Journal of Environmental Management 342 (2023) 118149

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Short communication

AI explainability framework for environmental management research

Mehrdad Arashpour

Department of Civil Engineering, Monash University, Melbourne, VIC, 3800, Australia

A R T I C L E I N F O

Handling Editor: Jason Michael Evans

Keywords:
Environmental crisis
Environmental management research
Explainable AI (XAI)
Management and valorization of solid waste
Multimodal and generative pre-trained transformers
Responsible and fair artificial intelligence
Vision-language deep learning models

A B S T R A C T

Deep learning networks powered by AI are essential predictive tools relying on image data availability and processing hardware advancements. However, little attention has been paid to explainable AI (XAI) in application fields, including environmental management. This study develops an explainability framework with a triadic structure to focus on input, AI model and output. The framework provides three main contributions. (1) A context-based augmentation of input data to maximize generalizability and minimize overfitting. (2) A direct monitoring of AI model layers and parameters to use leaner (lighter) networks suitable for edge device deployment, (3) An output explanation procedure focusing on interpretability and robustness of predictive decisions by AI networks. These contributions significantly advance state of the art in XAI for environmental management research, offering implications for improved understanding and utilization of AI networks in this field.

1. Introduction

The escalation of global populations, rapid urbanization, economic progress, and the subsequent surge in consumption patterns have collectively contributed to an unparalleled surge in waste generation on a global scale. Projections indicate that by the year 2050, the volume of waste produced will experience a staggering 73 per cent increase compared to the levels observed in 2020 (WorldBank, 2021). This trajectory highlights the urgent need to address the mounting waste crisis proactively and sustainably. (Papagiannis et al., 2021). Recent advancements in artificial intelligence (AI) and computer vision can facilitate the achievement of SDGs. In application fields, the power of AI networks is yet to be fully harnessed due to three shortcomings: weak generalizability and prevalent overfitting (Delanoë et al., 2023), high computation cost (Lundberg et al., 2020), and unknown reasons behind predictions (Sarabi et al., 2022). To address these shortcomings, this study develops an explainability framework with a triadic structure focusing on input, AI model and output. The framework aligns with the common AI network workflows (Fig. 1).

The explainability framework presented in this study establishes a strong link between AI theory and environmental management research, offering three significant contributions that enhance the field:

1. Improving AI network generalizability and minimizing overfitting. Classic AI networks are developed and trained for optimum performance on benchmark image datasets, including CIFAR (Krizhevsky and Hinton, 2009), Oxford Pets and Flowers (Nilsback and Zisserman, 2008; Parkhi et al., 2012) and ImageNet (original and cleaned-up labels) (Beyer et al., 2020). However, it is important to note that these benchmark datasets possess limited resemblance to real-world application data. This discrepancy becomes evident when considering waste images obtained from material recovery facilities (MRFs), which often exhibit contaminated objects and deformed boundaries (Bashkirova et al., 2022). In order to bridge this disparity, the first pillar of the framework focuses on systematic input data augmentation through the implementation of four primary strategy groups (Fig. 2). This rigorous augmentation approach ensures that AI networks exhibit maximum generalizability, aligning more closely with human expectations, while simultaneously minimizing the detrimental effects of overfitting.
2. Explaining the contributions made by model layers and related parameters to AI predictions. The developed explainability framework monitors kernel weight histograms to detect active contribution or redundancy of network layers to predictions. This is a non-trivial contribution towards deploying shallower (leaner) AI networks suitable for deployment on edge devices in application domains. XAI framework processes are continued by monitoring and optimizing model parameters with visualization of impact on AI prediction performance.
3. Explaining the interpretability and robustness of AI network outputs in the presence of noisy data. The developed explainability framework interprets AI network outputs by identifying important image

E-mail address: mehrdad.arashpour@monash.edu.

https://doi.org/10.1016/j.jenvman.2023.118149
Received 21 March 2023; Received in revised form 8 May 2023; Accepted 9 May 2023
Available online 13 May 2023

features that influence prediction. Moreover, robustness is monitored by generating counterfactual image examples using adversarial attacks. The existence of misleading images for state-of-the-art networks is a non-trivial observation considering wrong predictions are made with high confidence.

The framework's triadic structure creates a holistic perspective towards XAI by encapsulating the three main aspects of input, AI model and output.

2. Methods

Data. The utilization of appropriate data is paramount in illustrating the three fundamental pillars of the AI explainability framework. For this purpose, a publicly available dataset comprising 2100 meticulously annotated images depicting various categories of solid waste such as cardboard, glass, metal, paper, plastic, and trash was employed (accessible at https://github.com/garythung/trashnet). The image dataset was partitioned into training, validation and testing sets. The two training and validation sets were used in the development loop (Fig. 1), and the testing set was reserved for final inferencing. Preprocessing steps were taken to normalize images and resize them to the desired input size for deep AI networks.

Deep AI networks. PyTorch was chosen as the platform for the framework evaluation. Several factors were considered in selecting AI network architectures, including recency and popularity in environmental management research. PyTorch implementation of ResNet (He et al., 2016), ConvNext (Liu et al., 2022) and vision transformer (Dosovitskiy et al., 2020) and their pre-trained weights were obtained (https://pytorch.org/vision/stable/models.html). The networks are suitably regularized using batch normalization and dropout layers to prevent the coadaptation of neurons in the learning process. The building blocks of the networks to process inputs and predict target outputs are convolution (Eq. (1)) and encoder operations (Eq. (2)).

$$Conv^{f,s}(X)^{l}_{hv} = \sum_{r=0}^{f}\sum_{c=0}^{f} X[hs+r, vs+c].F^{l}_{hv}[r,c] + b^{l}_{hv} \quad (1)$$

where X is input, and *l, f* and *s* are layers, filters and strides. The horizontal *(h)* and vertical *(v)* positions of output are related to the column and row position of the filter matrix *(F)* and bias *(b)*.

$$h_t = f\left(W^{(hh)}_{h_{t-1}} + W^{(hx)}_{x_t}\right) \quad (2)$$

where $(h_1, \ldots, h_t)$ is the sequence of hidden states based on the input sequence $(x_1, \ldots, x_t)$.

For all three architectures, fine-tuning was required to classify six waste categories instead of 1000 classes on the ImageNet dataset. For this purpose, the fully connected layer in ResNet, the classifier layer in ConvNext and the head of vision transformer (ViT) were adjusted, and the standard processes of transfer learning were implemented. The deep learning networks were run using Cuda version 11.6 on an NVIDIA GPU (GeForce RTX 3080).

2.1. Input data augmentation to maximize AI network generalizability and minimize overfitting

The first requirement for AI networks as predictive tools is working on input data. Data preparation processes, including augmentation, effectively increase model generalizability and prevent overfitting (Tremblay et al., 2018). An indicator of overfitting is the gap between performance on training and validation runs (Sun et al., 2019). Significant gaps between the two imply that networks memorize prediction targets instead of learning essential data features. In the context of environmental management research, data augmentation should be undertaken so that the generalization of the AI network is explainable to and aligned with the expectations of subject matter experts. For example, in the waste classification problem, original and augmented images are expected to be classified in the same category.

Data augmentation is significant in application fields, including waste recycling, since AI networks are developed and trained using benchmark datasets in other fields. The uniqueness of waste image datasets regarding object contamination and deformations raises important questions about the generalizability of networks in this field. Fig. 2 illustrates the implementation of input data augmentation across four primary categories aimed at enhancing the generalizability of the AI network. The random affine technique is employed within the geometric augmentation strategies, involving a combination of image transformations such as rotation, translation, scale, and shear to modify the input data. The successful prediction of the network on augmented images demonstrates its capability to generalize effectively across diverse image variations. Additionally, the random crop augmentation method facilitates the precise resizing of input images to the desired dimensions, ensuring optimal compatibility for network processing (Arashpour et al., 2022). Optional padding can be applied if the cropped image is small in dimension. Using the random flip strategy, waste images are vertically or horizontally flipped by a given probability. Random perspective transformation is among geometric augmentations that distort input images to challenge AI networks to stimulate feature learning.

The second category of photometric augmentations is essential in preventing the network from memorizing target predictions based on the color patterns in the input data (Sen et al., 2023). Color jitter transforms images by randomly modifying the saturation, brightness, hue and contrast of input data. Notably, domain knowledge should inform augmentations, so that model behavior is explainable in application fields such as waste recycling. The second augmentation under the umbrella of photometric transformations randomly inverts image colors by a given probability. Random grayscale is a common strategy that minimizes the influence of color on target predictions. Random solarize is another example of photometric augmentations that randomly invert all pixel values above a user-specified threshold.

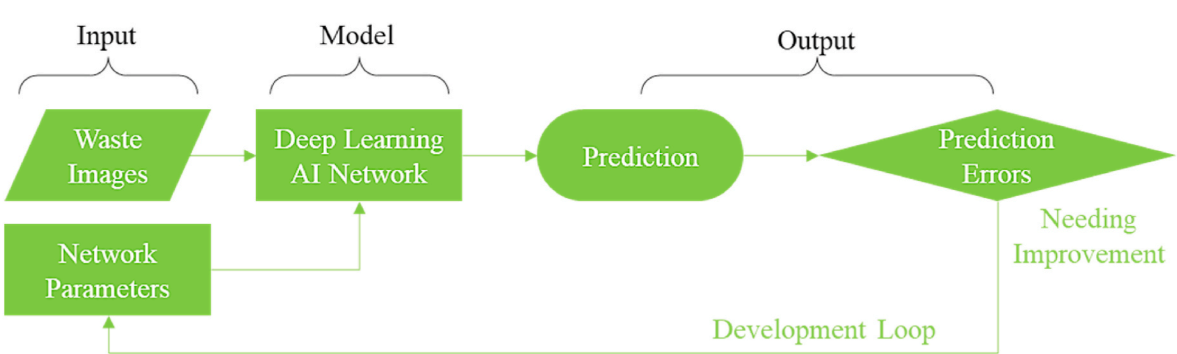

Fig. 1. Triadic structure of AI networks: input, model and output.

Considering the significant contamination of solid waste objects, photometric augmentations are very useful in increasing the generalizability of networks.

The third category of input data augmentation performs functional transformations, providing flexibility and control to field experts. For instance, equalize transformation creates a distribution of output grayscale values by nonlinear mapping of inputs. The morphology of waste images can be transformed, and a see-through effect is created using elastic augmentation. The augmentation strategy of posterize alters color channels by reducing the number of bits. Image borders can be added as another augmentation strategy belonging to functional transformations.

Smart augmentation strategies can be automatically searched to prevent overfitting and maximize network generalizability. For example, trivial augment (Müller and Hutter, 2021) selects an optimum set of transformations with low computation cost. Auto augment (Cubuk et al., 2018) is learnt from datasets such as ImageNet, SVHN and CIFAR10 and is considered a dataset-dependent strategy for augmenting input data. Rand augment (Cubuk et al., 2020) is the third smart strategy that reduces the search space to select and implement optimum transformations with desirable strength. AugMix (Hendrycks et al., 2019) applies a chain of augmentations on input images and then mixes results using elementwise convex combinations.

3. Results

Proper learning of essential image features by deep AI networks is critical to tackling the common problem of overfitting in application fields, including environmental management research. Input data augmentation prevents AI networks from memorizing target predictions and creating a gap between training and validation performances. Input data augmentation needs to be informed by expert knowledge, especially in environmental management, where solid objects are significantly deformed and contaminated. The focus on geometric and photometric augmentations in this field is consequently significant.

3.1. AI network explanations on the contribution of layers and parameters

Deep AI networks are powerful predictive tools to automate perception processes in various domains. The mainstream research on AI networks has focused on developing deep and complex networks to improve prediction performance (Prenafeta-Boldú and Kamilaris, 2019). The main challenge in application domains is the explainability of networks often initialized using random or pre-trained weights (Newman and Furbank, 2021). Those weights and network parameters influence outputs in AI perception, including classification, detection and segmentation tasks. Since 2012, deep AI networks have used a vast number of parameters, which was the case for AlexNet (Krizhevsky et al., 2012), with more than 60 million parameters. Later on, deeper networks such as ResNet 101 (He et al., 2016) became more accurate in predictions by using more than 80 million parameters. State-of-the-art networks, including vision transformers (ViT) use even more parameters (Dosovitskiy et al., 2020). An example is ViT-Large, which contains more than 300 million parameters.

In adopting AI networks in environmental management research, proper analysis is required to evaluate the contribution of layers and parameters. Using leaner models with acceptable predictive performance can increase the explainability of outputs for field experts. To demonstrate the three pillars of the AI explainability framework, state-of-the-art deep networks to classify 2100 images of solid waste were used (methods). Three networks of ResNet, ConvNext and ViT were implemented in PyTorch and kernel weight histograms were monitored. Notably, as depicted in Fig. 3, the weight histograms for all three

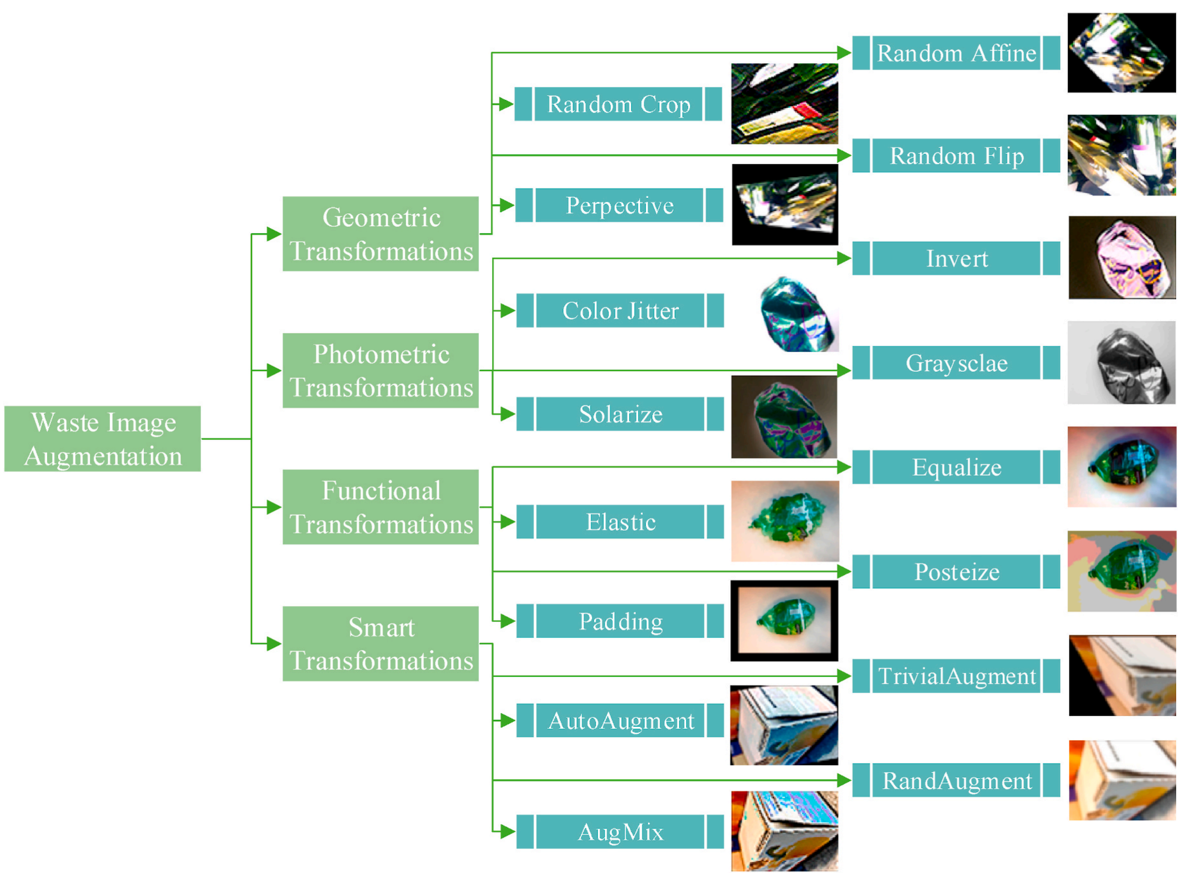

Fig. 2. Input augmentations to maximize AI network generalizability and minimize overfitting.

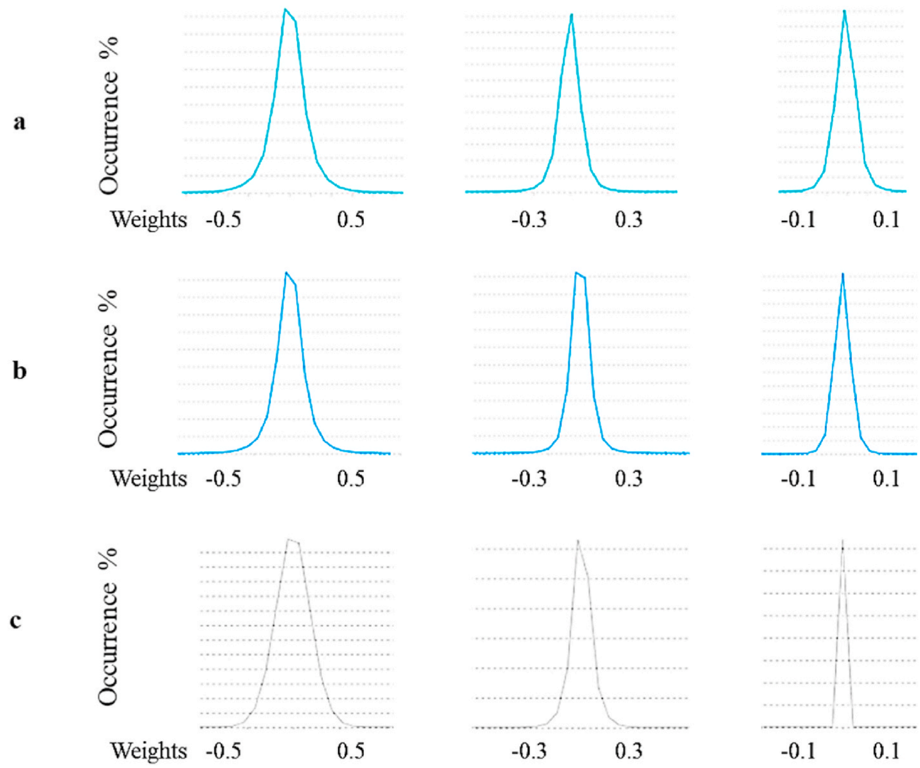

Fig. 3. Weight histograms in deep AI networks. a, ResNet weights related to shallow, intermediate and deep layers. b, ConvNext weights. c, Vision transformer weights.

networks exhibit a diminishing trend towards zero in deeper layers, indicating a minimal contribution to the overall output. In such instances, it is advisable to consider employing lighter variants of the state-of-the-art algorithms to strike a balance between accuracy and explainability, thus addressing the disparity between these two crucial aspects.

Weight histograms related to the experimentation on the waste image dataset show that ResNet has more than the necessary convolutional layers considering the classification scope (Fig. 3a). Similar patterns are observed in ConvNext histograms (Fig. 3b) and ViT histograms (Fig. 3c) where weights are centered around zero in deeper layers. The observations are consistent with the literature (Lee and Jhang, 2021; Shi et al., 2022), confirming that classic AI networks are often too deep for application fields. This is because AI networks are trained on extensive datasets, including ImageNet with 1000 object categories. Most problems in application fields, including waste recycling, have focused scopes with few target outputs.

A separate set of experiments is required to adjust the hyperparameters of deep AI networks. Following the above experiments on kernel weight histograms, leaner versions of the candidate networks were implemented, including ResNet 18, ConvNext-tiny and ViT-base. A grid search strategy was implemented to find optimum hyperparameters, followed by parallel coordinates visualization (Fig. 4).

The parallel coordinates plot in Fig. 4 shows the impact of changing the backbone architectures (ViT, ResNet, ConvNext) and associated learning rates on the loss and accuracy of training and validation. A nontrivial observation in Fig. 4 relates to lower validation than training loss values. This explains the effectiveness of input augmentation in the first

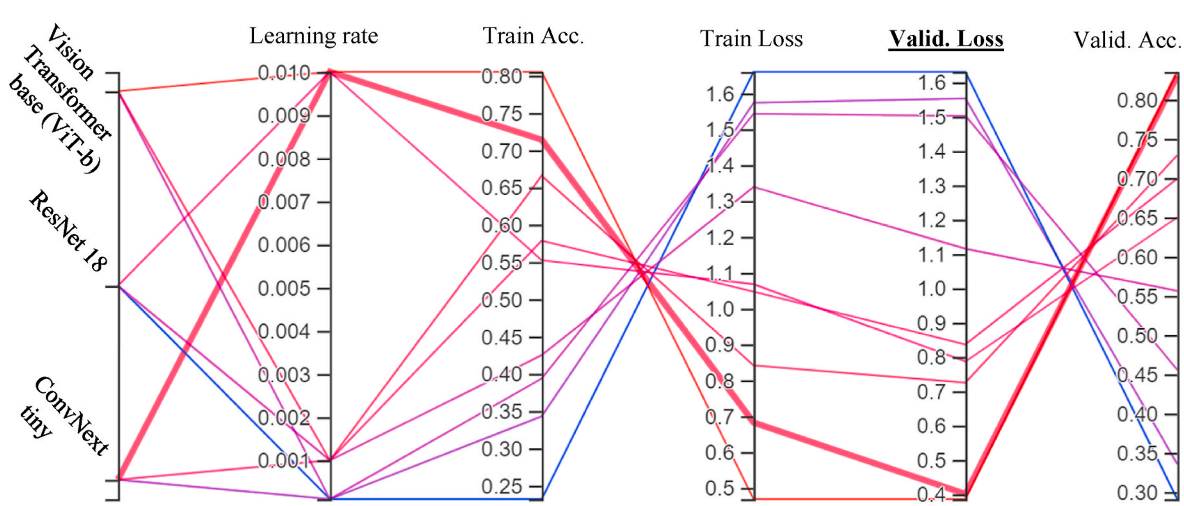

Fig. 4. Explaining network parameter impact on prediction performance.

stage of the explainability framework to overcome the common problem of overfitting in AI networks. The parallel coordinates plot clearly shows that the lowest validation losses belong to ViT-base and ConvNext-tiny when a learning rate of 0.010 is used. Considering the validation loss as the performance metric of choice in this waste sorting problem, ViT-base with a learning rate of 0.001 and ResNet 18 with a learning rate of 0.010 are the following high-performing classifiers.

3.2. AI network explanations on interpretability and robustness of outputs

When incorporating AI networks into environmental management research, it is crucial to prioritize the interpretability of outputs (Zhang et al., 2023). Understanding how network decisions are made contributes to establishing responsible AI in cross-disciplinary research fields, including environmental management. For example, a unique feature in waste images is the significant deformation of object boundaries (Bashkirova et al., 2022). To elucidate the importance of image pixels for decision-making by AI networks, specific methods such as integrated gradients with noise tunnel have been employed, as depicted in Fig. 5 (classification of waste paper in this experiment).

AI network outputs can be interpreted using local explanations, including Shapely Additive exPlanations or SHAP (Lundberg and Lee, 2017). As shown in Fig. 5, the importance of local features for network predictions can be found by introducing white noise to images and computing the output gradients. Occlusion is also used to mask critical regions of images to find the threshold for wrong decision-making (misclassification of waste plastic). Such interpretability experiments are instrumental in bridging the gap between model performance and explainability by visualizing significant contributing features to the outputs of AI networks (Chen et al., 2023).

Evaluation of output robustness is another aspect of the explainability framework. Perturbing images generate counterfactual (adversarial) examples, which result in wrong decisions with a high level of confidence. Fast Gradient Sign Method (FGSM) (Goodfellow et al., 2014) creates adversarial image examples by introducing noise with the same sign as the cost function of AI networks. Such images explain why state-of-the-art AI networks are misled in their predictions and highlight the importance of focusing on explainability in addition to accuracy. FGSM creates one-step attacks, but projected gradient descent (PGD) is an iterative attacker (Madry et al., 2017). PGD generates counterfactual image examples by taking multiple steps to analyze the robustness of AI network predictions. PGD example in Fig. 5 shows the adversarial example of cardboard waste, which is almost identical to the original image and yet results in misclassification by state-of-the-art AI networks.

AI network robustness can be analyzed by finding the minimum perturbation required for wrong decision-making. Fig. 5 shows the minimum blurring attack that results in the misclassification of metal waste. Another form of adversarial image generation is feature ablation that masks important regions to mislead the AI network. A feature mask can be passed to ablate top grouping features together. Overall, the more robust AI output predictions, the more intense adversarial perturbations are required to confuse state-of-the-art networks.

4. Discussion

Explainable AI (XAI) in environmental management research requires a holistic focus on three main aspects: input, AI model and output. Generalizable AI networks with expected behavior from the human perspective are required for application fields. Smart trash bins are examples of solutions to sort waste using AI perception power. Explainable networks can facilitate the robotization of waste sorting in material recovery facilities (Krechetov et al., 2018; Song et al., 2023). Moreover, XAI can be used to enforce environmental protection regulations and detect unlawful activities related to waste disposal and handling (Dong et al., 2022).

Input data augmentation is critical to AI network generalizability and prevention of overfitting as a common issue in application fields, including environmental management research. The first framework pillar is a systematic input data augmentation using four strategy groups to assist newcomers to the AI field and subject matter experts in environmental management. Augmentation strategies can be deployed individually or combined depending on available waste data and the use case. Smart augmentations are state-of-the-art image transformations that optimally combine geometric, photometric and functional transformations.

The second component of the explainability framework focuses on monitoring AI networks and their kernel weight histograms. This component is critical to environmental management research since complex and deep AI networks are developed and trained on large datasets containing many everyday object categories (Lin et al., 2021). Such networks impose unreasonably high computation costs to application fields outside computer science, in which real-world problems

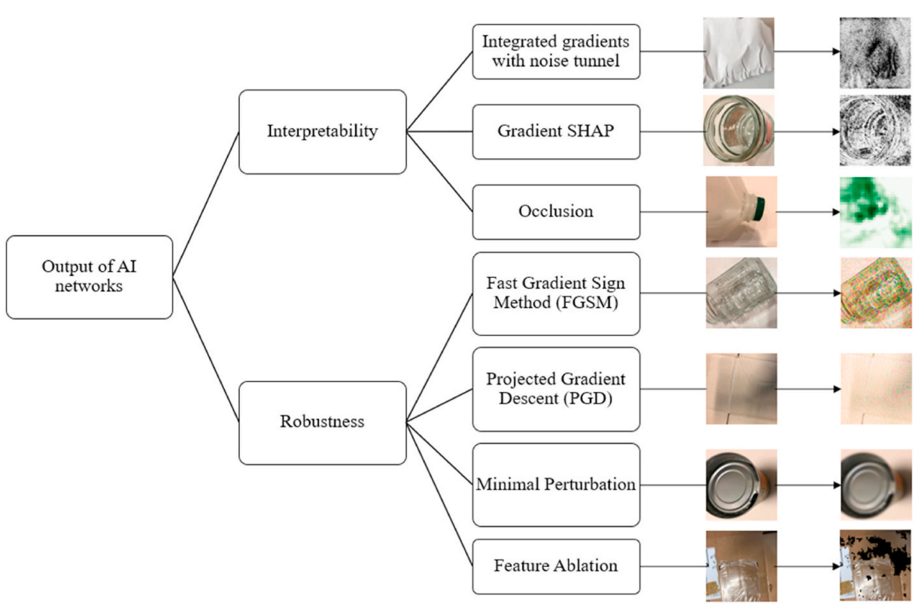

Fig. 5. Interpretability and robustness of AI network outputs.

have narrower but deeper scopes. Monitoring weight histograms can explain the contribution of network layers to the final prediction of AI networks (Dang et al., 2022). Observing the concentration of weights around zero indicates the possibility of implementing leaner variants of deep AI networks, which is preferable for deployment on edge devices. Monitoring AI networks can be continued by experimentation on optimal network parameters that explain fluctuations in predicting target outputs (Dulebenets, 2021; Zhao and Zhang, 2020).

The third pillar of AI network explainability is monitoring AI network outputs regarding interpretability and robustness. Critical features leading to output predictions are identified using noise tunnel, gradient SHAP and occlusion strategies. The robustness of AI network outputs is explained and monitored by generating counterfactual image examples that mislead state-of-the-art networks. Adversarial attacks, including FGSM, PGD, minimal perturbation and feature ablation, challenge networks to predict wrongly with high confidence levels. The three aspects of the proposed XAI framework contribute towards establishing responsible AI in cross-disciplinary fields, including environmental management research.

5. Conclusions

The implementation of AI networks in environmental management research is fast growing. However, the tension between performance and explainability remains, even for state-of-the-art networks. This study proposes an explainability framework with a triadic structure focusing on input, AI model and output. The framework is a non-trivial departure from the trend in existing XAI methods, including SHAP and its variants, which mainly emphasize output explanations.

There are ample opportunities for future research based on current limitations. Specialized input datasets for different aspects of environmental management research are required to generate context-specific augmentation strategies that maximize the generalizability of AI networks. Moreover, advanced visualization tools are required to clearly explain the contribution of network layers and related parameters to AI predictions. Finally, output explainability needs improvement to include more criteria than interpretability and robustness.

Credit author statement

MA: Conceptualization, Methodology, Software, Data curation, Writing the original draft, Writing review & editing.

Code availability

Codes for the AI networks in this paper are available at https://github.com/pytorch/vision/tree/main/references/classification#vit_b_16, https://github.com/pytorch/vision/tree/main/references/classification#convnext, and https://github.com/pytorch/vision/tree/main/references/classification#resnet.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The contributions and comments of ASCII Lab (https://www.monash.edu/ascii) members on this manuscript are greatly appreciated. This work is part of a research project (Project No IH18.04.1) funded by the Australian Research Council (ARC) Industrial Transformation Research Hub (ITRH) Scheme (Project ID: IH180100010).

References

Arashpour, M., Kamat, V., Heidarpour, A., Hosseini, M.R., Gill, P., 2022. Computer vision for anatomical analysis of equipment in civil infrastructure projects: Theorizing the development of regression-based deep neural networks. Autom. ConStruct. 137, 104193 https://doi.org/10.1016/j.autcon.2022.104193.
Bashkirova, D., Abdelfattah, M., Zhu, Z., Akl, J., Alladkani, F., Hu, P., Ablavsky, V., Calli, B., Bargal, S.A., Saenko, K., 2022. ZeroWaste dataset: towards deformable object segmentation in cluttered scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Beyer, L., Hénaff, O.J., Kolesnikov, A., Zhai, X., Oord, A.v. d., 2020. Are We Done with Imagenet? arXiv Preprint arXiv:2006.07159.
Chen, A., Jacob, M., Shoshani, G., Charter, M., 2023. Using computer vision, image analysis and UAVs for the automatic recognition and counting of common cranes (Grus grus). J. Environ. Manag. 328, 116948 https://doi.org/10.1016/j.jenvman.2022.116948.
Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V., 2018. Autoaugment: Learning Augmentation Policies from Data. arXiv. *preprint arXiv:1805.09501*.
Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V., 2020. Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
Dang, K.B., Dang, V.B., Ngo, V.L., Vu, K.C., Nguyen, H., Nguyen, D.A., Nguyen, T.D.L., Pham, T.P.N., Giang, T.L., Nguyen, H.D., Hieu Do, T., 2022. Application of deep learning models to detect coastlines and shorelines. J. Environ. Manag. 320, 115732 https://doi.org/10.1016/j.jenvman.2022.115732.
Delanoë, P., Tchuente, D., Colin, G., 2023. Method and evaluations of the effective gain of artificial intelligence models for reducing CO2 emissions. J. Environ. Manag. 331, 117261 https://doi.org/10.1016/j.jenvman.2023.117261.
Dong, Z., Chen, J., Lu, W., 2022. Computer vision to recognize construction waste compositions: a novel boundary-aware transformer (BAT) model. J. Environ. Manag. 305, 114405 https://doi.org/10.1016/j.jenvman.2021.114405.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv. *preprint arXiv:2010.11929*.
Dulebenets, M.A., 2021. An Adaptive Polyploid Memetic Algorithm for scheduling trucks at a cross-docking terminal. Inf. Sci. 565, 390–421.
Goodfellow, I.J., Shlens, J., Szegedy, C., 2014. Explaining And Harnessing Adversarial Examples. arXiv preprint arXiv:1412.6572.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
Hendrycks, D., Mu, N., Cubuk, E.D., Zoph, B., Gilmer, J., Lakshminarayanan, B., 2019. Augmix: A Simple Data Processing Method to Improve Robustness and Uncertainty. arXiv. *preprint arXiv:1912.02781*.
Krechetov, I.V., Skvortsov, A.A., Poselsky, I.A., Paltsev, S.A., Lavrikov, P.S., Korotkovs, V., 2018. Implementation of automated lines for sorting and recycling household waste as an important goal of environmental protection. J. Environ. Manag. Tourism 9 (8), 1805–1812. https://doi.org/10.14505/jemt.v9.8(32).21.
Krizhevsky, A., Hinton, G., 2009. Learning Multiple Layers of Features from Tiny Images.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst.
Lee, G., Jhang, K., 2021. Neural network analysis for microplastic segmentation. Sensors 21 (21), 7030.
Lin, K., Zhao, Y., Tian, L., Zhao, C., Zhang, M., Zhou, T., 2021. Estimation of Municipal Solid Waste Amount Based on One-Dimension Convolutional Neural Network and Long Short-Term Memory with Attention Mechanism Model: A Case Study of Shanghai, vol. 791. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2021.148088.
Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. https://doi.org/10.1038/s42256-019-0138-9.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A., 2017. Towards Deep Learning Models Resistant to Adversarial Attacks. arXiv. Preprint arXiv:1706.06083.
Müller, S.G., Hutter, F., 2021. Trivialaugment: tuning-free yet state-of-the-art data augmentation. Proc. IEEE/CVF Int. Conf. Comput. Vis.
Newman, S.J., Furbank, R.T., 2021. Explainable machine learning models of major crop traits from satellite-monitored continent-wide field trial data. Nat. Plants 7 (10), 1354–1363. https://doi.org/10.1038/s41477-021-01001-0.
Nilsback, M.-E., Zisserman, A., 2008. Automated Flower Classification over a Large Number of Classes. 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing.
Papagiannis, F., Gazzola, P., Burak, O., Pokutsa, I., 2021. A European household waste management approach: Intelligently clean Ukraine. J. Environ. Manag. 294, 113015 https://doi.org/10.1016/j.jenvman.2021.113015.
Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C., 2012. Cats and dogs. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition.

Prenafeta-Boldú, F.X., Kamilaris, A., 2019. AI assists in locating hidden farms. Nat. Sustain. 2 (4), 262–263. https://doi.org/10.1038/s41893-019-0264-8.

Sarabi, S., Han, Q., de Vries, B., Romme, A.G.L., Almassy, D., 2022. The nature-based solutions case-based system: a hybrid expert system. J. Environ. Manag. 324, 116413 https://doi.org/10.1016/j.jenvman.2022.116413.

Sen, S., Singh, K.P., Chakraborty, P., 2023. Dealing with imbalanced regression problem for large dataset using scalable Artificial Neural Network. N. Astron. 99, 101959 https://doi.org/10.1016/j.newast.2022.101959.

Shi, B., Patel, M., Yu, D., Yan, J., Li, Z., Petriw, D., Pruyn, T., Smyth, K., Passeport, E., Miller, R.J.D., Howe, J.Y., 2022. Automatic quantification and classification of microplastics in scanning electron micrographs via deep learning. Sci. Total Environ. 825 https://doi.org/10.1016/j.scitotenv.2022.153903.

Song, J., Chen, Y., Luan, F., 2023. Air pollution, water pollution, and robots: is technology the panacea. J. Environ. Manag. 330, 117170 https://doi.org/10.1016/j.jenvman.2022.117170.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S., 2018. Training deep networks with synthetic data: bridging the reality gap by domain randomization. Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops.

World Bank, 2021. Bridging the Gap in Solid Waste Management. World Bank Publications.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., Wen, H., 2023. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J. Environ. Manag. 332, 117357 https://doi.org/10.1016/j.jenvman.2023.117357.

Zhao, H., Zhang, C., 2020. An online-learning-based evolutionary many-objective algorithm. Inf. Sci. 509, 1–21.