI)^{1/\beta} \tag{7}$$

$$where\ TI = \frac{TP + \varepsilon}{TP + \alpha FN + (1 - \alpha)FP + \varepsilon} \tag{8}$$

We embed FTL loss into our models (+FTL) shown in Table 3 instead of their $Dice\_Loss$. In Equation (8), we calculate True Positive (TP), False Negative (FN), and False Positive (FP) pixels according to ground truth and prediction in $Dice\_Loss$ used in the model. If α is larger than 0.5, the loss penalizes FN, so the model will emphasize learning these challenging corrosion areas more. β further down-weights the contribution of corrosion that is easy to predict and focuses more on challenging corrosion during training. In this way, the model would learn to preserve sharp, minority, or challenging edges rather than avoid them, and as a result, this would make the prediction of edges smooth.

4.4 Experiments and Results

Experimentation. Using one Nvidia A100 GPU, we trained and evaluated all the models on the dataset consisting of 600 corrosion image pairs: the original images and their corresponding ground truth segmentation we provide in this work. To grid-search hyper-
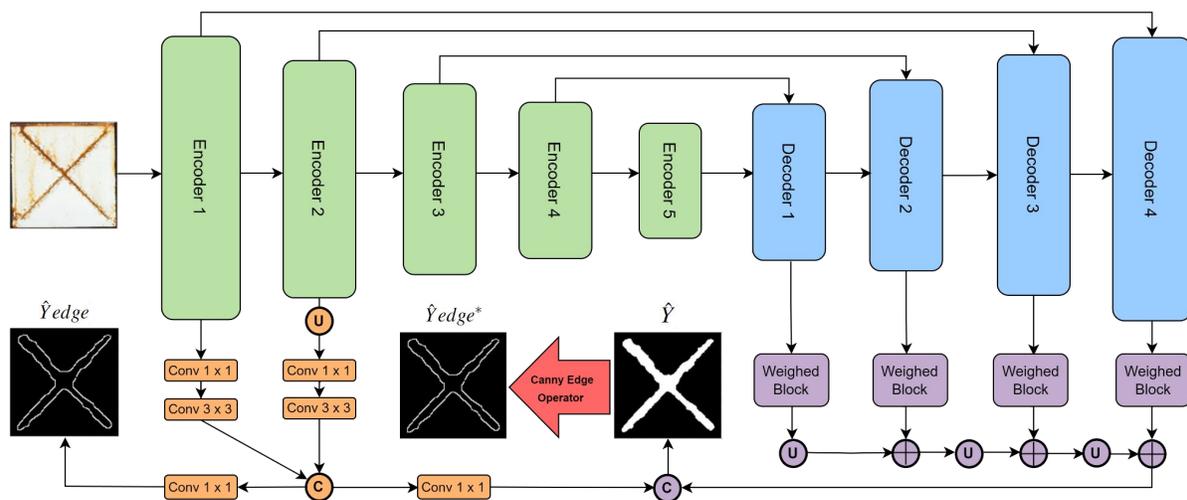
Figure 10: Guided ET-UNet: UNet encoder layers are in green, while UNet decoder layers are highlighted in blue. Edge Guidance Module [32] is in orange. Weighted Aggregation Module [32] is highlighted in purple. 'Conv', 'U', 'C', and '+' signify the convolutional layer, upsampling, concatenation, and addition layers, respectively. Red Arrow illustrates the deterministic canny edge operator. Our Strategy 1 develops a loss (Equation 5) to train this new decoder edge guidance with traditional encoder edge guidance loss (Equation 2) upon UNet loss (Equation 1) for segmentation. Our Strategy 2 proposes a novel loss (Equation 6) to regularize the proposed Guided ET-UNet loss so not to overfit to edges via weighting it with the segmentation loss (Equation 1). The decoder edge guidance in Guided ET-UNet is expected to gradually increase if it encourages the overall corrosion segmentation. However, this guidance will not be effective once the segmentation has been fully learned, as the weight will be 0 if the segmentation loss is smaller than a hyperparameter $theta$. In our Strategy 3 (similar to Strategies 1 and 2), we integrate a class imbalanced method (Equation 7) into the segmentation and related edge guidance methods for telling corrosion with challenging but critical boundaries.

parameters in UNet, we select learning rates from a set {1e-1, 1e-2, 1e-3, 1e-4, 1e-5} and batch size from a set {8, 16}. We found that all of the models work the best with learning rate 1e-3 and batch size 8. The best λ in ET-UNet (2) is 0.5 tuned from a set {0.001, 0.1, 0.5, 1.0, 1.5}. The best γ in Guided ET-UNet (5) based on the best ET-UNet is 1.0 tuned from a set {0.001, 0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0}. The best θ in S-Guided ET-UNet (6) based on the best ET-UNet is 3.0 tuned from a set {1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0}. For Proposed Strategy 3 on all the above best models, we found the best penalization weight α in FTL is 0.88 for ET-UNet + FTL and S-Guided ET-UNet + FTL models, and 0.90 for other models related – both showing a high false negative rate penalization is beneficial in the dataset. We also present an ablation study – showing our results are robust to hyperparameter choice in the proposed strategies.

We evaluate the model performance using Dice and IOU scores as broadly used in machine segmentation tasks [7]. To better observe the evaluation, we also define their variants other than scores – rank, increase, and weighted increase (W-Inc as shown in Table 4). The Dice or IOU rank takes the average rank performance of a model after calculating its rank compared with other models on every test image. The Dice or IOU increase takes the average score improvement of a model rather than the UNet baseline across the test set. The Dice or IOU weighted increase (W-Inc) takes the weighted average score improvement of a model rather than the UNet baseline after weighting each test image with its prediction rank using the UNet baseline. In this way, we evaluate how a model is or is not able to predict challenging corrosion of interest better. If W-Inc is larger than 0, it indicates that the model outperforms UNet in predicting corrosion segmentation, especially on challenging corrosion.

Shown in Table 4, our results indicate that all of our strategies can improve baseline models. S-Guided ET-UNet (+FTL) model performs the best based on its Dice, IOU, and corresponding ranks and increases, demonstrating its superior performance in solving the corrosion segmentation task compared to other strategies and baselines, especially

Model	Dice				IOU			
	Score ↑	Rank ↓	Increase ↑	W-Inc ↑	Score ↑	Rank ↓	Increase ↑	W-Inc ↑
UNet	0.8672 ± 0.0179	5.5167	0.0000	0.0000	0.7727 ± 0.0263	5.5167	0.0000	0.0000
ET-UNet	0.8670 ± 0.0181	5.6667	-0.0002	0.0014	0.7721 ± 0.0268	5.7000	-0.0006	0.0014
Guided ET-UNet	0.8687 ± 0.0176	5.3167	0.0015	0.0028	0.7749 ± 0.0260	5.2500	0.0022	0.0040
S-Guided ET-UNet	0.8696 ± 0.0199	4.6333	0.0024	0.0045	0.7763 ± 0.0288	4.6167	0.0036	0.0062
UNet + FTL	0.8738 ± 0.0152	4.0167	0.0066	0.0149	0.7823 ± 0.0226	4.0500	0.0096	0.0133
ET-UNet + FTL	0.8763 ± 0.0151	3.6833	0.0091	0.0094	0.7855 ± 0.0227	3.6833	0.0128	0.0214
Guided ET-UNet + FTL	0.8768 ± 0.0168	3.8500	0.0096	0.0155	0.7864 ± 0.0254	3.8667	0.0137	0.0232
S-Guided ET-UNet + FTL	0.8772 ± 0.0140	3.3167	0.0100	0.0166	0.7870 ± 0.0212	3.3167	0.0143	0.0241

Table 4: Test performance using 10-fold cross-validation: each cell shows the average test performance of a model under a metric. The best-performing cell in each metric column is highlighted with a red background.
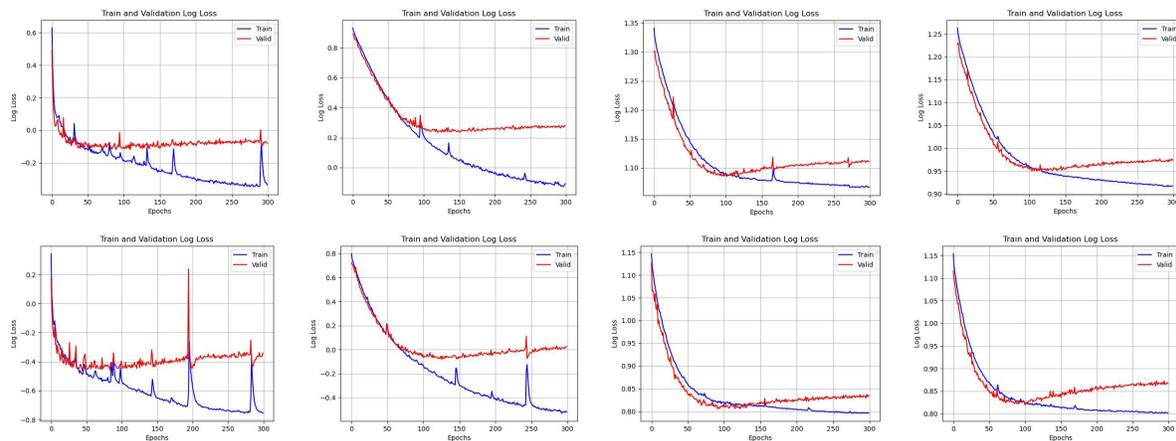
Figure 11: Loss plots on the best UNet validation fold: Top row, from left to right, shows UNet, ET-UNet, Guided ET-UNet, and S-Guided ET-UNet. The bottom row shows the corresponding models with FTL.

on challenging corrosion. Furthermore, incorporating any of our proposed strategies elevates ET-UNet performance over UNet – showing its effectiveness in edge-guidance. This is a notable improvement, considering that UNet is commonly used in the field of image segmentation with limited data. Moreover, from Weighted Increase performance for handling the difficult-to-segment corrosion, ET-UNet is able to outperform UNet without strategies, but they are indeed feasible to make this improvement larger. In addition, from the loss plots in Figure 11, the models with FTL tend to have smaller total loss values, and our best-performing model shows smooth training and validation losses – denoting the robustness.
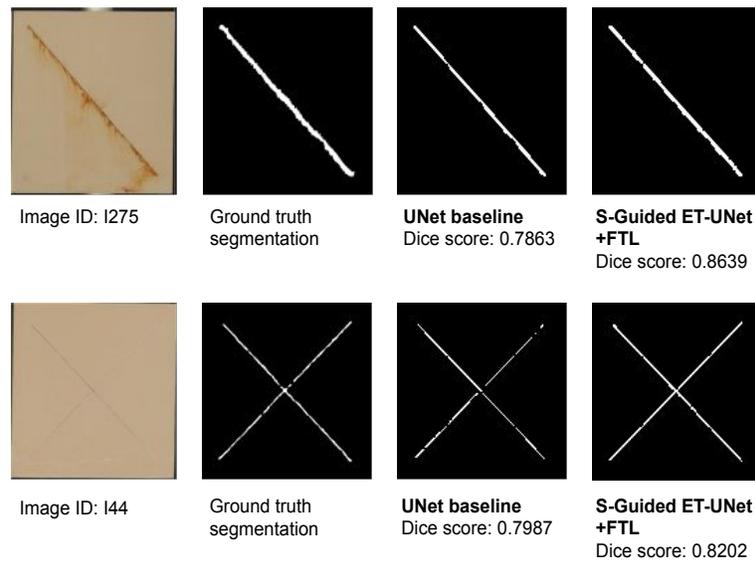
Figure 12: Case study: Challenging corrosion panels segmented by UNet and our proposed S-Guided ET-UNet + FTL model.

Model (+ FTL)	$\beta = 1.0$				$\beta = 1.5$				$\beta = 2.0$			
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 0.9$
UNet	0.7346	0.8322	0.8737	0.8632	0.7434	0.8223	0.8741	0.8717	0.7547	0.8072	0.8614	0.8669
ET-UNet	0.7891	0.8377	0.8740	0.8672	0.7513	0.8479	0.8711	0.8741	0.7927	0.8493	0.8572	0.8646
Guided ET-UNet	0.7506	0.8359	0.8763	0.8679	0.7461	0.8336	0.8721	0.8768	0.7582	0.8109	0.8557	0.8687
S-Guided ET-UNet	0.7346	0.8276	0.8758	0.8654	0.7706	0.8227	0.8755	0.8766	0.7777	0.8173	0.8659	0.8711

Table 5: Ablation Study to Proposed Strategy 3 – class-balanced loss in models. Each cell shows the test performance using Dice Score metric.

Ablation Study. From Table 4, the class-balanced loss, (7), shows a significant improvement to our task. This strategy forces the model to learn minority pixels better by penalizing false negatives or false positives while predicting segmentation or its edge. It helps us to solve the class imbalance challenge where corrosion of interest, in the minority, is visually similar to any other pixels such that it is hard to segment – via controlling α and β in the loss. We show an ablation study to these hyperparameters in Table 5. From the table, we conclude that our hyperparameter choice is robust.

Case study. Providing visual examples of the segmentation predictions from the baseline models and proposed strategies can be a useful way to explain why our strategies are necessary for corrosion segmentation when leveraging the popular UNet deep learning architecture. In Figure 12, we display test results of the baseline UNet using its best vali-

dation fold model and corresponding predictions from our best-performing model. This comparison highlights the benefits of our proposed strategies in challenging corrosion segmentation tasks. Overall, these visual examples provide an intuitive understanding of our proposed strategies and their potential impact on improving the performance of scientific corrosion segmentation. More examples can be viewed in our supplementary document.

4.5 Contributions

In summary, our key contributions in this work are as follows:

- Novel soft method enabling edge guidance deep learning loss enabling Decoder to learn edges by solving a general model issue from this Decoder rather than existing Encoder guidance – over-segmentation to edges.

- Uniquely apply the class-imbalanced loss to our novel edge guidance method learning segmentation with challenging edges.

- Our method's performance, particularly on images with challenging edges, is affirmed by the "Rank" metric.

- Expert-curated corrosion segmentation dataset in material science showing CV challenges.

- We exploit UNet models with the aid of ground-truth edges and a class-balanced loss – overcoming the identified challenges.

- We open rich research opportunities with our dataset and strategies for material science discovery, deep learning, and computer vision.

Part IV

GAN for Unseen Alloy Discovery

5 Promptable Generative Models to Discover Microstructure of Unseen Alloys Given Prior Chemical Knowledge (Task 4)
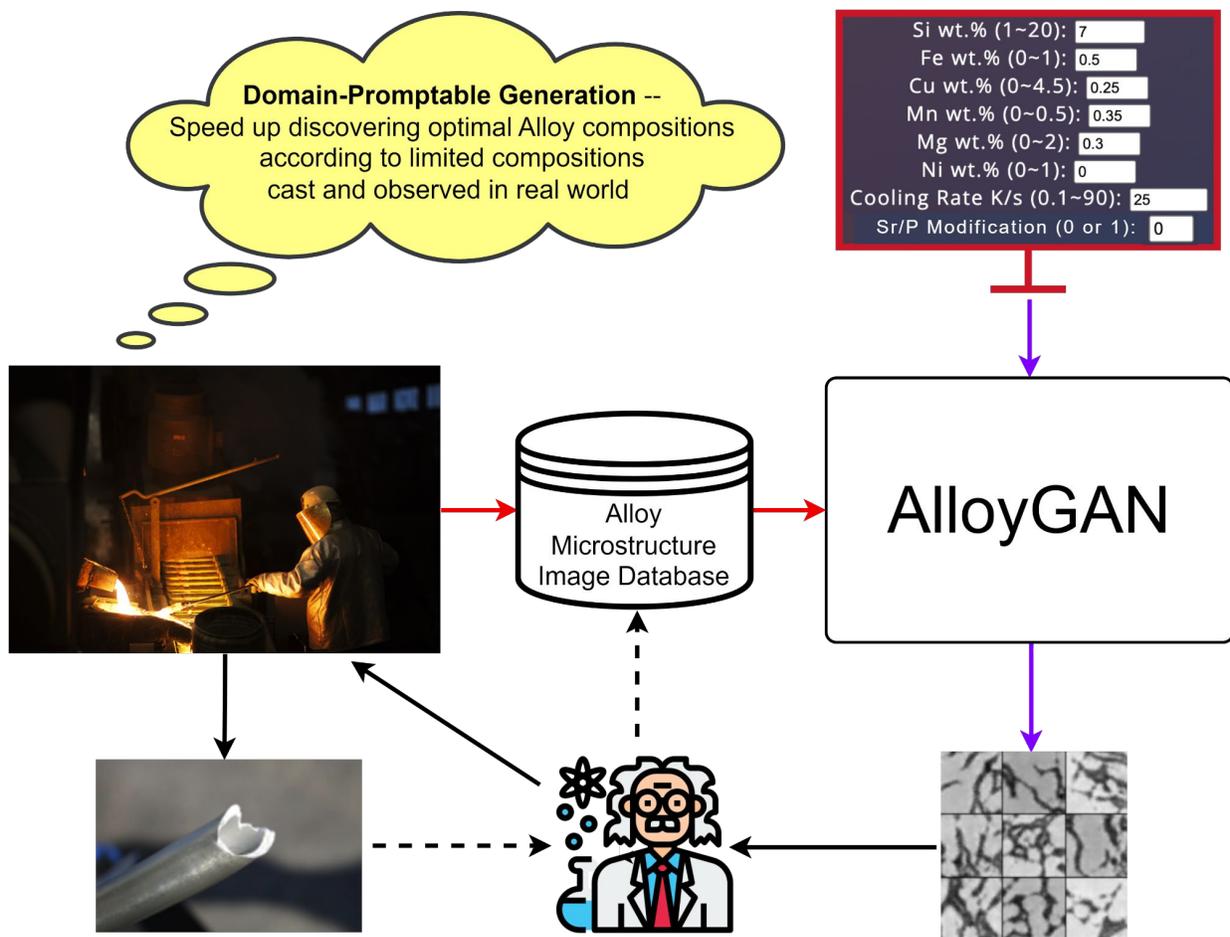
Figure 13: Our breakthrough material simulation model using deep neural networks, AlloyGAN, published at ICMLA 2023: Promptable generation algorithm serving a novel scientific alloy discovery platform built by an AIGC startup [2]. Red arrows indicate model training, purple for model inference, black solid for material science verification, and black dash for data collection.

5.1 Background

The global metal market forms a critical backbone for several economies, largely driven by the demand from sectors such as construction, automobiles, aerospace, electronics, and more. Central to this burgeoning market are alloys—a blend of two or more metals or a metal and another element—recognized for their diverse and enhanced properties. Alloy manufacturing, thereby, stands as a cornerstone in the metal market, which is projected to surpass a staggering $18.5 trillion by 2030 [34].

5.2 Motivation

Alloy production, while critical, faces substantial challenges, making it a subject of intense scrutiny and innovation. A striking pain point in the industry is the high rate of rejected metal casting products, which must be re-melted and re-cast due to various defects. Annually, the industry sees tens of millions of tons of metal casting products fall into this cycle [35]. This high rate of rejection and reprocessing not only introduces an alarming level of waste but also incurs prohibitive manufacturing costs, placing a considerable economic burden on the industry. Thus, there is tremendous value if efficient models that generate trustworthy results for accelerating scientific Alloy discovery and manufacturing could be found.

5.3 Challenge

Traditional numerical methods have struggled with the complex solidification process in alloy formation, characterized by vast nonlinear chemical and physical interactions [36–40]. This complexity challenges conventional mathematical modeling attempting to accurately simulate the final microstructure based on the basic Alloy compositions shown in Figure 13. Additionally, these models are intricate, computationally demanding, and require specific knowledge of material science, limiting their accessibility to the broader

research community [41–46].

5.4 Problem Definition

Given the limitations of traditional methods and the complexity of the problem at hand, a well-defined problem emerges, namely, there is an urgent need for a method that can effectively generate plenty of microstructure images of metal alloys based on initial processing conditions like chemical compositions. Such a method should overcome the limitations of traditional models, be capable of handling the intricacies of the solidification process, and require less intensive data collection efforts. It should also offer robust accurate results while being computationally efficient and capturing fundamental physical properties.

5.5 State-of-the-art

Conventional Alloy Microstructure Modelling Approach. Traditional alloy microstructure modeling demands significant resources and expertise in solving intricate Partial Differential Equations (PDEs) over analytical models and Finite Element Method (FEM) [36–40]. Phase Field modeling requires extensive computation – 768 GPUs for 12 days to compute a voxel cell [41], while the Monte Carlo method struggles with aligning simulation and physical time [42, 43]. Cellular Automation (CA) assists in simulating recrystallization and solidification [42, 43] but faces heavy domain challenges with complicated metallurgical processes and transition rules [46]. Despite their mathematical and physical clarity, these conventional methods may fall short in providing prompt, scientifically robust decisions in large-scale alloy discovery.

Deep learning in material generation. The application of deep learning methods, such as Variational Autoencoders (VAEs) and Convolutional Neural Networks (CNNs), in microstructural analysis is in its nascent stages [47–51]. Although the application of Generative Adversarial Networks (GANs) and Conditional Generative Adversarial Networks

(cGANs) in material science has gained some traction in modeling and augmenting microstructural images [52–59], their application is still largely limited in handling basic alloy compositions, like chemical elements, for creating scientifically valid microstructure images. While there has been an initial exploration of deep learning methodologies for material generation, these efforts have not yet effectively addressed the complexity of material science. This void emphasizes the need for innovative solutions capable of dealing with the complex, nonlinear dynamics of alloy microstructure formation and delivering robust generations to accelerate scientific Alloy discovery.

5.6 The Methodology: AlloyGAN

To bridge this gap, we propose AlloyGAN, a ground-breaking approach leveraging the prowess of deep learning to create scientifically valid alloy microstructure images from basic alloy compositions (See Figure 13). Augmented with unique adaptations that incorporate prior knowledge from solidification reaction factors, AlloyGAN, as a chemically-constrained cGAN architecture, successfully conditions the generation process related to the basic alloy compositions involved, thereby successfully simulating the complex solidification process to verify Alloy properties.

Illustrated in Figure 5.7, both the AlloyGAN Generator and Discriminator are equipped with a core module where prior chemical reaction factors can be applied to the basic compositions of alloy. This injects prior knowledge with domain-specific conditions into the Generator and Discriminator pair, providing AlloyGAN the capability to generate alloy microstructure images with rich informative chemical properties.

5.6.1 The AlloyGAN Generator

The AlloyGAN Generator concatenates two different types of inputs: a random noise tensor and a condition tensor. The noise vector introduces a degree of randomness into the generation process following a standard normal distribution. The condition tensor re-

Alloy ID	Si	Fe	Cu	Mn	Mg	Ni	Cooling Rate	Sr/P
A356	7	0.5	0.25	0.35	0.3	0	2.5, 10, 57, 143	Yes, No
A360	9.5	0.6	0.1	0.05	0.5	0	2.5, 10, 57, 143	No
A369	11.5	1	0.5	0.25	0.4	0.05	2.5, 10, 57, 143	Yes, No
A339	12	1.2	2	0.5	1	1	2.5, 10, 57, 143	No
A393	22	1.3	0.9	0.1	1	2.3	2.5, 10, 57, 143	Yes, No
A355	5	0.65	1.25	0.55	0.5	0	2.5, 10, 57, 143	No
A308	5.5	0.8	4.5	0.5	0.1	0	2.5, 10, 57, 143	No
A319	6	1	4	0.4	0.55	0.35	2.5, 10, 57, 143	No
A332	9.5	0.9	3	0.5	2.1	0.5	2.5, 10, 57, 143	No

Table 6: Summary of labels: Chemical composition with experienced manufacturing environments that together determine Alloy microstructure in our dataset.

ceives prompted user inputs of basic alloy compositions containing the amount of chemical elements and the manufacturing settings. Before concatenating, these conditions will be processed with prior deterministic factors from the chemical reaction of the alloy solidification process. To produce the desired realistic images, the processed conditions are randomly selected and concatenated with the noise tensor before passing through several transposed convolutional layers. This ensures the prompted conditions upon priors to better fool the Discriminator so that the generated images can be discriminative according to the conditions upon priors. The convolutional layers include batch normalization and leaky ReLU activation functions to stabilize the training process and prevent overfitting. The output of these layers is a 2-dimensional tensor in the same shape as our target binary microstructure image.

5.6.2 The AlloyGAN Discriminator

Illustrated in Figure 15, the Discriminator of AlloyGAN serves as a classifier that tries to distinguish between real and generated data. It takes in both real images with their corresponding conditions upon priors and the generated images produced by the Generator and assigns a probability that a generated image can be predicted as a real image.
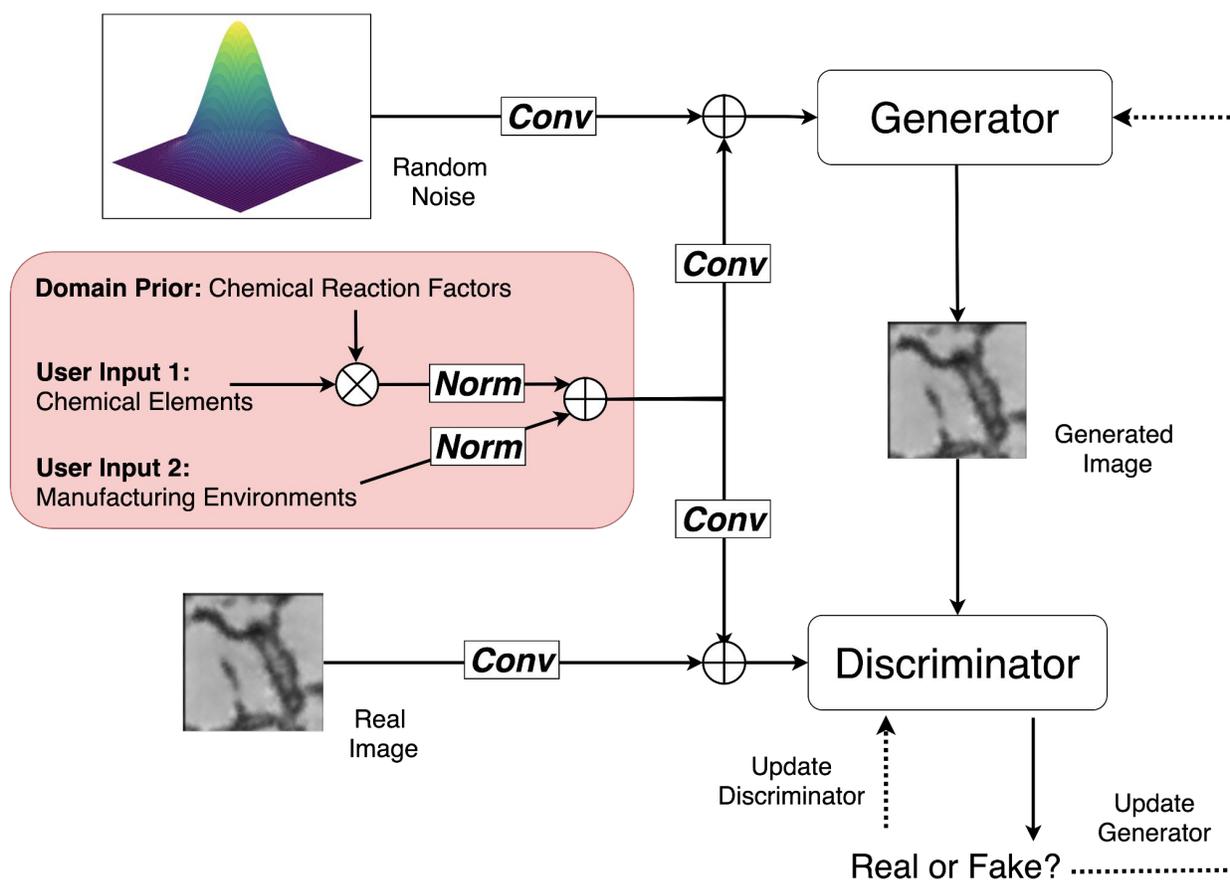
Figure 14: AlloyGAN: The pink-colored module illustrates how we incorporate the prompted inputs with the domain prior knowledge for cGAN enhancements for this domain-specific task. The resulting conditions are concatenated with the respective images as additional channels in the workflow. Conv indicates Convolutional layers, while Norm indicates Max-Min normalization on features.

5.6.3 Prior from Chemical Reaction in Alloy Solidification

Incorporating targeted domain-specific prior knowledge into the prompted conditions differentiates AlloyGAN from a standard cGAN. Illustrated in the pink subnet in Figure 5.7, the integration of the quantifiable prior with the prompted alloy composition conditions is formulated as a concatenation of two types of normalized features:

1. **Features of Chemical Elements:** This feature subset accounts for the influence of each chemical element on the generated alloy by multiplying its quantity with a conversion factor in Table 7. This factor represents the expected impact of each el-

ement on the solidification process critical for the resulting alloy properties. The features were then normalized with a unified scale from 0 to 1 for training stability using max-min normalization.

2. **Features of Manufacturing Environments:** This feature subset represents the conditions under which the alloy is manufactured. Key factors like the cooling rate and Sr/P modification are taken into account. Similar to the chemical features, these manufacturing features undergo a max-min normalization process. This step accounts for potential variations in manufacturing environments.

The above intricate formulations of the prompted alloy compositions with the chemical reaction prior factors capture domain-specific alloy chemical and manufacturing conditions that are critical for AlloyGAN to generate microstructure images taking the complex interplay among these conditions into consideration. This is why conditions not previously in our input data set could be inferred and thus meaningfully generated by our trained AlloyGAN with limited data.

Domain-driven Chemical Element Normalization. The standard cGAN individually normalizes each raw input feature shown in Table 6. However, this approach does not consider the relative contributions of each element to the alloy. For example, the Si element greatly changes the alloy microstructures, whereas other elements have less impact. This may be because their content is relatively low (less than 4.5 wt.) compared with Si. Moreover, the chemical reaction outcomes of these elements with aluminum need also be considered. Table 7 shows the outcome of chemical reactions of each element to aluminum and the factor of converting a unit of weight percentage of elements to the atomic percentage of outcomes. The conversion factor embodies an indicator of the impact of elements on the microstructure metallograph. This is what we thus propose to use as a prior to prompted chemical composition conditions.

Element	Si	Fe	Cu	Mn	Mg	Ni
Outcome	Si	Al3Fe	Al2Cu	Al12Mn	Mg2Si	Al3Ni
Factor	0.036	0.018	0.016	0.018	0.021	0.017

Table 7: Summary of chemical reaction outcome and factors in alloy manufacturing: a factor indicates the conversion factor of a unit of weight percentage of the Element to the amount of atomic percentage (at.) of the outcome – formulating the domain prior onto prompted conditions in AlloyGAN.

5.6.4 Training Process

AlloyGAN is trained using a two-step iterative process. In the first step, the generator creates a batch of synthetic microstructure images with the processed prior. The discriminator then evaluates the later. The generator's weights are updated to minimize the difference between the discriminator's output to distinguish the synthetic images from the real images. In the second step, the discriminator's weights are updated using a batch of real microstructure images from the training alloys and the synthetic images from the first step. This step aims to classify real versus generated images from random noise with conditions normalized with the prior.

This process is repeated for several epochs until the generator can produce synthetic microstructure images that the discriminator can no longer distinguish from real ones. The resulting trained model is then ready to generate new microstructure images for any given alloy composition — including those compositions that had previously not been unseen in the training process.

AlloyGAN training models a two-player minimax game with prior chemical reactions on the prompted alloy compositions for a robust, accurate, and fast generation. The Generator (G) tries to minimize this objective against an adversary D that tries to maximize it, i.e., $\min_G \max_D V(D, G)$. The prompted alloy compositions are represented in a conditioning vector y that is concatenated with G (as y_1 after $Conv$) or D (as y_2 after $Conv$) illustrated in Figure 5.7 and formulated as:

$$y = Concat\left(Norm(\mathbf{w} * \mathbf{I}_1), Norm(\mathbf{I}_2)\right) \tag{9}$$
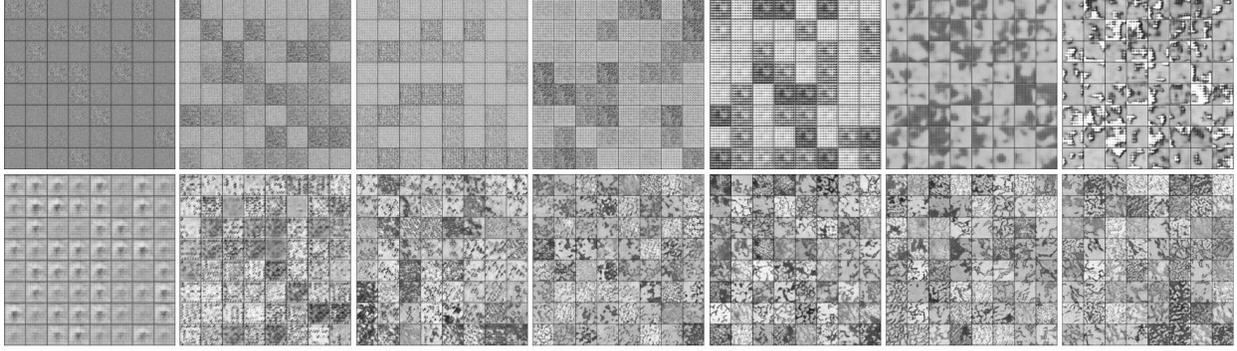
Figure 15: Example generated images in the training process. Top raw shows standard cGAN generations. The bottom row shows AlloyGAN generations. Each column shows the last batch of the epoch: 1st, 20th, 40th, 60th, 80th, 90th, and 100th. The random seed is the same for both models. Upon the same hyperparameters, including the random seed, AlloyGAN outperforms cGAN in generating informative microstructure images of alloys.

where Norm denotes min-max normalization, $\mathbf{I}_1$ User Input 1 to prompt the wt.% content of each chemical element, $\mathbf{I}_2$ User Input 2 to prompt manufacturing environments, and $\mathbf{w}$ Domain Prior from chemical reaction factors. The loss function of training AlloyGAN is:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x,y_1 \sim p_{\text{data}}(x,y_1)}[\log D(x|y_1)] +$$

$$\mathbb{E}_{z \sim p_z(z), y_2 \sim p_{\text{data}}(y_2)}[\log(1 - D(G(z|y_2)))] \quad (10)$$

In Eq. (1), x denotes the sample image, and y_1 is the conditioning vector capturing the domain prior illustrated as "Chemical Reaction Factors" in Figure 5.7. z is a noise vector. $D(x|y_1)$ corresponds to the discriminator's estimate of the probability that real alloy microstructure image x, given y_1. $G(z|y_2)$ is the fake alloy microstructure image by the generator, given z and y_2. The integration of the domain prior factors, $\mathbf{w}$, provides chances to further distinguish the importance of chemical composition prompts, $\mathbf{I}_1$, in AlloyGAN, thereby enhancing cGAN to generate applicable images with robust performance.

5.6.5 Evaluation Metrics

We evaluate AlloyGAN and cGAN performance in this study using FID [**?**], a widely-used machine learning metric. FID evaluates the performance of a generative model by looking at how close the generated data is to the real data in the embedding space of an object classification model (commonly using the InceptionV3 model in particular) [**?**]. We calculate FID between real and generated images on our test set so that we know how the model generates the microstructure images of previously not-trained alloys. To build the embedding space required for the calculation, we not only utilize the pre-trained InceptionV3 network upon the ImageNet dataset as a general space but also finetuned the network according to our domain train dataset as a domain-specific space for the evaluation.

We note, however, that general evaluation metrics in machine learning, such as FID, cannot measure the chemical properties of generated alloy microstructure images for scientific use. We thus also utilize three domain-science metrics to evaluate whether Alloy-GAN creates valid images that are feasible for fast material verification in alloy discovery.

Micrograph Evolution with Si Content. This metric measures whether the generated images show the trend that Si content in metallography increases with the addition of Si wt.% in alloys. The Si content in metallography can be quantifiable by calculating the area of fractions (black area after binarizing the image).

Effect of Cooling Rate on the Secondary Dendritic Arm Spacing. The theory of computational material science shows that there must be a correlation between the cooling rate and the Secondary Dendritic Arm Spacing in the images satisfying the following formula:

$$\lambda = k(CR)^{-n} \quad (n > 0), \tag{11}$$

where λ represents the secondary dendritic arm spacing measured by the distance be-

tween the neighborhood white areas separated by a black arm in the generated image and CR represents the cooling rate (K/s) in the manufacturing environment. For scientifically valid microstructure images of an alloy, there must be a strong linear relationship between $(CR)^{-n}$ and λ with a constant n and a coefficient k.

Modification of Strontium and Phosphor on the Alloy Microstructure. This metric measures whether Primary Al dendrites are branched after the modification. Sr/P is added to aluminum in very small amounts (<0.05 wt.%), yet they significantly impact the microstructures of aluminum alloys. Sr and P atoms act as "poisons" to the solidification of the Si phase, causing it to form more branched and rounded structures. Aluminum alloys with Sr/P modification typically exhibit microstructure images with dispersed dark particles. In contrast, alloys without Sr/P modification display larger and segregated dark Si particles.

5.7 Experimental Study

Our research results bear witness to the *transformative potential of AlloyGAN*. The technique has demonstrated the ability to quantify the impact of individual chemical elements on the microstructure's metallography. Notably, the accuracy and robustness of AlloyGAN's results stand on par with traditional computational material science methods. The key differentiator, however, lies in AlloyGAN's efficiency—our method accomplishes these outcomes with a significantly reduced computation time, offering a promising pathway for expedited material science research.

While evaluating AlloyGAN, we utilize *domain-specific metrics* assessing that the created images correctly show: 1) a trend that Si content in metallography increases with the addition of Si in alloys, 2) the effect of Cooling Rate on Secondary Dendritic Arm Spacing following a mathematical equation, and 3) the effects of Sr/P modification.

To compare AlloyGAN and standard cGAN performances, we evaluate their generated images on the same training hyperparameters default from standard cGAN. The

latent layer dimension was set to nz=100, allowing for a diverse range of features to be generated in the Generators. The Adam optimizer was utilized for both the generator and discriminator, with a beta value of 0.5. The training was conducted using a batch size of 64, a learning rate of 0.0002, and lasted for 100 epochs which are all recommended settings in standard cGAN. According to our dataset, the images used for training and testing are of size 128x128 pixels, with a single channel (nc=1). We use NVIDIA-A100 GPU with the same global random seed for all the GAN experiments. Moreover, in order to further tune and test the model performance, we tune AlloyGAN vs. standard cGAN by a combination of the epoch length (ep) from a set of {100, 1000} and latent layer dimension (nz) from a set of {30, 100}.

FID from Pretrained Inception Network ↓	ep = 100, nz = 30	ep = 100, nz = 100	ep = 1000, nz = 30	ep = 1000, nz = 100
AlloyGAN Overall test performance	804.08	746.15	1394.04	1550.11
cGAN Overall test performance	1948.17	1398.75	1455.63	1298.18
AlloyGAN Alloy-wise test performance	1022.15 ± 251.85	1001.22 ± 184.40	1650.96 ± 403.76	1769.98 ± 169.39
cGAN Alloy-wise test performance	2150.85 ± 226.71	1634.39 ± 231.64	1578.84 ± 288.56	1602.19 ± 151.22
FID from Finetuned Inception Network ↓	**ep = 100, nz = 30**	**ep = 100, nz = 100**	**ep = 1000, nz = 30**	**ep = 1000, nz = 100**
AlloyGAN Overall test performance	14.88	14.02	26.35	39.84
cGAN Overall test performance	68.01	38.00	38.18	46.82
AlloyGAN Alloy-wise test performance	29.83 ± 20.24	33.55 ± 22.32	43.35 ± 16.58	54.22 ± 22.80
cGAN Alloy-wise test performance	68.69 ± 21.61	43.51 ± 20.04	50.30 ± 10.90	60.49 ± 26.95

Table 8: We report one FID score over all test images as the Overall calculation method and mean $\pm$ standard deviation across different test alloys as the Alloy-wise calculation method. The FID score by the finetuned inception network shows AlloyGAN outperforms standard cGAN in every hyperparameter setting. The best AlloyGAN Alloy-wise test performance is under a small epoch with less latent dimension (ep = 100, nz = 30). The best AlloyGAN Alloy-wise test performance is under cGAN default hyperparameter setting (ep = 100, nz = 100). The best cGAN Overall or Alloy-wise test performance is under cGAN default hyperparameter setting (ep = 100, nz = 100).

AlloyGAN vs. cGAN. Figure 15 shows that AlloyGAN generates informative images that effectively represent the microstructure of the training alloys, unlike standard cGAN. We evaluate both models via machine learning metrics in Table 8 or material science metrics in Figures 16, 17, and 18.

In Table 8, the smaller FID scores, the better the generation images perform as real images. Since the inception network is pre-trained on ImageNet dataset which is distinctive

from our domain microstructure dataset, we finetuned this network by optimizing the alloy classification according to the train and validation microstructure images. Using the finetuned inception network as the feature extractor to calculate FID scores, Alloy-GAN outperforms standard cGAN on the test set with alloys previously unseen during the training process. The result remains consistent when calculating the score for each alloy and then aggregating the scores through methods like taking the mean or standard deviation. No matter the pretrained or fine-tuned networks as feature extractors, as a general metric, FID cannot evaluate whether the generated images are scientifically valid for alloy discovery. We thus also employ domain-science metrics to evaluate alloy foundational chemical properties of the generated images from the conditions of unseen alloys on our test set.

5.7.1 Micrograph Evolution with Si Content:

We evaluate whether generated images of AlloyGAN are able to reflect the proposed effect of Si content. Shown in Figure 16, generated images using AlloyGAN perform a clear trend of the area of fractions when Si content increases. This illustrates AlloyGAN has the capability to create valid microstructure images reflecting the Si effects, especially with alloys that previously had not been manufactured.

5.7.2 Effect of Cooling Rate:

We then evaluate whether AlloyGAN is able to handle the proposed effect of the Cooling Rate. Shown in Figure 17, its generated images exhibit a strong correlation between Cooling Rate and SDRS. This illustrates AlloyGAN has the capability to create valid microstructure images reflecting the cooling rate effects. This is of particular interest for generating alloys that previously had not been manufactured before.

5.7.3 Effect of Sr/P Modification:

Figure 18 shows the model successfully predicts the effects of Sr/P modification on A356 and A393 alloys. This illustrates that AlloyGAN can create valid microstructure images reflecting the Sr/P modification, especially with never-manufactured alloys.
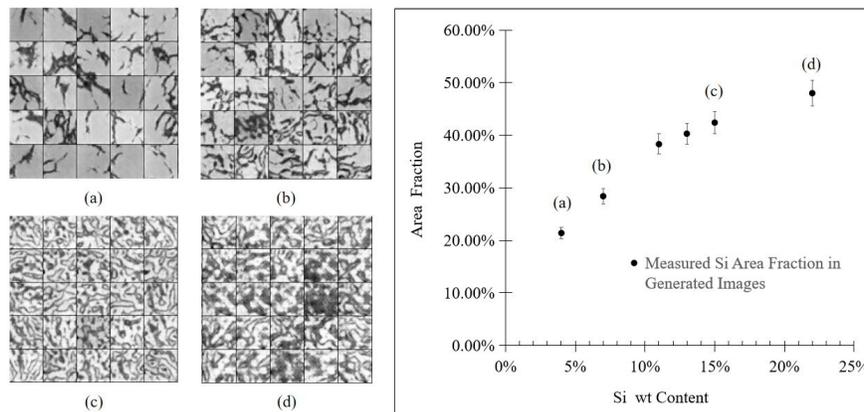
Figure 16: Area fraction of Si phase vs. Si amount in created images: Si wt. content (a) 4%, (b) 9%, (c) 15%, and (d) 22%.
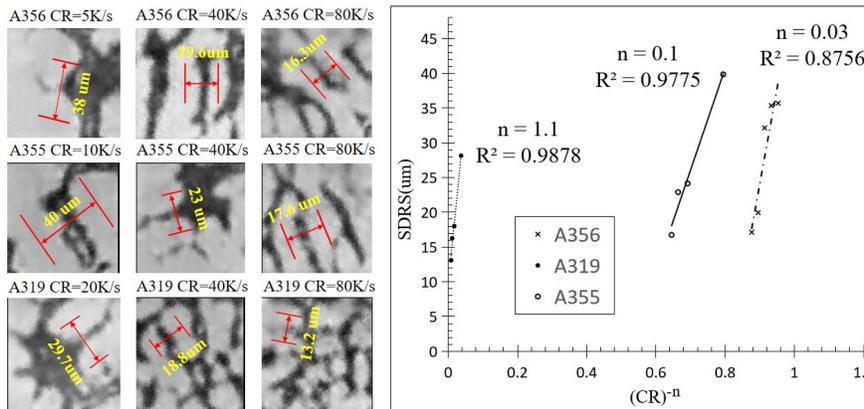
Figure 17: Effect of Cooling Rate: Red bars illustrate how we measure SDRS in each generated image in terms of cooling rates and manufactured alloys. R^2 indicates the fitness of linearity based on samples measured in each example alloy.
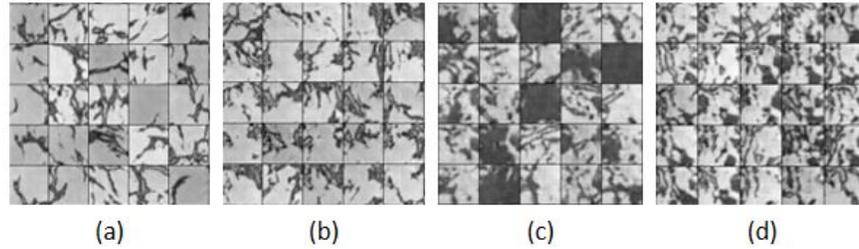
Figure 18: Example generations given different Sr/P conditions: (a) A356 alloys with Sr/P Non-Modified and (b) A356 alloys with Sr/P Modified; (c) A393 alloys with Sr/P Non-Modified and (b) A393 alloys with Sr/P Modified. All alloys are under a Cooling Rate of 10 K/S.

5.8 AlloyGAN Website

AlloyGAN is a product from a data science project team formulated by machine learning researchers, software engineers, and domain material scientists. We open a publicly accessible website to serve AlloyGAN interactive demo at https://deepalum.com/. The website is built by our collaborative startup. Users can create scientifically valid images given their text prompt to determining expected Alloy compositions within 1 second. Enhancements of AlloyGAN to support different types of materials continue to drive up its value to the material science community and customers.

6 Conclusion

This Ph.D. dissertation aims to revolutionize material discovery by leveraging the power of deep learning while addressing the challenges posed by small datasets in the field of material science. The overarching goal is to enhance the efficiency of material discovery processes through the integration of material science knowledge into deep neural networks. The research comprises two main tasks, with one already completed and the other proposed for investigation.

"Automating Scientific Corrosion Assessment on Existing Materials," began in collaboration with the Army Research Laboratory (ARL) in 2019. The research team developed the first-ever open corrosion image dataset, enriched with expert annotations. To tackle the limited accessibility of these sophisticated annotations to non-expert annotators, the team explored the effectiveness of Convolutional Neural Networks (CNNs) and Transformers in classifying corrosion. They further employed various techniques, including manual augmentation, transfer learning, and self-supervised learning, to improve performance. The cornerstone of this work was the development of a ground-truth edge guidance methodology on Decoder with innovative loss functions. The successful implementation of this approach on UNet architecture facilitated precise segmentation, enabling the automation of corrosion assessment. As a result, the team developed AI platforms, including an iPadOS APP, that revolutionized data collection and visualization for corrosion scientists. These tools have been adopted by prominent entities such as ARL, PPG, ASM, NASA, etc.

"Discovering Unknown from Known via Generative Models," explores how AI and deep learning can discover unseen materials guided by expert prompts and prior material science knowledge. We constructed promptable generative models capable of simulating the performance of unknown materials based on known parameters, such as basic chemical reactions. This rapid image simulation provides invaluable insights into unknown

material microstructure behavior within seconds.

The central focus of this dissertation is to address the challenges of applying deep learning techniques to small datasets in real-world scientific material discovery. By automating scientific corrosion assessment and developing promptable generative models, this research will significantly expedite material discovery processes and foster a seamless connection between material science and AI research. The proposed research has the potential to revolutionize material discovery, leading to technological advancements and saving time and effort in the quest for novel materials.

7 Future Work

Based on our experiment results, there are several avenues for future work that can be pursued to extend the findings and address the limitations identified in our studies.

7.1 Enhancement of Scientific Corrosion Segmentation Techniques

The current results of the domain of scientific corrosion segmentation have highlighted the effectiveness of DeepSC-Edge, our innovative UNet-based Decoder edge guidance, over MedTransformer. However, we could conduct and compare large language models (LLMs) for the task of corrosion detection and segmentation. Notably, we found that segmenting thin corrosion is not well-addressed by the Segment Anything Model (SAM), posing a significant opportunity for future research, such as expert-annotated ground-truth guidance in LLMs.

7.2 Integrating Generative Self-Supervised Learning with Multitask Framework for Long-tailed Corrosion Assessment

Future investigations will focus on a multitask learning framework that integrates two key tasks: ordinal regression and segmentation. The primary task involves employing Generative Self-Supervised Learning (GSSL), such as Masked Auto-Encoder that we used, for long-tailed ordinal regression specifically tailored for corrosion assessment. This task addresses the challenges of infrequent but critical corrosion phenomena, such as heavy corrosion in the testing field, by enhancing model training on underrepresented corrosion states, thereby improving the predictive accuracy for these vital scenarios.

The second task will augment the initial model by integrating advanced segmentation techniques aimed at the precise identification and characterization of different stages of corrosion. This addition is particularly focused on enhancing both corrosion regression and segmentation capabilities, especially for detecting and analyzing unseen anti-

corrosive materials. The integration of these tasks within a single framework enables cross-task knowledge transfer, where insights from segmentation directly refine the ordinal regression models, significantly boosting the overall predictive utility and specificity of the assessments.

This dual-task approach is expected to lead to substantial advancements in corrosion diagnostics, merging these methodologies to extend the analytical capabilities essential for material preservation and safety across various industrial applications. The focus on unseen anti-corrosive materials also prepares the model for future material innovations, ensuring its adaptability and long-term relevance.

7.3 Advancements in Alloy Microstructure Generation

The research into alloy microstructure generation using generative adversarial networks (GANs) has unveiled their potential in simulating the discriminative properties of alloys. Future studies should focus on further enhancing the capability of GANs or other conditional generative models like guided diffusion [60] to accurately replicate the complex physical and chemical properties of alloys. This includes the development of algorithms that can handle a wider range of material compositions, particularly those around outliers in the dataset. Improvements in the resolution and color fidelity of generated microstructures are also critical to better align with real-world materials.

Despite these advancements, a noticeable gap persists between AI and materials science [61], a divide our AlloyGAN design seeks to bridge by facilitating new discoveries, as depicted in Figure 13. Yet, AlloyGAN is not poised to replace materials scientists. The dataset used for AlloyGAN, derived from ASM-certified experiments, minimizes label bias and is chosen for its relevance to current market demands. However, it shows limitations in accurately modeling properties such as stability and grain boundaries in polycrystals [62]. My work with AlloyGAN highlights the significant role that GAN models can play in assisting materials science discovery. I am committed to advancing the inte-

gration of data science and materials expertise, and I am optimistic about the transformative potential of generative AI models in scientific material discovery, particularly as we strive to overcome existing limitations, inspired by the latest advancements in Large Language Models.

In summary, our future work aims to address the current limitations by innovating and refining the methodologies and models used in both domains. By focusing on these areas, researchers can significantly contribute to the fields of corrosion detection and materials science, pushing the boundaries of what is currently possible in scientific analysis and material generation.

8 Publication List

In this Ph.D. dissertation, I incorporate 5 papers for which I am the first author, as marked in italics below.

20. [In submission] Josselyn, N., **Yin, B.**, Considine, T., Kelley, J., Rinderspacher, B., Jensen, R., Snyder, J., Zhang, Z., Rundensteiner, E., Ensembles of Statistically Independent Models: An Embarrassingly Simple yet Strong Method for Domain Adaptation, In submission to a journal, 2024.

19. [In submission] Josselyn, N., **Yin, B.**, Considine, T., Kelley, J., Rinderspacher, B., Jensen, R., Snyder, J., Zhang, Z., Rundensteiner, E., Learning to Adapt Deep Corrosion Assessment Models from Indoor to Outdoor Image Domains, In submission to a journal, 2024.

18. [ARL Tech 2024] Considine, T., Jensen R., Josselyn N., **Yin, B.**, Zhang Z., Rundensteiner E., Automatic Corrosion Assessment Platform, DEVCOM Army Research Lab (ARL), 2024.

17. [ASM 2024] Enhanced Corrosion Data Analysis, Considine, T., Jensen R., Josselyn N., **Yin, B.**, Zhang Z., Rundensteiner E., ASM Handbook – AI for Materials, American Society of Materials (ASM), 2024.

16. *[TMS 2024]* ***Yin, B.***, *Fan, Y. Simulating Castable Aluminum Alloy Microstructures with AlloyGAN Deep Learning Model. In TMS 2024 Conference Proceedings.*

15. *[IEEE ICMLA 2023]* ***Yin, B.***, *Josselyn, N., Considine, T., Kelley, J., Rinderspacher, B., Jensen, R., Snyder, J., Zhang, Z., Rundensteiner, E. (2023). DeepSC-Edge: Scientific Corrosion Segmentation with Edge-Guided and Class-Balanced Losses. In 22nd IEEE International Conference on Machine Learning and Applications, 2023.*

14. *[IEEE ICMLA 2023]* ***Yin, B.***, *Fan, Y., Josselyn, N., Rundensteiner, E. (2023). AlloyGAN: Domain-Promptable Generative Adversarial Network for Generating Aluminum Alloy Microstructures. In 22nd IEEE International Conference on Machine Learning and Applications,*

2023.

13. [ACM CIKM 2023] **Yin, B.**, *Josselyn, N., Zhang, Z., Rundensteiner, E. A., Considine, T. A., Kelley, J. V., Rinderspacher, B. C., Jensen, R. E., Snyder, J. F. (2023). MOSS: AI Platform for Discovery of Corrosion-Resistant Materials. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (pp. 5128-5132), 2023.*

12. [ARL Tech 2023] Jensen, R., Brady, K., Parvatharaju, P., **Yin, B.**, Emdad, F., Rundensteiner, E. MIL-PRF 32662 Context-Aware Agile Platform, DEVCOM Army Research Lab (ARL), 2023.

11. [Issued Patent] Xu F., **Yin, B.**, Li Y., Liang H., Li X., and Ling Z. Automatic discriminant analysis of graphic problems based on tracked points, CN Patent CN201811352610.2 A., issued at 2022.

10. [IEEE ICMLA 2022] Josselyn, N., **Yin, B.**, Considine, T., Kelley, J., Rinderspacher, B., Jensen, R., Snyder, J., Zhang, Z., Rundensteiner, E. (2022, December). Transferring Indoor Corrosion Image Assessment Models to Outdoor Images via Domain Adaptation. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1386-1391), 2022.

9. [IEEE Big Data 2022] Josselyn, N., **Yin, B.**, Zhang, Z., Rundensteiner, E. (2022, December). An Empirical Study of Domain Adaptation: Are We Really Learning Transferable Representations?. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 5504-5513)., 2022.

8. [BMVC 2021] **Yin, B.**, *Josselyn, N., Considine, T., Kelley, J., Rinderspacher, B., Jensen, R., Synder, J., Zhang, Z., Rundensteiner, E. (2021, September). Corrosion image data set for automating scientific assessment of materials. In British Machine Vision Conference (BMVC) (pp. 1-15), 2021.*

7. [IEEE Big Data 2020] **Yin, B.**, Considine, T. A., Emdad, F., Kelley, J. V., Jensen, R. E., Rundensteiner, E. A. (2020, December). Corrosion assessment: Data mining for quantifying associations between indoor accelerated and outdoor natural tests. In 2020

IEEE International Conference on Big Data (Big Data) (pp. 2929-2936). 2020.

6. [ACL 2020] Sen, C., Hartvigsen, T., **Yin, B.**, Kong, X., Rundensteiner, E. (2020, July). Human attention maps for text classification: Do humans and neural networks focus on the same words?. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 4596-4608), 2020.

5. [Chinese Journal 2017] Xin S., **Yin, B.** Statistics Major in Big Data Era, Statistics and Strategy [J], 2017.

4. [ICEDM 2017] **Yin, B.**, Botelho, A. F., Patikorn, T., Heffernan, N. T., Zou, J. (2017). Causal Forest vs. Naïve Causal Forest in Detecting Personalization: An Empirical Study in ASSISTments. In the Tenth International Conference on Educational Data Mining, 2017.

3. [ACM L@S 2017] **Yin, B.**, et al. Observing Personalizations in Learning: Identifying Heterogeneous Treatment Effects Using Causal Trees. In the Fourth, ACM Conference on Learning @ Scale, MIT, MA

2. [ICEDM 2017] Adjei, S. A., Botelho, A. F., Patikorn, T., **Yin, B.**, Beck, J. E. (2017). Using "Snapshots" of Student Performance to Model Wheel Spinning. In the Tenth International Conference on Educational Data Mining, 2017.

1. [CodeCon 2016] Patikorn, T., Selent, D., Heffernan, N., **Yin, B.**, Botelho, A. F. (2016, March). Assistments dataset for a data mining competition to improve personalized learning. In Proceedings of CodeCon, MIT, MA., 2016.

References

[1] Hui Lin, Bin Li, Xinggang Wang, Yufeng Shu, and Shuanglong Niu. Automated defect inspection of led chip using deep convolutional neural network. *Journal of Intelligent Manufacturing*, 30(6):2525–2534, 2019. 11

[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv:1506.01497*, 2015. 11

[3] Young-Jin Cha, Wooram Choi, Gahyun Suh, Sadegh Mahmoudkhani, and Oral Büyüköztürk. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):731–747, 2018. 11

[4] Yongsheng Bai, Halil Sezen, and Alper Yilmaz. End-to-end deep learning methods for automated damage detection in extreme events at various scales. *arXiv preprint arXiv:2011.03098*, 2020. 11

[5] Daniel Vriesman, Alceu Britto Junior, Alessandro Zimmer, and Alessandro Lameiras Koerich. Texture cnn for thermoelectric metal pipe image classification. In *2019 IEEE 31st Intl. Conf. on Tools with Artificial Intelligence (ICTAI)*, pages 569–574. IEEE, 2019. 11

[6] Mantas Micikevičius. *Corrosion detection on steel panels using semantic segmentation models*. PhD thesis, Vilniaus universitetas, 2023. 11

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 12, 29, 30, 34

[8] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 12

[9] Qiang Zuo, Songyu Chen, and Zhifang Wang. R2au-net: attention recurrent residual convolutional neural network for multimodal medical image segmentation. *Security and Communication Networks*, 2021:1–10, 2021. 12

[10] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018. 12

[11] Guangqian Kong, Huan Tian, Xun Duan, and Huiyun Long. Adversarial edge-aware image colorization with semantic segmentation. *IEEE Access*, 9:28194–28203, 2021. 12

[12] Dong Wang, Wei Dai, Ding Tang, Yan Liang, Jing Ouyang, Huamiao Wang, and Yinghong Peng. Deep learning approach for bubble segmentation from hysteroscopic images. *Medical & Biological Engineering & Computing*, 60(6):1613–1626, 2022. 12

[13] Dongcai Cheng, Gaofeng Meng, Shiming Xiang, and Chunhong Pan. Fusionnet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12):5769–5783, 2017. 12

[14] Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Semeda: Enhancing segmentation precision with semantic edge aware loss. *Pattern Recognition*, 108:107557, 2020. 12

[15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 12, 32

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 25

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 25

[18] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 25

[19] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 25

[20] David M. Bastidas. Corrosion and protection of metals. *Metals*, 10(4), 2020. 27

[21] Gerhardus Koch. Cost of corrosion. *Trends in oil and gas corrosion research and technologies*, pages 3–30, 2017. 27

[22] R.A. Lane, C Fink, C Grethlein, and N Rome. Analysis of alternatives to hexavalent chromium: A program management guide to minimize the use of crvi in military systems. *Rome, NY: Advanced Materials, Manufacturing and Testing Information Analysis Center*, 2012. 27

[23] Christina Rudén and Sven Ove Hansson. Registration, evaluation, and authorization of chemicals (reach) is but the first step–how far will it take us? six further steps to improve the european chemicals legislation. *Environmental health perspectives*, 118(1):6–10, 2010. 27

[24] Noelle Easter C Co and James T Burns. Effects of macro-scale corrosion damage feature on fatigue crack initiation and fatigue behavior. *Intl. Journal of Fatigue*, 103:234–247, 2017. 27

[25] Thomas Considine, Daniel Braconnier, John Kelley, Thomas Braswell, Christopher Miller, Brian Placzankis, and Robert Jensen. Data analytic prediction and correlation visualization of corrosion assessment for DoD sustainment. *Technical Report of US. CDCC Army Research Lab*, 2018, Unpublished. 27

[26] L YING-YU and W QUI-DONG. Astm D1654: standard test method for evaluation of painted or coated specimens subjected to corrosive environments. *Annual Book of ASTM Standards. West Conshohocken*, 1992. 27, 28

[27] Jan Ivar Skar and Darryl Albright. Corrosion behavior of die-cast magnesium in astm b117 salt spray and gm9540p cyclic corrosion test. *2003 Magnesium Technology*, 1:59, 2003. 27

[28] ASTM Committee D-1 on Paint, Materials Related Coatings, and Applications. *Standard Test Method for Evaluation of Painted or Coated Specimens Subjected to Corrosive Environments*. ASTM International, 2008. 27

[29] Biao Yin, Nicholas Josselyn, Thomas Considine, John Kelley, Berend Rinderspacher, Robert Jensen, James Snyder, Ziming Zhang, and Elke Rundensteiner. Corrosion image data set for automating scientific assessment of materials. In *British Mach. Vis. Conf., BMVC 2021*. 28

[30] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. " grabcut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, 23(3):309–314, 2004. 29

[31] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *MICCAI 2021: 24th Intl. Conf., Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 36–46. Springer, 2021. 29

[32] Zhijie Zhang, Huazhu Fu, Hang Dai, Jianbing Shen, Yanwei Pang, and Ling Shao. Et-net: A generic edge-attention guidance network for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 442–450. Springer, 2019. 30, 33

[33] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 683–687. IEEE, 2019. 32

[34] Allied Market Research. Metal & metal manufactured products market by metal type (lithium, magnesium, aluminum, chromium, manganese, iron, cobalt, cop-

per, zinc, molybdenum, silver, and others), product type (wires cables, jewelry ornaments, bars rebar, sheets, rolls, pipe fixture fittings, pipes, molded components, batteries, paints coatings, and others), and end-use industry (automotive transportation, aerospace defense, marine, consumer goods, electrical electronics, manufacturing, energy power, construction, packaging, pharmaceuticals, and others): Global opportunity analysis and industry forecast, 2021-2030. https://www.alliedmarketresearch.com/metal-and-metal-manufactured-products-market, 2022. Accessed: April, 2022. 39

[35] Abdel Salam Hamdy Makhlouf and Mahmood Aliofkhazraei. *Handbook of materials failure analysis with case studies from the aerospace and automotive industries*. Butterworth-Heinemann, 2015. 39

[36] Ryo Kobayashi. Modeling and numerical simulations of dendritic crystal growth. *Physica D: Nonlinear Phenomena*, 63(3-4):410–423, 1993. 39, 40

[37] J Eiken. Dendritic growth texture evolution in mg-based alloys investigated by phase-field simulation. *International Journal of Cast Metals Research*, 22(1-4):86–89, 2009. 39, 40

[38] I Steinbach and M Apel. The influence of lattice strain on pearlite formation in fe–c. *Acta Materialia*, 55(14):4817–4822, 2007. 39, 40

[39] N Zhou, C Shen, PM Sarosi, MJ Mills, T Pollock, and Y Wang. γ rafting in single crystal blade alloys: a simulation study. *Materials Science and Technology*, 25(2):205–212, 2009. 39, 40

[40] YL Li and LQ Chen. Temperature-strain phase diagram for ba ti o 3 thin films. *Applied physics letters*, 88(7):072905, 2006. 39, 40

[41] Tomohiro Takaki, Takashi Shimokawabe, Munekazu Ohno, Akinori Yamanaka, and Takayuki Aoki. Unexpected selection of growing dendrites by very-large-scale phase-field simulation. *Journal of Crystal Growth*, 382:21–25, 2013. 40

[42] MP Anderson, GS Grest, and DJ Srolovitz. Grain growth in three dimensions: a lattice model. *Scr. Metall.;(United States)*, 19(2), 1985. 40

[43] AD Rollett, David J Srolovitz, RD Doherty, and MP Anderson. Computer simulation of recrystallization in non-uniformly deformed metals. *Acta Metallurgica*, 37(2):627–639, 1989. 40

[44] Dierk Raabe. Cellular automata in materials science with particular reference to recrystallization simulation. *Annual review of materials research*, 32(1):53–76, 2002. 40

[45] K Reuther and M Rettenmayr. Perspectives for cellular automata for the simulation of dendritic solidification–a review. *Computational materials science*, 95:213–220, 2014. 40

[46] Jinhu Zhang, Xuexiong Li, Dongsheng Xu, and Rui Yang. Recent progress in the simulation of microstructure evolution in titanium alloys. *Progress in Natural Science: Materials International*, 29(3):295–304, 2019. 40

[47] Elizabeth A Holm, Ryan Cohn, Nan Gao, Andrew R Kitahara, Thomas P Matson, Bo Lei, and Srujana Rao Yarasi. Overview: Computer vision and machine learning for microstructural characterization and analysis. *Metallurgical and Materials Transactions A*, 51:5985–5999, 2020. 40

[48] Juwon Na, Gyuwon Kim, Seong-Hoon Kang, Se-Jong Kim, and Seungchul Lee. Deep learning-based discriminative refocusing of scanning electron microscopy images for materials science. *Acta Materialia*, 214:116987, 2021. 40

[49] Seyed Majid Azimi, Dominik Britz, Michael Engstler, Mario Fritz, and Frank Mücklich. Advanced steel microstructural classification by deep learning methods. *Scientific reports*, 8(1):2128, 2018. 40

[50] In Yong Moon, Ho Won Lee, Se-Jong Kim, Young-Seok Oh, Jaimyun Jung, and Seong-Hoon Kang. Analysis of the region of interest according to cnn structure in hierarchical pattern surface inspection using cam. *Materials*, 14(9):2095, 2021. 40

[51] Yongju Kim, Hyung Keun Park, Jaimyun Jung, Peyman Asghari-Rad, Seungchul Lee, Jin You Kim, Hwan Gyo Jung, and Hyoung Seop Kim. Exploration of optimal microstructure and mechanical properties in continuous microstructure space using a variational autoencoder. *Materials & Design*, 202:109544, 2021. 40

[52] Jianan Tang, Xiao Geng, Dongsheng Li, Yunfeng Shi, Jianhua Tong, Hai Xiao, and Fei Peng. Machine learning-based microstructure prediction during laser sintering of alumina. *Scientific Reports*, 11(1):1–10, 2021. 41

[53] Wufei Ma, Elizabeth J Kautz, Arun Baskaran, Aritra Chowdhury, Vineet Joshi, Bulent Yener, and Daniel J Lewis. Image-driven discriminative and generative machine learning algorithms for establishing microstructure–processing relationships. *Journal of Applied Physics*, 128(13):134901, 2020. 41

[54] Kang-Hyun Lee and Gun Jin Yun. Microstructure reconstruction using diffusion-based generative models. *Mechanics of Advanced Materials and Structures*, pages 1–19, 2023. 41

[55] Xiaolin Li, Yichi Zhang, He Zhao, Craig Burkhart, L Catherine Brinson, and Wei Chen. A transfer learning approach for microstructure reconstruction and structure-property predictions. *Scientific reports*, 8(1):13461, 2018. 41

[56] Zijiang Yang, Xiaolin Li, L Catherine Brinson, Alok N Choudhary, Wei Chen, and Ankit Agrawal. Microstructural materials design via deep adversarial learning methodology. *Journal of Mechanical Design*, 140(11), 2018. 41

[57] Nimel Sworna Ross, C Sherin Shibi, M Sithara, Munish K Gupta, Mehmet Erdi Korkmaz, Vishal S Sharma, and Z Li. Measuring surface characteristics in sustainable machining of titanium alloys using deep learning based image processing. *IEEE Sensors Journal*, 2023. 41

[58] Zhiyuan Yang, Shu Li, Shuai Li, Jia Yang, and Dongrong Liu. A two-step data augmentation method based on generative adversarial network for hardness prediction of high entropy alloy. *Computational Materials Science*, 220:112064, 2023. 41

[59] Ivan Ferreira, Luis Ochoa, and Ardiansyah Koeshidayatullah. On the generation of realistic synthetic petrographic datasets using a style-based gan. *Scientific Reports*, 12(1):12845, 2022. 41

[60] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 56

[61] Akshay Subramanian, Wenhao Gao, Regina Barzilay, Jeffrey C Grossman, Tommi Jaakkola, Stefanie Jegelka, Mingda Li, Ju Li, Wojciech Matusik, Elsa Olivetti, et al. Closing the execution gap in generative ai for chemicals and materials: Freeways or safeguards. 2024. 56

[62] Malik Wagih, Peter M Larsen, and Christopher A Schuh. Learning grain boundary segregation energy spectra in polycrystals. *Nature communications*, 11(1):6376, 2020. 56