2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-6654-6283-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMLA55696.2022.00219

Transferring Indoor Corrosion Image Assessment Models to Outdoor Images via Domain Adaptation

Nicholas Josselyn[1], Biao Yin[1], Thomas Considine[4], John Kelley[4], Berend Rinderspacher[4], Robert Jensen[5], James Snyder[6], Ziming Zhang[1,3], Elke Rundensteiner[1,2]

[1]Department of Data Science, Worcester Polytechnic Institute, Worcester, MA
[2]Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA
[3]Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA
[4]ARD, DEVCOM Army Research Laboratory, Aberdeen Proving Ground, MD, USA
[5]ARD, DEVCOM Army Research Laboratory, Northeast Regional Extended Site Burlington, MA, USA
[6]ARD, DEVCOM Army Research Laboratory, Adelphi, MD, USA

***Abstract*—Corrosion of materials impacts critical economic sectors from infrastructure, transportation, defense, health, to the environment. The development of safe anti-corrosive materials is thus an important area of study in materials science. Corrosion science of preparing materials and then monitoring their corrosion under adverse conditions is labor intensive, time consuming, and extremely costly. While deep learning has become popular in automating various engineering tasks, the development of deep models for corrosion assessment is lacking. We are the first to study deep domain adaptation (DA) models for the automated assessment of the corrosion status of anti-corrosive materials. Corrosion data, i.e., photographic images of treated corroding materials, is abundant when produced in artificially controlled laboratory settings, while corrosion image data sets from rich natural outdoor environments are more challenging to produce and thus much smaller. We leverage the more readily available indoor corrosion data to train a classifier and then transfer it via deep domain adaptation to also perform well on the small yet more realistic outdoor corrosion image data set – without requiring target labels. We empirically compare 5 popular domain adaptation models on real-world corrosion image data sets. Our study finds that DA achieves 27% improvement in test accuracy compared to the performance of the no-DA baseline for classifying real-world outdoor corrosion data.**

***Index Terms*—Materials Science, Domain Adaptation**

I. Introduction

Corrosion is a High-Stake Problem. Corrosion is the gradual degradation of a metal over time due to chemical interactions with its environment. The economic burden corrosion inflicts equates to over 4% of the gross domestic product (GDP) loss in developed countries [1]. The risk associated with corrosion is tremendous - resulting in major safety implications in regards to infrastructure (bridges collapsing), transportation (automobile durability and safety), household appliances (washers, dryers, furnaces), the environment (leakage of chemicals into bodies of water), and even societal health (cancer causing chemical side-products). There is thus active, high-stake research in materials science focused on the design and testing of new anti-corrosive materials that are both effective and non-hazardous [2]–[5].

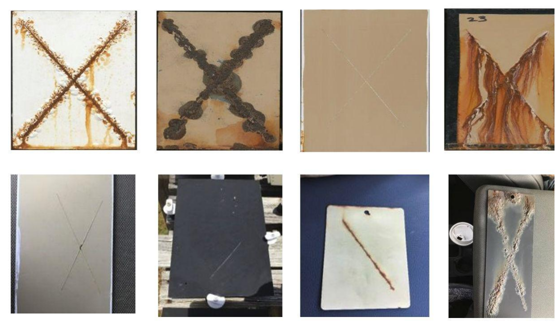

Figure 1: Corrosion test image panels. Top: Indoor laboratory environment tests. Bottom: Outdoor natural environment tests.

Corrosion Assessment. Validation and subsequent assessment of new materials for corrosion fall into 2 categories seen in Figure 1: (1) indoor tests subject to an accelerated, controlled laboratory setting, and (2) outdoor tests subject to the natural elements of the environment (rain, sun, heat, etc). In both tests, a stack-up of material coatings is created as a panel, placed in a laboratory or outdoor environment, and rated according to ASTM standards [4]. 12 equally spaced measurements are made across the "X" shape on the panels, averaged, divided by 2, and then associated with a rating between 0 and 10; with 0 meaning heavily corroded, and 10 meaning no corrosion.

Corrosion Challenges. This corrosion assessment process to research, develop, and validate these new anti-corrosive materials is extremely challenging. It is difficult to observe and assess corrosion progression in a quick, highly accurate, practically relevant and financially feasible manner – impeding the discovery of new materials. First, the costs to manufacture and test new material samples with respect to labor, resources, and validation can amount to 10s to 100s of millions of dollars [6]. Second, testing may involve the use of hazardous chemicals, resulting in additional environmental and safety concerns. Third, skilled professionals are required to collect

978-1-6654-6283-9/22/$31.00 ©2022 IEEE
DOI 10.1109/ICMLA55696.2022.00219

and analyze appropriate experimental testing materials, driving up costs and exacerbating the human professional scarcity issue. Fourth, observation times of corrosion progression may be up to several years for material validation, especially, in real-world environments. And finally fifth, more readily available testing options in laboratory environments are also not sufficient for final material validation [2], [3].

State-of-the-Art: Machine Learning for Corrosion Assessment. To overcome some of the above challenges, machine learning solutions have been applied to automate tedious engineering tasks, such as defect detection [7]–[10] and corroded pipe detection [11]. Recent works, one with corrosion sensor data [12] and the other focused on indoor corrosion data [13], show the promise of applying deep learning to support corrosion assessment experimentation.

Our Proposed Approach. In this work, we observe that the state-of-the-art in machine learning model development is focused on indoor laboratory tests [13]. This is in part due to limited access to and difficulty in obtaining outdoor data. Thus, outdoor datasets tend to be small, while the training of deep learning models typically requires a substantial volume of labeled data. Therefore, we study the problem of how to leverage the larger indoor experimental datasets to boost the performance of models on the data we desire to assess, namely, realistic outdoor data. In doing so, we aim to overcome the challenges associated with changes across datasets such as changes in image pose, background changes, lighting changes, different image capture methods, and dimensionality changes of the images. These changes between domains are known as *domains shifts*. Given these domain shift challenges, we investigate the utility of *domain adaptation* (DA), a subset of transfer learning, for corrosion assessment.

Domain adaptation utilizes two or more similar but distinct sets of data (domains) called *source* domain and *target* domain, where the task for each is the same (e.g. corrosion rating classification) and the aim is to learn to jointly perform this task well on both domains of data. Correspondingly, we therefore frame our problem of corrosion assessment as a domain adaptation problem where the more readily available well-labeled indoor data is the source domain, and the more scarce and typically missing-label outdoor data is the target domain. In short, we utilize the information and availability of data from the indoor domain to adapt a classifier to perform well on the desirable outdoor domain.

Contributions. We make the following contributions:

- We are the first to formulate experimental corrosion assessment as a domain adaptation problem.
- We design an experimental study for evaluating the ability of five state-of-the-art (SOTA) DA models of transferring knowledge from the indoor (source) to the outdoor (target) domain images in this application domain.
- We conduct this study on real-world experimental data sets produced and collected by materials scientists, namely, on 600 images from indoor laboratory tests and 191 images from outdoor natural environment tests [13].
- We derive interesting experimental findings with implications for next steps in corrosion assessment, namely, DA shows performance improvement for domain transfer from indoor to outdoor across several evaluation metrics – calling for additional exploration into this line of research.

II. Scientific Corrosion Assessment

Corrosion scientists follow material standards along the entire manufacturing and assessment procedure. Material panels are manufactured for assessment as a layered stack up consisting of a topcoat, primer, pretreatment, surface profile, and substrate layer. When assessing the panel, the scientists consider the condition of the surface and the composition of the five layers in the stack up. Commercial names of materials are omitted in this paper and dataset for proprietary reasons.

Figure 2: Indoor (left) and outdoor (right) corrosion testing environments.

Scientific corrosion experiments come from either indoor laboratory-based tests or outdoor natural environment tests. For each scenario, a panel that is exposed to these respective environments (rain, sun, heat, etc.) is manually observed and rated with respect to the amount of visual corrosion by corrosion scientists at regular intervals throughout testing. The example testing environments are depicted in Figure 2.

A. Indoor Laboratory Tests

Indoor corrosion tests are performed in a controlled laboratory setting. They are primarily used as a quality control element of sample production to reduce the labor investment needed for outdoor tests. The advantage of conducting indoor laboratory-based tests is that artificial conditions can be introduced and corrosion can this way be accelerated. Thus, panels can be evaluated on daily or weekly timescales, resulting in easier collection of data and more rapid insights about material performance. Images are collected on flatbed scanners to provide high quality images. The images can be easily traced in in-house labs. However, indoor tests are expensive to operate. Also, laboratory experiments rely on artificial standardized environments that may not have a direct real-world equivalent. In that sense, it is a simulation and not the real-world gold standard that the outdoor environment offers for material evaluation. Since indoor scientific corrosion assessment data for machine learning use has been released, we leverage this open image data set [13].

B. Outdoor Field Tests

Outdoor corrosion tests are conducted by placing panels directly into natural environments. These sites are typically selected as areas of "worst-case scenario expectation", such as coastal regions associated with an expected higher rate of corrosion. Outdoor panels are observed and rated by corrosion scientists at 3-month intervals. One major disadvantage to outdoor tests is the difficulty and cost with obtaining data from these remote environments. A large burden is placed on corrosion scientists to go out into the field and collect data at these remote sites. Panels are only observed every 3 months (or longer), leading to rather small sample sizes in such data collected. Further, as seen in Figure 1, the quality of outdoor images as compared to indoor images can be worse. Outdoor images are often taken using digital cameras or cellphones at multiple angles, heights, and with various objects present in the background of the images. However, outdoor experiments are the gold standard experiments used by corrosion scientists to evaluate realistic material performance. Thus they are a vital means by which materials are evaluated, developed, and produced for vendors. We work with outdoor corrosion data from materials similar to those for the indoor corrosion data.

III. Related Work

It is important to establish a strong relationship between indoor and outdoor corrosion assessment for material scientists. Although statistical methods were studied [14] and corrosion sensor data has been used [12] [15], the capability of deep learning to solve this domain adaptation task has not yet been explored in the literature.

In the past several years, some deep learning models with respect to the detection of material defects have been explored, including LEDNet [7], Faster Region-based CNN [8] [9], fully connected networks [10], and Texture CNN – with the latter on 150 raw corroded pipe images [11]. However, these works aim to detect defects on pipes or buildings. This however is not the focus of materials discovery. Nor, do they tackle a scenario where domain adaptation can be a potential solution.

Although a large body of literature on domain adaptation techniques in general exists, to our best knowledge, domain adaptation has not yet been applied to scientific corrosion assessment. These works reflect diverse areas of domain adaptation, such as single-source closed-set DA [16]–[19], multi-source DA [20]–[23], multi-target DA [24], open-set DA [25], and partial DA [26] [27]. In recent work [28] [29], general transfer learning is studied to overcome the lack of supervised corrosion information. They search for the best-fit source models and transfer them by training on target data – however without applying domain adaptation techniques.

IV. Evaluated Models

In this section, we introduce the 5 domain adaptation models we have carefully selected from the literature for this study. All 5 models are unsupervised domain adaptation (UDA) models, meaning target data labels are *not* made available to the model during training. This is an active area of research in the DA field. It naturally applies to our problem of corrosion assessment because labeled target data (outdoor images) is difficult to obtain in our application. The 5 models have been introduced over the past 6 years and are: (1) recent or established models in the literature with high citation counts, (2) models frequently evaluated against in a large majority of the DA literature, and (3) models that match our corrosion assessment scenario of being single-source, single-target, closed-set DA where class labels match in both source and target domains (the 5 corrosion ratings). To implement these models, we use a public DA library [30] as a starting code base. Updated codes are provided for reproducibility [1].

DANN. Domain-Adversarial Training of Neural Networks [18] is a highly cited UDA model introduced in 2016. DANN famously incorporates a gradient reversal layer to promote alignment of the source and target domains with respect to a domain discriminator. DANN consists of a label predictor that predicts the class label of the source data during training and a domain discriminator that discriminates between source and target data during training. The training procedure aims to minimize the loss of the label predictor and maximize the domain confusion loss of the discriminator via an adversarial approach with the gradient reversal layer.

JAN. Deep Transfer Learning with Joint Adaptation Networks [16], a UDA model introduced in 2017, aims to reduce the shifts in joint distributions across domains. JAN advances its discrepancy alignment-based predecessors that primarily use only the Maximum Mean Discrepancy [32] to align source and target domains by introducing a function which also aligns the residual joint distributions between domains across multiple CNN layers.

CDAN. Conditional Adversarial Domain Adaptation [19], a UDA model introduced in 2017, was influenced by the concurrent advances in conditional generative adversarial networks (CGANs). CGANs make use of discriminative features between real and fake data and incorporate them into the generator and discriminator networks. In CDAN, the authors condition the domain discriminator with the cross-covariance of domain-specific feature representations and classifier predictions. Additionally, the discriminator is conditioned based on the uncertainty of the classifier predictions – thus allowing the discriminator to prioritize easy-to-transfer examples.

AFN. Larger Norm More Transferable: An Adaptive Feature Norm Approach for Unsupervised Domain Adaptation [17], a UDA model introduced in 2019, proposes a novel method for aligning domains based on statistical criterion different from the literature. That is, they suggest finding a shared, average feature norm (length of the feature vector) between the two domains. They state that the target domain feature norms are typically much smaller than the source feature norms and hypothesize that this complicates adaptation. By adapting the feature norms of both domains to a large range of scalars, they

[1] https://github.com/njosselyn13/Empirical-Study-Domain-Adaptation

	Baseline	DANN [18]	JAN [16]	CDAN [19]	AFN [17]	MCC [31]
Learning Rate	0.001	0.1	0.1	0.1	0.01	0.1
Weight Decay	0.001	0.1	0.001	0.1	0.01	0.001
Loss Trade-off	-	1.0	1.0	10.0	0.05	2.0

Table I: Tuned hyperparamters for baseline and the 5 DA models.

Corrosion Rating	5	6	7	8	9
No. Indoor Images	120	120	120	120	120
No. Outdoor Images	25	39	39	31	57

Table II: Indoor and outdoor data summary table. Subdivided for each corrosion rating class 5-9.

expect they can achieve better adaptation.

MCC. Minimum Class Confusion for Versatile Domain Adaptation [31], a UDA model introduced in 2020, differs from other DA models as it does not aim to explicitly align two domain feature spaces. Instead, it aims to reduce class confusion in the label space. MCC is versatile because its approach is suitable for a variety of DA settings such as: closed-set, partial-set, multi-source, and multi-target DA. MCC claims to outperform prior models, including the models we study here.

V. Experimental Study Design

A. Experimental Corrosion Data

Corrosion assessment data is crucial to the verification process of material reliability for use in production and also for the discovery of new materials that are resistant to corrosion. In this section we outline the data collection procedure for indoor and outdoor corrosion image data.

Ground Truth Rating. The assessment and ground-truth ratings, for both indoor and outdoor tests, are done by corrosion scientists in accordance with standards defined in ASTM D1654 [4]. These standards define the standard practice to visually evaluate the amount of scribe corrosion for a panel. Scribe corrosion is referred to as corrosion creep, which is emanating out from a deliberately cut area in the panel. Analysis is done using an optical magnifying tool to measure the amount of corrosion emanating from the scribe at 12 equally spaced points along the ”X” cross-scribe; 6 points along one scribe direction and 6 along the second direction (6 points are used if the panel only has a single scribe), as seen in Figure 1. The 12 measurements are averaged, divided by 2, and correlated to a discrete rating label between 0 and 10. A rating of 0 signifies a large amount of corrosion, whereas a rating of 10 signifies no corrosion present and is generally only observed at the start of the testing process.

Indoor Data. For indoor tests, we directly utilize open data in this field, including indoor corrosion panel images and ground-truth ratings. The indoor test data set consists of 600 images of corroded panels that underwent accelerated, indoor corrosion experiments. Each panel received a rating from 5-9, and these rating classes are balanced with 120 images in each class as seen in Table II [13]. This data had been split into 10 cross-validation folds with a held-out test set of data. We use this released split of data in our work.

Outdoor Data. Our outdoor tests consists of panels prepared for the same materials as the indoor tests. Similarly, a corrosion science expert annotated the images three times for thorough verification. The outdoor corrosion data set contains images of corroded panels with a rating from 0-10 using the ASTM D1654 standard [33]. We focus on a subset of 191 images ranging from scribe corrosion ratings 5 to 9; as this allows for simplicity in assessment performance comparison between indoor and outdoor experiments as they both have a similar distribution of samples in rating classes 5 to 9 (see Table II). We split the 191 outdoor images into 20% test data. The remaining 80% we split into 10 stratified, cross-validation folds to be comparable to the publicly released indoor data.

B. Domain Adaptation Setup

Baseline. We first establish a baseline performance of outdoor classification without involving domain adaptation. We train a traditional classifier on labeled *indoor* data only but then test it on only outdoor data. This scenario, with respect to the application domain, is realistic as it is possible to collect larger amounts of indoor data and we should be able to make use of it in the assessment of outdoor data. The simplest approach would be that a trained classifier on a labeled and larger indoor dataset could be used for the similar task on outdoor data. Therefore, we use this indoor to outdoor assessment as our baseline. Note, we optimize learning rate and weight decay hyperparameters, and experiment with values of: 0.1, 0.01, and 0.001. See Table I for details on final hyperparameters.

Model Comparisons. We train each of the 5 domain adaptation models on the same splits of data. We select the indoor images as the source domain for each model. Labels are made available to the model at train time. We select the outdoor images as the target domain for each model, with the labels not available to the model during training. The target domain outdoor data for the corresponding validation and test sets are used for tuning and testing, respectively. Code for reproducibility is made available.

For each model we tune hyperparameters such as learning rate, weight decay, and loss trade-off. The loss trade-off weighs how much to focus on classification loss vs each model's respective transfer loss. This trade-off is seen in Equation 1.

$$loss = cls\_loss + transfer\_loss * trade\_off \qquad (1)$$

We experiment with learning rate and weight decay values of: 0.1, 0.01, and 0.001. Loss trade-off values are: 0.05 (AFN default), 0.25, 0.5, 1.0, 2.0, 5.0, and 10.0.

	Baseline	DANN [18]	JAN [16]	CDAN [19]	AFN [17]	MCC [31]
Accuracy	0.22 ± 0.05	0.17 ± 0.06	0.19 ± 0.06	0.21 ± 0.08	0.22 ± 0.06	**0.28 ± 0.09**
F1	0.23 ± 0.02	0.24 ± 0.05	0.21 ± 0.06	**0.26 ± 0.04**	0.25 ± 0.05	0.25 ± 0.08
Balanced Accuracy	0.21 ± 0.03	0.2 ± 0.04	0.22 ± 0.08	0.19 ± 0.02	0.22 ± 0.04	**0.26 ± 0.1**
F1 (relaxed)	0.58 ± 0.02	0.45 ± 0.07	0.55 ± 0.07	0.47 ± 0.04	**0.59 ± 0.04**	0.55 ± 0.09
Balanced Acc. (relaxed)	0.61 ± 0.02	0.44 ± 0.07	0.58 ± 0.07	0.44 ± 0.06	**0.62 ± 0.04**	0.56 ± 0.1

Table III: Mean and standard deviation values over 10 cross-validation folds. Test accuracy and normal and relaxed F1-score and balanced accuracy metrics for classification of outdoor data for baseline and 5 DA models.

We then compare each domain adaptation model performance against our baseline and discuss the utility of domain adaptation models for the critical task of corrosion assessment.

Evaluation Metrics. We treat corrosion ratings as discrete rating labels and therefore utilize classification loss. However, the corrosion rating scale considered in this work is converted from a continuous millimeter measurement [4], which leads us to consider the evaluation under a less strict scenario where neighboring rating labels may also be acceptable predictions. To accomplish this less strict assessment, we not only evaluate models by traditional classification metrics but also customize them for a relaxed evaluation. Under this relaxed evaluation, a prediction can be true even if it is within a range of 1 rating higher or lower than its ground-truth rating (e.g. a ground truth rating of 5 when predicted as 4 or 6 is acceptable).

We implement general metrics using Sklearn [34], and further customize each to realize the relaxed evaluation. We provide in Table III overall test accuracy along with F1-score with micro-averaging and balanced accuracy – reported under normal and relaxed classification settings. F1 with micro averaging allows us to assess performance with respect to both the number of false positives and false negatives with micro-averaging addressing any data imbalance. Balanced accuracy compliments this and takes into account the number of true negatives and thus provides more information on model performance. Since the outdoor assessment tends to be naturally imbalanced, this balanced accuracy metric would be a more trustworthy approach to evaluating model performance.

VI. Results and Discussion

Next, we discuss our results presented in Table III and how our work impacts the corrosion science domain.

Baseline. In Table III we observe that the baseline classifier with no domain adaptation never, on average, beats any of the domain adaptation models on any of the evaluation metrics. However, we do note overlapping confidence intervals between baseline and best performing DA models across evaluated metrics, meaning there is uncertainty with respect to how much DA improves outdoor classification. With this, however, DA still shows some promise for improvement on outdoor data classification by utilizing the indoor data.

Further, we note that random chance for our scenario is 0.20 given the 5 possible classes, and as shown, domain adaptation (MCC) provides the greatest increase over this random chance.

In practice, this baseline is suitable as materials scientists are able to more readily collect and label indoor data. However, they ideally want to assess material performance under outdoor environmental conditions. Therefore, we use this baseline to compare domain adaptation model performance against.

Model Comparisons. We observe all 5 DA models generally perform comparably across all metrics compared to the baseline when considering standard deviations. Although, at least one DA model, on average, beats the baseline on every evaluation metric, as seen in Table III. In particular, CDAN outperforms the baseline and all other models with respect to F1-score. AFN outperforms the baseline and all other models with respect to relaxed F1 and relaxed balanced accuracy. MCC outperforms the baseline and all other models with respect to balanced accuracy and overall test accuracy. Given these results, the MCC and AFN models appear to be performing best for our corrosion assessment task, with domain adaptation overall offering a promising future on the assessment of the critical and challenging outdoor corrosion data. All models were tuned for the corrosion data, optimized for overall accuracy, and given a fair chance to perform their best. Final hyperparameters for each model are in Table I.

We note that reported metric values show improvement over baseline, yet are overall fairly low values. We attribute these overall smaller values to the inherent difficulty of the task and outdoor data. We have a small amount of outdoor data (and a small amount of indoor data). Additionally, the outdoor data, as seen in Figure 1, presents unique challenges such as being captured via varying angles, different lighting scenarios, and objects such as coffee cups present in the background.

To our knowledge, we are the first to attempt domain adaptation in the scientific assessment of corrosion for material discovery. The importance that DA models can have in the corrosion science field is shown in this work to have a positive impact. In this work we show the benefit DA models have on boosting outdoor corrosion classification and motivate the use and development of DA methods for corrosion science. Demonstrating the utility of DA is a step towards bigger goals for revolutionizing how new materials are discovered.

Vision for Material Discovery. As an ultimate objective in corrosion science, it is desirable for one to reliably conduct corrosion assessment exclusively on indoor tests only – effectively substituting the more expensive and extremely slow outdoor tests by the rapid indoor tests. If material scientists were able to determine how newly designed materials can be expected to perform in natural environments without having to actually conduct these longitudinal outdoor tests, this would be a game changer in the advancement of material discovery. Deep learning methods such as DA explored by our work are indeed based on leveraging indoor assessment, along with the design of a "transfer" model that transfers this trained model from indoor to outdoor assessment. Better yet, DA aims to utilize only a small amount of outdoor data without requiring labels for this outdoor data. In this sense, this work is a first step into an exciting new direction for technological advances in support of material discovery.

VII. Conclusion and Future Work

We are the first to evaluate the utility of domain adaptation for tackling the important materials science application of corrosion assessment. We evaluate 5 SOTA domain adaptation models on achieving the indoor to outdoor corrosion assessment task. We compare them against a baseline indoor to outdoor assessment task with no domain adaptation. We observe that at least one domain adaptation model across several evaluation metrics outperforms the baseline model. In particular, we note the AFN and MCC models as promising in this task. Given these initial results, our work has shown that domain adaptation holds potential promise in addressing this important problem.

However, future work is needed to explore optimization of these methodologies to further enhance their performance for this domain problem. Future work could also focus on other advanced domain adaptation and deep learning models such as few-shot learning to further address the limited data issue present in this application context.

Acknowledgments

This research is supported by the US ARMY, ACC-APG-RTP, Cooperative Agreement W911NF-19-2-0112, US. CDCC Army Research Lab and NSF grants CCF-2006738 and NRT-HDR-2021871.

References

[1] D. M. Bastidas, "Corrosion and protection of metals," *Metals*, vol. 10, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2075-4701/10/4/458

[2] N. E. C. Co and J. T. Burns, "Effects of macro-scale corrosion damage feature on fatigue crack initiation and fatigue behavior," *Int. Journal of Fatigue*, vol. 103, pp. 234–247, 2017.

[3] T. Considine *et al.*, "Data analytic prediction and correlation visualization of corrosion assessment for DoD sustainment," *Tech. Report of US. CDCC Army Res. Lab*, 2018, Unpublished.

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

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

[6] A. Wolverton, A. E. Ferris, and N. B. Simon, "Retrospective evaluation of the costs of complying with light-duty vehicle surface coating requirements," *Journal of Benefit-cost Anal.*, vol. 10, no. 1, pp. 39–64, 2019.

[7] H. Lin, B. Li, X. Wang, Y. Shu, and S. Niu, "Automated defect inspection of led chip using deep convolutional neural network," *Journal of Intell. Manuf.*, vol. 30, no. 6, pp. 2525–2534, 2019.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.

[9] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Comput.-Aided Civil and Infrastructure Eng.*, vol. 33, no. 9, pp. 731–747, 2018.

[10] Y. Bai, H. Sezen, and A. Yilmaz, "End-to-end deep learning methods for automated damage detection in extreme events at various scales," *arXiv preprint arXiv:2011.03098*, 2020.

[11] D. Vriesman, A. B. Junior, A. Zimmer, and A. L. Koerich, "Texture cnn for thermoelectric metal pipe image classification," in *2019 IEEE 31st Int. Conf. on Tools with Artificial Intell. (ICTAI)*. IEEE, 2019, pp. 569–574.

[12] T. Okura, N. Kasai, H. Minowa, and S. Okazaki, "Application of machine learning for data with an atmospheric corrosion monitoring sensor based on strain measurements," *Metals*, vol. 12, no. 7, p. 1179, 2022.

[13] B. Yin *et al.*, "Corrosion image data set for automating scientific assessment of materials," in *British Mach. Vis. Conf., BMVC 2021*.

[14] B. Yin *et al.*, "Corrosion assessment: Data mining for quantifying associations between indoor accelerated and outdoor natural tests," in *IEEE Int. Conf. on Big Data*, 2020, pp. 2929–2936.

[15] M. Drozda and A. Miszczyk, "Selection of organic coating systems for corrosion protection of industrial equipment," *Coatings*, vol. 12, no. 4, p. 523, 2022.

[16] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Intl. Conf. on Mach. Learning*. PMLR, 2017, pp. 2208–2217.

[17] R. Xu, G. Li, J. Yang, and L. Lin, "Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation," in *IEEE/CVF Intl. Conf. on Comput. Vis.*, 2019, pp. 1426–1435.

[18] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *The Journal of Mach. Learning Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[19] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," *arXiv:1705.10667*, 2017.

[20] X. Peng *et al.*, "Moment matching for multi-source domain adaptation," in *IEEE/CVF Intl. Conf. on Comput. Vis.*, 2019, pp. 1406–1415.

[21] Y. Li, L. Yuan, Y. Chen, P. Wang, and N. Vasconcelos, "Dynamic transfer for multi-source domain adaptation," in *IEEE/CVF Conf. on Comput. Vis. and Pattern Recog.*, 2021, pp. 10 998–11 007.

[22] H. Zhao *et al.*, "Adversarial multiple source domain adaptation," *Advances in Neural Inf. Process. Syst.*, vol. 31, 2018.

[23] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift," in *IEEE Conf. on Comput. Vis. and Pattern Recog.*, 2018, pp. 3964–3973.

[24] X. Peng, Z. Huang, X. Sun, and K. Saenko, "Domain agnostic learning with disentangled representations," in *Intl. Conf. on Mach. Learning*. PMLR, 2019, pp. 5102–5112.

[25] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *IEEE Intl. Conf. on Comput. Vis.*, 2017, pp. 754–763.

[26] Z. Cao, L. Ma, M. Long, and J. Wang, "Partial adversarial domain adaptation," in *Eur. Conf. on Comput. Vis. (ECCV)*, 2018, pp. 135–150.

[27] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, "Importance weighted adversarial nets for partial domain adaptation," in *IEEE Conf. on Comput. Vis. and Pattern Recog.*, 2018, pp. 8156–8164.

[28] G. Canonaco *et al.*, "A transfer-learning approach for corrosion prediction in pipeline infrastructures," *Appl. Intell.*, vol. 52, no. 7, pp. 7622–7637, 2022.

[29] A. H. B. Aqel, "Corrosion detection using transfer learning-based modeling for image classification," Master's thesis, 2019.

[30] J. Jiang, B. Chen, B. Fu, and M. Long, "Transfer-learning-library," https://github.com/thuml/Transfer-Learning-Library, 2020.

[31] Y. Jin, X. Wang, M. Long, and J. Wang, "Minimum class confusion for versatile domain adaptation," in *Eur. Conf. on Comput. Vis.* Springer, 2020, pp. 464–480.

[32] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *The Journal of Mach. Learning Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[33] American Society for Testing and Materials, "Standard test methods for cyclic (reversed) load test for shear resistance of vertical elements of the lateral force resisting systems for buildings," *ASTM Int. (1997 Edition)*, 2011.

[34] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *the Journal of Mach. Learning Res.*, vol. 12, pp. 2825–2830, 2011.