



HAL
open science

ICPR 2024 Competition on Domain Adaptation and GEneralization for Character Classification (DAGECC)

Sofia Marino, Jennifer Vandoni, Emanuel Aldea, Ichraq Lemghari, Sylvie Le Hégarat-Mascle, Frédéric Jurie

► **To cite this version:**

Sofia Marino, Jennifer Vandoni, Emanuel Aldea, Ichraq Lemghari, Sylvie Le Hégarat-Mascle, et al.. ICPR 2024 Competition on Domain Adaptation and GEneralization for Character Classification (DAGECC). International Conference on Pattern Recognition 2024, Dec 2024, Kolkata, India. pp.161-172, 10.1007/978-3-031-80139-6_12 . hal-04854883

HAL Id: hal-04854883

<https://hal.science/hal-04854883v1>

Submitted on 23 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ICPR 2024 Competition on Domain Adaptation and GEneralization for Character Classification (DAGECC)

Sofia Marino¹, Jennifer Vandoni¹, Emanuel Aldea², Ichraq Lemghari^{1,2}, Sylvie Le Hégarat-Masclé², and Frédéric Jurie³

¹ Safran Tech, Digital Sciences & Technologies Department, Magny-Les-Hameaux, France

² SATIE CNRS UMR 8029, Paris-Saclay University, Gif-sur-Yvette, France

³ University of Caen Normandie, France

Abstract. In this companion paper for the DAGECC (Domain Adaptation and GEneralization for Character Classification) competition organized within the frame of the ICPR 2024 conference, we present the general context of the tasks we proposed to the community, we introduce the data that were prepared for the competition and we provide a summary of the results along with a description of the top three winning entries. The competition was centered around domain adaptation and generalization, and our core aim is to foster interest and facilitate advancement on these topics by providing a high-quality, lightweight, real world dataset able to support fast prototyping and validation of novel ideas.

Keywords: Domain Adaptation · Domain Generalization · Competition · Image Classification.

1 Introduction

In the rapidly evolving field of computer vision, the ability to develop robust and accurate models that can perform well across diverse scenarios is of paramount importance. Two critical concepts that have emerged as cornerstones in achieving this objective are domain adaptation and domain generalization.

Domain adaptation is a transfer learning technique that aims to improve the performance of models in a target domain by leveraging knowledge from a different but related source domain. This approach is crucial because traditional machine learning models often struggle to generalize well when the data distribution changes between the training (source) and testing (target) domains. In real-world applications, such as autonomous driving [24][31][17], surveillance systems [14][5][18] or industrial quality control [20][27][30], this shift in data distribution is a common occurrence due to factors like varying weather conditions, lighting, or camera angles. Domain adaptation, therefore, plays a vital role in enhancing the model's resilience and adaptability to these changes, ensuring consistent performance and reliability.

On the other hand, domain generalization is a technique that seeks to develop models that can generalize well to any unseen target domain, without requiring access to any target domain data during training. This is a more challenging task than domain adaptation, as it requires the model to learn a universal representation that can capture the underlying data structure across multiple domains. The importance of domain generalization lies in its potential to create models that can truly adapt to any new environment, making them highly valuable for real-world applications where the target domain may be unknown or constantly changing [4][19][3][13].

The importance of datasets in the context of domain adaptation and domain generalization cannot be overstated. Diverse and representative datasets are crucial for training models that can effectively learn and adapt to different domains. For domain adaptation, the availability of well-labeled source and labeled or unlabeled target domain datasets is essential, as it allows the model to learn the underlying data distributions and map the knowledge from the source to the target domain. On the other hand, domain generalization typically relies on multi-domain datasets, where the model is trained on data from several different but related domains. This exposure to a variety of data distributions during training helps the model to learn a more universal representation, enabling it to generalize better to unseen target domains. Moreover, the quality and diversity of datasets can significantly influence the performance of these techniques, making the curation and selection of appropriate datasets a critical aspect of their success. Thus, the development and availability of comprehensive and diverse datasets are vital for advancing research and applications in domain adaptation and domain generalization.

The competition on Domain Adaptation and GEneralization for Character Classification (DAGECC) aims to foster advancements in domain adaptation and domain generalization techniques by introducing a novel dataset comprising real images of digits and characters captured from serial numbers on manufactured objects. In particular, we focus on an image classification task for industrial serial number recognition. In industrial environments, serial part numbers play a critical role in ensuring traceability and streamlining the management of diverse components (e.g., preventive maintenance, analytics, etc.). However, the manual process of reading and recording these serial numbers is both labor-intensive and prone to errors. As a result, the development of an automated serial number recognition system that can work seamlessly across multiple parts holds tremendous potential for improving operational efficiency in industrial environments.

To address this challenge, we are introducing Safran-MNIST, a brand new dataset suite that comprises images of serial numbers extracted from diverse avionic parts manufactured by Safran, an international high-technology group and world leader operating in the aviation (propulsion, equipment and interiors), defense and space markets. The content resembles that of the well-known MNIST dataset of handwritten digits [16], hence the name, but with a focus on industrial contexts. The Safran-MNIST dataset suite offers a realistic representation of industrial serial number images, encompassing variations in lighting

conditions, orientations, writing styles and surface textures. Thus, it represents a good compromise between 1) requiring a low computational cost for processing, in line with other MNIST variants [16][10], and 2) being highly relevant for real-world applications.

2 Competition overview

The competition encompasses two primary tasks: (i) Domain Generalization and (ii) Unsupervised Domain Adaptation. The central objective of this competition is to investigate and foster progress in domain adaptation and domain generalization methodologies, specifically in the context of character recognition.

2.1 Datasets

DAGECC competition is based on the brand new Safran-MNIST dataset suite, specially crafted for the application of character recognition. Characters have been extracted from larger images of Safran’s avionic parts acquired from aircraft engines returned from flights, that have to be periodically inspected for predictive maintenance and security reasons. They correspond to product references or serial numbers engraved on metal parts with various techniques which are widely used for traceability (laser, pencil, or micro percussion engraving).

The Safran-MNIST dataset suite comprises two datasets, namely Safran-MNIST-D⁴[28] and Safran-MNIST-DLS⁵[21], which are used in the two tasks respectively. Specifically:

Dataset for Task 1 - Domain Generalization: Target data for this task consists of real-world Safran-MNIST-D dataset containing RGB images of size 128×192 pixels representing digits ranging from 0 to 9. Each image in this dataset belongs to a distinct class. We annotated 1684 images for testing and 421 images for validation. Figure 1 shows the class distribution for each set. Since target data is not to be used in this task, the validation and testing datasets and associated ground-truth are released at the end of the competition. Sample examples from this dataset are shown in Figure 3a.

Dataset for Task 2 - Unsupervised Domain Adaptation: Target data for this task consists of real-world Safran-MNIST-DLS dataset containing variable size⁶ gray-scale images of 32 characters. Specifically, this dataset includes 10 digits ranging from 0 to 9, 20 letters (A, B, C, D, E, F, G, H, J, K, L, M, N, P, R, S, T, U, W and Y) and 2 symbols (/ and .). Each image in this dataset belongs to a distinct class. We manually annotated 3448 images for testing and 862 for validation. Furthermore, we provided 9314 unlabeled images for training. Figure 2 shows the class distribution for each set. The validation and testing datasets are released at the end of the competition as well as the ground-truth associated to all of the sets. Sample examples from this dataset are shown in Figure 3b.

⁴ <https://zenodo.org/records/13320997>

⁵ <https://zenodo.org/records/11093441>

⁶ from 18×30 pixels to 86×79 pixels

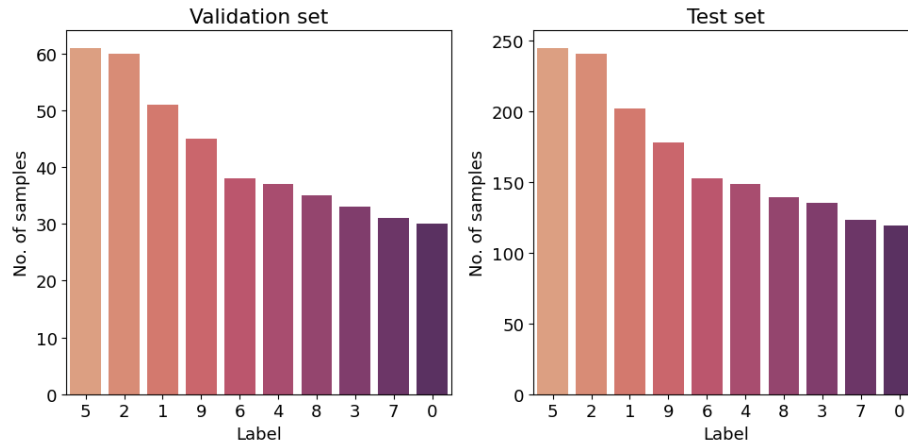


Fig. 1: Distribution of class samples in Safran-MNIST-D by sets.

An overview of both datasets is shown in Table 1.

Table 1: Overview of Safran-MNIST-D and Safran-MNIST-DLS datasets.

Dataset	Number of Images			Data Type	Number of Classes
	Training	Validation	Test		
Safran-MNIST-D	-	421	1684	RGB	10
Safran-MNIST-DLS	9314	862	3448	Gray-scale	32

2.2 Description of tasks

Task 1 - Domain Generalization: The aim of this task is to develop models that can generalize well to an unseen target domain, without requiring access to any target domain data during training. Thus, participants could not use any data from the target domain, which as the new Safran-MNIST-D dataset that contains images of numbers ranging from 0 to 9. In addition, participants were granted the freedom to use publicly available data or generated data from various source domains, such as MNIST [16], MNIST-M [11], SVHN [22], HASyV2 [26], DIGITS [1], EMNIST [6]. This allowed participants to leverage existing datasets or generate realistic synthetic data to train their models without compromising the requirement of no access to the target domain. Note that we deliberately not imposed any dataset for the source domain in order to let participants find and explore relevant datasets (or combinations/derivations of datasets) adapted to the task, provided that these are *publicly available*. Proprietary data were not allowed in order to facilitate the reproducibility of results.

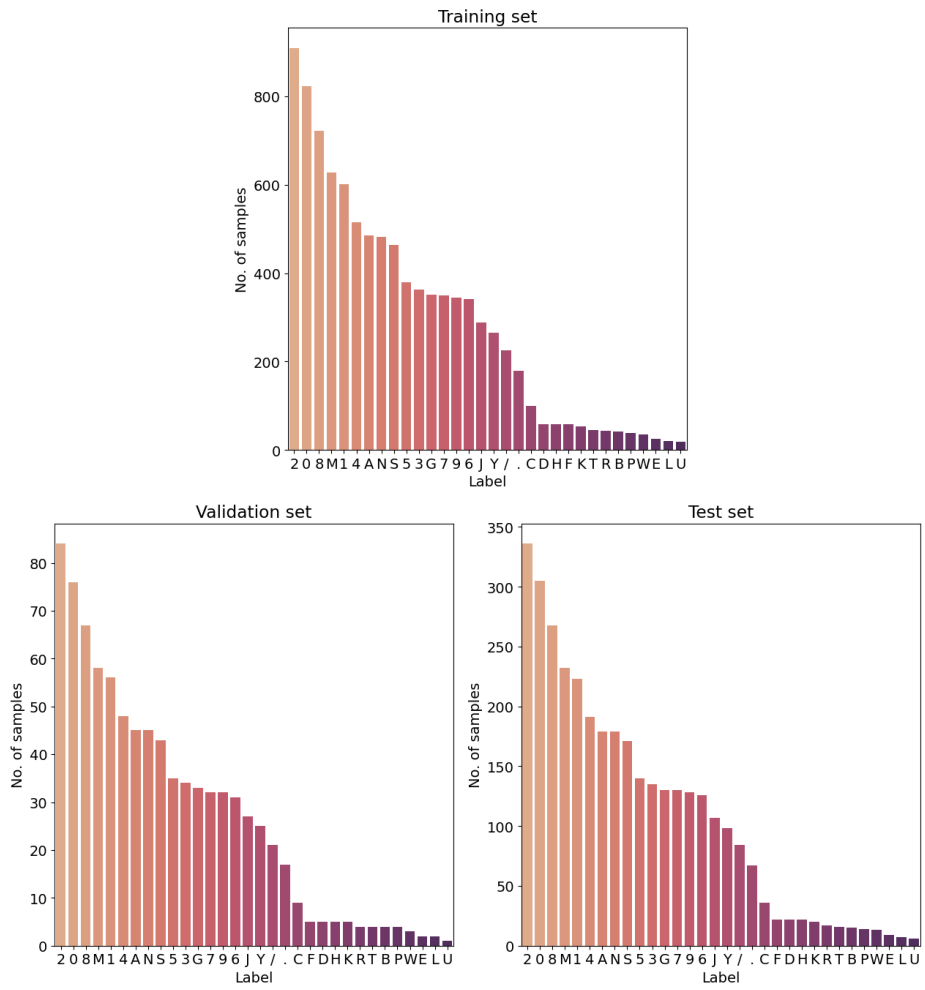


Fig. 2: Distribution of class samples in Safran-MNIST-DLS by sets.

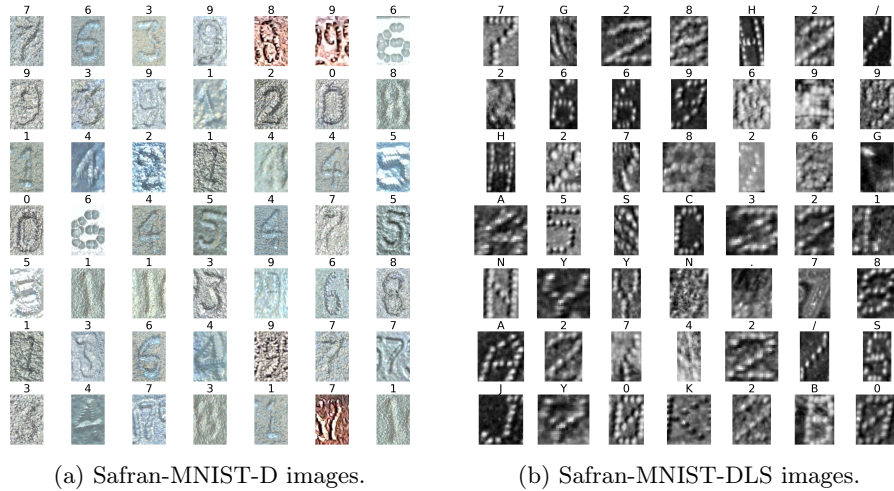


Fig. 3: Images extracted from Safran-MNIST dataset suite.

Task 2 - Unsupervised Domain Adaptation: This task is focused on unsupervised domain adaptation methods, in which we provided unlabeled data from a target domain: the new Safran-MNIST-DLS dataset, which comprises images of 32 classes depicting numbers, alphabetic characters, and symbols. Participants had access to the unlabeled target data during training. Furthermore, participants were tasked with sourcing a suitable dataset that can serve as the *source data* for domain adaptation. This could involve either finding an existing publicly available dataset or generating a synthetic dataset that aligns with the problem domain, using traditional image processing techniques or generative AI provided that the generative model has been trained on public data.

For both tasks, participants could use any publicly available and appropriately licensed data to pre-train their models.

Note that data from one task of the competition could not be exploited for the other task.

2.3 Evaluation metric

Macro-averaged F_1 -score was selected as evaluation metric in this competition to take into account the imbalanced nature of the dataset. This metric is described as follows:

$$F_1^{Macro} = \frac{\sum_{k=1}^K F_1^k}{K}. \quad (1)$$

For class k ,

$$F_1^k = \frac{2TP^k}{2TP^k + FP^k + FN^k}, \quad (2)$$

where TP^k is the number of true positives for class k , FP^k is false positives of class k and FN^k is false negatives of class k .

2.4 Submission system

Codabench [29] was used for submissions. Codabench is an open source platform that efficiently facilitates the organization of competitions. Participants could try out their methods, get real-time feedback and results on a competitive leaderboard. Each task had its individual competition website where participants had the opportunity to make submissions during two distinct phases. The first of these was the development phase, where the models were evaluated on the unseen validation set. In this phase, each teams had a quota of six submissions per day. This was followed by the final phase, during which the best final model selected by the teams was evaluated only once on the unseen test set.

3 Competition Results and Analysis

A total of 28 teams registered for the competition using the competition website⁷ and six of them made submissions on Codabench (including for both tasks). A total of 125 submissions were recorded for Task 1, compared to 181 for Task 2, indicating strong participant activity. Figure 4 summarizes the information of active teams.

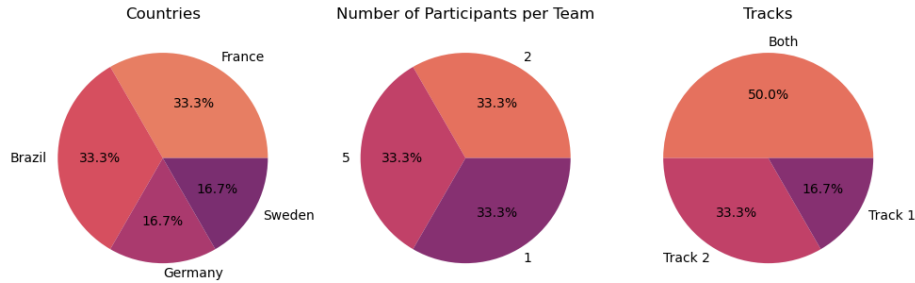


Fig. 4: Breakdown of competition participation.

3.1 Task 1: Domain Generalization

Results for Task 1 are shown in Table 2. Here we briefly describe the top three winners of this task.

⁷ <https://dagecc-challenge.github.io/icpr2024/>

1. *Team Deng*: constituted of two people, Victor Deng, student from ENS Paris in France, and Erwin Deng, student from CentraleSupélec also situated in France. The winning solution relies on the ResNet50 [12] architecture, initialized with pre-trained weights on the ImageNet [8] dataset. Several specific training techniques were used. The model was fine-tuned using a custom synthetic dataset, along with MNIST, SVHN, and MNIST-M. Due to the imbalanced nature of the dataset, the team used the WeightedRandomSampler provided by PyTorch [2] during training. To generate the custom dataset, they proposed the following method. Firstly, the background was generated using a random color, a random noise, followed by an embossing effect with custom filters, and a Gaussian blur, mimicking real-world imperfections. Secondly, many open-source fonts were used to generate digits from 0 to 9. Additionally, geometric transformations were applied to these digits (random rotations, scaling, translations). Finally, digits and backgrounds were combined with a random intensity followed by embossings effects, blurring and more.
2. *Fraunhofer IIS DEAL*: Anne Klier, Sai Rahul Kaminwar, Gabriel Dax, Sina Emami and Mohamed Hesham Ibrahim Abdalla from Fraunhofer IIS DEAL in Germany. The team used a GoogLeNet [25] pretrained on ImageNet and fine-tuned on various public datasets (MNIST, Chars74K, HASYv2, SVHN, USPS, SYN NUMBERS, MNIST-M). In addition, a synthetic dataset of 500 images per digit was generated using a pretrained stable diffusion model [9] which was required to provide weathered images of digits embossed or engraved on a metallic surface.
3. *JasonMendoza2008*: Romain Lhotte, registered as an individual participant, is a student from CentraleSupélec, located in France. His solution is based on gathering as much data from publicly available datasets, until reaching 200,000 images. Then, prediction was achieved by a weighted mean of the predictions of five distinct networks. These networks are based on AlexNet [15] and a baseline CNN architecture that was provided by organizers as example. The latter contains two convolutional layers: the first layer containing 16 channels with a kernel size of 5, and a second layer containing 32 channels with a kernel size of 5. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function along with a MaxPooling layer.

Analyzing the top-performing solutions for Task 1 - Domain Generalization reveals common strategies among the participating teams that likely contributed to their success. Each team leveraged pretrained deep learning architectures, such as ResNet50, GoogLeNet, and AlexNet, initialized with ImageNet weights. In an interesting differentiation strategy, the top two teams turned to synthetic datasets. Each of them crafted unique data generation methods: the winner utilized conventional image processing techniques, while the runner-up chose to leverage generative AI methods. Underscoring the importance of vast and diverse datasets, all teams focused on amassing as much data as possible from

various sources. Lastly, only the winning team deployed a technique to mitigate the challenge of imbalanced datasets.

Table 2: Competition results (list of all successful submissions on the final phase) for Task 1: Domain Generalization.

Rank	Team name	Codabench username	Macro F1-score
1	Team Deng	eternel22	0.82
2	Fraunhofer IIS DEAL	Fraunhofer IIS DEAL	0.74
3	JasonMendoza2008	JasonMendoza2008	0.50
4	Deep Unsupervised Trouble	heitor	0.44

3.2 Task 2: Domain Adaptation

The results for Task 2 are shown in Table 3. We describe briefly the top three winners of this task:

1. *Team Deng*: this is the same team that won Task 1 of competition. The proposed winning solution for this track bears similarity to the solution suggested for Track 1. Specifically, the pre-trained ResNet50 was employed as the architecture for classifying 32 different classes, including digits, letters, and symbols. The distinguishing factor lies in the datasets used for this track, which include EMNIST and custom-generated data. The procedure for generating this latter dataset closely aligns with the approach used in Track 1. Available unlabeled Safran-MNIST-DLS data was not used.
2. *Deep Unsupervised Trouble*: Artur Jordão Lima Correia, Bruno Lopes Yamamoto, Heitor Gama Ribeiro, Leandro Giusti Mugnaini, Victor Sasaki Venzel from the Escola Politécnica of the University of São Paulo. EMNIST [6] was used as the primary training dataset, along with additional samples generated using the Consola font in order to mimic the dotted style of the original data. Then, the competitors applied a preprocessing step based on Otsu thresholding [23], which was used to increase the training data size, as follows. From a single input image, multiple character versions were created by using binarization thresholds varying around the optimal value provided by Otsu’s algorithm, and then entropy and contrast filters were applied. A standard ResNet18 [12] architecture was used with the augmented dataset.
3. *Raul*: Raul Cavalcante, from Instituto de Matemática e Estatística da Universidade de São Paulo. The competing team generated a large amount of synthetic data (approximately 240k images) by creating 3D-rendered characters engraved into metallic plates using Blender [7]. Data diversity was obtained by varying programmatically parameters such as material properties, camera angles, deletion of random dots and applying random noise to dot locations. Then, a CNN architecture with residual connections was used for training.

In analyzing the methods employed by the top three teams for Task 2: Domain Adaptation, several prevailing trends emerge. Firstly, the top two teams leveraged pretrained deep learning architectures, such as ResNet50 and ResNet18, initialized with ImageNet weights. Secondly, it is interesting to note that none of the top three teams took advantage of the Safran-MNIST-DLS unlabeled dataset. Instead, the participating teams opted to generate their own datasets or utilize widely recognized datasets such as EMNIST.

Although the best performing methods relied on synthetic data generation, we commend the effort of the Machine Learning Group LTU team (Simon Corbillé and Elisa H. Barney Smith, Luleå University of Technology, Sweden) which decided to focus more than the competitors on the unlabeled corpus. The team proposed a learning process that trains a ResNet-34 model with labeled data, then cycles through a 1) prediction and pseudo label generation on the unlabeled data, 2) majority voting and 3) retraining using real labels and pseudo labels in order to maintain stability. We argue that, beyond the low hanging fruit of data generation, the most promising perspective for further improving the task performance is to better exploit the information contained in the unlabeled corpus.

Table 3: Competition results (list of all successful submissions on the final phase) for Task 2: Domain Adaptation.

Rank	Team name	Codabench username	Macro F1-score
1	Team Deng	eternel22	0.65
2	Deep Unsupervised Trouble	heitor	0.6
3	Raul	raulcd	0.52
4	Machine Learning Group LTU	simcor	0.44
5	JasonMendoza2008	JasonMendoza2008	0.14

4 Conclusion

In this paper, we have presented the competition on Domain Adaptation and Generalization for Character Recognition. We have introduced a new real-world dataset suite, Safran-MNIST, which comprises two datasets, Safran-MNIST-D and Safran-MNIST-DLS. A total of six teams submitted results on two different tasks, namely Domain Adaptation and Domain Generalization. The winner of both tracks was Team Deng from CentraleSupélec and ENS Paris, both institutions located in France. As expected given the lightweight format of the data, the architectures used by the participants were relatively compact, and their efforts focused mainly on scraping and generating synthetic data samples akin to the target distribution. We hope that the interest generated by the competition and the ease of use of the proposed datasets will sustain the improvement efforts in the future, and benefit the worldwide research community and the industry.

Acknowledgements A special thank to SAFRAN, especially Basile Musquer (SAFRAN Aircraft Engines) and Thierry Arsaut (SAFRAN Helicopter Engines) for participating in the acquisition of the images, the creation of the dataset and for allowing the public release of the data. We are also grateful to Codabench, and specifically to Adrien Pavão, for their great support.

References

1. Alpaydin, E., Kaynak, C.: Optical recognition of handwritten digits data set. UCI Machine Learning Repository **64**(5620) (1998)
2. Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C., Maher, B., Pan, Y., Puhersch, C., Reso, M., Saroufim, M., Siraichi, M.Y., Suk, H., Suo, M., Tillet, P., Wang, E., Wang, X., Wen, W., Zhang, S., Zhao, X., Zhou, K., Zou, R., Mathews, A., Chanan, G., Wu, P., Chintala, S.: PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In: 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24). ACM (Apr 2024). <https://doi.org/10.1145/3620665.3640366>, <https://pytorch.org/assets/pytorch2-2.pdf>
3. Chen, X., Liu, M., Niu, Y., Wang, X., Wu, Y.C.: Deep-learning-based lithium battery defect detection via cross-domain generalization. IEEE Access (2024)
4. Choi, S., Jung, S., Yun, H., Kim, J.T., Kim, S., Choo, J.: Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11580–11590 (2021)
5. Ciampi, L., Santiago, C., Costeira, J.P., Falchi, F., Gennaro, C., Amato, G.: Unsupervised domain adaptation for video violence detection in the wild. In: IMPROVE. pp. 37–46 (2023)
6. Cohen, G., Afshar, S., Tapson, J., Van Schaik, A.: Emnist: Extending mnist to handwritten letters. In: 2017 international joint conference on neural networks (IJCNN). pp. 2921–2926. IEEE (2017)
7. Community, B.O.: Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam, <http://www.blender.org>
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
9. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: Forty-first International Conference on Machine Learning (2024)
10. Ganin, Y., Lempitsky, V.S.: Unsupervised domain adaptation by backpropagation. In: Bach, F.R., Blei, D.M. (eds.) Proceedings of the 32nd International Conference

- on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015. JMLR Workshop and Conference Proceedings, vol. 37, pp. 1180–1189. JMLR.org (2015), <http://proceedings.mlr.press/v37/ganin15.html>
11. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. *Journal of machine learning research* **17**(59), 1–35 (2016)
 12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
 13. Hémadou, L., Vorobieva, H., Kijak, E., Jurie, F.: Beyond internet images: Evaluating vision-language models for domain generalization on synthetic-to-real industrial datasets. In: *Synthetic Data for Computer Vision Workshop @ CVPR 2024* (2024), <https://openreview.net/forum?id=BgpApqspGw>
 14. Himeur, Y., Al-Maadeed, S., Kheddar, H., Al-Maadeed, N., Abualsaud, K., Mohamed, A., Khattab, T.: Video surveillance using deep transfer learning and deep domain adaptation: Towards better generalization. *Engineering Applications of Artificial Intelligence* **119**, 105698 (2023)
 15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25** (2012)
 16. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
 17. Li, J., Xu, R., Ma, J., Zou, Q., Ma, J., Yu, H.: Domain adaptive object detection for autonomous driving under foggy weather. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 612–622 (2023)
 18. Li, S., Li, F., Li, J., Li, H., Zhang, B., Tao, D., Gao, X.: Logical relation inference and multiview information interaction for domain adaptation person re-identification. *IEEE Transactions on Neural Networks and Learning Systems* (2023)
 19. Liu, Y., Xu, Y., Zou, Z., Wang, Z., Zhang, B., Wu, L., Guo, Z., He, Z.: Adversarial domain generalization for surveillance face anti-spoofing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 6352–6360 (2023)
 20. Marino, S., Beuseroy, P., Smolarz, A.: Unsupervised adversarial deep domain adaptation method for potato defects classification. *Computers and Electronics in Agriculture* **174**, 105501 (2020)
 21. Marino, S., Vandoni, J., Lemghari, I., Musquer, B., Arsaut, T.: Safran-mnist-dls (0.0.1) [data set]. Zenodo (2024). <https://doi.org/https://doi.org/10.5281/zenodo.11093441>
 22. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning. In: *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011* (2011), http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf
 23. Otsu, N., et al.: A threshold selection method from gray-level histograms. *Automatica* **11**(285-296), 23–27 (1975)
 24. Shan, Y., Lu, W.F., Chew, C.M.: Pixel and feature level based domain adaptation for object detection in autonomous driving. *Neurocomputing* **367**, 31–38 (2019)
 25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1–9 (2015)

26. Thoma, M.: The hasyv2 dataset. arXiv preprint arXiv:1701.08380 (2017)
27. Thota, M., Kollias, S., Swainson, M., Leontidis, G.: Multi-source domain adaptation for quality control in retail food packaging. *Computers in Industry* **123**, 103293 (2020)
28. Vandoni, J., Marino, S., Lemghari, I., Arsaut, T., Musquer, B.: Safran-mnist-d (0.0.1) [data set]. Zenodo (2024). <https://doi.org/https://doi.org/10.5281/zenodo.13320997>
29. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. *Patterns* **3**(7), 100543 (2022). <https://doi.org/https://doi.org/10.1016/j.patter.2022.100543>, <https://www.sciencedirect.com/science/article/pii/S2666389922001465>
30. Zhang, S., Zhang, Q., Gu, J., Su, L., Li, K., Pecht, M.: Visual inspection of steel surface defects based on domain adaptation and adaptive convolutional neural network. *Mechanical Systems and Signal Processing* **153**, 107541 (2021)
31. Zhang, Y., David, P., Gong, B.: Curriculum domain adaptation for semantic segmentation of urban scenes. In: Proceedings of the IEEE international conference on computer vision. pp. 2020–2030 (2017)