Towards Engram Memory Machines. A Feasibility Study of Image Classification Using Warping and Low-dimensional Class Prototypes.

F Colecchia

College of Engineering, Design and Physical Sciences, Brunel University of London, Kingston Lane, Uxbridge UB8 3PH, UK

Email: federico.colecchia@brunel.ac.uk

Abstract. Against the backdrop of ongoing green AI research to reduce the environmental footprint of artificial intelligence technology, a new supervised machine learning approach is presented with potential to contribute to the development of artificial systems capable of learning incrementally from new data. This article reports results from a feasibility study on image classification using the Modified National Institute of Standards and Technology data set. Training consists of producing class-level image prototypes, referred to as engrams, using a single-pass method. At inference stage, previously unseen images are warped to match the engrams individually, using a continuous transformation including a tuneable rigidity parameter. The images are classified by maximising similarity to the engrams across classes following warping. Unlike alternative approaches recording binary feature coincidences in high-dimensional spaces, this method operates on low-dimensional data representations. In addition to enhancing interpretability, this has the advantage of circumventing memory scaling bottlenecks. A test accuracy above 90% was observed in this study, which is a promising result considering that the engrams exclusively encoded average data patterns. Future work will focus on reducing inference time, enhancing classification accuracy and estimating the performance of the method on additional datasets.

Keywords: Green AI. Single-pass supervised learning. Image classification. Image warping. Image registration.

1. Introduction

The concept of green AI has received significant attention in recent years in an effort to reduce the environmental footprint of artificial intelligence (AI) technology [1]. As far as supervised training algorithms are concerned, single-pass methods are noteworthy, as they circumvent the need for iterative refinement of artificial neural network parameters during training and can therefore significantly reduce training time. Examples include data stream learning algorithms such as incremental trees, online classification algorithms, Online Support Vector Machines, Sparse Binary Coincidence (SBC), one-pass online learning, Budgeted Stochastic Gradient Descent (BSGD) and Fourier Online Gradient Descent [2-8].

Supervised training of traditional neural networks normally occurs via iterative modification of parameters at the level of individual neurons, in such a way that previously unseen input data, when

processed through the network at inference stage, result in the activation of target output neurons. The backpropagation algorithm [9] is the mainstream option for supervised neural network training; its adoption typically results in global changes to the network, usually in the form of modifications to artificial neuron weights and biases, when new data are processed or when the network is trained on a new task. Catastrophic forgetting, i.e. degradation in classification accuracy at inference stage due to parameter overwriting during subsequent training phases, is a key challenge when engineering artificial systems capable of incremental learning [10].

Different methods of mitigating catastrophic forgetting have been proposed including regularisation, rehearsal and parameter allocation [11-15]. Such methods are normally used in conjunction with deep learning techniques that are typically associated with significant memory footprints. Sparse Binary Coincidence [5], a recent algorithm relying on counting binary feature coincidences in a sparse high-dimensional space, has also been reported to mitigate catastrophic forgetting. However, its memory footprint can increase significantly outside simpler image data sets.

This feasibility study on an image classification task is a first step towards the development of a new single-pass supervised learning method with a competitively low memory footprint, likely to scale well to more complex data sets, and designed to address catastrophic forgetting.

2. Methods

With a theoretical and methodological underpinning in variational approaches to image registration [16-20], the research builds on a combination of established and more recent ideas including K-Nearest Neighbors (k-NN), memory-augmented neural networks [21], prototype-based models [22], and morphological techniques developed for medical image registration [23]. As opposed to traditional approaches to neural network training in which artificial neuron weights and biases are iteratively modified following exposure of the system to training instances, the training phase of the proposed method consists of recording low-dimensional data patterns in the form of class-level image prototypes, referred to as engrams in the following. In the computer science literature, the term ‘engram’ has often been used to describe a sparse structural alteration within an artificial system allowing storage and retrieval of information. With reference to artificial neural networks, the term has been used in relation to autoencoders as well as specialised architectures [24-26]. In this discussion, engrams are data structures encoding class-level low-dimensional data patterns, used for classifying images at inference stage -- e.g., envisaged as stored within a memory unit distinct from the neural network.

In the context of this study, engrams were obtained by averaging pixel intensities across training images following a signal integration step described below. During inference, images were warped to match the engrams individually and then classified by maximising a similarity score across classes. The process can be formulated as the minimisation of an energy functional consisting of a dissimilarity and a regularisation term, with a tuneable rigidity parameter to accommodate image transformations along a spectrum from elastic to rigid.

2.1. Preprocessing

The steps listed below were executed on all images in the Modified National Institute of Standards and Technology (MNIST) training set.

1. Artefact removal. The largest connected component was extracted from each image. Connected components were defined based on 8-connectivity, i.e. two pixels were considered belonging to the same component if and only if they shared an edge or a diagonal corner.
2. Signal integration. Pixel intensities were summed using 2x2 filters with stride 2, resulting in 14x14 images (formula 1). This step contributed to reducing the memory footprint by reducing image size.
3. Normalisation. Pixel intensities were normalised to [0, 1].

Denoting the α-th training image of class c by X_α^c, pixel intensity following signal integration is given by:

$$S_\alpha^c(x,y) = \sum_{u=-1}^{1} \sum_{v=-1}^{1} X_\alpha^c(2x+u, 2y+v) \quad (1)$$

The quantities x and y are integers labelling a N_x x N_y 2D grid with $N_x = N_y = 28$, $0 \leq x \leq N_x - 1$ and $0 \leq y \leq N_y - 1$.

The 2x2 filter size was selected to reflect the spatial characteristic scale of MNIST digits along the horizontal and vertical axis ($\sigma_x = \sigma_y \simeq 2$ pixels). When applying this method to higher-resolution images, the filter size will be selected based on the relevant scale as opposed to a predetermined number of pixels.

2.2. Training

For each class c, engram $\mathcal{E}^c$ was calculated as:

$$\mathcal{E}^c(x,y) = \langle S_\alpha^c(x,y) \rangle_{\alpha \in \mathcal{T}} \quad (2)$$

where $\mathcal{T}$ denotes the training set.

2.3. Inference

Let Ω denote the domain over which the images (X) are defined, which in this case is a N_x x N_y grid, and let $\varphi: \Omega \rightarrow \Omega$ be a continuous function. An energy functional can be defined as follows:

$$E(\varphi) = D(X \circ \varphi, \mathcal{E}) + \lambda \cdot R(\varphi) \quad (3)$$

where X and $\mathcal{E}$ denote a generic image and a generic engram, respectively, D is a dissimilarity function and R is a regularisation term. The rigidity parameter $\lambda > 0$ controls a trade-off between image alignment to achieve a smaller dissimilarity score and the smoothness of the transformation φ. Values of $\lambda \approx 0$ correspond to elastic deformations that can more easily warp images to match any template, whereas higher λ results in transformations closer to rigid.

Image warping, i.e. the derivation of geometric mappings between distinct image coordinate spaces with a view to maximising a similarity score, has been a subject of intense research well documented in the computer vision and medical image registration literature [23;27-28]. Transformations include rigid, affine and elastic mappings. With reference to medical image registration, local variations in morphology are often detected using elastic models, diffeomorphic frameworks relying on invertible, topology-preserving vector fields, machine learning-augmented methods and others.

Each image X in the MNIST test set was warped to match the engrams individually following preprocessing. For each $(X, \mathcal{E})$ pair and for each value of λ, a transformation $\hat{\varphi}$ was obtained by minimising the above energy functional:

$$\hat{\varphi} = argmin_\varphi E(\varphi) \quad (4)$$

In this study, D was defined as the pixel-by-pixel absolute intensity difference between the warped image and the engram, integrated over the image domain:

$$D(X \circ \hat{\varphi}, \mathcal{E}) = \int_\Omega |(X \circ \hat{\varphi})(\mathbf{x}) - \mathcal{E}(\mathbf{x})| d\mathbf{x} \quad (5)$$

Given all engrams $\{\mathcal{E}^c\}_c$, the inferred class label is given by:

$$\hat{c} = argmin_c\, D(X \circ \hat{\varphi}, \mathcal{E}^c) \tag{6}$$

It is worth noting that the algorithm is stable with respect to small image deformations thanks to the signal integration and image warping steps.

For each $(X, \mathcal{E}^c)$ pair, $\hat{\varphi}$ was obtained using the Stationary Velocity Field (SVF) method [29] minimising the following energy functional:

$$E_\lambda^{SVF} = \frac{1}{2} \cdot \int_\Omega \left[X\big(\mathbf{x} + \exp\big(\mathbf{v}(\mathbf{x})\big)\big) - \mathcal{E}^c\,(\mathbf{x})\right]^2 d\mathbf{x} + \frac{\lambda}{2} \cdot \int_\Omega \|\nabla^2 \mathbf{v}\|_2\, d\mathbf{x} \tag{7}$$

The regularisation term in equation (7) is proportional to the integral of the L2 norm of the Laplacian of a 2D velocity field defined on the domain of the image. A gradient descent optimisation loop was implemented warping X to match $\mathcal{E}^c$ while minimising $\mathcal{E}_\lambda^{SVF}$. Values of $\lambda \lesssim 0.2$ were selected with a view to avoiding the functional being dominated by the penalty term and the gradient descent step size becoming too small. The pseudocode is reported in figure 1.

Algorithm 1 Stationary Velocity Field (SVF) Alignment via Scaling and Squaring

function INTEGRATEVELOCITY(v_x, v_y, steps $= 2$)
  $H, W \leftarrow$ dimensions of v_x
  $X, Y \leftarrow$ identity coordinate grids of size $H \times W$
  $u_x \leftarrow v_x / 2^{\text{steps}}$
  $u_y \leftarrow v_y / 2^{\text{steps}}$
  for $s = 1$ **to** steps **do**
    $T_x \leftarrow X + u_x$
    $T_y \leftarrow Y + u_y$
    $u_x \leftarrow u_x + \text{Resample}(u_x, T_x, T_y)$ ▷ Bilinear interpolation with border replication
    $u_y \leftarrow u_y + \text{Resample}(u_y, T_x, T_y)$
  end for
  return u_x, u_y
end function

function SVFALIGNMENT($I_{\text{source}}, I_{\text{target}}, \lambda = 0.1$, iterations $= 100$)
  $H, W \leftarrow$ dimensions of I_{target}
  $v_x, v_y \leftarrow \mathbf{0}_{H \times W}$ ▷ Initialize stationary velocity fields
  $X, Y \leftarrow$ identity coordinate grids of size $H \times W$
  $\gamma \leftarrow 0.3/(1.0 + 10.0 \cdot \lambda)$ ▷ Compute adaptive step size
  $\epsilon \leftarrow 0.1$
  for $i = 1$ **to** iterations **do**

    ▷ **Phase 1: Exponentiation**
    $u_x, u_y \leftarrow$ INTEGRATEVELOCITY(v_x, v_y, steps $= 2$)

    ▷ **Phase 2: Compute Data Force**
    $I_{\text{warped}} \leftarrow \text{Resample}(I_{\text{source}}, X + u_x, Y + u_y)$
    $D \leftarrow I_{\text{warped}} - I_{\text{target}}$
    $\nabla_x I_{\text{warped}}, \nabla_y I_{\text{warped}} \leftarrow \text{Gradient}(I_{\text{warped}})$
    $\text{Denom} \leftarrow (\nabla_x I_{\text{warped}})^2 + (\nabla_y I_{\text{warped}})^2 + \epsilon$
    $du_{\text{data}} \leftarrow (D \cdot \nabla_x I_{\text{warped}})/\text{Denom}$
    $dv_{\text{data}} \leftarrow (D \cdot \nabla_y I_{\text{warped}})/\text{Denom}$

    ▷ **Phase 3: Laplacian Regularisation & Update**
    $v_{x,\text{smooth}} \leftarrow \text{GaussianBlur}(v_x, \text{kernel\_size} = 5)$
    $v_{y,\text{smooth}} \leftarrow \text{GaussianBlur}(v_y, \text{kernel\_size} = 5)$
    ▷ Compute negative Laplacian approximation via diffusion difference
    $\Delta v_x \leftarrow v_x - v_{x,\text{smooth}}$
    $\Delta v_y \leftarrow v_y - v_{y,\text{smooth}}$
    ▷ Update velocity fields with data forces and regulariser penalty
    $v_x \leftarrow v_x - \gamma \cdot (du_{\text{data}} + \lambda \cdot \Delta v_x)$
    $v_y \leftarrow v_y - \gamma \cdot (dv_{\text{data}} + \lambda \cdot \Delta v_y)$
  end for
  ▷ Final exponentiation to get the final displacement fields
  $u_x, u_y \leftarrow$ INTEGRATEVELOCITY(v_x, v_y, steps $= 2$)
  return $u_x, u_y, I_{\text{warped}}$
end function

Figure 1. Pseudocode for SVF image warping.

3. Results and discussion

The MNIST engrams are displayed in figure 2. Defining the engrams as averages over the training set reduces training time. However, it can also affect classification performance considering the limited information about intra-class variability encoded in such engrams.

The algorithm was calibrated by maximising test accuracy over λ on the MNIST training set using 10-fold cross-validation. The dependence of cross-validation test accuracy on λ is shown in figure 3. Classification accuracy degrades rapidly as λ approaches 0, as expected considering that the algorithm can more easily warp digits to match any template. A value of $\lambda = 0.1$ was selected for the rest of the analysis. Slowly varying test accuracy was observed for values of λ between 0.05 and 0.15, suggesting stability with respect to small changes to λ around the optimal value.

Figure 4 shows MNIST digits warped to match the engram of a wrong class for different values of λ: a '5' digit was warped to match the '1' engram corresponding to $\lambda = 0$, 0.05, 0.1, 0.15, 0.2 (left to right). As expected, the warped image becomes morphologically more similar to the target as λ approaches 0 and tends to retain its initial morphology at higher λ.

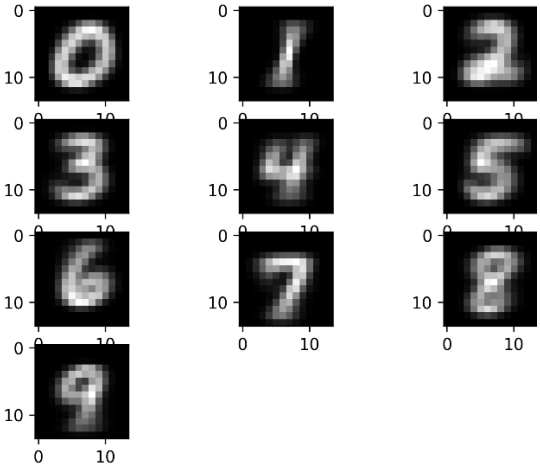

Figure 2. MNIST engrams.

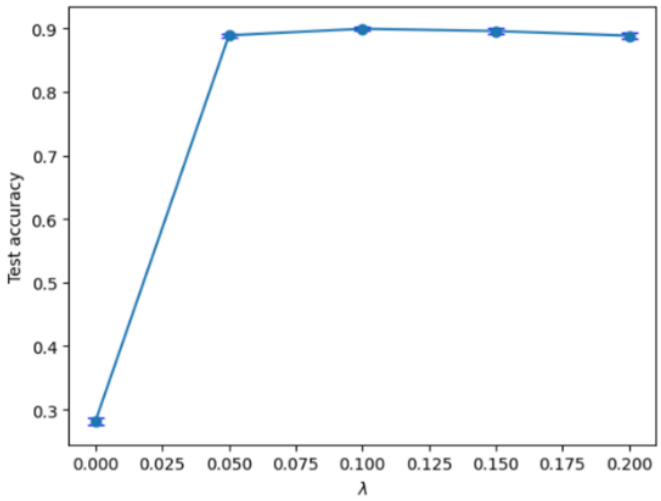

Figure 3. Calibration plot: test accuracy as a function of λ using 10-fold cross-validation on the MNIST training set.

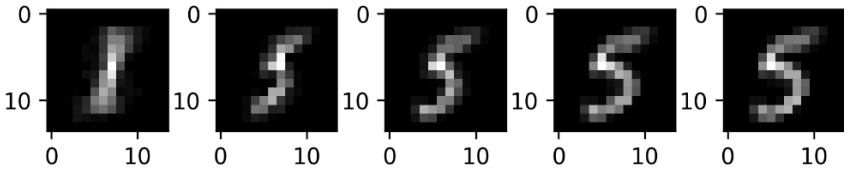

Figure 4. Results from warping a '5' digit to match a '1' engram at $\lambda = 0, 0.05, 0.1, 0.15, 0.2$ (left to right).

Figure 5 shows the confusion matrix on the MNIST test set ($\lambda = 0.1$), corresponding to an observed total accuracy of 90.4%. An ablation experiment was performed by replacing SVF with the identity operator: figure 6 shows the resulting confusion matrix, corresponding to a total accuracy of 75%. This confirms the importance of the role played by image warping in the context of this method.

The results are encouraging although below state-of-the-art performance. The algorithm was benchmarked against LeNet-5 and alternative single-pass methods (LOL, i-LOL, BSGD, FOGD and

SBC). The results are shown in table 1 in terms of classification accuracy on the MNIST test set. It should be noted that the test accuracy reported for some of the algorithms is lower than what documented in the original publication; this is likely due to the parameter configurations used in this study and will be investigated in more detail as part of future work.

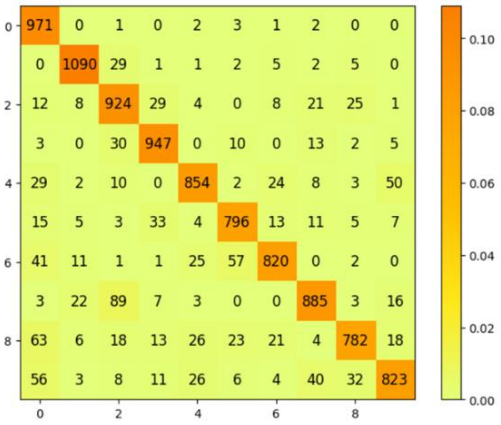

Figure 5. Confusion matrix on the MNIST test set ($\lambda = 0.1$).

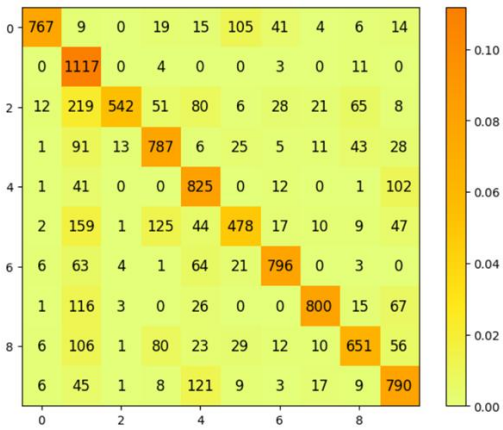

Figure 6. Confusion matrix on the MNIST test set following SVF ablation, showing degraded classification performance.

Table 1. MNIST test accuracy obtained using this method ($\lambda = 0.1$) compared to alternative algorithms.

Model/Algorithm	Test Accuracy (Mean $\pm$ Standard Deviation)
LeNet-5	(98.1 $\pm$ 0.2) %
i-LOL	(97.2 $\pm$ 0.2) %
LOL	(93.3 $\pm$ 0.2) %
BSGD	(93.1 $\pm$ 0.3) %
FOGD	(92.6 $\pm$ 0.0)%
SBC	(90.9 $\pm$ 0.3) %
This study	(90.4 $\pm$ 0.0) %

The estimated memory footprint of this method is of approximately 4.3 kB per MNIST image[1], notably smaller than several MB or higher as reported for SBC and other algorithms. This is ultimately due to the use of a simple pooling-like signal integration step taking advantage of the continuity of the input signal as opposed to relying on thousands of Address Decoder Elements [5]. Moreover, since there is no need for high-dimensional binary feature coincidences to be recorded, this method is anticipated to scale more favourably than SBC outside simpler image data sets. The execution time of the algorithm at inference stage was (19 $\pm$ 4) ms per MNIST image (CPU only; AMD Ryzen 7 @ 2 GHz; 16 GB RAM), the primary bottleneck being image warping. If this method is to be considered for real-time applications, ways of speeding up inference will therefore have to be identified.

[1] Total memory footprint (bytes) = 28·28 (input image) + 14·14·10 (engrams) + 14·14·4·2 (image warping, assuming 32-bit floats are used for the 2D velocity field).

4. Conclusions

This study has shown the feasibility of a new single-pass supervised learning method with reference to an image classification task, with potential to help address catastrophic forgetting in continual learning systems. At inference stage, images were warped to match class-level prototypes, referred to as engrams, encoding low-dimensional data patterns. For each test image, class membership was inferred by maximising a similarity score across classes following warping. The potential of this method to address catastrophic forgetting is highlighted by the possibility of using separate memory locations for storing different engrams. The scope of this study has been restricted to showing the feasibility of the core technique with reference to an image classification task. Further research will rely on sequential training experiments to validate this approach in a continual learning setting. The estimated memory footprint of 4.3 kB per MNIST image is remarkably smaller than the footprint of alternative algorithms. Memory requirements are expected to scale favourably with increasing image complexity, considering that recording of high-dimensional binary feature coincidences is not required. A test accuracy over 90% was observed on MNIST. This is a promising result, considering that a simple version of the algorithm was used with engrams exclusively encoding average data patterns. The replacement of the current implementation of the engrams with more advanced data structures such as wider sets of categorical metadata features will be a subject of follow-on studies. Whereas target variance encoding is an interesting option, the selection of an encoding not relying on deep learning techniques and capable of preserving morphological and structural image properties – and therefore of capturing the intraclass variability patterns observed in the data – requires additional research. Future research will focus on reducing inference time with a view to targeting real-time use cases, on increasing classification accuracy using more advanced implementations, and on estimating the performance of the method under noise. The scaling properties of this method in terms of memory footprint across data sets will also be a subject of further investigation, as will the potential of the engrams to enhance interpretability. Finally, it should be noted that the selection of SVF in this study was driven by computational convenience, and alternative methods will be investigated.

5. References

[1] Verdecchia R, Sallou J and Cruz L 2023 A systematic review of Green AI *WIREs Data Min. Knowl. Discov.* **13** e1507 (https://doi.org/10.1002/widm.1507)

[2] Read J and Zliobaite I 2025 Supervised Learning from Data Streams: An Overview and Update *ACM Comput. Surv.* **57** Article 307 (https://doi.org/10.1145/3737279)

[3] Raman V and Tewari A 2024 Online classification with predictions *Proc. 38th Int. Conf. on Neural Information Processing Systems (NeurIPS 2024)* (Article 1777) (https://doi.org/10.52202/079017-1777)

[4] Hu L, Hu C, Huo Z, Jiang X and Wang S 2022 Online Support Vector Machine with a Single Pass for Streaming Data *Mathematics* **10** 3113 (https://doi.org/10.3390/math10173113)

[5] Hopkins M, Fil J, Jones E G and Furber S 2023 BitBrain and Sparse Binary Coincidence (SBC) memories: Fast, robust learning and inference for neuromorphic architectures *Front. Neuroinform.* **17** 1125844 (https://doi.org/10.3389/fninf.2023.1125844)

[6] Zhou H, Yin H, Wang B and Liao C 2025 One-pass online learning under evolving feature data streams: A non-parametric model *Pattern Recognition* **168** 111719 (https://doi.org/10.1016/j.patcog.2025.111719)

[7] Wang Z, Crammer K and Vucetic S 2012 Breaking the curse of kernelization: Budgeted stochastic gradient descent for large-scale SVM training *J. Mach. Learn. Res.* **13** 3103 (https://dl.acm.org/doi/10.5555/2503308.2503341)

[8] Ruiz C, Alaíz C M and Dorronsoro J R 2024 A survey on kernel-based multi-task learning *Neurocomputing* **577** 127255 (https://doi.org/10.1016/j.neucom.2024.127255)

[9] Chauvin Y and Rumelhart D E (eds) 2013 *Backpropagation: Theory, Architectures, and Applications* (New York: Taylor & Francis) (https://doi.org/10.4324/9780203763247)

[10] van de Ven G M, Soures N and Kudithipudi D 2025 Continual Learning and Catastrophic Forgetting *Learning and Memory: A Comprehensive Reference* vol 1 3rd edn ed J Wixted (London: Academic Press) pp 153–68 (https://doi.org/10.1016/B978-0-443-15754-7.00073-0)
[11] Zhou S and Li Q 2026 Mitigating catastrophic forgetting in lifelong learning: a hybrid architecture integrating neural ordinary differential equations with memory-augmented transformers *Sci. Rep.* **16** 2012 (https://doi.org/10.1038/s41598-025-31685-9)
[12] Bonnet D, Cottart K, Hirtzlin T, Januel T, Dalgaty T, Vianello E and Querlioz D 2025 Bayesian continual learning and forgetting in neural networks *Nat. Commun.* **16** 9614 (https://doi.org/10.1038/s41467-025-64601-w)
[13] Yang M, Zhan Z, Chen Y, Li H, Wang K, Zheng K, Wang Y, Wang Q, Gao J and Wu J 2026 *Collaborative Parameter Learning: Mitigating Forgetting via Parameter-Level Gradient Analysis* (arXiv:2601.21577) (https://doi.org/10.48550/arXiv.2601.21577)
[14] Lai S, Ma C, Zhu F, Zhao Z, Lin X, Meng G and Zhang Q 2025 Gradient-Guided Epsilon Constraint Method for Online Continual Learning *Adv. Neural Inf. Process. Syst.* **38** 26972–84 (https://proceedings.neurips.cc/paper_files/paper/2025/hash/b3c2854d9e94282a373d8fa58b567b27-Abstract-Conference.html)
[15] Pietroń M, Żurek D, Faber K and Corizzo R 2023 Ada-QPacknet – Multi-Task Forget-Free Continual Learning with Quantization Driven Adaptive Pruning *Proc. 26th Eur. Conf. Artificial Intelligence (ECAI 2023) (Frontiers in Artificial Intelligence and Applications* vol **372**) ed K Inokuchi *et al* (Amsterdam: IOS Press) pp 1795–802 (https://doi.org/10.3233/FAIA230477)
[16] Li Y, Chen K, Chen C and Zhang J 2024 A bi-variant variational model for diffeomorphic image registration with relaxed Jacobian determinant constraints *Appl. Math. Model.* **130** 66–93 (https://doi.org/10.1016/j.apm.2024.02.033)
[17] Tu Z, Xie W, Zhang D, Poppe R, Veltkamp R C, Li B and Yuan J 2019 A survey of variational and CNN-based optical flow techniques *Signal Process. Image Commun.* **72** 9–24 (https://doi.org/10.1016/j.image.2018.12.002)
[18] Chen K and Han H 2024 Three-stage approach for 2D/3D diffeomorphic multimodality image registration with textural control *SIAM J. Imaging Sci.* **17** 1690–728 (https://doi.org/10.1137/23M1583971)
[19] Maleelai K, Watchararuangwit C and Chumchob N 2022 An improved numerical method for variational image registration model with intensity correction *J. Interdiscip. Math.* **25** 1005–22 (https://doi.org/10.1080/09720502.2021.1889781)
[20] Zhang J and Li Y 2021 Diffeomorphic Image Registration with an Optimal Control Relaxation and Its Implementation *SIAM J. Imaging Sci.* **14** 1890–931 (https://doi.org/10.1137/21M1391274)
[21] Ataeiasad F, Elizondo D, Ramírez S C, Greenfield S and Deka L 2024 Out-of-Distribution Detection with Memory-Augmented Variational Autoencoder *Mathematics* **12** 3153 (https://doi.org/10.3390/math12193153)
[22] Ma C, Donnelly J, Liu W, Vosoughi S, Rudin C and Chen C 2024 Interpretable Image Classification with Adaptive Prototype-based Vision Transformers *Adv. Neural Inf. Process. Syst.* **37** 1312 (https://doi.org/10.52202/079017-1312)
[23] Nenoff L, Amstutz F, Murr M, Archibald-Heeren B, Fusella M, Hussein M, Lechner W, Zhang Y, Sharp G and Vasquez Osorio E 2023 Review and recommendations on deformable image registration uncertainties for radiotherapy applications *Phys. Med. Biol.* **68** 24TR01 (https://doi.org/10.1088/1361-6560/ad0d8a)
[24] Szelogowski D 2025 *Engram Memory Encoding and Retrieval: A Neurocomputational Perspective* (arXiv:2506.01659) (https://doi.org/10.48550/arXiv.2506.01659)
[25] de Lucas J M 2023 *Implementing engrams from a machine learning perspective: matching for prediction* (arXiv:2303.01253) (https://doi.org/10.48550/arXiv.2303.01253)

[26] Aguilar I, Herbozo Contreras L F and Kavehei O 2025 *Stochastic Engrams for Efficient Continual Learning with Binarized Neural Networks* (arXiv:2503.21436) (https://doi.org/10.48550/arXiv.2503.21436)
[27] Hernández M and Ramón Julvez U 2024 Insights into traditional Large Deformation Diffeomorphic Metric Mapping and unsupervised deep-learning for diffeomorphic registration and their evaluation *Comput. Biol. Med.* **178** 108761 (https://doi.org/10.1016/j.compbiomed.2024.108761)
[28] Zhu W and Myronenko A 2020 NeurReg: Neural Registration and Its Application to Image Segmentation *2020 IEEE Winter Conf. Applications of Computer Vision (WACV)* pp 110–118 (https://doi.org/10.1109/WACV45572.2020.9093506)
[29] Bostelmann J, Gildemeister O and Lellmann J 2026 Stationary Velocity Fields on Matrix Groups for Deformable Image Registration *J. Math. Imaging Vis.* **68** 9 (https://doi.org/10.1007/s10851-026-01284-y)

Acknowledgments
This study was conducted in the absence of financial support. Free generative AI tools were used to facilitate the implementation of some parts the algorithm, with emphasis on utility functions. The pseudocode rendering in figure 1 was generated using a free AI tool.