

---

# Physics-Aware Image-to-Image Translation to Explore Long-Life All-Solid-State Batteries

---

Masaki Adachi<sup>1,2,\*</sup>, Ignas Budvytis<sup>1</sup>, Caterina Ducati<sup>2</sup> & Roberto Cipolla<sup>1</sup>

<sup>1</sup> Department of Engineering  
University of Cambridge  
ma921@cam.ac.uk

<sup>2</sup> Department of Materials Science & Metallurgy  
University of Cambridge

## Abstract

We introduce an image-to-image translation model, which uses physics-aware representations. These representations were obtained through a pretext task using simulations for self-supervised curriculum learning. During this pretraining phase, the generative model is trained to acquire generalised discriminative features to predict battery performances, classify the manufacturing conditions, and generate interpolated images. This model succeeded in organising the degradation patterns in all-solid-state batteries by visualising latent variables similarity. The latent space allows us to explore a battery design for a greatly extended lifespan. The proposed design predicts improved capacity retention from 86.72% to 96.2% after 300 cycles.

## 1 Introduction

All-solid-state batteries (ASSBs) have been considered one of the most promising candidates revolutionising the electric vehicle. ASSBs are not only high-powered but also high-capacity [4], which promises to extend the cruising range and allow quicker acceleration than liquid-type batteries. In contrast, ASSBs have difficulty making contact interfaces between solid particles. Repeated charge-discharge cycles cause disconnects in inter-particle interfaces, resulting in shortening battery lifespan. This degradation phenomenon is derived from a crack propagation in a heterogeneous powder mixture, making it challenging to accurately predict using the existing simulation methods. The problems of interfacial contact and preventing degradation are some of the most significant barriers to the practical application of ASSBs.

Recently, a data-driven approach succeeded in accurately predicting liquid-type battery life in various charge-discharge patterns [2, 11, 15, 22], which implies this technique is more effective than a simulation in predicting battery degradation. However, whilst these studies provide optimal control methods for suppressing identical battery degradation, they do not provide a way to explore a better battery design. As such, a new data-driven method was investigated in order to organise the relationship between microstructures and lifespan. 17,700 scanning electron microscope (SEM) images were acquired from various battery designs with relevant multimodal data (manufacturing conditions, battery lifespan, and output power). This study will discuss the results of recognised degradation patterns acquired from the use of an image translator between two domains paired in the fresh and degraded states.

## 2 Context-aware image-to-image translation

Conventionally, CycleGAN [23] has been commonly used for an image-to-image translation. This model re-converts the converted image to the original domain and minimises the difference between

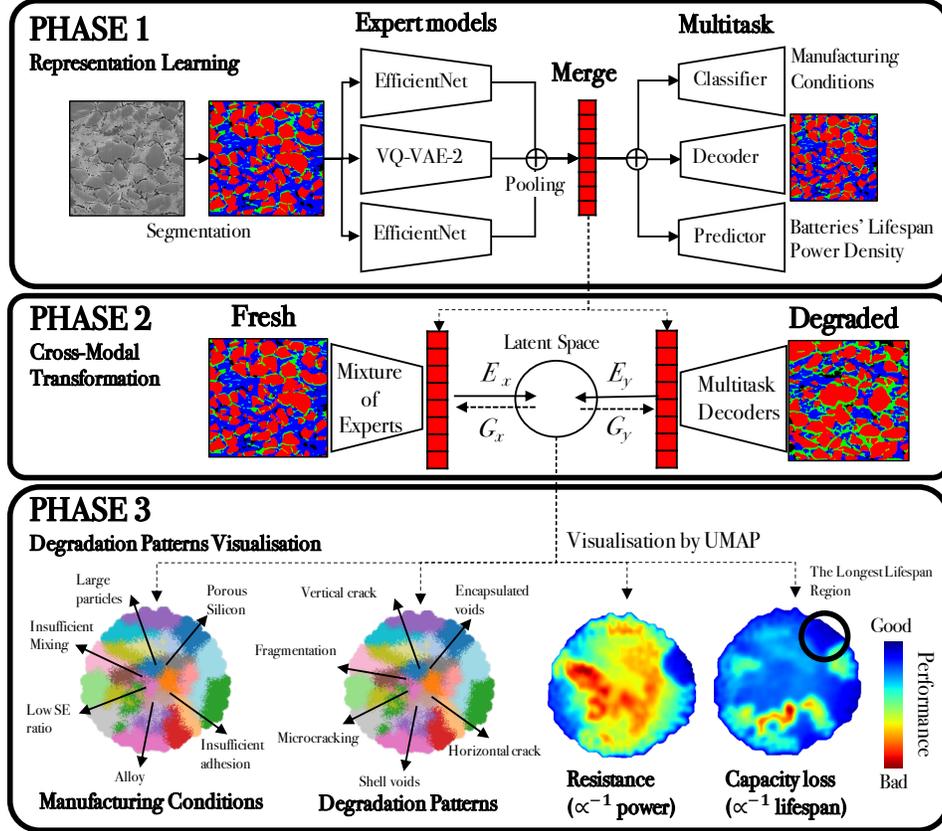


Figure 1: The proposed context-aware cross-domain transformation model.

the converted and original images to satisfy cycle-consistency. This constraint successfully achieves natural image translation in many applications. However, it lacks a context-aware translation; it does not change the conversion manner corresponding to the original images but instead tends to be an averaged manner [3]. Additionally, Generative Adversarial Networks (GAN) have a problem of mode collapse [1], which could be a barrier for context-aware translation.

## 2.1 Architecture

The proposed context-aware image-to-image translation is described in Figure 1. In order to eliminate the influence of imaging conditions, the raw SEM images were preprocessed into segmented images for each of the four constituent materials by the fine-tuned Feature Pyramid Network [7] from the ImageNet pretrained model [20]. In phase 1, multitask representation learning is used to obtain a context-aware representation by merging multimodal data. In phase 2, the resulting representations are cross-domain transformed by Variational Canonical Correlation Analysis (VCCA) [18] with cycle-consistency loss added to the original model. In phase 3, the degradation patterns were visualised by embedding the inter-domain latent variables by UMAP [8].

The multitask representation learning comprises of the following three tasks:

- (task 1) classifying manufacturing conditions from images
- (task 2) reconstructing images
- (task 3) predicting battery performance from images

We prepared the expert model for each task, which are EfficientNet-b1 + AdaCos [19, 21], Vector-Quantized Variational Autoencoder (VQ-VAE) [13], and EfficientNet-b1 + L1 regularisation, respectively. AdaCos can learn an optimal distance metric to measure the similarity among the images of

an unspecified number of classes. In addition, VQ-VAE can generate high-fidelity images equivalent to state-of-the-art GAN, and it is also effective in preventing mode collapse [12].

## 2.2 Training Procedure

Firstly, the expert models for three distinct tasks were pretrained separately. Subsequently, the pooling layer was introduced to connect the last layer of all expert models. The combined model was trained by multitasking, using uncertainty to weigh losses [5]. Furthermore, self-supervised curriculum learning (SSCL) was introduced to alleviate the conflict between reconstruction and classification [9]. Simulated values were used as a pretext task before the prediction of experimental data. The simulated values are the battery power density and capacity simulated by the Doyle-Fuller-Newman model [16], the ionic path tortuosity simulated by the random walk [17], and the other essential shape descriptors [6], which are taught as a curriculum in the order of ease of prediction.

Secondly, thusly trained multitasking representation learning architecture was used as a feature extractor and multitask decoder. The set of fresh and degraded images was converted into feature vectors using multitask encoders that output the combined features extracted from each expert models' encoders. Using the set of feature vectors, the VCCA model was trained to convert each modal reversibly. The VCCA model has two encoders and decoders specialised for each modal, but the encoders' bottleneck layer is shared as a latent space. Each encoder-decoder was trained to minimise the binary cross-entropy between the original and converted vectors. Furthermore, this model was set to satisfy the cycle-consistency. Similar to the first training phase, uncertainties were utilised to weigh losses: namely, cycle-consistency loss and Kullback–Leibler divergence loss.

## 3 Results

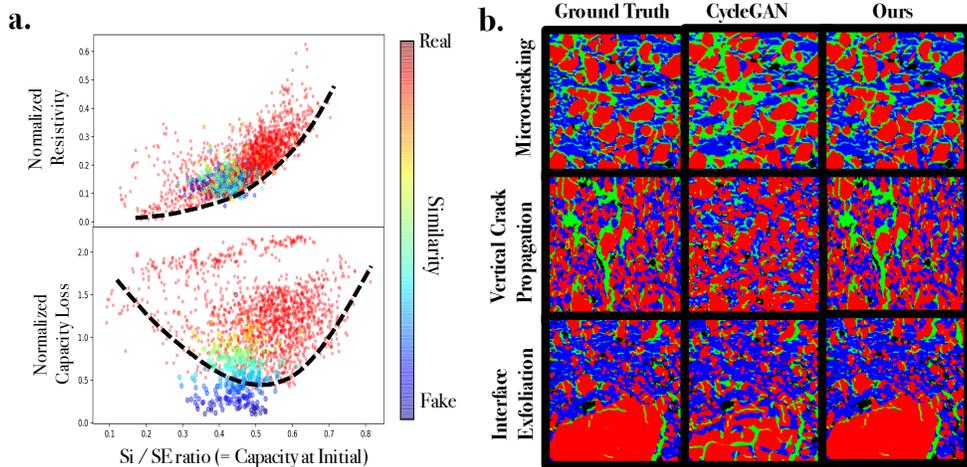
### 3.1 Visualisation of Degradation Patterns

The learnt representations were reduced to two dimensions by UMAP, as shown in Figure 1. Each point represents the corresponding image. The coordinate is set so as to place similar degradation patterns adjacent to each other. The four scatter plots elaborate on the visually understandable correlation between the quad-modal data (manufacturing conditions, degradation patterns, resistance, and capacity loss). Each plot shares a coordinate commonly. In the two left plots, the colour represents the corresponding sample number, and the annotations describe the common characteristics among the samples. In the two plots on the right, the darker blue region indicates an improvement in each performance metric. As is clearly shown in the rightmost plot, the upper right region achieves the most extended lifespan. Moreover, by comparing the other plots' corresponding regions, the manufacturing methodology that causes this improvement can be immediately reasoned to be determined; the most desirable degradation pattern is concentrated voids between silicon, and the manufacturing condition that permits this is to employ porous silicon.

### 3.2 Exploration for better microstructural design

A better battery design was explored in the latent space through the iteration to minimise the predefined loss function. Herein, the loss function is the sum of three target variables - resistance, capacity loss, and synthesizability, where the distributions of each are standardised by z-score. Weighing each loss, the definition of a better battery design is modifiable.

In this study, synthesizability was defined by the nearest distance between real and generated data in the space embedded by AdaCos. Obviously, this definition is indicative of the degree of similarity between images and is not entirely consistent with the possibility of synthesis. Strictly speaking, however, the microstructure of an all-solid-state battery can be synthesised in any structure. For example, it is possible to manufacture them in a programmable way, such as 3D printing, no matter how dissimilar they are to the input data. This heuristic definition was adopted because it is empirically known that similar images suggest that their designs can be manufactured in a similar manner and require less effort to establish a manufacturing method. In the future, it is hoped that further improvements to the definition are developed, either through actual experimental prototyping or by specifying more detailed parameters for manufacturing.



**Figure 2:** (a) Generative exploration of microstructural design to extend lifespan with balancing synthesizability. (b) Qualitative comparison of the transformed images by CycleGAN and ours.

**Table 1:** Qualitative evaluation and ablation study

Method	After Conversion		
	Average	Classification accuracy@top1	Prediction $R^2$ score
<b>Ours</b>	<b>90.33</b>	96.75	83.90
Ours w/o SSCL	48.46	96.51	0.40
Ours w/o classification	48.77	24.83	72.7
Ours w/o prediction	49.06	97.81	0.30
Only classification		<b>98.27</b>	
Only prediction			<b>90.5</b>
CycleGAN	10.97	19.84	2.10

The result of the exploration is shown in Figure 2(a). The horizontal axis represents the ratio of silicon to solid electrolyte mixtures (Si/SE ratio), which is the dominant parameter in determining battery capacity. The vertical axis shows the resistivity (top) and capacity loss (bottom). The synthesizability is indicated by the colour, with red indicating an exact match with the input and blue indicating less similarity to the input. The minimum input data points were shown as dotted lines. Data points below the minimum line were not found in the resistivity but were found in the capacity loss while sacrificing synthesizability.

The generated design with the most extended lifespan suggests an improvement in capacity retention from 86.72% to 96.2% after 300 cycles. The corresponding generated images show fewer voids between the solid electrolytes and more voids encapsulated in silicon than conventional designs. The predicted manufacturing conditions suggest changing the mixing process from the traditional method of homogenising the suspension into kneading a stiff mixture. Similar suggestions were separately reported by battery experts [10, 14], which supports that this prediction is accurate.

## 4 Discussion

The qualitative comparison of the transformed images by CycleGAN and those from this study is illustrated in Figure 2(b). While the current model successfully transformed images taking into account the intra-domain variety of degradation patterns, CycleGAN fails to recognise the intra-domain context.

Next, the quantitative evaluation of the proposed model was performed. As the metric, the averaged accuracies of both classification and lifespan prediction tasks were adopted, after converting from the

fresh to the degraded state. The classification accuracy is defined here as the percentage of correct answers of the prediction label in the top 1, and the prediction accuracy is based on  $R^2$  scores between the predicted and measured values. While CycleGAN recorded an average accuracy of 10.97, the proposed model achieved 90.33. The cause of low classification accuracy for CycleGAN is likely to be a lack of the intra-domain context.

Lastly, an ablation study was conducted. While the classification accuracy of the model without SSCL was 96.51, the prediction accuracy plummeted to 0.4. This is possibly due to the fact that the model fails to alleviate the conflict between classification and reconstruction. Likewise, in the model without sample classification pretraining, the classification accuracy dropped to 24.83. These results indicate that multitasking and SSCL are essential parts of the proposed model.

## 5 Conclusion

An automatic degradation analysis and microstructural design exploration were investigated based on a new dataset for designing ASSBs. The proposed model successfully organised the degradation patterns based on their multimodal similarity. A new structure and manufacturing method are proposed to extend the ASSBs' lives from 86.72% to 96.2% by exploring the latent space. This model could be applied to new materials discovery and potential future research by converting the degradation patterns, a common bottleneck in materials development, into human-interpretable and explorable latent variables.

### Broader Impact

In recent years, attempts have been made to accelerate new materials discovery using ab-initio calculations and machine learning. However, it is challenging to predict performance after degradation, even though the fresh state performance is predictable. In fact, most of the found materials have good initial performance but degrade quickly. Consequently, the degradation prediction is of crucial importance for a practical application of a data-driven approach.

Meanwhile, degradation is a complex process, making it challenging to extract common patterns in experimental datasets. Thus, the degradation patterns are not systematised, and the data often discards before thorough exploitation. Consequently, the degradation analysis model plays a significant role in experimental materials science.

### Who may benefit from this research

Therefore, our model can be beneficial for both an analytical model for experimental scientists and a degradation prediction model for materials explorers using theoretical calculations. Furthermore, we believe that our proposed model is applicable to all material degradation phenomena, not only to batteries. In other words, we believe that it will be of benefit to all those who are developing materials for specific applications.

### Who may be put at disadvantage from this research

If anyone were disadvantaged, it would be those who were doing empirical-based degradation prediction research. As data-driven research progresses, human-proposed simple physics models will become less important, as the data itself is more valuable than the models.

### What are the consequences of failure of the system

An expert can distinguish when the system fails, as the system produces unrealistic images. Experiments take much more time and effort than calculations, so it is hard to imagine not checking the calculation results before the experiment.

### Whether the task/method leverages biases in the data

Regarding the use of data bias, it should be verified by cross-dataset tests, but this is difficult to do as our dataset is currently the first battery design dataset. First, the cross-validation method ensured no

overfitting or dataset bias, unlike the hold out method. Secondly, we visualised the CNN last layer features by GradCAM [15] to check if the features were selected according to the task. As a result, we confirmed that the model was successfully trained to focus on the ionic path connectivity and the active material-solid electrolyte interface, which were reasonable given the battery’s physical model. On the other hand, the model without SSCL did not recognize the differences between the elements. Therefore, there would be no data bias, as the SSCL using simulation as a pretask contributes to the acquisition of general features.

## References

- [1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *ICML*, 2017. URL <https://arxiv.org/abs/1701.07875>.
- [2] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y. H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, and W. C. Chueh. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(397), 2020. URL <https://www.nature.com/articles/s41586-020-1994-5>.
- [3] J. P. Cohen, M. Luck, and S. Honari. Distribution matching losses can hallucinate features in medical image translation. In *MICCAI*, 2018. URL <https://arxiv.org/abs/1805.08841>.
- [4] Y. Kato, S. Hori, T. Saito, K. Suzuki, M. Hirayama, A. Mitsui, M. Yonemura, H. Iba, and R. Kanno. High-power all-solid-state batteries using sulfide superionic conductors. *Nature Energy*, 1(16020), 2016. URL <https://www.nature.com/articles/nenergy201630>.
- [5] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *CVPR*, 2018. URL <https://arxiv.org/abs/1705.07115>.
- [6] M. F. Lagadeec, R. Zahn, S. Müller, and V. Wood. Topological and network analysis of lithium ion battery components: the importance of pore space connectivity for cell operation. *Energy & Environmental Science*, 11(3194), 2018. URL <https://doi.org/10.1039/C8EE00875B>.
- [7] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017. URL <https://arxiv.org/abs/1612.03144>.
- [8] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. In *arXiv*, 2018. URL <https://arxiv.org/abs/1802.03426>.
- [9] N. Mrabah, M. Bouguessa, and R. Ksantini. Adversarial deep embedded clustering: on a better trade-off between feature randomness and feature drift. In *arXiv*, 2019. URL <https://arxiv.org/abs/1909.11832>.
- [10] Y. J. Nam, D. Y. Oh, S. H. Jung, and Y. S. Jung. Toward practical all-solid-state lithium-ion batteries with high energy density and safety: Comparative study for electrodes fabricated by dry- and slurry-mixing processes. *J. Power Sources*, 375(93), 2018. URL <https://www.sciencedirect.com/science/article/pii/S0378775317314908>.
- [11] M. F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh. Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*, 2(161), 2020. URL <https://www.nature.com/articles/s42256-020-0156-7>.
- [12] S. Ravuri and O. Vinyals. Classification accuracy score for conditional generative models. In *neurIPS*, 2019. URL <https://arxiv.org/abs/1905.10887>.
- [13] A. Razavi, A. V. D. Oord, and O. Vinyal. Generating diverse high-fidelity images with vq-vae-2. In *ICLR*, 2019. URL <https://arxiv.org/abs/1906.00446>.
- [14] J. Sakabe, N. Ohta, T. Ohnishi, K. Mitsuishi, and K. Takada. Porous amorphous silicon film anodes for high-capacity and stable all-solid-state lithium batteries. *Communications Chemistry*, 1(24), 2018. URL <https://www.nature.com/articles/s42004-018-0026-y>.
- [15] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggadakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(383), 2019. URL <https://www.nature.com/articles/s41560-019-0356-8>.

- [16] V. Sulzer, S. G. Marquis, R. Timms, M. Robinson, and S. JonChapma. Python battery mathematical modelling (pybamm). In *ECSSarXiv*, 2020. URL <https://ecsarxiv.org/67ckj/>.
- [17] T. G. Tranter, M. D. R. Kok, and J. T. Gostick. pytrax: A simple and efficient random walk implementation for calculating the directional tortuosity of images. *SoftwareX*, 10(100277), 2019. URL <https://www.sciencedirect.com/science/article/pii/S2352711019302286>.
- [18] W. Wang, X. Yan, H. Lee, and K. Livescu. Deep variational canonical correlation analysis. In *arXiv*, 2017. URL <https://arxiv.org/abs/1610.03454>.
- [19] Q. Xie, M. T. Luong, E. Hovy, and Q. V. Le. Self-training with noisy student improves imagenet classification. In *CVPR*, 2020. URL <https://arxiv.org/abs/1911.04252>.
- [20] P. Yakubovskiy. Segmentation models pytorch. In *GitHub repository*, 2020. URL [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch).
- [21] X. Zhang, R. Zhao, Y. Qiao, X. Wang, and H. Li. Adacos: Adaptively scaling cosine logits for effectively learning deep face representations. 2019. URL <https://arxiv.org/abs/1905.00292>.
- [22] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications*, 11(1706), 2020. URL <https://www.nature.com/articles/s41467-020-15235-7>.
- [23] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. URL <https://arxiv.org/abs/1703.10593>.

## 6 Appendix

### 6.1 Dataset Description

#### 6.1.1 Details of Data Collection

Table 2: Dataset description

degradation cycle	cell designs	images per cell	labelled images	raw images
0	83	100	215	8,300
300	94	100	215	9,400
		SUM	430	17,700

Table 2 displays the details of the data collection. Data points were acquired using laminate all-solid-state lithium-ion batteries made in the laboratory. One hundred electron microscope images were acquired from random positions on the anode of each sample. Degradation tests were prepared by charging and discharging 300 cycles at 2 C-rate. Eighty-three battery designs were imaged with pairwise combinations of pre- and post-degradation, and 11 data points were acquired for post-degradation only. Furthermore, there are 430 expert-labelled teacher data points for raw grayscale images. Direct Current Internal Resistance (DCIR) and capacity retention were also measured for each design as the corresponding performance measures.

#### 6.1.2 Details of Data Distribution

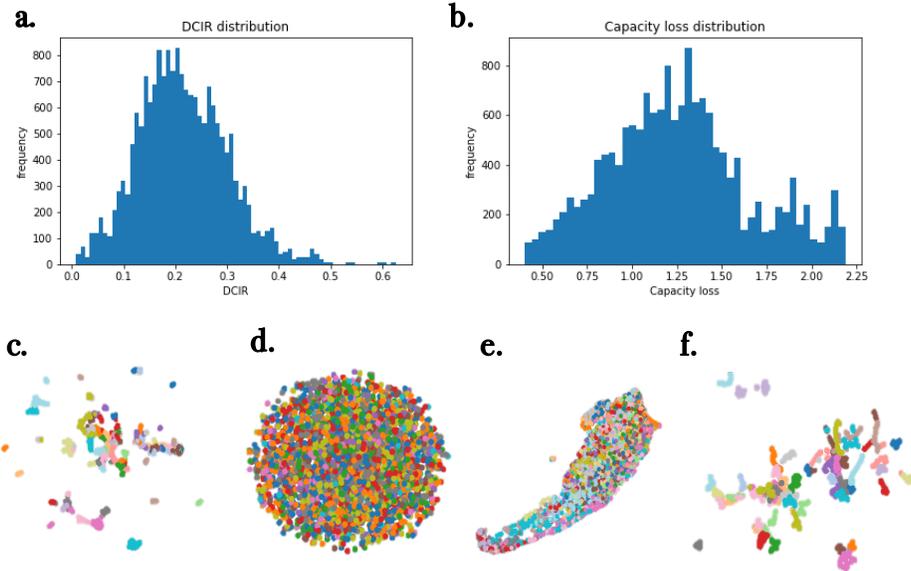


Figure 3: Distribution of objective variables; DCIR (a) and capacity loss (b). Distribution of feature vectors extracted from the last layers of; manufacturing condition classifier (c), image generator (d), performance predictor (e), and merged multitasking encoders (f).

Figure 3 illustrates the details of the data distributions. UMAP visualised the distribution of images in (c) to (f). In (c), a discrete cluster distribution was observed. Conversely, in (d), a confused distribution was obtained. Similarly, in the performance prediction, images were embedded in the performance order, which was not discriminative. In the multitasking representation, the distribution was discriminative as well as performance predictive.

#### 6.1.3 Details of Data Splitting

The data was divided into three parts: the ratios for training, testing, and validation were 0.7, 0.15, and 0.15, respectively. The ratios were set to be preserved for each cell design. Of the 100 images in each sample, 70 were used for training, 15 for validation, and 15 for testing. Moreover, the extracted 12,390 training data points, 2,655

validation data points, and 2,655 test data points were extracted to have the same distribution. As the metric to evaluate the similarity of distribution of three datasets, Kullback-Leibler divergence was adopted using the simulated DCIR values corresponding to each image. Three sets of images were iteratively randomly extracted until reaching the minimum value. The optimised image sets were used in common for the entire learning process.

## 6.2 Training Details

A detailed learning process of the expert model will be described. First, for EfficientNet, which was used to classify the manufacturing conditions, the B1 structure with the highest classification accuracy was adopted. A general method of data augmentation was also used, and the image size was cropped at 256 pix. A cycle scheduler trained the model with a batch size of 64 and a learning rate of 0.01. The initial weights were fine-tuned using the model pretrained by ImageNet. After training EfficientNet-B1, the final layer was changed to AdaCos and trained again. EfficientNet-B1 was also trained for performance prediction under the same conditions. However, L1 regularisation was added and the final layer was changed to a simple multilayer perceptron to predict a single objective value. The VQ-VAE-2 was also trained with an image size of 256 pix, a learning rate of  $3 \times 10^{-4}$ , and a cycle scheduler. Afterwards, the pixelSNAIL models of the top and bottom layers were trained with a batch size of 32, a learning rate of  $3 \times 10^{-4}$ , and a cycle scheduler.

A detailed learning process of the multitasking representation will be described: when merging the three trained expert models, EfficientNet combined the final layer of 1,280 dimensions, and VQ-VAE combined both the final layer  $128 \times 16 \times 16$  encoded in the top layer and the quantised  $4 \times 16 \times 16$ . The bottleneck layer was then convolved with Convolutional Neural Network until it was  $8 \times 8$  in size and the bottleneck layer was  $512 \times 8 \times 8$ . As such, the dimension of the multitask representation feature vector is  $512 \times 8 \times 8$ . It was trained with a batch size of 64, a learning rate of  $3 \times 10^{-4}$ , and a cycle scheduler. Finally, the learnt representation was used to train the VCCA. The structure of the VCCA was VCCA-private, with 20 dimensions for the shared latent space and 30 dimensions for the private latent space. It was trained with a batch size of 128, a learning rate of  $1 \times 10^{-4}$ , and a cycle scheduler. UMAP visualised the 20 dimensions of the shared latent space. The hyperparameters of nneighbors, mindist, ncomponents, and metric were 15, 0.1, 1, and euclidean, respectively.