JMI Journal of Medical Imaging

RESEARCH PAPER

Attention-gated 3D CapsNet for robust hippocampal segmentation

Clement Poiret,[a,b] Antoine Bouyeure,[a,b] Sandesh Patil,[a,b] Cécile Boniteau,[a,b]
Edouard Duchesnay,[a] Antoine Grigis,[a] Frederic Lemaitre,[c,d]
and Marion Noulhiane[a,b,*]

[a]UNIACT, NeuroSpin, Institut Joliot, CEA Paris-Saclay, Gif-sur-Yvette, France
[b]Université Paris Cité, InDEV team, U1141 NeuroDiderot, Inserm, Paris, France
[c]Université de Rouen, CETAPS EA 3832, Rouen, France
[d]CRIOBE, UAR 3278, CNRS-EPHE-UPVD, Mooréa, Polynésie Française

ABSTRACT. **Purpose:** The hippocampus is organized in subfields (HSF) involved in learning and memory processes and widely implicated in pathologies at different ages of life, from neonatal hypoxia to temporal lobe epilepsy or Alzheimer's disease. Getting a highly accurate and robust delineation of sub-millimetric regions such as HSF to investigate anatomo-functional hypotheses is a challenge. One of the main difficulties encountered by those methodologies is related to the small size and anatomical variability of HSF, resulting in the scarcity of manual data labeling. Recently introduced, capsule networks solve analogous problems in medical imaging, providing deep learning architectures with rotational equivariance. Nonetheless, capsule networks are still two-dimensional and unassessed for the segmentation of HSF.

Approach: We released a public 3D Capsule Network (3D-AGSCaps, https://github.com/clementpoiret/3D-AGSCaps) and compared it to equivalent architectures using classical convolutions on the automatic segmentation of HSF on small and atypical datasets (incomplete hippocampal inversion, IHI). We tested 3D-AGSCaps on three datasets with manually labeled hippocampi.

Results: Our main results were: (1) 3D-AGSCaps produced segmentations with a better Dice Coefficient compared to CNNs on rotated hippocampi ($p = 0.004$, cohen's $d = 0.179$); (2) on typical subjects, 3D-AGSCaps produced segmentations with a Dice coefficient similar to CNNs while having 15 times fewer parameters (2.285M versus 35.069M). This may greatly facilitate the study of atypical subjects, including healthy and pathological cases like those presenting an IHI.

Conclusion: We expect our newly introduced 3D-AGSCaps to allow a more accurate and fully automated segmentation on atypical populations, small datasets, as well as on and large cohorts where manual segmentations are nearly intractable.

Keywords: hippocampal subfields; convolutional neural networks; deep learning; equivariance; MRI

Paper 22308GRR received Nov. 14, 2022; revised Nov. 18, 2023; accepted Dec. 4, 2023; published Jan. 2, 2024.

1 Introduction

The hippocampus, located in the medial temporal lobe, plays a crucial role in learning and memory processes.[1] The hippocampus is also implicated in diverse neuropathologies with high

*Address all correspondence to Marion Noulhiane, marion.noulhiane@cea.fr

2329-4302/2024/$28.00 © 2024 SPIE

prevalence across the lifespan, from neonatal hypoxia to Alzheimer's disease or medial-temporal lobe epilepsy.[2] Recent research has focused on the distinct roles of hippocampal subfields (HSF) in memory and disease progression: the dentate gyrus (DG), four parts of the cornu ammonis (CA4 to CA1), and the subiculum (Sub).

1.1 Segmenting the Hippocampus, Stakes, and Methods

Such research involves an accurate delineation (or segmentation) of the HSF, which consists of assigning a class to every voxel of a given image. In the context of MRI segmentation, bilateral regions of interest (ROI) like the HSF are assigned the same labels, to divide an image into a set of semantically meaningful, homogeneous, and non-overlapping regions of similar attributes, such as intensity, depth, color, or texture.[3] This delineation process enables the study of structural patterns which may in fine lead to a better comprehension, diagnosis and prognosis of such diseases, as segmentation allows one to easily derive the geometry, shape, and size of a given ROI. To date, the segmentation of the hippocampus can capture anatomical variability, such as the incomplete hippocampal inversion (IHI),[4] a developmental abnormality occurring in consequent subsets of the healthy or pathological population, where the hippocampal body and the collateral sulcus can be rotated up to 90 deg.[5] However, this methodology remains so time-consuming that it cannot be considered as routine clinical practice. While distinct techniques of various complexities have been developed to segment HSF on MRIs,[6–8] the field suffers from labeled data scarcity as manual segmentation is a time-consuming and error-prone process partially caused by inconsistent guidelines. Because manual segmentation is the only way to gather labeled datasets to train neural networks, the aforementioned difficulties greatly limit the size of available datasets, thus reducing the probability of learning in specific cases where IHI is found, such as in temporal lobe epilepsy.

Nowadays, segmentation tasks are now almost exclusively handled through specific and supervised convolutional neural networks (CNN), an architecture called UNet,[9] leveraging the properties of an auto-encoder architecture to quickly achieve a segmentation with an expert-level accuracy and small sample size. Nevertheless, those models suffer from several pitfalls. It has been shown that the performances of such computer-vision models are prone to image corruptions, such as noise or rotations.[10–12] Albeit recent works validated deep learning as a great candidate for automated segmentation of HSF,[13,14] IHI, which are unassessed in recent automated segmentation methods, may cause troubles to most conventional CNNs. With a transformation g, an image x, and a model f, equivariance is defined as $g(f(x)) = f(g(x))$. Similarly, invariance is achieved if and only if $f(g(x)) = f(x)$. CNNs are efficient in modeling structural patterns in a given image, especially due to their built-in translation equivariance: a translation of a specific pattern in the input image, shifts the output of the convolutional layer. As previously found, transformations such as rotations can impair CNNs performances.[15] On the one hand, a standard approach to approximate rotational equivariance would be to use data augmentation by providing multiple rotated versions of the training set. However, it involves learning redundant parameters corresponding to similar patterns at varying angles.[16] In addition, data augmentation may increase overfitting risks,[17,18] meaning the improvement on standard CNNs would only be marginal on small training sets and may sometimes lead to a drop in accuracy on unperturbed images.[10] Partly to solve this issue, recent works introduced Capsule Networks (CapsNets), a novel kind of neural network designed to benefit from natural or augmented variability more efficiently.[19,20]

1.2 Capsule Networks

CapsNets are replacing standard convolutions with capsules (Fig. 1). A capsule aims to replace scalar activation values with vectors (of which the number of dimensions constituting its space is sometimes referred to as the number of atoms). The L2-Norm of a given vector is equivalent to the activation of a standard convolution, but now a network can encode information into the orientation of the vector. This intriguing characteristic promotes the emergence of a key property: the theoretical ability to learn equivariances to features (sometimes called "instantiation parameters"), such as rotations, local deformations,[20] or even to more subtle features such as sphericity, lobulation, or textures.[21] Intuitively, whereas convolutions learn multiple kernels to detect different versions of the same object (e.g., a rotated hippocampus), capsules embed those different

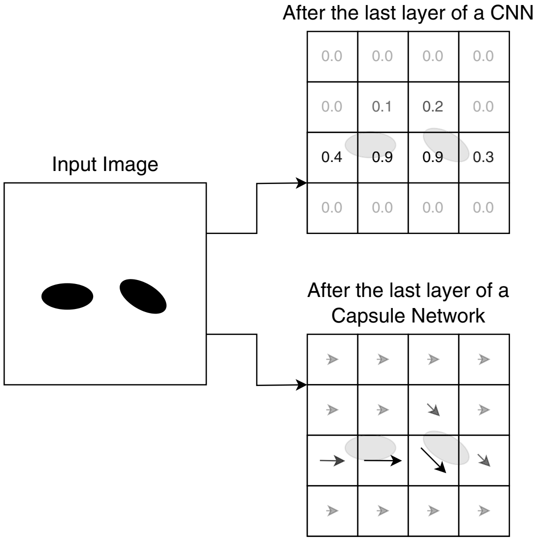

Fig. 1 Capsule Networks versus CNNs. The schematic represents the output of both networks on a toy example. A CNN will output, for each voxel, the probability of presence of the object, while a Capsule Network will output a vector, where its norm is the probability of presence, and its orientation encodes additional properties (here, the rotation of the object).

versions under the same weights through vectors' orientations, leading to less redundancy in the network, and greater expressive power for the same number of parameters. Moreover, the usual feedforward pass is altered by using a routing-by-agreement between capsule layers.[20] This routing-by-agreement recurrently weights the feedforward pass by selectively passing information from capsules in the layer l to the layer $l + 1$. Each capsule in l will vote for the potential output of capsules in $l + 1$. Then, activations between all capsules in l and a specific capsule in $l + 1$ are weighted by their L2-Norm to the centroid of the predictions: similar predictions are likely to be sent to a single parent capsule. In two-dimensional (2D), CapsNets have shown improved robustness against physical alterations, such as rotations,[22] placing themselves as possible candidates to explore for MRI processing. To date, CapsNets are yet to be publicly implemented and benchmarked on a 3D segmentation task. For example, if a capsule represents a high-level object, such as a hippocampus, for every patch of a given image the vector's norm represents the probability of presence of the object. Then, its direction encodes relevant instantiation parameters. On the other hand, their activation (L2-norm) stays invariant. This behavior leads to an increased expression power and a higher sample efficiency.[23]

CapsNets have already been implemented in the biomedical field with promising results, where the authors were able to overcome the shortcomings of CNNs on a brain tumor and lung nodule type classification tasks.[24,25] By processing MRIs in a slice-by-slice manner, they achieved a classification accuracy of 78% against 61.97% for a CNNs of comparable architecture. To date, CapsNets for image segmentation are poorly investigated. The authors of the SegCaps model were the first ones to successfully perform segmentation with capsule layers.[21] They made this possible by building a deeper model than the original implementation using locally constrained routing and transformation matrix sharing to reduce the number of parameters and memory consumption. To build a UNet-shaped model, they also introduced transposed capsules. Following this segmentation paradigm, recent works handled coronary artery segmentation from intravascular optical coherence tomography in a slice-by-slice manner.[26] While they did not reach the accuracy of the best model in the state-of-the-art (SotA accuracy) on their dataset, they managed to get honorable segmentations with a model of nearly 5M parameters, while SotA models were between 30M and 40M parameters. This may suggest that capsules can effectively benefit from learned equivariances in segmentation tasks. Nevertheless, both implementations act in a 2D space, on binary segmentation tasks. CapsNets able to perform 3D segmentation tasks are yet to be implemented. If 3D segmentation is not yet handled by CapsNets, several works experimented with 3D capsules in other tasks. Newer developments used 3D CapsNet to perform object recognition with a shallow architecture,[27] where they found that 3D CapsNets were more data efficient than analogous CNN architectures. Additional works found a consistent improvement of 3D capsules over SotA models for 3D point set classification, especially for noisy observations.[28,29] Finally, 3D CapsNets were applied with success in the biomedical field for lung nodule malignancy prediction with a highly competitive accuracy.[24]

Thus, 3D capsules come out as relevant candidates for tasks where the number of observations is limited, variable or noisy, and where models are operating in resource-constrained environments. Notwithstanding the fact that 3D CapsNets are an active research area, no implementation is currently publicly accessible.

1.3 Attention-Gated Networks

The idea behind attention gates (AG) is to allow a CNN to implicitly learn how to suppress or highlight specific regions in an input image, with minimal computational overhead.[30] Initially developed as an extension to the standard U-Net model, it generates through additive attention a soft-attention grid, composed of gating coefficients $\alpha_i \in [0,1]$. Finally, those gating coefficients are multiplied by the input feature map. The original study reported significant improvements related to their additive AG on their segmentation task.[30] Later, similar improvements were obtained for 3D coronary computed tomography (CT) angiography segmentation[31] or liver CT image segmentation.[32] While the attention mechanism has been implemented in the routing-by-agreement algorithm,[33–36] attention-gated CapsNets are yet to be assessed. The original routing-by-agreement algorithm aims at weighting information sent from a layer l to a layer $l + 1$. We think that it could potentially work synergistically with AG by modulating on-the-fly the activation of a capsule layer.

The aim was to extend CapsNets for segmentation tasks in three-dimensional spaces applied to MRI segmentation of the hippocampus to investigate the robustness of our new model, namely 3D attention-gated SegCaps (3D-AGCaps) on developmental particularities, such as the IHI. In this aim, our approach was as follows: (1) we validated 3D-AGSCaps on hippocampal segmentation against the equivalent architectures using classical convolutions in-place of capsule layers; (2) we investigated the robustness of 3D-AGSCaps to various random rotational perturbations of the MRI acquisitions, simulating IHI.

Considering three metrics, the Dice coefficient, the Hausdorff distance, and the volumetric similarity, we hypothesized that

1. On typical MRIs, 3D-AGSCaps will not be statistically different from analogous CNN architectures, both architectures producing near-optimal segmentations;
2. On atypical MRIs with rotational perturbations replicating atypical conditions, such as the IHI, 3D-AGSCaps will produce segmentations closer to manual segmentations due to their implicit ability to learn equivariances over various instantiation parameters.

2 Methods

2.1 Datasets Description

We used two public and one in-house datasets with manually labeled hippocampi by expert raters (Table 1). The three datasets were manually labeled by experts, from which the first one is an anteroposterior hippocampal segmentation of 263 hippocampi, while the other two are 50 hippocampi segmented in subfields. The Kulaga–Yoskovitz dataset has been segmented from head to tail according to an in-house segmentation protocol. MemoDev has hippocampal bodies manually segmented (AB, SP, MN, CP) following.[40] Examples of both types of hippocampal segmentation are shown in Fig. 2.

Data acquisition for our in-house dataset, MemoDev, was performed under the regulations of an appropriate Ethical Committee board (CPP 2011-A00058-33).

2.2 From 2D to 3D Capsule Networks in MRI Segmentation of the Hippocampus

CapsNets are compute-intensives, both in terms of computational complexity and memory requirements.[21] If they are solving issues inherent to CNNs, this is the major drawback for the adoption of capsules. Therefore, we used the public 2D SegCaps implementation of Ref. 21 to

Table 1 Description of the datasets used. DG, dentate gyrus; CA, cornu ammoni; Sub: subiculum.

Name	N	Acquisition parameters	Segmentation	Ref.

Simpson	263	3T;
		3D T1-weighted MPRAGE sequence;
		TI/TR/TE, 860/8.0/3.7 ms;
		170 sagittal slices;
		voxel size, 1.0 mm^3;
			–Anterior
			–Posterior
				37

Kulaga-Yoskovitz	50	3T;
		3D T1-weighted MPRAGE sequence;
		TI/TR/TE, 1500/3000/4.32 ms;
		176 sagittal slices;
		Voxel size, 1.0 mm^3;
			–DG
			–CA
			–Sub
				38

MemoDev	50	3T;
		Coro-T2-weighted TSE sequence;
		TR/TE, 3970/89 ms;
		46 coronal slices;
		Voxel size, 0.4 $*$ 0.4 $*$ 1.2 mm;
			–DG
			–CA1
			–CA2/3
			–Sub
			—
				39

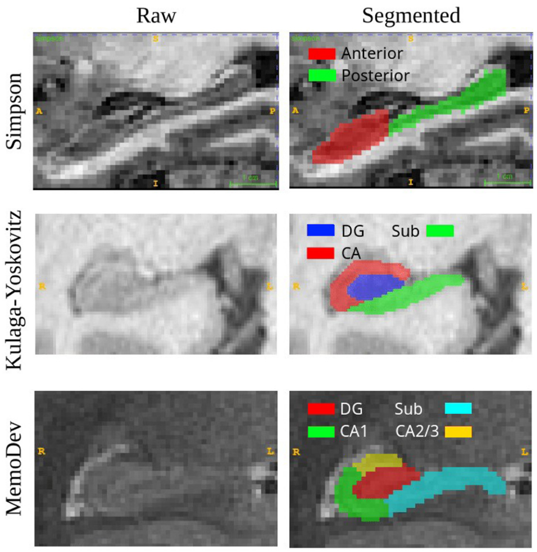

Fig. 2 Random segmentation examples of the three datasets: Simpson (axial slice), Kulaga–Yoskovitz (coronal slice), and MemoDev (coronal slice). Letters indicate spatial directions: left (L), right (R), anterior (A), posterior (P), superior (S), and inferior (I).

migrate the architecture from 2D to 3D in PyTorch. Their implementation offers important addons to reduce the number of parameters of CapsNets, such as (de)convolutional capsules, and a locally constrained routing.

In addition to the reimplementation by adding a spatial dimension, we introduced a novel activation function to handle multiclass classification tasks. The Squash function[20] has been originally introduced to rescale the L2-norm of the capsules to $[0;1]$ without changing their directions, with s_j the output vector of the capsule j such as

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}.$$

To segment multiple classes, we want the L2-norm of each capsule c to represent a probability distribution over the brain regions. Therefore, we introduced the Softmax-Squash (or SMSquash) function that we used for our last capsule layer, defined as

$$v_j = \frac{e^{\|s_j\|}}{\sum e^{\|s_j^c\|}} \frac{s_j}{\|s_j\|}.$$

However, migrating SegCaps from 2D to 3D worsened the computational burden of CapsNets. This led us to introduce attention-gated capsules to route the information with greater precision while reducing the number of routing-by-agreement iterations to improve the efficiency of our model.

2.3 3D Attention-Gated SegCaps (3D-AGSCaps)

To complement the routing-by-agreement algorithm, we introduced a variation of the AG (Fig. 3) coming from Ref. 30, which helps the network to focus on the target brain structures. Our AG, implemented at the concatenation of the volumes of the downsampling and the upsampling path, aims to modulate the L2-norm of the capsules.

The gating signal g from the layer $l - 1$ is upsampled using transposed capsules.[21] Then, g and x of the corresponding layer l of the downsampling path are combined to form an attention grid of which the size matches the number of atoms. Information coming from the downsampling path is then multiplied to the attention grid to modulate capsules' L2-norms. Convolutions are followed by SwitchNorm layers.[41]

Our final model, 3D-AGSCaps is shown in Fig. 4. It retakes a UNet-like architecture where we used our AG in-place of the concatenation of both downsampling and upsampling paths and assessed its efficacy through an ablation study. The resulting implementation in PyTorch is publicly available under an MIT license available in GitHub repository at: https://github.com/clementpoiret/3d-agscaps.

To perform ablation studies, we tested multiple variations of our proposed model to analyze the impact of our AG.

2.4 Implementation Details and Evaluation Metrics

The models have been implemented in Python, using PyTorch. The training is done with PyTorch Lightning. Data augmentation is handled with TorchIO. We trained our models with automatic mixed precision (16-bit) and validated them using a 10-fold cross-validation. We kept a hold-out

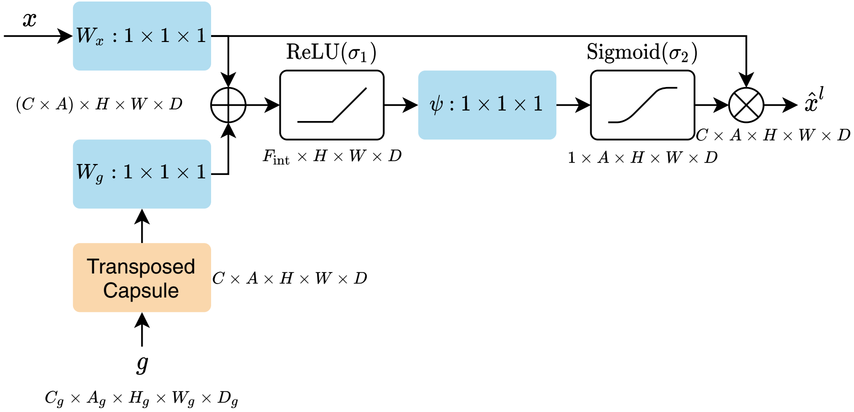
x
$W_x : 1 \times 1 \times 1$
ReLU(σ_1)
Sigmoid(σ_2)
$(C \times A) \times H \times W \times D$
$\psi : 1 \times 1 \times 1$
$\hat{x}^l$
$C \times A \times H \times W \times D$
$F_{\text{int}} \times H \times W \times D$
$1 \times A \times H \times W \times D$
$W_g : 1 \times 1 \times 1$
Transposed Capsule
$C \times A \times H \times W \times D$
g
$C_g \times A_g \times H_g \times W_g \times D_g$

Fig. 3 Our proposed AG. The input x comes from the downsampling path. The gating signal g comes from the layer $l - 1$ in the upsampling path. g passes through a transposed capsule to match x's size. Blue rectangles represent 3D convolutions and SwitchNorm3D. We indicated the shape of each object below each operation, with H, W, and D the height, width, and depth of the cube, C the number of capsules, and A the number of atoms, i.e., the number of elements in the capsules' vectors.

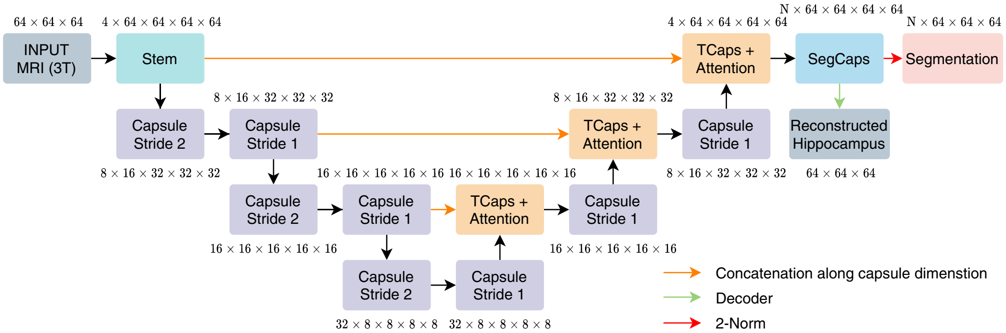

Fig. 4 Attention-gated SegCaps for volumetric segmentation (3D-AGSCaps). AG are implemented after our transposed capsules (orange rectangles) to ensure both inputs are of the same size. Our network takes an MRI as input (of size 64^3 in this example) and outputs a reconstruction of the original hippocampus (without the background class) alongside the segmentation.

test set for further analysis to keep samples free from potential unwanted tuning. Complete implementation for training and validation details is listed in Table 2.

Following our cross-validation protocol to validate our models, we assessed the effect of rotational data augmentation by training 10 times all the models with a different maximum amounts of random rotations of the training set. Finally, to assess the behavior of the models in the presence of atypical hippocampi, we randomly rotated our test sets (Table 2) with an increasing amount of maximum amplitude (from 0 deg to 180 deg). As this process involves random perturbations of the test sets, we repeated this process 10 times to better estimate the

Table 2 Implementation and validation details for each dataset. Given a fixed (hold-out) test set, training, and validation sets are defined using a standard 10-fold cross-validation on the remaining samples. (*) denote an "either/or" scheme, i.e., affine and elastic transformations cannot be applied at the same time, but one of them is always applied.

	Simpson ($N = 263$)	Kulaga–Yoskovitz ($N = 50$)	MemoDev ($N = 50$)
Training	$N = 225$	$N = 36$	$N = 36$
		64 epochs, batch size 8	
		AdamW, learning rate 1e-3	
		Cosine annealing scheduler (no restart, no warm-up)	
		Stochastic weight averaging	
Validation	$N = 25$	$N = 4$	$N = 4$
		10-fold cross-validation	
Test	$N = 13$	$N = 10$	$N = 10$
		Hold-out test set	
Preprocessing		1. Crop/pad around hippocampus	
		2. Z-normalization	
Augmentation		1. Left/right flips ($p = 0.5$)	
		2a. Affine transformations ($p = 0.8$)*	
		2b. Elastic deformations ($p = 0.2$)*	
		3. Gaussian noise ($p = 0.5$)	
		4. Random contrast ($p = 0.5$)	

variability induced by our artificial alterations of the MRIs. Segmentations are assessed with the Dice coefficient (DC), the volumetric similarity (VS), and the Hausdorff distance (HD) computed with PyMia. Given a manual segmentation y_m and a predicted segmentation y_p:

- the DC is an overlap metric ranging from 0 (no overlap), to 1 (full overlap) defined as $\mathrm{DC} = \frac{2|y_m \cap y_p|}{|y_m| + |y_p|}$,
- the HD is a metric of surfacic distance ranging from 0 to +inf. With the directed Hausdorff distance between two sets X and Y, such as $hd(X, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|_2$, the HD is defined as $\mathrm{HD}(y_m, y_p) = \max(hd(y_m, y_p), hd(y_p, y_m))$,
- the VS is a comparison between volumes of two segmentations ranging from 0 (complete dissimilarity between volumes) to 1 (exact match between volumes). With S_m and S_p the volumes of a region S given a manual and a predicted segmentation, respectively, it is defined as $VS = 2\frac{|S_m \cap S_p|}{|S_m + S_p|}.100\%$.

We computed each metric on a per-channel basis to assess the quality of each class, then averaged across classes to get a general score. In order to evaluate our hypotheses, we performed a two-way ANOVA with model types and rotation angles of the test images as independent variables. p-values are corrected using a Benjamini–Hochberg false discovery rate. While CNNs were trained using a focal Tversky loss L_s.[42] Given an input x, our segmentation loss L_s is defined with TP and TN the true positives and negatives, FN and FP the false positives and negatives, and $\alpha = 0.3$, $\beta = 0.7$, $\gamma = 3/4$ such as

$$L_s = \left(1 - \frac{\mathrm{TP}}{\mathrm{TP} + \beta\mathrm{FN} + \alpha\mathrm{FP}}\right)^{\gamma}.$$

Values of the hyper-parameters α and β are following the recommendations of their original paper: weighing more the FN enhanced convergence by shifting the focus on minimizing the FN. According to the authors, it helped balance precision–recall scores and gave better DC on a similar architecture of ours. 3D-AGSCaps uses a combination of L_s and the mean squared error of the reconstruction of the hippocampus $\hat{x}$. Thus, our loss L for our 3D-AGSCaps is defined as

$$L = L_s + \frac{1}{n}\sum (x_i - \hat{x}_i)^2.$$

2.5 Comparison with Analogous Convolutional Models

We benchmarked 3D-AGSCaps against the best-known models used in the hippocampal segmentation literature in a 3D approach:

- UNet (16.3M),[43] the baseline of most segmentation models, consisting of an auto-encoder architecture with skip connection between layers of the same depth,
- Residual UNet (35.0M),[44,45] grouping every couple of convolutions with the aim to stabilize the training of deeper networks,
- and their counterparts DUNet (16.7M) and residual DUNet (35.5M),[14] replacing the second to last skip connection with a dilated dense network of convolutions to improve the information flow between the encoder and the decoder.

3 Experimental Results

3.1 Ablation Studies

Results of ablation studies (Table 3) across the three datasets revealed that the best overlap (DC) between HSF was obtained by models with the AG (e.g., 0.872 ± 0.028 versus 0.834 ± 0.058 for Kulaga–Yoskovitz). The 10-fold cross-validation results are reported in Table 3. For the quality of the reconstruction (Table 4), we showed a significant impact of the reconstruction ($p = 0.034$, $T = 5.262$, $BF10 = 3.196$, Cohen's $d = 0.032$). An example of the reconstruction is shown in Fig. 4. However, we found no significant differences regarding the value of α ($p > 0.05$, Table 4).

Table 3 Ablation study of our AG. Baseline models are capsule networks without AG. DC, Dice coefficient; HD, Hausdorff distance; VS, volumetric similarity; and MSE, mean squared error assessing the reconstruction head. Results are presented in mean ± standard deviation.

Dataset	Model	DC	HD	VS	MSE
Kulaga–Yoskovitz	Baseline (2.3M)	0.834 ± 0.058	22.173 ± 17.975	0.921 ± 0.071	1.006 ± 0.004
	AG (2.4M)	0.872 ± 0.028	15.350 ± 15.516	0.962 ± 0.034	1.005 ± 0.003
MemoDev	Baseline (2.3M)	0.659 ± 0.163	17.182 ± 26.403	0.835 ± 0.165	0.896 ± 0.061
	AG (2.4M)	0.654 ± 0.175	10.502 ± 13.288	0.828 ± 0.181	0.897 ± 0.059
Simpson	Baseline (2.3M)	0.877 ± 0.037	3.160 ± 2.234	0.937 ± 0.037	0.238 ± 0.031
	AG (2.4M)	0.880 ± 0.037	2.875 ± 1.874	0.941 ± 0.036	0.236 ± 0.028

Table 4 Impact of the reconstruction module. Evolution of segmentation with different weights α of the loss across all three datasets, where DC is the Dice coefficient, HD is the Hausdorff distance, and VS is the volumetric similarity. Results are presented in mean ± standard deviation.

Dataset	α	DC	HD	VS
Kulaga–Yoskovitz	0.0	0.869 ± 0.029	11.784 ± 11.831	0.960 ± 0.031
	0.1	0.870 ± 0.029	10.143 ± 9.661	0.959 ± 0.031
	1.0	0.869 ± 0.030	11.167 ± 10.622	0.960 ± 0.031
	10.0	0.870 ± 0.029	9.843 ± 10.476	0.960 ± 0.030
MemoDev	0.0	0.664 ± 0.160	17.011 ± 22.268	0.865 ± 0.141
	0.1	0.664 ± 0.161	12.970 ± 17.867	0.866 ± 0.139
	1.0	0.662 ± 0.165	13.775 ± 20.456	0.869 ± 0.144
	10.0	0.667 ± 0.161	14.848 ± 19.280	0.859 ± 0.149
Simpson	0.0	0.878 ± 0.037	3.977 ± 4.088	0.944 ± 0.036
	0.1	0.879 ± 0.037	3.993 ± 4.339	0.944 ± 0.036
	1.0	0.879 ± 0.036	4.237 ± 4.909	0.944 ± 0.036
	10.0	0.878 ± 0.036	3.794 ± 3.592	0.943 ± 0.036

3.2 3D-AGSCaps: Comparison with Analogous Convolutional Models on Typical MRIs

We started by comparing 3D-AGSCaps against different data-augmentation strategies. Best results on test sets are obtained with little (15 deg) to no rotational augmentation: 3D-AGSCaps and both dilated models were showing better segmentation quality without training-time rotational augmentation, while simpler models (UNet and Residual UNet) slightly benefited from 15 deg maximum augmentation. On typical MRIs, we noted an overall superiority of residual models (namely 3D-AGSCaps, Residual UNet, and Residual DUnet) compared to the single ones (Table 5). However, among residual models, we failed to show a significant difference ($p > 0.05$) on DC, but 3D-AGSCaps showed a higher HD and VS (Table 5, with a qualitative comparison Fig. 5). We additionally monitored the computational resources required during the training phases (Table 6).

3.3 3D-AGSCaps: Robustness to Random Rotational Perturbations

After assessing segmentation quality on typical MRIs, we evaluated generalization on randomly rotated MRIs of our test set. An example of an MRI comprised in our test set is shown in Fig. 6,

Table 5 Effect of rotations as data augmentation during training time on segmentation quality on a test set. Each model is trained multiple times with a varying amounts of random rotations as part of the data augmentation pipeline, and then evaluated on (a) the Dice coefficient, (b) the Hausdorff distance, and the (c) volumetric similarity on an unseen test set. Results are presented as mean ± std. Bold results highlight the maximum amplitude of random rotations leading to the best performances.

	AGSCaps	UNet	Residual UNet	DUNet	Residual DUNet
(a). DC					
0 deg	**0.839 ± 0.098**	0.664 ± 0.289	0.815 ± 0.138	**0.712 ± 0.280**	**0.849 ± 0.096**
15 deg	0.811 ± 0.120	**0.676 ± 0.284**	**0.840 ± 0.105**	0.593 ± 0.345	0.819 ± 0.126
45 deg	0.669 ± 0.218	0.415 ± 0.317	0.681 ± 0.269	0.461 ± 0.281	0.711 ± 0.234
90 deg	0.460 ± 0.342	0.331 ± 0.328	0.482 ± 0.369	0.283 ± 0.315	0.481 ± 0.361
180 deg°	0.453 ± 0.342	0.274 ± 0.372	0.368 ± 0.347	0.252 ± 0.343	0.306 ± 0.357
(b) HD					
0 deg	**11.376 ± 9.045**	26.316 ± 19.955	5.699 ± 2.694	**19.107 ± 16.640**	**4.934 ± 2.113**
15 deg	18.502 ± 11.598	**14.390 ± 16.215**	**5.029 ± 2.575**	21.393 ± 15.304	5.991 ± 2.637
45 deg	29.278 ± 14.429	28.512 ± 15.197	16.382 ± 10.921	27.881 ± 15.240	13.209 ± 8.453
90 deg	25.868 ± 12.671	29.960 ± 14.665	24.034 ± 14.328	30.375 ± 13.002	26.238 ± 13.827
180 deg	26.912 ± 13.513	30.750 ± 11.788	27.891 ± 13.320	29.926 ± 12.436	31.018 ± 12.752
(c) VS					
0 deg	0.953 ± 0.043	0.739 ± 0.302	0.928 ± 0.113	**0.815 ± 0.281**	**0.961 ± 0.040**
15 deg	**0.971 ± 0.030**	**0.754 ± 0.308**	**0.963 ± 0.044**	0.683 ± 0.376	0.929 ± 0.088
45 deg	0.888 ± 0.112	0.543 ± 0.350	0.834 ± 0.244	0.600 ± 0.318	0.843 ± 0.177
90 deg	0.660 ± 0.282	0.501 ± 0.352	0.618 ± 0.356	0.374 ± 0.324	0.587 ± 0.388
180 deg	0.772 ± 0.227	0.445 ± 0.372	0.525 ± 0.367	0.382 ± 0.328	0.447 ± 0.360

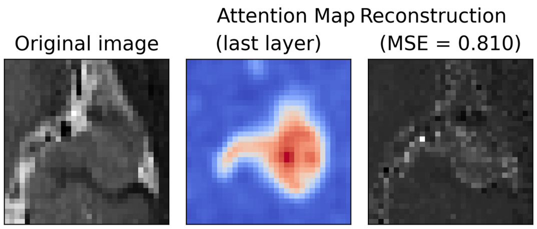

Fig. 5 Example of attention map given an input x and its reconstruction $\hat{x}$ (Kulaga–Yoskovitz dataset). The attention map comes from the very last layer just before entering the reconstruction decoder and the last two capsule layers.

Table 6 Computational comparisons. Timings were monitored on a Nvidia RTX8000 and are given only for information purposes.

	AGSCaps	UNet	Residual UNet	DUNet	Residual DUNet
# Parameters	2.285M	16.318M	35.069M	16.723M	35.475M
# FLOPS	0.493T	0.148T	0.180T	0.215T	0.246T
Epoch duration (s)	59.063	17.813	21.563	19.688	23.438
Training time (min)	63	19	23	21	25

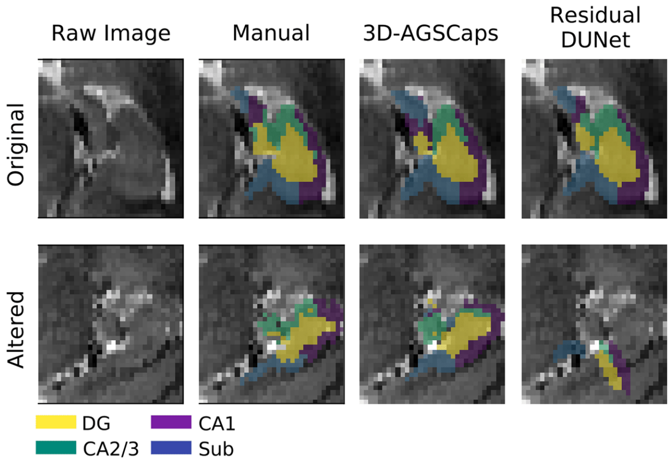

Fig. 6 Comparison between automatic segmentations against manual labeling. A single unseen hippocampi is segmented when left unaltered (original), and when altered with a random rotation (45 deg of maximum amplitude in a random axis). We present a manual segmentation, a segmentation from our model 3D-AGSCaps, and a segmentation produced by a residual DUNet from Ref. 14.

with and without deformation. Across our three datasets, an ANOVA showed no evidence of significant differences between all segmentation models for observations with little (15 deg) to no rotation ($p > 0.05$). However, for rotations greater or equal to 45 deg, segmentation models start to differentiate (Fig. 7). 3D-AGSCaps showed a higher DC ($p = 0.004$, $BF10 = 12.588$, Cohen's $d = 0.179$) than its CNN counterparts, a lower HD ($p = 0.001$, $BF10 = 1.638$, Cohen's $d = -0.120$), and a higher volumetric similarity ($p < 0.001$, $BF10 = 1e + 15$, Cohen's $d = 0.356$).

4 Discussion

The aim was to validate a public implementation of 3D-AGSCaps which offers a more accurate and fully automated segmentation on atypical populations and small datasets. We showed that 1/ 3D-AGSCaps challenged analogous convolutional architectures on hippocampal segmentation on typical MRI that is especially relevant in clinical population, and 2/ that 3D-AGSCaps exhibited robustness to random rotational perturbations (replicating atypical conditions, such as IHI) due to their implicit ability to learn equivariances over various instantiation parameters. On the one hand, because 3D-AGSCaps has been on-par with all other convolutional networks, we confirmed its ability to perform hippocampal segmentation on T1w and T2w MRIs. On the other hand, we showed that with an increasing quantity of random rotational perturbations, 3D-AGSCaps provided better segmentations than CNNs. Therefore, 3D-AGSCaps exhibits interesting properties in clinical settings: a better robustness to atypical images even when trained on small cohorts with only few patients.

4.1 3D-AGSCaps: Implementation and Ablation

We implemented a 3D SegCaps on a segmentation task and showed that 3D-AGSCaps is capable of hippocampal segmentation, with up to 15 times fewer parameters (35.5M parameters for a residual DUNet, against 2.3M for 3D-AGSCaps, Table 6). During our experiments, we found that a single iteration of the routing-by-agreement algorithm leads to the best results. This is a known effect, as previous works reported that usual routing-by-agreement algorithms may not behave as expected, unaffecting classification results, and often producing worse results than baseline algorithms.[46]

Most CapsNets use an atypical regularization and explanation technique, although the recent alternative approach from Ref. 47. developed after the start of this work, removed it. In addition to outputting the results of our task of interest, they output a reconstruction of the original input, optimized through a specific term in the loss function (Fig. 4). The relative weight of this loss term in the final loss functions remains unclear in the literature. Interestingly, we found that the reconstruction did not yield any significant enhancement of the segmentation, even if it reduced outliers produced in the MemoDev dataset (Table 4). However, if regularization is the main goal of the reconstruction module, other more efficient techniques should provide the same benefits without forcing the use of additional layers. If its goal is to achieve some sort of explainability of capsules' atoms, tuning the coefficient defined in the loss lead to no significant improvement in reconstruction quality. Qualitatively, reconstructions were of a relatively poor quality (e.g., Fig. 4), but the object of interest is recognizable enough for the sake of explanations. This is certainly caused by the MSE term of the loss function that assumes pixel independence without accounting for spatial relationships. MSE has been shown to produce low-quality reconstructions compared to more recent and specific loss functions, such as the SSIM or LWSSIM.[48] It has to be noted that not all implementations use such reconstruction modules.

To deal with the problem of the exponentially increasing number of parameters when switching from a 2D to a 3D space, we introduced AG (Fig. 3) with the aim of improving the efficiency of information routing. Our results showed an enhancement in segmentation quality, with a better overlap (mean DC increased by 0.012), fewer outlier voxels (mean HD reduced by 4.713), and a better VS (increased by 0.012).

4.2 3D-AGSCaps: Comparison with Analogous Convolutional Models both on Typical and Atypical MRIs

As a traditional approach to achieve rotational equivariance would be to use data augmentation, we assessed the effect of data augmentation on both our network (encoding equivariances of visual patterns through the orientation of the capsules) and SotA networks with training-time rotational augmentation (learning equivariance by learning the same pattern multiple times for different angles). Interestingly, we found that rotational augmentation mostly deteriorated segmentation quality (Table 5). At first sight, this fact may seem counterintuitive, as data augmentation should improve the generalization of deep learning models. However, this is coherent with part of the literature stating that data augmentation may increase overfitting risks,[17,18] leading to a marginal improvement on small training sets[47] or even lead to a drop in accuracy on unperturbed images.[10] This highlights pieces of evidence that efficient and robust segmentations on small training sets will benefit from networks showing built-in capacities to handle equivariances.

Given the best amount of training-time data augmentation for each model, 3D-AGSCaps did not show a significant improvement for typical MRIs on DC (Table 5) compared with architecturally equivalent models,[14] but showed a higher HD and VS. Overall, all models handled our segmentation task equally well with most DC superior to 0.8, but it is worth noting that we did not gather any clear evidence for statistically significant differences between models introduced in Ref. 14 and classic residual UNet models. Simpler and non-residual models (UNet and DUnet) were consistently left behind.

4.3 3D-AGSCaps: Robustness to Random Rotational Perturbations

Finally, we assessed our 3D-AGSCaps against SotA models regarding behaviors facing plausible alterations of the hippocampus, such as the IHI. Therefore, we monitored segmentation quality with an increasing maximum angle of random rotations up to 90 deg, following a realistic range (Fig. 7) on test observations. Alongside the previously discussed lack of differences for typical hippocampi (i.e., no rotation) heterogeneity was highlighted by increasing the amount of rotation. For angles as small as 45 deg, 3D-AGSCaps stood out, giving a better DC with a higher VS, followed by the two models introduced by Zhu et al. and then the UNet and Residual UNet baselines. Those evidences support our hypothesis stating that CapsNets can segment the hippocampus with more robustness toward alterations such as rotations, which is beneficial when working with clinical settings affected by data-scarcity issues.

It should be noted, however, that this improved robustness comes at the cost of computational requirements. This cost is mainly driven by the storage of activation values as vectors instead of scalars. As of now, CapsNets suffers from scalability issues, consuming up to 10 times the amount of GPU memory compared to CNNs with analogous architectures. This computational overhead also increased training time, from 19 min with a batch-size of 8 for a UNet, to 1 h for 3D-AGSCaps on an Nvidia RTX 8000 (Table 6). In order to process MRIs with a CapsNet

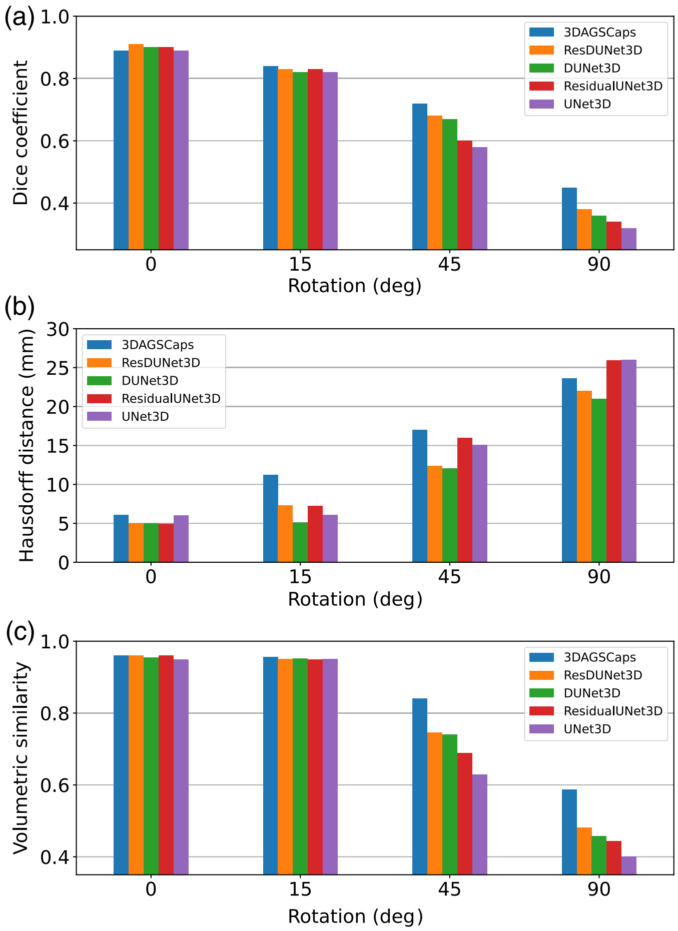

Fig. 7 Mean evolution of segmentation quality with respect to an increasing degree of random rotations. Comparisons of (a) the Dice coefficient, (b) Hausdorff distance, and (c) volumetric similarity between our model, 3DAGSCaps, and CNNs of similar architecture from the literature, ResDUNet and DUNet from Ref. 14, and simpler UNets with and without residual blocks.

such as 3D-AGSCaps, the use of a specific preprocessing of the input has to be performed, such as automatic detection of an enclosing box of both hippocampi to crop the MRI and reduce its memory footprint. This is the reason why we also published a third-party tool called ROILoc[49] (available in a GitHub repository at: https://github.com/clementpoiret/ROILoc), as a modest solution to this limit. In addition, recently introduced approaches, such as Ref., 47, optimized capsule networks by removing parts of the network from the original implementation, such as the reconstruction module.[20] As their implementation only works for binary classification tasks and because they did not replace the routing-by-agreement with an attention mechanism, we suggest the merging our two main contributions as a future interesting experiment: their fast implementation with our AG and the use of our SMSquash function to enable multiclass segmentation workflows.

Therefore, the use of CapsNets for MRI processing has to be justified by an underlying hypothesis such as the presence of IHI or hippocampal sclerosis in a pathological population. While this type of architecture seems promising, we believe it would be important to further investigate the computational efficiency of CapsNets to find ways to address these limitations. For example, it could be interesting to explore the GLOM architecture,[50] although still prototypical, but introduced specifically to solve some of the difficulties posed by the capsule design. Alternatively, as one of the main issues to solve this complex problem lies in the scarcity of labeled datasets, other tracks might be interesting to explore. In this way, self-supervised pretraining could help with the relative uselessness of rotational data augmentation and semi-supervised training, such as the recently introduced annotation-efficient deep learning (AIDE) framework,[51] seems to provide a simple way to handle segmentation tasks with scarce and noisy labeling.

To date, the segmentation of the hippocampus can capture anatomical variability, such as the IHI,[4] a developmental abnormality occurring in consequent subsets of the healthy or pathological population, such as in temporal lobe epilepsy or hippocampal sclerosis. The IHI is gradual, locally impacting shapes of the hippocampus. Because capsules in our architecture have a kernel size of 3, they can encode instantiation parameters as finely as a local cube of 3^3 voxels. Therefore by construction, 3D-AGSCaps can handle the naturally occurring variations of the IHI going up to 90 deg rotations, by modeling the verticality and roundness of the hippocampal body and the collateral sulcus (local and global rotational statistical equivariance of capsules) or the medial positioning of the hippocampus (translation equivariance of convolutions).[5] However, our experiments relied on relatively small datasets with limited labeled training data. For example, the MemoDev dataset only contained 50 labeled hippocampal volumes. While our results demonstrate 3D-AGSCaps can achieve robust segmentation even when trained on small datasets, validating the method on larger cohorts will be an important next step. Expanding the training data to encompass more anatomical variability could further improve segmentation accuracy. Finally, this work focused solely on hippocampal segmentation as a proof of concept for 3D capsules. Applying 3D-AGSCaps to other anatomical structures and segmentation tasks, such as brain lesions in multiple sclerosis or tumor detection, represents a promising direction for future research. Overall, this work provides initial evidence that 3D capsule networks can achieve robust medical image segmentation, but larger-scale validation and expanded applications will help realize the full potential of this approach.

The robustness of 3D-AGSCaps to rotational variations demonstrated in this study suggests the method may have clinically useful applications beyond hippocampal segmentation. For example, the hippocampus is affected by pathologies, such as hippocampal sclerosis and Alzheimer's disease, that can alter its shape and orientation. By learning robust equivariant representations, 3D-AGSCaps could enable more accurate segmentation and quantification of hippocampal subfields in these disease states compared to conventional CNNs. This in turn could shed light on how specific subregions are impacted over disease progression. In addition, other brain structures prone to rotational variations during development or disease may benefit from analysis with 3D-AGSCaps. More broadly, the capsule architecture's inherent robustness could prove useful for segmenting obscured or partially imaged structures in cases of traumatic injury or lesions. While further validation on diverse clinical datasets is needed, this work provides initial evidence that 3D-AGSCaps can achieve robust segmentation even when anatomical variation poses challenges for conventional deep learning methods.

5 Conclusion

With respect to not-so-rare atypical variations of the hippocampus, we assessed the usefulness of Capsule Networks in hippocampal segmentation both in an anteroposterior and in subfields manner. With our newly introduced architecture, 3D-AGSCaps, we validated a public implementation of 3D capsules. On the one hand, we confirmed the ability to perform hippocampal segmentation on T1w and T2w MRIs with 3D-AGSCaps, even if we found no evidence of superior segmentation quality for typical hippocampi (i.e., without rotation). On the other hand, we demonstrated that with an increasing quantity of random rotational perturbations, 3D-AGSCaps provided better segmentations than analogous CNNs due to their implicit ability to learn equivariances over various instantiation parameters. Unfortunately, we also found capsules to be extremely demanding for GPU memory, which is the main drawback of this methodology. This concern raises the need for further investigations to bring back scalability into this promising methodology offering enhanced robustness, especially given that the hippocampus is a small brain region demanding for higher resolution MRIs.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Code and Data Availability

The code employed in the current work has been published at https://github.com/clementpoiret/3D-AGSCaps. The Kulaga-Yoskovitz and Simpson datasets are publicly available in their original papers. The MemoDev dataset is available upon request to the corresponding author as the participants did not emitted consent for the public release of their data.

Acknowledgments

The authors declare not to have any conflicts of interest. This work has been partly funded by the Fondation de France. Clément Poiret has a PhD funding from Paris Cité University.

References

1. L. R. Squire et al., “Role of the hippocampus in remembering the past and imagining the future,” *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19044–19048 (2010).
2. T. Toda et al., “The role of adult hippocampal neurogenesis in brain health and disease,” *Mol. Psychiatry* **24**, 67–87 (2019).
3. I. Despotović, B. Goossens, and W. Philips, “MRI segmentation of the human brain: challenges, methods, and applications,” *Comput. Math. Methods Med.* **2015**, 450341 (2015).
4. R. Raininko and D. Bajic, ““Hippocampal Malrotation”: no real malrotation and not rare,” *Am. J. Neuroradiol.* **31**, E39–E39 (2010).
5. C. Cury et al., “Incomplete hippocampal inversion: a comprehensive MRI study of over 2000 subjects,” *Front. Neuroanat.* **9** (2015).
6. J. E. Iglesias et al., “A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of *in vivo* MRI,” *NeuroImage* **115**, 117–137 (2015).
7. J. E. Romero, P. Coupé, and J. V. Manjón, “HIPS: a new hippocampus subfield segmentation method,” *NeuroImage* **163**, 286–295 (2017).
8. P. A. Yushkevich et al., “Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment: automatic morphometry of MTL subfields in MCI,” *Hum. Brain Mapp.* **36**, 258–287 (2015).
9. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” https://doi.org/10.48550/arXiv.1505.04597 (2015).
10. L. Engstrom et al., “Exploring the landscape of spatial robustness,” https://doi.org/10.48550/arXiv.1712.02779 (2019).
11. C. Kanbak, S.-M. Moosavi-Dezfooli, and P. Frossard, “Geometric robustness of deep networks: analysis and improvement,” https://doi.org/10.48550/arXiv.1711.09115 (2017).
12. C. Xiao et al., “Spatially transformed adversarial examples,” https://doi.org/10.48550/arXiv.1801.02612 (2018).
13. Q. Qiu et al., “Feasibility of automatic segmentation of hippocampus based on deep learning in hippocampus-sparing radiotherapy,” *Int. J. Radiat. Oncol. Biol. Phys.* **105**, E137–E138 (2019).
14. H. Zhu et al., “Dilated dense U-Net for infant hippocampus subfield segmentation,” *Front. Neuroinf.* **13**, 30 (2019).
15. A. Punjabi, J. Schmid, and A. K. Katsaggelos, “Examining the benefits of capsule neural networks,” https://doi.org/10.48550/arXiv.2001.10964 (2020).
16. D. Marcos, M. Volpi, and D. Tuia, “Learning rotation invariant convolutional filters for texture classification,” in *23rd Int. Conf. Pattern Recognit. (ICPR)*, pp. 2012–2017 (2016).
17. S. O’Gara and K. McGuinness, “Comparing data augmentation strategies for deep image classification,” in *IMVIP 2019: Irish Mach. Vision & Image Process.*, p. 9 (2019).
18. C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *J. Big Data* **6**, 60 (2019).
19. G. E. Hinton, A. Krizhevsky, and S. D. Wang, “Transforming auto-encoders,” *Lect. Notes Comput. Sci.* **6791**, 44–51 (2011).
20. S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” https://doi.org/10.48550/arXiv.1710.09829 (2017).
21. R. LaLonde et al., “Capsules for biomedical image segmentation,” *Med. Image Anal.* **68**, 101889 (2020).
22. D. Li et al., “Robustness comparison between the capsule network and the convolutional network for facial expression recognition,” *Appl. Intell.* **51**, 2269–2278 (2021).
23. M. Kwabena Patrick et al., “Capsule networks: a survey,” *J. King Saud Univ. – Comput. Inf. Sci.* **34**, 1295–1310 (2019).
24. P. Afshar et al., “3D-MCN: a 3D multi-scale capsule network for lung nodule malignancy prediction,” *Sci. Rep.* **10**, 7948 (2020).

25. P. Afshar, A. Mohammadi, and K. N. Plataniotis, “Brain tumor type classification via capsule networks,” https://doi.org/10.48550/arXiv.1802.10200 (2018).
26. A. Balaji et al., “Coronary artery segmentation from intravascular optical coherence tomography using deep capsules,” *Artif. Intell. Med.* **116**, 102072 (2021).
27. A. Ahmad, B. Kakillioglu, and S. Velipasalar, “3D capsule networks for object classification from 3D model data,” in *52nd Asilomar Conf. Signals, Syst., and Comput.*, IEEE, pp. 2225–2229 (2018).
28. A. Cheraghian and L. Petersson, “3D capsule: extending the capsule architecture to classify 3D point clouds,” in *IEEE Winter Conf. Appl. Comput. Vision (WACV)*, IEEE, pp. 1194–1202 (2019).
29. B. Kakillioglu et al., “3D capsule networks for object classification with weight pruning,” *IEEE Access* **8**, 27393–27405 (2020).
30. O. Oktay et al., “Attention U-Net: learning where to look for the pancreas,” https://doi.org/10.48550/arXiv.1804.03999 (2018).
31. Y. Shen et al., “Coronary arteries segmentation based on 3D FCN with attention gate and level set function,” *IEEE Access* **7**, 42826–42835 (2019).
32. C. Li et al., “Application of U-shaped convolutional neural network based on attention mechanism in liver CT image segmentation,” *Lect. Notes Comput. Sci.* **633**, 198–206 (2020).
33. J. Choi et al., “Attention routing between capsules,” https://doi.org/10.48550/arXiv.1907.01750 (2019).
34. W. Huang and F. Zhou, “DA-CapsNet: dual attention mechanism capsule network,” *Sci. Rep.* **10**, 11383 (2020).
35. V. Mazzia, F. Salvetti, and M. Chiaberge, “Efficient-CapsNet: capsule network with self-attention routing,” *Sci. Rep.* **11**(1), 14634 (2021).
36. Y.-H. H. Tsai et al., “Capsules with inverted dot-product attention routing,” https://doi.org/10.48550/arXiv.2002.04764 (2020).
37. A. L. Simpson et al., “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” https://doi.org/10.48550/arXiv.1902.09063 (2019).
38. J. Kulaga-Yoskovitz et al., “Multi-contrast submillimetric 3 Tesla hippocampal subfield segmentation protocol and dataset,” *Sci. Data* **2**, 150059 (2015).
39. A. Bouyeure et al., “Hippocampal subfield volumes and memory discrimination in the developing brain,” *Hippocampus* **31**, 1202–1214 (2021).
40. M. A. Dalton et al., “Segmenting subregions of the human hippocampus on structural magnetic resonance image scans: an illustrated tutorial,” *Brain Neurosci. Adv.* **1**, 239821281770144 (2017).
41. P. Luo et al., “Differentiable learning-to-normalize via switchable normalization,” https://doi.org/10.48550/arXiv.1806.10779 (2019).
42. N. Abraham and N. M. Khan, “A novel focal Tversky loss function with improved attention U-Net for lesion segmentation,” https://doi.org/10.48550/arXiv.1810.07842 (2018).
43. Ö. Çiçek et al., “3D U-Net: learning dense volumetric segmentation from sparse annotation,” https://doi.org/10.48550/arXiv.1606.06650 (2016).
44. M. Bhalerao and S. Thakur, “Brain tumor segmentation based on 3D residual U-Net,” *Lect. Notes Comput. Sci.* **11993**, 218–225 (2020).
45. A. G. Rassadin, “Deep residual 3D U-Net for joint segmentation and texture classification of nodules in lung,” Vol. **12132**, pp. 419–427, http://arxiv.org/abs/2006.14215 (2020).
46. I. Paik, T. Kwak, and I. Kim, “Capsule networks need an improved routing algorithm,” https://doi.org/10.48550/arXiv.1907.13327 (2019).
47. A. Avesta et al., “3D capsule networks for brain image segmentation,” *Am. J. Neuroradiol.* **44**(5), 562–568 (2023).
48. Y. Lu, “The level weighted structural similarity loss: a step away from the MSE,” https://doi.org/10.48550/arXiv.1904.13362 (2019).
49. C. Poiret et al., “A fast and robust hippocampal subfields segmentation: HSF revealing lifespan volumetric dynamics,” *Front. Neuroinform.* **17**, 1130845 (2023).
50. G. Hinton, “How to represent part-whole hierarchies in a neural network,” https://doi.org/10.48550/arXiv.2102.12627 (2021).
51. S. Wang et al., “Annotation-efficient deep learning for automatic medical image segmentation,” *Nat. Commun.* **12**, 5915 (2021).

Clement Poiret received his PhD in neuroscience from the University of Paris. He is specialized in deep learning and computer science applications in neuroimaging. He received his PhD from Neurospin, CEA Saclay with INSERM U1141.

Antoine Bouyeure is a postdoctoral researcher in neurosciences, Ruhr University Bochum. He is an expert in neuroimaging of episodic memory and fear conditioning/extinction. He received his PhD from NeuroSpin, CEA Saclay with INSERM U1141.

Sandesh Patil was an engineer at Neurospin, CEA Saclay with INSERM U1141. Currently, he is a research and development engineer at INRIA, working on workflow and data management for medical imaging platforms.

Cécile Boniteau was a research intern at Neurospin. Currently, she is a machine learning engineer passionate about AI applications in medical imaging and healthcare.

Edouard Duchesnay is a research director in data science at NeuroSpin, CEA Paris-Saclay University. He has been designing machine learning models for neuroimaging signatures of mental disorders since 2003.

Antoine Grigis is a researcher at NeuroSpin CEA-Saclay. He received his PhD in medical image processing from the University of Strasbourg in 2013. He manages neurospin analysis platform. His interests include computational neuroscience and high-performance computing applications.

Frederic Lemaitre is an assistant professor at the Université de Rouen, specialized in physiology, health, rehabilitation, obesity, cardiology, and exercise science research.

Marion Noulhiane, PhD, HDR, is a researcher at NeuroSpin CEA-Saclay, in anatomofunctional maturation of hippocampus using neuroimagery methodology (volumetric analysis, deep learning…) and its role in memory and timing. She supervises the HippoMnesis research group, and the thesis of Clement Poiret.