

Visual Microfossil Identification via Deep Metric Learning

Tayfun Karaderi^{1,3}, Tilo Burghardt¹, Allison Y. Hsiang²,
 Jacob Ramaer¹, Daniela N. Schmidt³

¹Department of Computer Science, University of Bristol, UK

²Institutionen för geologiska vetenskaper, Stockholm University, Sweden

³School of Earth Sciences, University of Bristol, UK

Abstract. We apply deep metric learning for the first time to the problem of classifying planktic foraminifer shells on microscopic images. This species recognition task is an important information source and scientific pillar for reconstructing past climates. All foraminifer CNN recognition pipelines in the literature produce black-box classifiers that lack visualisation options for human experts and cannot be applied to open set problems. Here, we benchmark metric learning against these pipelines, produce the first scientific visualisation of the phenotypic planktic foraminifer morphology space, and demonstrate that metric learning can be used to cluster species unseen during training. We show that metric learning outperforms all published CNN-based state-of-the-art benchmarks in this domain. We evaluate our approach on the 34,640 expert-annotated images of the Endless Forams public library of 35 modern planktic foraminifera species. Our results on this data show leading 92% accuracy (at 0.84 F1-score) in reproducing expert labels on withheld test data, and 66.5% accuracy (at 0.70 F1-score) when clustering species never encountered in training. We conclude that metric learning is highly effective for this domain and serves as an important tool towards expert-in-the-loop automation of microfossil identification. Key code, network weights, and data splits are published with this paper for full reproducibility.

Keywords: applied computer vision · planktic foraminifers · deep learning · animal biometrics · paleobiology · climate science.

1 Introduction

Motivation. Planktic foraminifers are an invaluable source of information for reconstructing past climate records [25]. Estimating ocean temperature, salinity, and pH using foraminifers involves quantifying the species composition of shell assemblages or picking individual specimens, often at low abundance, out of thousands of specimens. Identification of specimens to the species level is necessary as species-specific vital effects can result in different isotopic fractionation values [34]. Foraminifers grow their calcium carbonate shells by adding chambers in a spiral, where the main gross morphological traits used for taxonomic identification are chamber form and arrangement, the size and position of an aperture, and other features [15] (see Figs. 1 and 2). The differences between species are often plastic along sliding morphological change [35] and human identifiers manipulate the specimen under the microscope to aid recognition. In contrast, single-view static image recognition confidence can be restricted by acquisition artefacts, imaging quality, and the viewpoint-dependent visibility limitations of traits. These factors pose a significant challenge with regard to the use of computer vision systems for automating single image identification.

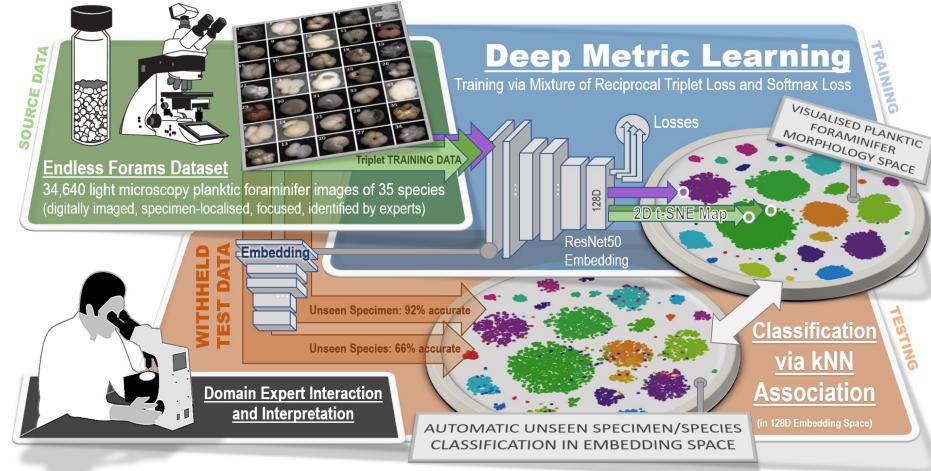


Fig. 1. Deep Metric Learning of Foraminifer Spaces. Schematic overview of the approach. Expert-annotated planktic foraminifer imagery (green) is used to metrically learn (blue) embeddings of specimen visuals in a 128D morphological appearance space via a Resnet50-based deep network using hybrid reciprocal triplet and Softmax loss minimisation. The built space naturally allows for a t-SNE visualisation of specimens and species including their geometric and topological relations. Unseen specimens and even unseen species not used for training can be projected into this space for automatic classification (red). Domain experts can now for the first time see the location of specimens in the morphological space and interpret results taxonomically.

Paper Contributions. Inspired by classic machine learning (ML) applications for microfossils [4,5,48], recent pioneering works [19,30] started to evaluate convolutional neural networks (CNNs) for visual foraminifer recognition. However, none of the published pipelines so far allow for meaningful visualisations and expert interactions with the learned space of specimens, application to unseen species, or utilisation of contrastive gradients. In this paper, we address the above limitations for the first time via deep metric learning [2,40] in the domain. Fig. 1 illustrates our approach visually. After a review of background and methodology, we will experimentally demonstrate that metric learning can outperform the current state-of-the-art (SOTA) in foraminifer classification benchmarks. In addition, we will show that our approach can generate grey-box foraminifer classification models with which domain experts can interact. Finally, we will explore how far the projection of unseen foraminifer species into this space (zero-shot scenario) can generalise learning beyond trained foraminifer classes.

2 Background

Manual Microfossil Identification. Despite the importance of foraminifers in paleoclimatology, few aggregated public resources exist to train people in the task of distinguishing morphologically similar taxa. Further, diverging views on species concepts and boundaries (e.g., ‘clumpers’ vs. ‘splitters’) result in conflicting taxonomies in the literature. As a result, the taxonomic agreement is some-

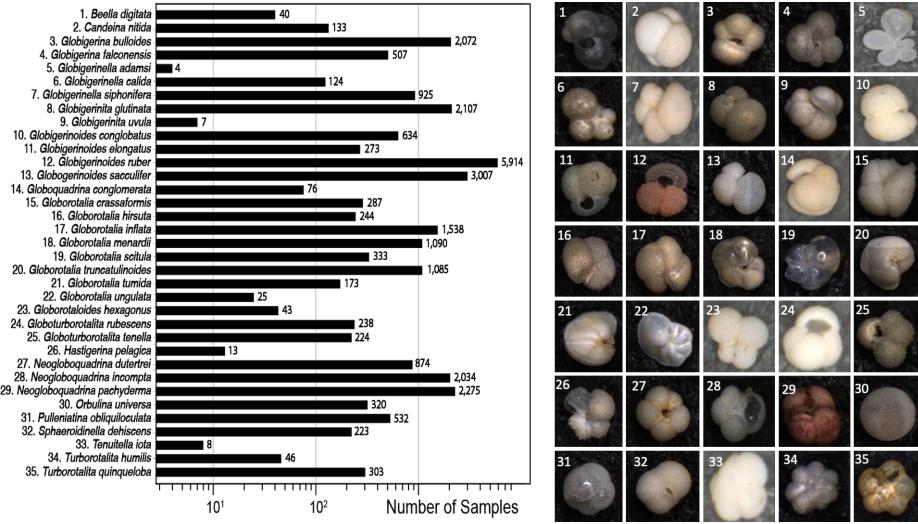


Fig. 2. Foraminifer Microscope Imagery. The Endless Forams public image library is used as our data source of modern planktonic foraminifera. (*left*) The distribution of the species classes varies strongly with a mean of 795 and a standard deviation of 1171. (*right*) Visualisation of a sample from each of the 35 species in the dataset.

times only $\sim 75\%$ [1] for planktic foraminifera since they exhibit near-continuous morphological gradations between closely related taxa [11] (see Fig. 2). In some cases, morphological variation is unrelated to genetic differentiation; however in others genetic analysis has revealed the existence of some pseudo-cryptic species between morphological endmembers [20]. These difficulties resulted in different species concepts over time, which are particularly prominent in self-trained taxonomists [1]. Some of these challenges might be removed by growing databases of expert-classified images [7], opening opportunities for machine-driven classification that limit subjective biases of human classifiers.

Current Machine Vision for Microfossil Classification. Modern semantic image classification frameworks are almost without exception grounded in feed-forward CNNs, introduced in their earliest form by the ground-breaking work of Krizhevsky et al. [24] and leading to further milestone architectures including VGG16 [42], Inception [44], ResNet [14], ResNeXt [50], and deep transformer networks [47]. Taxonomic computer vision (CV) applications of these techniques for microfossils are still very rare in the literature, despite steep advances in general animal biometrics [26,46]. Nevertheless, ML-based species identification [8,33] has been applied to several microscopic taxa, including coccoliths [43], pollen [45], and phytoplankton [49]. While there have been early attempts for automatic classification on marine microfossils [4,48] the most successful focused on coccoliths which are predominantly flat [5,6]. There exist only very few papers which investigate the use of modern deep learning techniques on planktic foraminifers [19,30,32], all of which put forward traditional non-contrastive CNN architectures optimising for prediction correctness via SoftMax cross-entropy.

Metric Latent Spaces. Metric learning [2,40] moves away from focusing learning on optimising prediction correctness only; instead, a mapping into a class-distinctive latent space is constructed where maps to the same class naturally cluster together and distances directly relate to input similarity under the training task. A simple way of building a latent space of this form is to pass two inputs through an embedding function and then use a contrastive loss L_C [13]:

$$L_C = (1 - Y)0.5d(x_1, x_2) + 0.5Y\max(0, \alpha - d(x_1, x_2)), \quad (1)$$

where x_1 and x_2 are the embedded input vectors, Y is a binary label denoting class equivalence/difference for the two inputs, and $d(\cdot, \cdot)$ is the Euclidean distance between two embeddings. However, this formulation cannot put similarities and dissimilarities between different pairs of embeddings in relation. A triplet loss formulation [41] instead utilises embeddings x_a , x_p and x_n denoting an anchor, a positive example of the same class, and a negative example of a different class, respectively. Minimising the distance between the same-class pair and maximising the distance between the different-class pair can be achieved by:

$$L_{TL} = \max(0; d(x_a, x_p) - d(x_a, x_n) + \alpha), \quad (2)$$

where α is the margin hyper-parameter. Reciprocal triplet loss removes the need for this parameter [31] and accounts for large margins far away from the anchor:

$$L_{RTL} = d(x_a, x_p) + 1/d(x_a, x_n). \quad (3)$$

Including a SoftMax term in this loss can improve performance, as shown by recent work [27,17]. Thus, the SoftMax and reciprocal triplet losses can be combined into the standard formulation first published in [2] used in this paper:

$$L = \frac{-\log(e^{x_{class}})}{\sum_i e^{x_i}} + \lambda L_{RTL}, \quad (4)$$

where λ is a mixing hyper-parameter. For the foraminifer classification problem in particular this allows both relative inter-species difference information captured by the reciprocal triplet loss component as well as overall species information captured by the SoftMax term to be used as backpropagation gradients.

Latent Space Partitioning for Classification. Once an embedding function has been learned for a metric space, any new sample – even one of the unseen classes – can be projected into this space. However, assigning a class label to this sample based on its position in the space eventually requires a partitioning of the latent space. Direct maps, as well as hierarchical [22] and partitional [12] clustering algorithms, have been used for this. We will experiment with a wide variety of partitioning options in Tab. 2(*top right*) including Gaussian Mixture Models (GMMs) [36], Logistic Regression, Support Vector Machines (SVMs) [9], Multi-layer Perceptrons (MLPs) [8], and k-Nearest Neighbours [18]. Once partitioned, metric spaces expose model structure plus outlier information, can be visualised by mapping into lower-dimensional spaces, and often capture properties of the target domain beyond the specific samples used for training. Before using these techniques on the problem of foraminifer classification, we will first introduce the dataset and experimental settings.

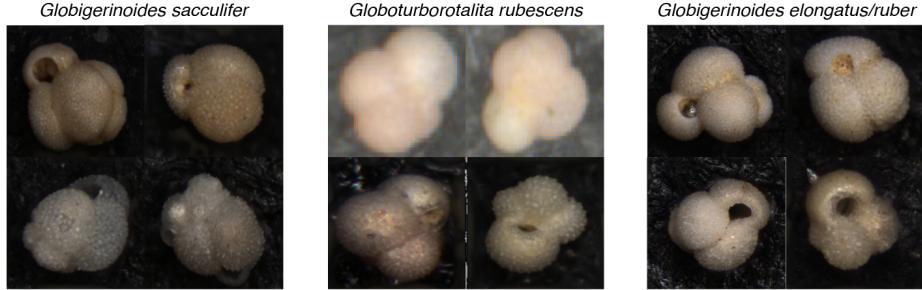


Fig. 3. Imaging Details and Classification Challenges. Microscopic imaging of 3D microfossils often obfuscates key taxonomic features due to variable viewpoint and acquisition conditions. (*left*) Within-class 3D viewpoint variability for *Globigerinoides sacculifer* yields 2D images with different visible chamber numbers rendering this key feature less informative. (*middle*) Out-of-focus acquisition can remove species-characteristic texture information, here of *Globoturborotalita rubescens*, blurred in the top images. Light microscopy images are also unable to capture fine grain wall texture details that would be useful for taxonomic identification. (*right*) Morphologically similar species such as *Globigerinoides elongatus* (top) and *G. ruber* (bottom) are difficult to distinguish from single-view static visuals alone. Ground truth labeling often requires additional information to establish secure taxonomic classification.

3 Dataset

Endless Forams. We use the Endless Forams image library [21] for all experiments. It is one of the largest datasets of its kind and publically available (at endlessforams.org). It contains 34,640 labelled images of 35 different foraminiferal species as detailed and exemplified in Fig. 2. This dataset was built based on a subset of foraminifer samples from the Yale Peabody Museum (YPM) Coretop Collection [10] and the Natural History Museum, London (NHM) Henry A. Buckley Collection [37]. The dataset is also associated with a taxonomic training portal hosted on the citizen science platform Zooniverse (zooniverse.org/projects/ahsiang/endless-forams). Species classification in this dataset is truly challenging compared to many other computer vision tasks since: 1) planktonic foraminifers exhibit significant intra-class variability (see Fig. 3(*left*)); 2) critical morphological properties are not consistent across 3D viewpoint and acquisition conditions (see Fig. 3(*left*), (*middle*)); and 3) visual intra-species differences are barely apparent between some taxa (see Fig. 3(*right*)).

4 Experimental Setup

Implementation Details. For all experiments, our PyTorch-implemented metric learning architecture extends a ResNet50 backbone pre-trained with ImageNet. The network is set to optimise the loss specified in Eq. 4 combining SoftMax and reciprocal triplet loss components with the mixing parameter $\lambda = 0.01$ as suggested in [17]. Training progresses via the SGD [38] optimiser for 100 epochs as quantitatively illustrated in Fig. 4. For full testing comparability

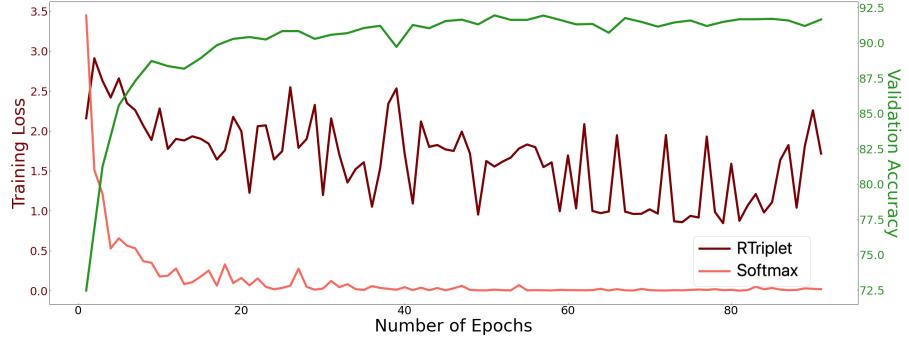


Fig. 4. Metric Learning Training Details. Curves quantifying the evolution of training losses and validation accuracy across all 100 epochs of training. Note that the two loss components are plotted separately showing the early conversion of the Softmax dimension, whilst RTriplet loss continues to alter the metric embedding space at higher magnitudes albeit at down-scaled weighting given $\lambda = 0.01$ in Eq. 4.

with [21] we utilised their withheld test set for performance stipulation, whilst using the remaining 27,731 images augmented via rotations, scale, and Gaussian noise transforms for training (reported as **Ours**). In a second workflow and for full compatibility with [30], we also produced results via 5-fold cross-validation on random train-test data splits (reported as **Ours***). For sample selection during training of all workflows, we follow the ‘batch hard’ mining approach [16] where triplets are mined within mini-batches. This yields overall moderate triplets, i.e. training with the hardest examples within each mini-batch. The published source code [23] provides full details regarding all of the above for full result reproducibility. Training takes approx. 48 hours on a P100 GPU system with 12GB RAM. We obtain classifications for each test sample projected into the metric space via running kNN [18] with $n=5$ over the projected training samples.

5 Results.

Baseline Comparisons. The simplest baselines for the foraminifer species classification problem are given by using transfer learning via ImageNet-initialised off-the-shelf CNN architectures [30]. Tab. 1(*rows 01-06*) compares key architectures and their performance against our setup using 5-fold cross-validation at standard resolution of 224×224 . These results show that metric learning dominates such baselines without exception. Operating at 160×160 pixels, the original 2019 VGG-based benchmark by Hsiang et al. [21] achieved 87.4% accuracy on a withheld test set. Using the same test set, our metric learning approach improves on this performance significantly by 4.5% as shown in Tab. 1(*rows 07-08*).

Improving the State-of-the-Art. CycleNet used by Marchant et al. [30] claims current state-of-the-art performance on the data at an accuracy of 90.3% when using 5-fold cross-validation at 128×128 . Tab. 1(*rows 09-12*) compare their key results against our metric learning approach under this regime, showing improved metric learning accuracy of 91%. When using their best-performing

	Row	Method	Resolution (pixels)	Acc (weighted)	Prec (macro)	Rec (macro)	F1S (macro)
A	01	VGG19* [30]	224 × 224	77.1	70.5	64.9	66.9
	02	MobileNetV2* [30]	224 × 224	77.7	70.0	65.2	66.8
	03	InceptionV2* [30]	224 × 224	77.7	69.5	64.8	66.4
	04	DenseNet121* [30]	224 × 224	80.2	75.2	69.2	71.3
	05	ResNet50* [30]	224 × 224	81.8	76.7	71.4	73.4
	06	Ours*	224 × 224	91.6	88.2	78.3	81.3
B	07	VGG16 [21]	160 × 160	87.4	72.4	69.8	70.0
	08	Ours	160 × 160	91.9	91.3	81.9	84.6
C	09	ResNet18Full* [30]	128 × 128	88.5	84.1	77.8	79.9
	10	ResNet50CycleFull* [30]	128 × 128	90.1	85.1	78.7	80.8
	11	BaseCycleFull* [30]	128 × 128	90.3	84.9	78.4	80.5
	12	Ours*	128 × 128	91.0	87.2	78.9	81.1
D	13	BaseCycleFull [30]	224 × 224	90.5	84.5	79.6	81.5
	14	Ours	224 × 224	91.7	88.1	77.0	80.7
			416 × 416	92.0	89.0	81.5	84.2

Table 1. Species Classification Results. Performance of our metric learning approach against SOTA techniques [30,21] quantified via accuracy, precision, recall and F1 score. Testing regime and input resolution were made to match the settings of previous works exactly where * indicates 5-fold cross-validation (see Sec. 4). **(A)** Metric learning supremacy against various off-the-shelf CNNs at 224 × 224 initialised via ImageNet as described in [30]. **(B)** Metric learning increases accuracy of [21] at 160 × 160 by 4.5%. **(C)** Critically, our approach outperforms SOTA CycleNet [30] at their published 128 × 128 resolution. **(D)** Finally, re-running CycleNet [30] under the regime of [21] and at 224 × 224 shows metric learning also dominating peaking at 92% at 416 × 416.

CycleNet [30], scaling it up to a standard 224 × 224 resolution and using the testing regime of [21] their results improve slightly. However, our approach outperforms this new benchmark by another 1.2% as shown in Tab. 1 (rows 13-14). Metric learning reaches top accuracy of 92% at resolution 416 × 416 as shown in Tab. 1 (row 14). Fig. 5 (right) depicts a detailed confusion matrix for this setting.

Visualisation of Foraminifer Space. In contrast to basic CNN approaches, we can now visualise the learned 128D metric space, revealing an appearance-based distribution of planktic foraminifers. Fig. 5 (left) shows this first scientific visualisation of phenotypic appearance space by projecting training and testing sets into the metric space before using t-SNE [28] to reduce dimensionality to 2.

Open Set Performance. By withholding some species¹ from training altogether, we can start to evaluate metric learning potential beyond seen training species, i.e. learning about the planktic foraminifer domain more generally. To do this, we project the unused training data of withheld classes during testing into the metric space together with the training data of used classes. We then utilise kNN with all these data points to measure how far test sets of seen or unseen classes are classified correctly. For the unseen species classes at 224 × 224, accuracy drops to 66.5% (at 0.70 F1-score) in our experiment. This result is remarkable since about 2/3 of never-seen species specimens can still be associated correctly amongst all 35 species. We thus conclude that metric learning does indeed capture important general features relevant for planktic foraminifer classification beyond any particular species appearance.

¹ Tail classes 1, 5, 9, 14, 22, 23, 26, 29, 33, and 34 were chosen as our open set to have maximum specimen counts available during training.

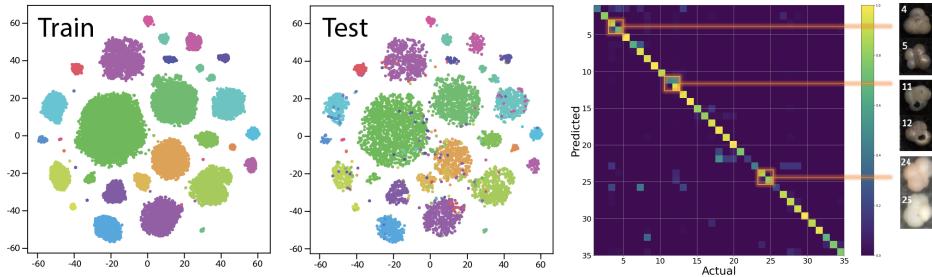


Fig. 5. Learned Metric Space Visualisation and Confusion Matrix. (left) 2D t-SNE visualisation of the training (RandIndex=99.9) and testing (RandIndex=87.0) sets projected into the metric space; and (right) Confusion matrix detailing test performance where taxonomic similarities of prominent mix-ups are visually highlighted.

Ablations, Resolution, and Augmentation. Tab. 2(*top*) details ablation experiments that demonstrate that the chosen metric learning loss functions and clustering algorithms are effective, contribute to performance, and indeed outperform other tested setups. An analysis of the dependency of our approach on resolution is presented in Tab. 2(*bottom left*). It outlines that above 160×160 pixels performance gains are widely diminished and flat-line, supporting the choices taken in [21]. Whilst SOTA competitor networks [30] also augment their data, augmentation can be essential for metric learning techniques generally [40]. Tab. 2(*bottom right*) quantifies this fact and shows that performance in the foraminifer classification domain is indeed intricately linked to augmentation, with rotation variations producing the strongest component impact.

6 Taxonomic Reflection.

Taxonomic Interpretation of Performance. We observe that species that are taxonomically related to each other are indeed the ones that are confused most often by the machine. For example, species from the same genus (e.g. 4 vs 5; 11 vs 12; 24 vs 25) are often misclassified symmetrically as highlighted in Fig. 5(*right*). This alignment is consistent with human classification difficulties and suggests that metric space distances are indeed related to visual properties humans use to differentiate species. Generally, we suggest three main reasons for lower classification performances, all of which would also present challenges for human classifiers. First, phylogenetic closeness, i.e. closely related sister taxa, can result in many shared taxonomic features that make differentiation difficult. Classes 4 and 5, for example, share many large-scale morphological features and are only distinguished by a lip at the aperture [29]. Secondly, some classes (e.g. 11 and 12) are not even typically distinguished in palaeontological studies, but are genetically distinct [3] and hence separated in the database. The distinguishing visual feature is a sliding scale of radial chamber compression which makes systematic labelling challenging. Thirdly, not all planktic foraminiferal species reach the same maximum size in the modern ocean [39], and size is an important feature used for taxonomic classification by human experts. As all the images have the same number of pixels, size information is lost in the database used and

Loss	Acc	Prec	Rec	F1S	RI	Alg	Acc	Prec	Rec	F1S	RI
Ours ($L_{RTL} + SM$)	91.9	91.3	81.9	84.6	87.1	Ours (kNN)	91.9	91.3	81.9	84.6	87.1
SoftMax+ L_{TL}	91.4	89.6	81.3	84.3	86.2	SVM	91.8	88.3	78.8	81.6	86.6
RTriplet (L_{RTL})	89.3	79.2	71.0	73.9	82.9	MLP	90.9	80.9	80.8	78.9	86.2
Triplet (L_{TL})	22.2	12	9.3	9.5	4.7	LR	91.5	84.6	74.7	78.0	86.4
SoftMax	91.2	89.9	81.9	84.2	86.5	GMM	-	-	-	-	63.8

Resolution	Acc	Prec	Rec	F1S	RI	Transform	Acc	Prec	Rec	F1S	RI
416 × 416	92.0	89.0	81.5	84.2	87.3	Ours (R+S+G)	91.9	91.3	81.9	84.6	87.1
224 × 224	91.7	88.1	77.0	80.7	86.8	R	91.2	86.8	79.3	81.5	85.8
160 × 160	91.9	91.3	81.9	84.6	87.1	S	88.4	87.6	76.3	80.0	81.1
120 × 120	91.0	88.4	79.2	82.1	86.3	G	85.0	75.0	66.2	69.1	76.3
80 × 80	90.0	84.8	78.1	80.2	84.1	-	82.7	75.2	66.2	69.0	72.4

Table 2. Design Ablations, Resolution and Augmentation. Benchmarks use the testing regime of [21] and resolution 160×160 as standard; RI denotes the Rand Index. (*top left*) Performance of different loss functions confirm the superiority of L . (*top right*) Benchmark of various clustering approaches. (*bottom left*) Performance increases saturate around 160×160 providing only marginal gains if input resolution is increased beyond this point. (*bottom right*) Metric learning performance in this domain is intricately dependent on sufficient data augmentation. Rotation augmentation (R) in particular benefits accuracy, whilst scale augmentation (S), and Gaussian noise addition (G) have smaller effects. Applying all three (R+S+G) is most beneficial.

the amount of information per pixel will vary strongly between a species which is up to 5 times larger than another. For the smallest specimens, imaging often reaches the limits of typical optics, resulting in blur and other effects, as shown in Fig. 3(*middle*) for class 24.

7 Conclusion.

The accurate and efficient visual recognition of planktic foraminifers within shell assemblages in light microscopy imagery is an important pillar required for unlocking climatic archives. We have shown here for the first time that deep metric learning can be effectively applied to this task. We documented how deep metric learning outperforms all published state-of-the-art benchmarks in this domain when tested on one of the largest public, expert-annotated image libraries covering modern planktic foraminifera species. We further provided detailed result evaluation and ablation studies. Based on the metrically-learned feature spaces, we also produced the first scientific visualisation of a phenotypic planktic foraminifer appearance domain. Finally, we demonstrated that metric learning can be used to recognise specimens of species unseen during training by utilising the metric space embedding capabilities. We conclude that metric learning is highly effective for this domain and can form an important tool towards data-driven, expert-in-the-loop automation of microfossil identification.

Acknowledgements.

TK was supported by the UKRI CDT in Interactive Artificial Intelligence under the grant EP/S022937/1. AYH was supported by VR grant 2020-03515. DNS was supported by NERC grant NE/P019439/1. We thank R Marchant and his team for making available source code and testing regime details to compare to [30]. Thanks to M Lagunes-Fortiz and W Andrew for permitting use and adaptation of source code related to metric learning.

References

1. Al-Sabouni, N., Fenton, I., Telford, R., Kučera, M.: Reproducibility of species recognition in modern planktonic foraminifera and its implications for analyses of community structure. *J. Micropalaeontol.* **37**, 519–534 (2018)
2. Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A.W., Burghardt, T.: Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Computers and Electronics in Agriculture* **185**, 106133 (2021)
3. Aurahs, R., Treis, Y., Darling, K., Kucera, M.: A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence. *Marine Micropaleontology* **79**, 1–14 (2011)
4. Balfoort, H., Snoek, J., Smiths, J., Breedveld, L., Hofstraat, J., Ringelberg, J.: Automatic identification of algae: neural network analysis of flow cytometric data. *Journal of Plankton Research* **14**, 575–589 (1992)
5. Beaufort, L., Dollfus, D.: Automatic recognition of coccoliths by dynamical neural networks. *Marine Micropaleontology* **51**, 57–73 (2004)
6. Beaufort, L., Probert, I., de Garidel-Thoron, T., Bendif, E., Ruiz-Pino, D., Metzl, N., Goyet, C., Buchet, N., Coupel, P., Grelaud, M., Rost, B., Rickaby, R., de Vargas, C.: Sensitivity of coccolithophores to carbonate chemistry and ocean acidification. *Nature* **476**, 80–83 (2011)
7. Bown, P., Huber, B., Wade, B., Young, J.: pforams@mikrotax - introduction (2020), link: <https://www.mikrotax.org/pforams/>, last accessed: 01/05/2021
8. Bozinovski, S., Fulgosi, A.: The use of artificial neural networks to classify primate vocalizations: a pilot study on black lemurs. *American Journal of Primatology* pp. 1098–2345 (2009)
9. Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* **20**(3), 273–297 (1995)
10. Elder, L.E., Hsiang, A.Y., Nelson, K., Strotz, L.C., Kahanamoku, S.S., Hull, P.M.: Sixty-one thousand recent planktonic foraminifera from the Atlantic Ocean. *Scientific Data* **5** (2018)
11. Franz, N.: On the lack of good scientific reasons for the growing phylogeny/classification gap. *Cladistics* **21**, 495–500 (2005)
12. Gandhi, G., Srivastava, R.: Review paper: A comparative study on partitioning techniques of clustering algorithms. *International Journal of Computer Applications* **87** (01 2014)
13. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* **2**, 1735–1742 (2006)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
15. Hemleben, C., Spindler, M., Anderson, O.: Modern planktonic foraminifera. Springer p. 363 (1989)
16. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification (03 2017)
17. Hodan, T., Haluza, P., Obdrzalek, S., Matas, J., Lourakis, M., Zabulis, X.: T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects. *Winter Conference on Applications of Computer Vision (WACV)*, IEEE **0**, 880–888 (2017)
18. Hodges, L.: Nonparametric discrimination: Consistency properties. *USAF School of Aviation Medicine* (1951)

19. Hsiang, A.Y., et al.: Endless Forams: > 34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology* **34**, 1157–1177 (2019)
20. Huber, B., Bijma, J., Darling, K.: Cryptic speciation in the living planktonic foraminifer *Globigerinella siphoniphera* (d'Orbigny). *Paleobiology* **23**, 33–62 (1997)
21. Hull, P.M., Hsiang, A.Y.: Endless Forams Most Beautiful (2020), link: <http://endlessforams.org>, last accessed: 01/05/2021
22. Johnson, S.C.: Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967)
23. Karaderi, T.: Visual microfossil identification via metric learning (2022), <https://github.com/TayfunKaraderi/ICPRAI-2022-Visual-Microfossil-Identification-via-Deep-Metric-Learning>, last accessed: 19/03/2022
24. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* pp. 1097–1105 (2012)
25. Kucera, M.: Planktonic foraminifera as tracers of past oceanic environments. C. Hillaire-Marcel and A. De Vernal (Editors) pp. 213–262 (2007)
26. Kühl, H.S., Burghardt, T.: Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution* **28**(7), 432–441 (2013)
27. Lagunes-Fortiz, M., Damen, D., Mayol-Cuevas, W.: Learning discriminative embeddings for object recognition on-the-y. 2019 International Conference on Robotics and Automation (ICRA), IEEE **0**, 2932–2938 (2019)
28. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008)
29. Malmgren, B., Kennett, J.: Biometric analysis of phenotypic variation in recent globigerina bulloides d'orbigny in the southern indian ocean. *Marine Micropaleontology* **1**, 2–25 (1976)
30. Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., de Garidel-Thoron, T.: Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology* **39**(2), 183–202 (2020)
31. Masullo, A., Burghardt, T., Damen, D., Perrett, T., Mirmehdi, M.: Who goes there? Exploiting silhouettes and wearable signals for subject identification in multi-person environments. *Proceedings of the IEEE International Conference on Computer Vision Workshops* pp. 1599–1607 (2019)
32. Mitra, R., Marchitto, T., Ge, Q., Zhong, B., Kanakiya, B., Cook, M., Fehrenbacher, J., Ortiz, J., Tripathi, A., Lobaton, E.: Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. *Marine Micropaleontology* **147**, 16–24 (2019)
33. Pacifico, L.D.S., Macario, V., Oliveira, J.F.L.: Plant classification using artificial neural networks. *IJCNN* pp. 1–6 (2018)
34. Ravelo, A.C., Hillaire-Marcel, C.: The use of oxygen and carbon isotopes of foraminifera in paleoceanography, vol. 1. Chapter in *Developments in Marine Geology*, Elsevier (2007)
35. Renaud, S., Schmidt, D.: Habitat tracking as a response of the planktic foraminifer *Globorotalia truncatulinoides* to environmental fluctuations during the last 140 kyr. *Marine Micropaleontology* **49**, 97–122 (2003)
36. Reynolds, D.A.: Gaussian mixture models. *Encyclopedia Biometrics* (2009)
37. Rillo, M.C., Whittaker, J., Ezard, T.H., Purvis, A., Henderson, A., Stukins, S., Miller, C.: The unknown planktonic foraminiferal pioneer Henry A. Buckley and his collection at The Natural History Museum. *Journal of Micropalaeontology* **36**, 191–194 (2016)

38. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
39. Schmidt, D., Renaud, S., Bollmann, J., Schiebel, R., Thierstein, H.: Size distribution of Holocene planktic foraminifer assemblages: biogeography, ecology and adaptation. *Marine Micropaleontology* **50**, 319–338 (2004)
40. Schneider, S., Taylor, G.W., Linquist, S.S., Kremer, S.C.: Similarity learning networks for animal individual re-identification - beyond the capabilities of a human observer. *WACV* **abs/1902.09324** (2019), <http://arxiv.org/abs/1902.09324>
41. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition* **2**, 815–823 (2015)
42. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)
43. Sluys, R.: The unappreciated, fundamentally analytical nature of taxonomy and the implications for the inventory of biodiversity. *Biodiversity and Conservation* **22**, 1095–1105 (2013)
44. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1–9 (2015)
45. Tan, D., Ang, Y., Lim, G., Ismail, M., Meier, R.: From ‘cryptic species’ to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (sepsidae: Diptera). *Zoologica Scripta* **39**, 51–61 (2010)
46. Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risso, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., et al.: Seeing biodiversity: perspectives in machine learning for wildlife conservation. arXiv preprint arXiv:2110.12951 (2021)
47. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *CoRR* **1706.03762** (2017)
48. Weller, A., Harris, A., Ware, J.: Two supervised neural networks for classification of sedimentary organic matter images from palynological preparations. *Mathematical Geology* **39**, 657–671 (2007)
49. Wägele, H., Klussmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., Wägele, J.: The taxonomist – an endangered race. A practical proposal for its survival. *Frontiers in Zoology* **8**, 25 (2011)
50. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks (2017)