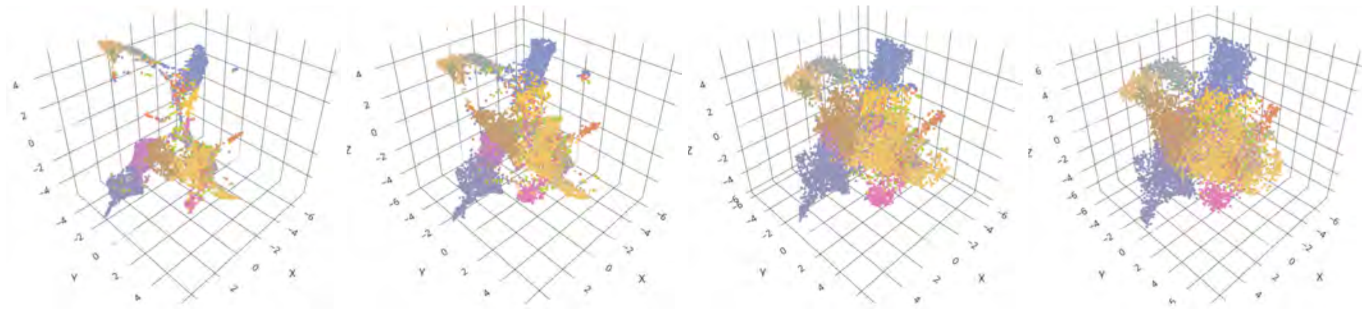
EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2024)
M. Corsini, D. Ferdani, A. Kuijper, and H. Kutlu (Editors)

Short Paper

AI-Driven Classification of a Design Photographic Archive

Karina Rodriguez Echavarria[†1] and Myrsini Samaroudi[1]

[1]University of Brighton, Brighton, UK



(a) min_dist*: 0.1* **(b)** min_dist*: 0.3* **(c)** min_dist*: 0.6* **(d)** min_dist*: 0.9*

Figure 1: *UMAP projection to 3D points of AI-driven classification of Photographic Collection under 40 different classes plotted for parameter* min_dist *of 0.1, 0.3, 0.6, 0.9.*

Abstract

The paper presents a workflow for deploying an Artificial Intelligence (AI) classification of a previously unclassified photographic collection, the Design Archive's glass plate negatives. This involved fine-tuning the DinoV2 self-supervised image retrieval system with a domain-expert taxonomy to classify approximately 10K images within 40 classes. As such, it addresses challenges relevant to the curation, analysis and discovery of large-scale visual collections. A 3D visualisation was implemented for users to access the outputs presenting images as data points using the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) to project the embeddings of the neural network. The paper demonstrates the advantages of this approach and reflects how users can participate in the AI processes making them more transparent and trustable.

Keywords: Information Systems, Cultural Heritage Collections, Information Discovery

CCS Concepts

• **Information systems** → **Information extraction;** • **Computing methodologies** → **Graphics systems and interfaces;** • **Applied computing** → *Fine arts;* • **Human-centered computing** → **Visualization techniques;**

1. Introduction

The aim of the research is to develop FAIR (Findable, Accessible, Interoperable and Reusable) compliant workflows to generate metadata for large-scale collections of visual content, in ways which facilitate the discovery, use and reuse of data. Thus, the paper describes an AI-driven workflow applied to a photographic collection which is part of the Design Archives at the University of Brighton.

The research is timely as it addresses a key need to improve the interaction with and comprehension of large-scale datasets and complex data analytical tasks driven by AI algorithms. The automatic classification of artworks based on criteria such as artist, style, or genre is a well-known challenge of computational art analysis [CS22]. Although AI has great potential to support data automation and analytical tasks, there is still a lack of effective interactive visual approaches to enable potent communication between humans and AI.

The research contribution of this paper is a workflow for deploying and providing interactive access to large-scale classified data.

† email: K.Rodriguez@brighton.ac.uk

DOI: 10.2312/gch.20241258

The paper is organised as follows: Section 2 presents the visual collection, the Design Archive's *glass plate negatives*, the research engages with. Section 3 describes the implementation of the AI-classification algorithm based on the DinoV2 Foundation Model and Section 4 presents a visualisation to interact with the results of the classification. Finally, Section 5 discusses the conclusions and further research.

2. Design Council Photographic Library

The research engages with the Design Archives' Photographic Library of the Council of Industrial Design [Mor00]. Historically, the Photographic Library played a key role in using photography to promote design in the 20th century. It now provides material for understanding the activities and operation of the UK Council of Industrial Design from 1944 to 1971 and the Design Council from 1972 to the present. Users of the archive include scholars and students across the arts and humanities and beyond, who are researching the designed environment, the design profession, and design practice.

The photographic material relates to the journal *Design*, from the first issue in 1949 until the 1980s, and photographic material relating to the Council's various award schemes from 1957 onwards. The library comprises 1) photographs of objects and environments including architecture, ceramics, clothing, domestic appliances, medical, musical instruments, textiles, tools and toys amongst other types of design; as well as 2) slides and negatives, including a large quantity of glass plate negatives. Examples of digitised images of the negatives from the library can be found in Figure 2.

Figure 2: *Top-left) original image with number U-360-178-O; Top-right) original image with number ODD-to-4-O; Bottom-left) original image with number GB-1837-DES-DCA-30-2-412.14-O; and Bottom-right) original image with number GB-1837-DES-DCA-30-2-467.2 (©CoID) Design Council Archive, University of Brighton Design Archive.*

The glass plates were received by the Design Archives in their original boxes along with their corresponding accession books. Through ongoing digitisation efforts, high-resolution scanned images of part of the photographs and the glass negatives' collections have been made in the last few years. The digital files are labelled with identifiers available in the physical medium. To this date, ∼10,000 out of ∼100,000 photographs in physical form and ∼9,300 of ∼60,000 glass plates in physical form have been digitised. The digitisation and cataloguing tasks already represent a significant challenge for the archival team. An additional hurdle arises from the fact that the glass plates' digital images are not classified or fully documented neither matched to their corresponding photographs.

Similarly to other cultural heritage collections, cataloguing this visual material is not an easy task as there are multiple interpretations for its subject classification. Moriarty [Mor00] explored how these classifications evolved within the Design Council through the years, as categories were added and deleted to reflect wider policy changes within the Council. For instance in 1950, among other categories, photographs were classified as: Radio; Woodware; Pottery; Sideboards and Dressers; Glass; Clocks; Easy Chairs; Knives, Spoons, Forks; and Home Lighting.

By the time the Design Archives received the material, the following high-level classifications were used: Textiles; Souvenirs; Tableware; Interior design; Jewellery; Lighting; Ceramics; Other; Graphics; Furniture; Wallpaper; Building accessories; Telecommunications; Tools; Clocks; Design theory; Smoking accessories; Portraits; Travel goods; Ornaments; Clothing; Carpets; Toys; Industrial and manufacturing processes; Domestic appliances; Musical instruments; Heating; Scientific equipment; Engineering; Transport; Religious objects; Office equipment; Architecture; Optical; Photography; Drawing and painting equipment and accessories; Sanitary equipment; and Materials.

Early on it was identified that, by themselves, these non-static and pluralist classifications could provide multiple pathways to explore the Photographic Library. Similarly to other photographic collections established to serve the needs of public bodies across the world, the potential to create customised classifications and groupings to suit the researchers' exploration needs can opening access points and hence enable further interpretations of the collection.

3. Deploying AI-based Classification of Visual Archives

Generating useful metadata is often a time-consuming but critical step for enabling the use and reuse of visual collections. The research focused mostly on the *glass plate negatives* dataset which contain no metadata with the exception of their filenames. With an estimated human time of approximately 3 minutes to access the record, open the relevant image and enter a textual classification in the appropriate field, the effort of manually classifying the ∼9,300 images will be equivalent to 28K minutes, or 62 working days. This makes the classification a labour-intensive endeavour, highlighting the need for Artificial Intelligence (AI)-based approaches.

Existing workflows and tools for *feature* extraction are largely based on *Foundation Models (FMs)*. Also known as 'general-purpose AI' systems, these models are trained on generalised data generally with unsupervised learning. They can be used as a starting point to enable building further systems or tools, a process known as 'fine-tuning' the model. Popular FMs for image tasks include YOLO (You Only Look Once) [RDGF15] and DinoV2

[Oea24] which are pre-trained models on large curated data. DinoV2, in particular, has properties such as an understanding of object parts and scene geometry regardless of the image domains.

For the research, we deployed the high-level classifications which were used by the Design Council at the time the Design Archives received the Photographic Library (see Section 2). Previous to this research, the uncatalogued images were organised sequentially, as they were received from the Design Council. DinoV2 [Oea24] was used for fine-tuning a classifier. Although the FM is presented as a generic solution for image classification, it is recognised that expert-specific scenarios require the FM to be fine-tuned for accuracy purposes. The process of fine-tuning involved the following steps.

1. Manual labelling process: Sample images for each classification category were selected. In total, a sample of ∼2K images covering the taxonomy were used. Additional images were gathered from heritage-related web sources, including Europeana, for the classes which were under-represented.

2. Training phase: Using the labelled examples, an algorithm for training was implemented. This algorithm feeds the labelled images to the pre-trained DinoV2 model from Torch's model hub. Images are resized to 518 by 518 pixels, and rotated randomly. The output vectors, the vector embeddings, encapsulate the compressed representation of the images. The embeddings are used for various tasks, including image classification. These are used as input for a compact two-layered sequential neural network to produce a score for the feature classification task. The training loop was set to 50 EPOCHS. The machine used for the training has an Intel Core i9, 16MB, 8 Cores CPU, with 32GB (2x16GB) DDR4 and a GeForce RTX 3080 10 GB. The achieved accuracy is 73%, whereas, in 88% of the cases, one of the top 3 predictions is accurate.

3. Classifying the dataset: The fine-tuned model is used to predict the three top categories and confidence levels shown as percentages for all images in the dataset. Figure 3 illustrates the total number of images which were classified under each category for the Glass Plate Negatives dataset. The predictions, along with the embeddings, are then stored in a CSV file and the images are stored in separate folders across their top-predicted classification.

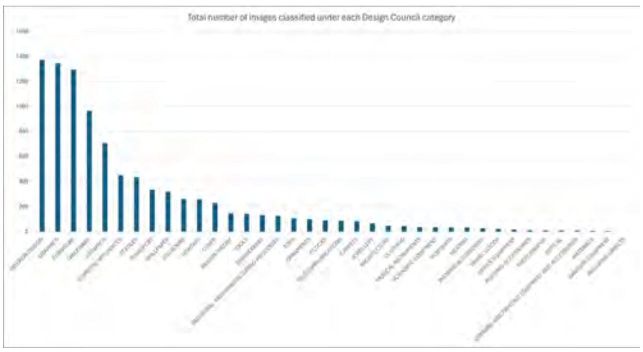


Figure 3: *Graph illustrating the number of images classified for each Design Council class/category.*

4. Visualisation of AI-driven Classification

The output of the AI-driven classification process includes labels with the predicted category/ies of the image along with their confidence value. Thereafter, it is the task of the user to use this information to support a decision-making process on whether to use or not such labels. Given the AI task is scaled up to thousands of images, making sense of these outputs becomes challenging for users as the required cognitive load increases. For this, the user needs to systematically explore the outputs of the classification, including predictions and confidence values. Such processes can be further supported by additional outputs, such as the *embeddings* of the neural network. In ML, embeddings, also known as encodings, are mathematical representations in the form of a vector of numbers in a high dimensional space that captures/represents the underlying relationships and patterns of input data [HKPC19].

Visualisation of the embeddings requires of dimensionality reduction techniques, including t-SNE and Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP). These techniques can map the embeddings of a large-scale dataset in two or three dimensions. For instance, Smilkov et al. [STN*16] introduce a tool for interactive visualisation and interpretation of embeddings, the Embedding Projector, which is available both as a standalone tool and as an integrated tool into the TensorFlow platform. Some works have studied how PCA and t-SNE differ, mathematically and visually, and suggest the need for improving the way semantic similarities are mapped [LBT*18]. Jiazhi et al. [XHL*23] address this challenge by enhancing the visual cluster separation of embeddings through a dimensionality reduction approach.

A visualisation tool based on Shiny [Pos24] was deployed. Shiny is an R package for developing web applications for the embeddings. The user requirement of allowing for basic interactivity to inspect visually the images and their predictions was considered. As shown in Figure 5, the main components of the graphical user interface (GUI) include a side and a top menu. These allow the user to customise the visualisation to facilitate access and interpretability. The visualisations include 1) a 3D visualisation of the embeddings (*3D View*), which can be customised as described below, and 2) a bar chart (*Chart Overview*) presenting the total number of images classified for each category as well as the total of images used for training for each category. The latter might provide an overall idea of biases in the training data towards certain features.

The 3D visualisation is implemented via Plotly R graphing library [Plo24]. It uses three distinct visual elements to facilitate interaction and comprehension of the large-scale dataset. Through a dimensionality reduction approach, the embeddings are plotted as 3-dimensional data points, X, Y, Z, in the Euclidean space. For this, the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) for dimensionality reduction library [MHM20] is deployed. UMAP was selected as it ensures that local structure is preserved in balance with global structure; as well as being faster and scaling well when compared to other methods. The two most used parameters for UMAP include the number of neighbours, or *n_neighbours* and minimum distance, or *min_dist*.

The *n_neighbours* is used by UMAP to balance local versus global structure in the data by constraining the size of the local neighbourhood. The other parameter, the *min_dist*, controls how tightly UMAP is allowed to pack points together. something which can be useful for clustering. Lower values of *min_dist* will pack points together, while large values will focus on the preservation of the broad topological structure instead. Figure 1 shows a comparison of the points projected by UMAP or various *min_dist* values when the *n_neighbours* is set to 10, ranging from 0.1, 0.3, 0.6, and 0.9. As illustrated by the graphs, this value affects the global shape of the graph, allowing the user to perceive clusters in the data. It is also faster to compute for lower *n_neighbours* values. Therefore, the interface allows the user to modify these values.

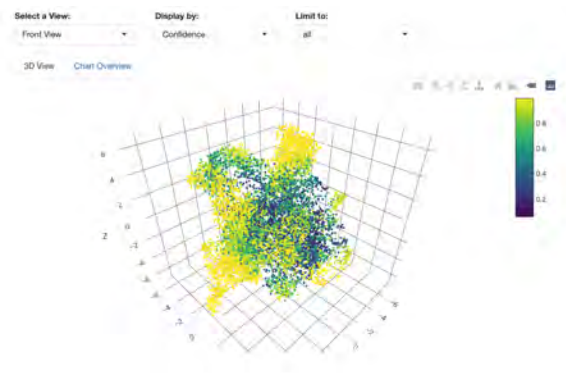


Figure 4: *3D Visualisation of confidence values for AI-driven Classifications of Design Council Photographic Collection.*

The functionality to click on points allows the user to see the image, along with others in the near vicinity. The images are queried via the IIIF image API and the top 3 predictions with their levels of confidence are presented in a table (see Figure 5).

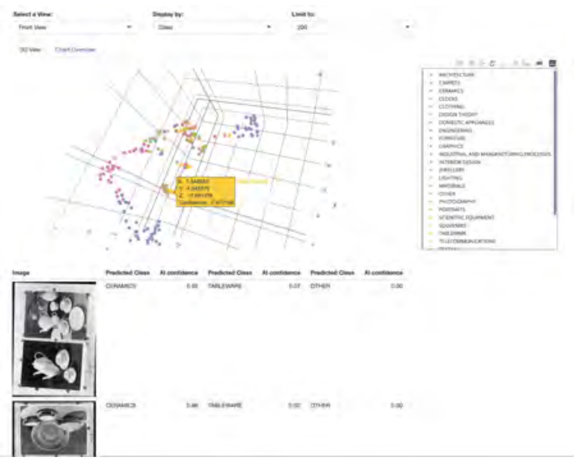


Figure 5: *Interactive 3D Visualisation for AI-driven Classifications of the Design Council Photographic Collection allowing users to inspect the images and their predictions via the IIIF image API.*

5. Conclusions and Further work

This paper presented research on workflows and tools for Artificial intelligence-driven methods for the classification of a photographic collection. The research fine-tuned the DinoV2 foundational model to classify ∼10K images according to the Design Council taxonomy. Thereafter, the research deployed a 3D visualisation using the Shiny framework and an implementation of the UMAP dimensionality reduction algorithm to map the *embeddings* in a lower 3D dimension. Through the interface, the user can modify in real-time UMAP parameters such as *min_dist* to support the interpretability of the mapped data. Furthermore, an IIIF image service enables access to the images. Further work includes a comprehensive evaluation to understand how interactive visualisations can affect the cognitive load and interpretability of AI-driven classifications.

6. Acknowledgments

We thank colleagues at the University of Brighton's Design Archives, Sue Breakell and Jen Grasso, for their contributions to the research in making the data available and providing access to their expertise. We would also like to acknowledge Vladislav Vaculin and Andrei Sobo who contributed to the developments.

References

[CS22] Cetinic E., She J.: Understanding and creating art with ai: Review and outlook. *ACM Trans. Multimedia Comput. Commun. Appl. 18*, 2 (feb 2022). URL: https://doi.org/10.1145/3475799, doi:10.1145/3475799. 1

[HKPC19] Hohman F., Kahng M., Pienta R., Chau D. H.: Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics 25*, 8 (2019), 2674–2693. doi:10.1109/TVCG.2018.2843369. 3

[LBT*18] Liu S., Bremer P.-T., Thiagarajan J. J., Srikumar V., Wang B., Livnat Y., Pascucci V.: Visual exploration of semantic relationships in neural word embeddings. *IEEE Transactions on Visualization and Computer Graphics 24* (2018), 553–562. URL: https://api.semanticscholar.org/CorpusID:24412850. 3

[MHM20] McInnes L., Healy J., Melville J.: Umap: Uniform manifold approximation and projection for dimension reduction, 2020. arXiv:1802.03426. 3

[Mor00] Moriarty C.: A backroom service?: The photographic library of the council of industrial design, 1945-1965. *Journal of Design History 13*, 1 (2000), 39–57. URL: http://www.jstor.org/stable/3527186. 2

[Oea24] Oquab M., et al.: Dinov2: Learning robust visual features without supervision, 2024. arXiv:2304.07193. 3

[Plo24] Plotly: 3d scatter plots in r, 2024. URL: https://plotly.com/r/3d-scatter-plots/. 3

[Pos24] Posit: Shiny, 2024. URL: https://shiny.posit.co/. 3

[RDGF15] Redmon J., Divvala S. K., Girshick R. B., Farhadi A.: You only look once: Unified, real-time object detection. *CoRR abs/1506.02640* (2015). URL: http://arxiv.org/abs/1506.02640, arXiv:1506.02640. 2

[STN*16] Smilkov D., Thorat N., Nicholson C., Reif E., Viégas F. B., Wattenberg M.: Embedding projector: Interactive visualization and interpretation of embeddings, 2016. arXiv:1611.05469. 3

[XHL*23] Xia J., Huang L., Lin W., Zhao X., Wu J., Chen Y., Zhao Y., Chen W.: Interactive visual cluster analysis by contrastive dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2023), 734–744. doi:10.1109/TVCG.2022.3209423. 3