

OMVMLP: 3D Texture Classification by Orthographic Multi-View Data Characterization and MLP (short paper)

Ahmed Hazem Youssef¹, Halim Benhabiles¹, Adnane Cabani², Ruiwen He³ and Karim Hammoudi⁴

¹IMT Nord Europe, Institut Mines-Télécom, Université de Lille, Center for Digital Systems, F-59000 Lille, France

²Univ Rouen Normandie, ESIGELEC, Normandie Univ, IRSEEM UR 4353, F-76000 Rouen, France

³De Vinci Higher Education, De Vinci Research Center, Paris, France

⁴Department of Computer Science, IRIMAS, Université de Haute-Alsace, 68093 Mulhouse, France

Abstract

Relief pattern retrieval is a task that allows us to extract and analyze visual patterns on 3D objects which gives insight about these objects and their cultural value. However, this task is challenging due to the variability of textures on the same 3D object, that poses a tedious multi-label problem, especially in large datasets. In this work we propose our method OMVMLP (Orthographic Multi-View data characterization and MLP-based classification), which uses six orthographic projections of a given 3D object to extract the relief patterns using 2D deep learning techniques. These projections are then processed with our CNN model which uses a VGG19 backbone finetuned to extract the texture features. These features are then concatenated and fed into a Multi-Layer Perceptron (MLP) trained for multi-label classification. The goal of the work is to define the 3D objects that share a relief in common with a given query relief and generate a membership matrix between them. Indeed, classifying object reliefs is a critical and essential task towards facilitating the reassembly fragmented artifacts (e.g. for archeological object restauration). The mentioned 3D objects have different relief categories, 15 of which are predefined and we classify them using the MLP. As for the unseen classes we proposed two complementary scenarios: (i) feature similarity estimation using Euclidean distance, and (ii) classifier adaptation via fine-tuning on the reliefs from the query set. Experimental results show that our model has strong validation performance with an F-1 score of 0.809, but it is yet limited regarding the generalization on unseen classes, resulting in a membership matrix AUC of 0.523. Despite this gap in performance, this demonstrates the potential of view-based approached in the analysis of 3D models, specifically pattern retrieval, additionally it highlights the challenges in recognizing unseen categories during training. Our code is publicly available at <https://github.com/AhmedHazemYoussef/OMVMLP-3D-Relief-Recognition>.

Keywords

3D shape retrieval, Cultural heritage, 3D cultural objects, 3D artwork, Relief pattern recognition, Multi-view learning, Deep learning, Multi-label classification.

1. Introduction and motivation

Throughout history different civilizations have always expressed themselves through art, leaving us antiques that represent their culture, creativity and innovation. Analyzing the different art works represents a way to see through time and understand the cultural heritage of these civilizations. As mentioned in the report [1], the analysis of reliefs plays a fundamental role in the restoration and reassembly of fragmented artifacts and engravings [2–6]. In this article we represent our work done on the classification of relief patterns extracted from 3D objects which would allow us to classify the artwork based on their textures and surface patterns. Our work has been conducted in the context of the international SHREC challenge, which involves researchers in 3D shape analysis towards solving complex problems [7]. Among the tracks proposed in the 2025 edition, we can emphasize the track on recognition and segmentation of multiple relief patterns organized as part of the Eurographics 3DOR

VIPERC2025: 4th International Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 3-4 December 2025

✉ ahmed-hazem.youssef@imt-nord-europe.fr (A. H. Youssef); halim.benhables@imt-nord-europe.fr (H. Benhabiles); adnane.cabani@esigelec.fr (A. Cabani); ruiwen.he@devinci.fr (R. He); karim.hammoudi@uha.fr (K. Hammoudi)

ORCID 0009-0008-4320-4747 (A. H. Youssef); 0000-0003-2950-5312 (H. Benhabiles); 0000-0001-5948-8950 (A. Cabani); 0000-0002-5381-6068 (R. He); 0000-0002-4804-4796 (K. Hammoudi)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

project [1]. In this latter, two tasks have been described [1]. The first one is to build a membership matrix between the 3D objects and the query ones addressing the retrieval part of the problem, each value in the matrix represents the probability that both the object and the query contain the same relief pattern measuring the similarity between them. A relief pattern could be understood as the surface patterns (e.g., engraving, motif, or texture) that's independent from the overall 3D shape [8]. The second task was the segmentation of the relief regions, determining the exact areas where a certain relief occurs. We determined that a simple and effective approach is using multiple orthographic views of the 3D object and analyze them using 2D convolutional neural networks (CNN) [9, 10]. This offers several advantages regarding the complexity of the problem and computational efficiency [11], as the 3D objects in the dataset are particularly large (on average 7.2 MB and about 400,000 faces per mesh), making the application of deep learning methods designed for 3D objects like PointNet [12] very costly in terms of memory and computational time [13]. Our 2D view-based approach in its current form is focused on the retrieval part of the project.

2. Proposed methodology

2.1. Designed classification model

We propose our method OMVMLP (Orthographic Multi-View data characterization and MLP-based image classification), for retrieving the relief patterns of the 3D object using its corresponding 2D orthographic projections inspired by Benhabiles and Tabia [9]. Our approach extracts six orthographic views per object to ensure capturing the different textures and reliefs present on the 3D object (Fig. 1). Each view is then processed by VGG backbone [14] [15] to obtain the feature vector of each view which are then concatenated into a global feature. This global representation is then passed into an MLP to predict the presence of certain surface embedded textures on the object.

Our classification pipeline is composed of both a VGG backbone and a MLP based classifier which are designed specifically for the 15 relief classes provided in the training dataset. For feature extraction we applied a customized finetuning strategy on an ImageNet-pretrained VGG19 backbone, where we freeze all the layers except for the last three to ensure capturing texture related features instead of shape related ones. This feature extraction is applied on the six views and the produced feature vectors each of a size of 512 are then concatenated into a single 3072 global feature vector.

This global feature vector is then passed to our MLP which consists of a linear projection into a 512-dimensional feature vector, followed by a ReLU activation and a final prediction layer of 15 classes that supports multi-label prediction as we applied Sigmoid activation function instead of SoftMax [16]. Additionally, we applied dropout regularization to avoid overfitting.

2.2. Adopted evaluation protocol

In addition to the 15 relief classes provided in the training data set, there were 4 additional classes that were only defined in the query set. To handle these unseen classes, we proposed two different scenarios: i) to use our classifier pipeline except for the final classification layer to extract a 512-vector representation of those reliefs, ii) to replace the output layer with a 4-class classification head, and fine tune the classifier on those classes.

To train the model we used an 80/20 train validation split, a batch size of 16 for 70 epochs and applied categorical cross-entropy loss which handles multi-label classification. Additionally, we applied cosine annealing scheduler [17] to gradually reduce the learning rate from 10^{-3} to 10^{-5} to ensure training stability. After each epoch the model was evaluated, and only the one with the highest F-1 score was retained.

The membership matrix represents the similarity between the relief queries and the 3D objects in the retrieval set, where each line in the matrix represents a certain query and each value represents the probability that the given object contains the same relief as the query. If the query is one of the 15 known classes in the training set, we apply our classification pipeline on the objects and test if the

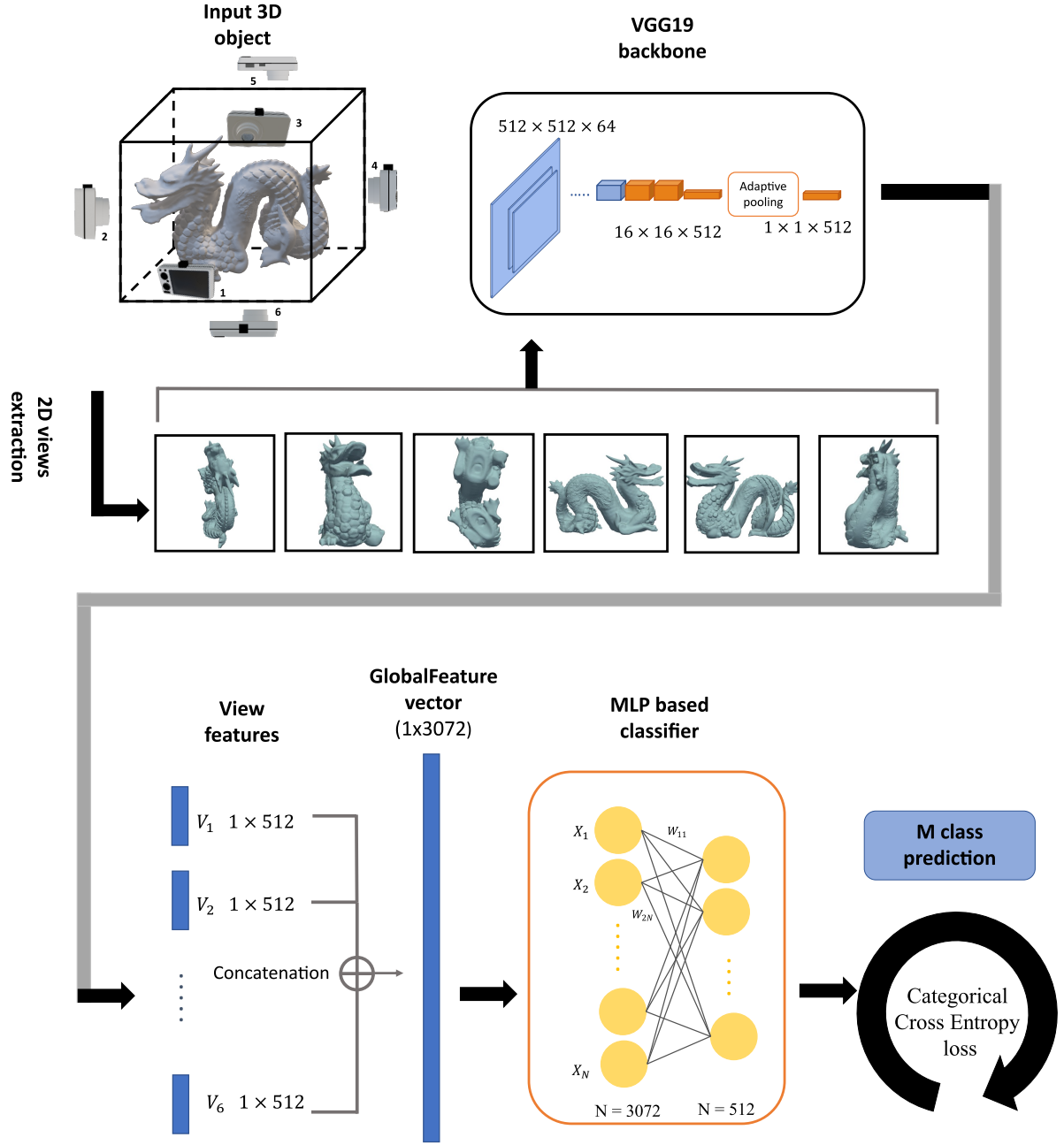


Figure 1: Processing pipeline for predicting relief patterns from a 3D object.

probability of each object being of this certain class is higher than a certain threshold (0.5). If so, we enter this value in the membership matrix, otherwise we enter zero. As for the case when the queries are not from those 15 classes, we used both the scenarios proposed in Section 2.2, generating two membership matrices, one for each scenario.

3. Experimental results and evaluation

Although five teams registered for the track, we were the only team that submitted results, having only one team among the others submitting preliminary experiments while nothing was submitted from the remaining teams. This is largely due to the difficulty of the task, particularly due to the scale and complexity of the dataset.

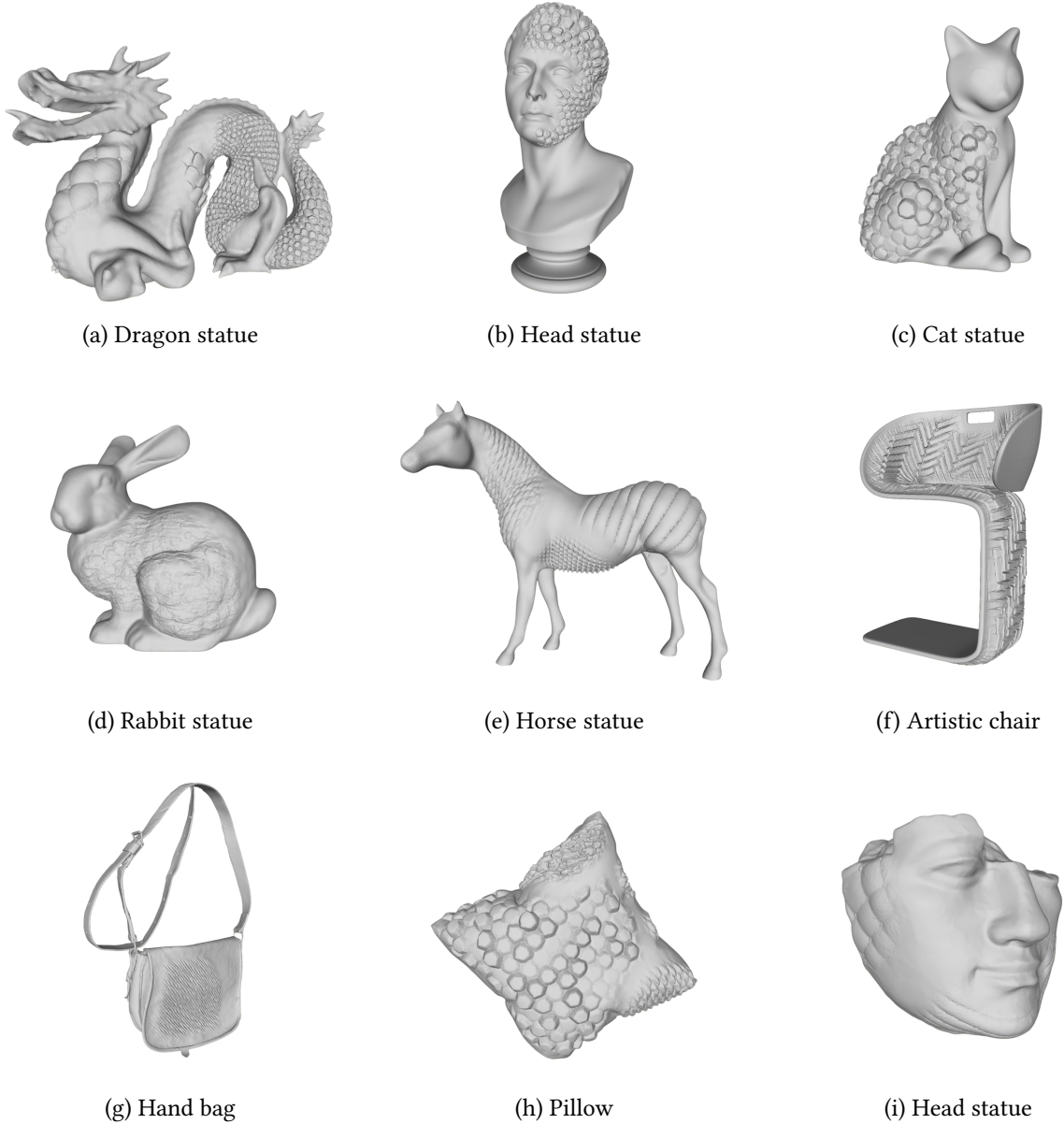


Figure 2: Variety of 3D decorative or functional objects including artistic textures with relief (sample of the considered dataset [18]).

3.1. Challenge dataset

The dataset provided in the context of this track is divided into three datasets: queries, retrieval and training [18]. The query set contains 54 3D objects each textured by one or two relief patterns, while the training set consists of 700 objects each having one to three different reliefs and each face of the training object (triangle mesh) is annotated with its corresponding relief class allowing for supervised training, additionally it contains only 15 reliefs, where there are reliefs that exist in both queries and retrieval objects but unseen to the training data. Finally, the retrieval set contains 300 objects of different geometries but with more challenging texture sizes and positioning. The 3D objects are composed of synthetic 3D models with geometric textures applied on them [18]. Fig. 2 shows examples of these objects, we can see that they contain variable art works that could be attributed to different cultures across history, where the texture analysis in real-world scenario would allow us to understand the cultural value of these artworks.

3.2. Data preparation

To facilitate the processing of the proposed raw dataset, we implemented an algorithm that allows us to map the original class indexation (which was randomized) into an ordered set of indexed relief categories, which relies on the pseudocode 1.

Pseudocode 1: Reassignment of Sparse Indexed Files into Successive Classes.

Input: A set of files $\mathcal{F} = \{f_i\}$, each named as $x_y.ply$, where $x, y \in \mathbb{N}$ represent category identifiers.

Output: A set of class directories containing reindexed and reassigned files (with duplicates when applicable).

Initialization:

Create an empty mapping table $T \leftarrow \emptyset$

Create an empty dictionary of classes $C \leftarrow \emptyset$

Step 1: Extract unique categories

foreach file $f \in \mathcal{F}$ **do**

- └ Extract (x, y) from the filename f
- └ Add x and y to the set of unique categories $\mathcal{X}$

Step 2: Reindex categories

Sort $\mathcal{X}$ in ascending order

Assign to each category $x_i \in \mathcal{X}$ a new consecutive index i : $T[x_i] \leftarrow i$

Step 3: Reassign files to classes

foreach file $f \in \mathcal{F}$ **do**

- └ Extract (x, y) from the filename f
- └ $i \leftarrow T[x], j \leftarrow T[y]$
- └ Copy f to class directory C_i
- └ Copy f to class directory C_j (creating a duplicate)

Step 4: Output

Return the set of class directories $\{C_i\}$ containing all reassigned files.

3.3. Pre-submission evaluation

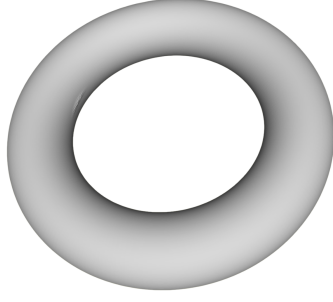
To evaluate our classifier, we used three standard metrics [19]:

- **Precision:** percentage of correct predictions among those classified as positive.
- **Recall:** proportion of true positive elements correctly identified.
- **F1-score:** harmonic mean of precision and recall, balancing them.

It is worth mentioning that the challenges faced regarding the dataset were the following:

- The large size of the meshes, averaging 7.2 MB and 393,543 faces per mesh in the training dataset, along with the constraint of providing a label for each face, makes the application of deep learning techniques very challenging in terms of computational resources and memory space. This is one of the main reasons why we proposed a different approach: generating a membership matrix instead of assigning a label to each face.
- The size of the relief patterns in some of the retrieval meshes, such as meshes 108 and 110, shown in Figure 3 are small and almost imperceptible to the naked eye. For these objects, the classification of the reliefs is highly complicated.

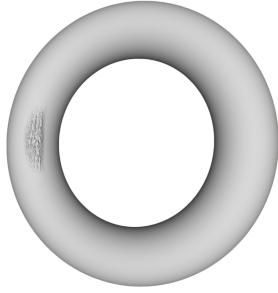
On the validation split, our model achieved the results shown in Table 2. The high precision means that the model is usually correct when it predicts a relief. The lower recall means that the model



(a) Object 108.



(b) its corresponding relief.



(c) Object 110.



(d) its corresponding relief.

Figure 3: Example of two objects with subtle surface embedded textures.

Table 1

Summary of the participating teams and their proposed methods with respect to the report [1].

Team	Proposed Method	Experimented Techniques	Evaluation
Team 1–3	Not submitted	N.A.	N.A.
Team 4	KU-3DSeg	LSCM + ResNet50	N.A.
Team 5 (Ours)	OMVMLP	VGG19 + MLP	See Table 3

sometimes misses reliefs that are present on the object [20]. Meaning it is careful not to make false predictions at the expense of overlooking certain existing patterns.

In our case, the training required approximately 56 minutes on Intel Xeon Bronze 3204 @ 1.90GHz CPU with an NVIDIA RTX 4000 Ada GPU, while the inference (view extraction, feature computation, classification) required 1.5 seconds per object.

3.4. Post-submission evaluation

Table 1 shows an overview of participants’ feedback on the challenge. It is shown that only two teams proposed solutions for the challenge. Our group’s methodology is explained in Section 2. As for Team 4, they proposed their method *KU-3DSeg*, in which they generate 3-channel 2D images that represent local geometric variations of randomly selected triangular meshes by applying the Least Squares Conformal Map (LSCM) algorithm. They proceed by fine-tuning a ResNet50 on classifying these images to detect the surface embedded textures [1]. However, they did not produce experimental results for their approach.

In our case, we submitted the two membership matrices produced to the track organizer for evaluation. However, the evaluation was limited to the queries that share reliefs with models in the retrieval set [1]. The organizer applied a set of information retrieval metrics to assess the quality of the membership matrix alongside the ranking of the retrieved objects [21]. These metrics include:

Table 2

Classification results on the validation set.

Precision	Recall	F1-score
0.885	0.744	0.809

Table 3

Results obtained for the evaluation of the membership matrix, reported from [1].

NN	FT	ST	mAP	nDCG	e	AUC
0.313	0.219	0.430	0.218	0.655	0.364	0.523

- **Nearest Neighbor (NN)**: indicates whether the first retrieved object (the one with the highest precision) is of the same class as the query, which is a strict measure of top-rank accuracy.
- **First Tier (FT)**: measures the percentage of relevant objects (the same class as the query) found in the top k results, where k is the number of relevant objects for that query.
- **Second Tier (ST)**: similar to First Tier but considers the top $2k$ results, providing a more tolerant evaluation of retrieval.
- **mean Average Precision (mAP)**: computes the average precision at each ranking where a relevant object appears, rewarding relevant objects appearing earlier in the ranking.
- **normalized Discounted Cumulative Gain (nDCG)**: emphasizes the ranking quality by giving higher weight to relevant objects ranked earlier and normalizes scores to account for different query sizes (number of relevant objects per query).
- **E-measure**: combines precision and recall into a single score adapted for information retrieval.
- **Area Under the ROC Curve (AUC)**: measures the model’s ability to distinguish relevant from non-relevant objects across all classes; a value close to 0.5 indicates near-random performance, while 1 indicates perfect discrimination.

Since the evaluation was done by the organizers, we report their results directly. Despite the promising results obtained on the validation split, the performance of the membership matrix was low. For example, the AUC was only 0.523 (Fig. 4), indicating limited discriminative ability [22]. The other metrics (Table 3) reinforce this observation: low mAP and nDCG indicate that relevant objects were often placed lower in the ranked list of retrieved 3D objects for each query, which shows that the model is yet limited in regards to generalization on unseen reliefs.

4. Conclusion and future work

In this paper, we presented our method OMVMLP, a view-based method for the retrieval of relief patterns in 3D objects allowing us to classify them based on their textures only, which would enable us to differentiate between the artworks of different civilizations and understand their cultural value. Our method which is built upon the pottery retrieval approach [9], provides additional insights towards analyzing multi-structured objects requiring finer-scale surface characterization. Our method combines capturing orthographic multi-view 2D image analysis with an MLP classifier which provides encouraging results on our validation set, demonstrating its ability to capture reliefs on 3D objects of different shapes. Hence it establishes our method as a preliminary descriptor of 3D objects with substructures. Additionally, although we had limited results that reveal the difficulty of generalization on unseen classes, we have been the only group among five international teams that was able to submit results regarding this challenge [1], which shows the capacity of our approach to perform an end-to-end 3D model analysis regardless of their size or complexity. Hence, it can be seen as a promising baseline towards conducting experiments in the reassembly of fragmented artifacts.

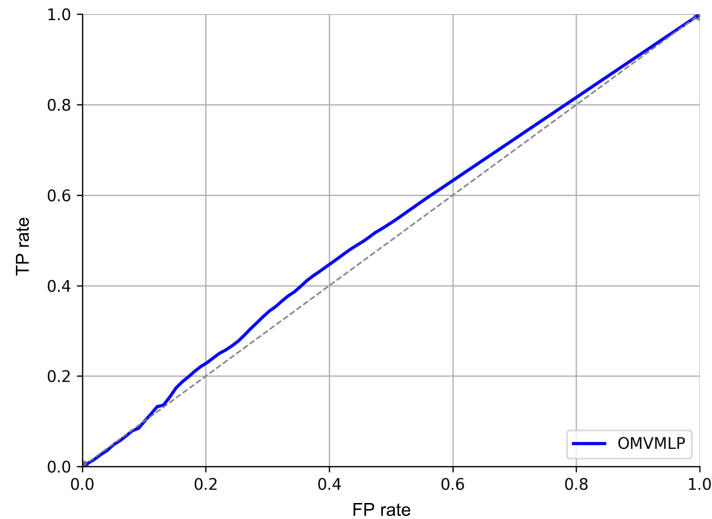


Figure 4: ROC curve for our method’s execution. The dotted diagonal represents the baseline performance of a random classifier, reported from [1].

To address the presence of unseen relief classes we applied two strategies, namely feature similarity calculation, and classifier fine-tuning, but further work is still required to integrate these approaches into a single pipeline that’s both accurate and efficient. Further directions include segmentation of the reliefs and investigating domain adaptation techniques to better handle unseen classes. Such advancement could help bridge the gap between controlled environment validation performance and real-world applications like the classification of different artworks.

Declaration on generative AI

The authors have not employed any Generative AI tools.

References

- [1] G. Paolini, C. Tortorici, S. Berretti, A. H. Youssef, H. Benhabiles, A. Cabani, R. He, K. Hammoudi, I. I. Ganapathi, S. S. Ali, D. Velayudhan, M. Assefa, N. Werghi, Shrec’25 track on multiple relief patterns: Report and analysis, 2025. doi:10.48550/ARXIV.2508.09909.
- [2] K. Lawonn, E. Trostmann, B. Preim, K. Hildebrandt, Visualization and Extraction of Carvings for Heritage Conservation, *IEEE Transactions on Visualization and Computer Graphics* 23 (2017) 801–810. doi:10.1109/tvcg.2016.2598603.
- [3] Y.-C. Choi, S. Murtala, B.-C. Jeong, K.-S. Choi, Deep Learning-Based Engraving Segmentation of 3-D Inscriptions Extracted From the Rough Surface of Ancient Stelae, *IEEE Access* 9 (2021) 153199–153212. doi:10.1109/access.2021.3127229.
- [4] M. Stamatopoulos, C. N. Anagnostopoulos, Digital modelling of ceramic sherds by means of photogrammetry and macrophotography: Uncertainty calculations and measurement errors (2023). doi:10.5281/ZENODO.7978086.
- [5] M. F. C. Pinho, G. L. A. Mota, G. A. O. P. da Costa, A Deep Learning Approach to Assist in Pottery Reconstruction from Its Sherds, *Heritage* 8 (2025) 167. doi:10.3390/heritage8050167.
- [6] J.-B. Barreau, T. Nicolas, G. Bruniaux, E. Petit, Q. Petit, Y. Bernard, R. Gaugne, V. Gouranton, Ceramics Fragments Digitization by Photogrammetry, Reconstructions and Applications, in: *International Conference on Cultural Heritage, EuroMed, 2014, Lemessos, Cyprus, 2014*. URL: <https://hal.science/hal-01090145>.

- [7] A. Godil, Z. Lian, H. Dutagaci, R. Fang, V. T. P., C. P. Cheung, Benchmarks, Performance Evaluation and Contests for 3D Shape Retrieval, 2011. doi:10.48550/ARXIV.1105.3685.
- [8] A. Giachetti, R. C. Giachetti, Effective Characterization of Relief Patterns, Computer Graphics Forum 37 (2018) 125–135. URL: https://www.researchgate.net/publication/327118382_Effective_Characterization_of_Relief_Patterns.
- [9] H. Benhabiles, H. Tabia, Convolutional neural network for pottery retrieval, Journal of Electronic Imaging 26 (2016) 011005. doi:10.1117/1.jei.26.1.011005.
- [10] Y. Feng, Y. Feng, H. You, X. Zhao, Y. Gao, MeshNet: Mesh Neural Network for 3D Shape Representation, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 8379–8386. arXiv:10.1609/aaai.v33i01.33018379.
- [11] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view Convolutional Neural Networks for 3D Shape Recognition, 2015. doi:10.48550/ARXIV.1505.00880.
- [12] C. R. Qi, H. Su, K. Mo, L. J. Guibas, PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, 2016. doi:10.48550/ARXIV.1612.00593.
- [13] M. Gao, N. Ruan, J. Shi, W. Zhou, Deep Neural Network for 3D Shape Classification Based on Mesh Feature, Sensors 22 (2022) 7040. doi:10.3390/s22187040.
- [14] K. Weiss, T. M. Khoshgoftaar, D. Wang, A Survey on Transfer Learning in Deep Neural Networks, IEEE Transactions on Neural Networks and Learning Systems 32 (2021) 4420–4436. arXiv:10.1109/TNNLS.2020.3024059.
- [15] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition (2014). doi:10.48550/ARXIV.1409.1556. arXiv:1409.1556.
- [16] X. Ouyang, X. Zhou, X. Li, H. Liu, Q. Li, K. Li, D. Xu, X. Ying, Deep Learning Architectures for Multi-Label Classification of Intelligent Health Risk Prediction, BMC Bioinformatics 18 (2017) 542. doi:10.1186/s12859-017-1898-z. arXiv:10.1186/s12859-017-1898-z.
- [17] I. Loshchilov, F. Hutter, SGDR: Stochastic Gradient Descent with Warm Restarts, in: International Conference on Learning Representations (ICLR), 2017. arXiv:1608.03983.
- [18] G. Paolini, C. Tortorici, S. Berretti, SHREC’25 track: Retrieval and Segmentation of Multiple Relief Patterns (2025). doi:10.2312/3DOR.20251202.
- [19] T. Fawcett, An Introduction to ROC Analysis, Pattern Recognition Letters 27 (2006) 861–874. URL: <https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X>. doi:10.1016/j.patrec.2005.10.010.
- [20] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing & Management 45 (2009) 427–437. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0306457309000259>.
- [21] C. D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008. URL: https://www.researchgate.net/publication/238833393_Introduction_to_Information_Retrieval_Christopher_D_Manning_Prabhakar_Raghavan_and_Hinrich_Schutze_Stanford_University_Yahoo_Research_and_University_of_Stuttgart_Cambridge_Cambridge_University_Press_2.
- [22] A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition 30 (1997) 1145–1159. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0031320396001422>.