

Developing a brain atlas through deep learning

Asim Iqbal^{1,2}, Romesa Khan³ and Theofanis Karayannis^{1,2*}

Neuroscientists have devoted substantial effort to the creation of standard brain reference atlases for high-throughput registration of anatomical regions of interest. However, the variability in brain size and form across individuals poses a considerable challenge for such reference atlases. To overcome these limitations, we introduce a fully automated deep neural network-based method (named SeBRe) for segmenting brain regions of interest with minimal human supervision. We demonstrate the validity of our method on brain images from different developmental time points of mice, across a range of neuronal markers and imaging modalities. We further assess the performance of our method on images of magnetic resonance-scanned human brains. Our registration method can accelerate brain-wide exploration of region-specific changes in brain development and, by easily segmenting brain regions of interest for high-throughput brain-wide analysis, offer an alternative to existing complex brain registration techniques.

The development and efficient implementation of various methods in neuroscience for labelling specific populations of brain cells *in situ*—such as the genetic modification of animals, or the more conventional immunocytochemistry and mRNA *in situ* hybridization (ISH) of brain tissue—have enabled neuroscientists to label and track various neuronal types in different brain regions¹. These cellular expression patterns are captured through a variety of high-throughput imaging techniques (for example, whole-tissue light-sheet microscopy, or bright wide-field and fluorescent microscopy using high-resolution confocal microscopes and slide scanners), which allow for their exploration at a mesoscopic level. However, quantitative analysis of these brain datasets remains a great challenge in the field of neuroscience, with a major issue being the complexity of registering mouse brain sections against a standard reference atlas.

A number of efforts are underway to develop high-throughput image registration frameworks for analysis of such large-scale brain datasets^{2,3}. Most of these frameworks are semi-manual; that is, the user is either expected to set a certain range of parameters (such as the intensity threshold, background contrast and so on) for every brain section, or completely transform their brain datasets into a framework-readable format. Nevertheless, the ability of these methods to register every brain section against a reference atlas is limited, and, hence, they lack the generalizability required for analysis of a variety of datasets. Furthermore, there are no reference atlases for most of the developing age groups.

The development of deep learning-based techniques is now delivering state-of-the-art results in real-world object classification⁴, localization⁵ and segmentation⁶ tasks. Although deep learning methods have been applied to different organs of the body (such as the heart and bones) for segmentation⁷, they have rarely been used in the analysis of complex brain image datasets^{8,9} and have not been applied to atlas registration and region annotation (classification). Among the many challenges, this artificial intelligence-based approach would require the generation of a sufficiently large dataset of brain sections that is labelled with brain regions in reference to a standard atlas, and captures the variability of various regions in different sections across the brain.

We propose that the task of registering images of brain sections against a standard reference atlas can also be achieved by using a deep neural network (DNN) to segment regions in the brain through

feature engineering. Therefore, we introduce a different approach to the classical problem of brain image registration, which relies on segmentation. This approach deploys a fully automated DNN-based method to segment and annotate various regions in images of the mouse brain (as well as the human brain) at different ages by optimizing the Mask R-CNN architecture⁶ and applying transfer learning onto a network pre-trained on the Microsoft common objects in context (MS COCO) dataset¹⁰. To train and test the performance of the network, we generated two distinct human-annotated mouse brain region datasets using the Allen Brain Institute online public resource (OPR)¹¹, as well as a set of publicly available human magnetic resonance (MR) images¹². By comparing the performance of our network with human-annotated ground-truth data and traditional brain registration and segmentation approaches, we achieve high average precision scores and thus demonstrate the power of our method. In essence, SeBRe automatically generates the brain reference atlas corresponding to any input brain section. This approach marks a paradigm shift in dealing with large-scale brain datasets, because it is independent of the time-intensive step of manual selection, followed by the computationally expensive step of registration, of the putative reference atlas onto the input brain sections. We propose that our approach of segmenting brain regions could become a launch pad for replacing the classical methods for registering brain sections against a standard reference atlas in the future.

Results

The performance of SeBRe on mouse brain images. Our approach is presented as a block diagram in Fig. 1, where an input brain image from a developing mouse brain is fed into the DNN and is passed through a series of feature processing stages, in which brain regions are proposed, classified and segmented. Step-by-step processing of a sample input brain section is demonstrated in Fig. 2, where an input brain section is passed through a series of image-processing stages (we demonstrate the generation of the ground-truth dataset in Supplementary Figs. 1–3, which is explained in the Methods). To facilitate access to our DNN among the neuroscience community, we make our pre-trained DNN on the mouse and human datasets, along with the complete dataset, open-source.

We tested the performance of SeBRe by designing two kinds of experiments: the first involved training and testing the DNN

¹Laboratory of Neural Circuit Assembly, Brain Research Institute (HiFo), UZH, Zurich, Switzerland. ²Neuroscience Center Zurich (ZNZ), UZH/ETH Zurich, Zurich, Switzerland. ³Department of Biology (D-BIOL), ETH Zurich, Zurich, Switzerland. *e-mail: karayannis@hifo.uzh.ch

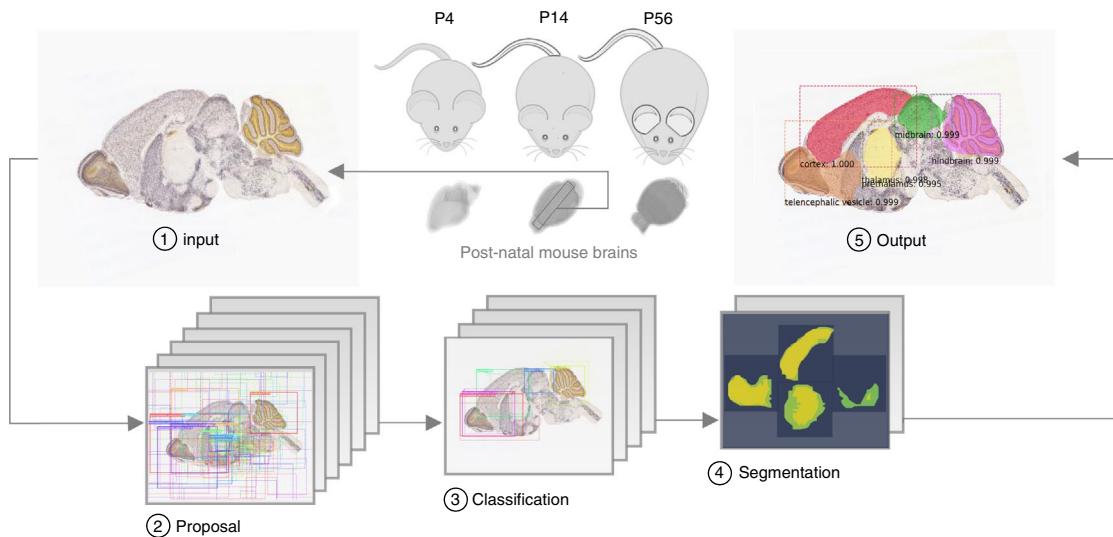


Fig. 1 | Architecture of the SeBRe deep learning pipeline. A medial section (top left) of a P14 GAD1 mouse brain is fed into the network and the output (top right) shows the segmented brain regions overlaid on the input brain section after the proposal, classification and segmentation stages.

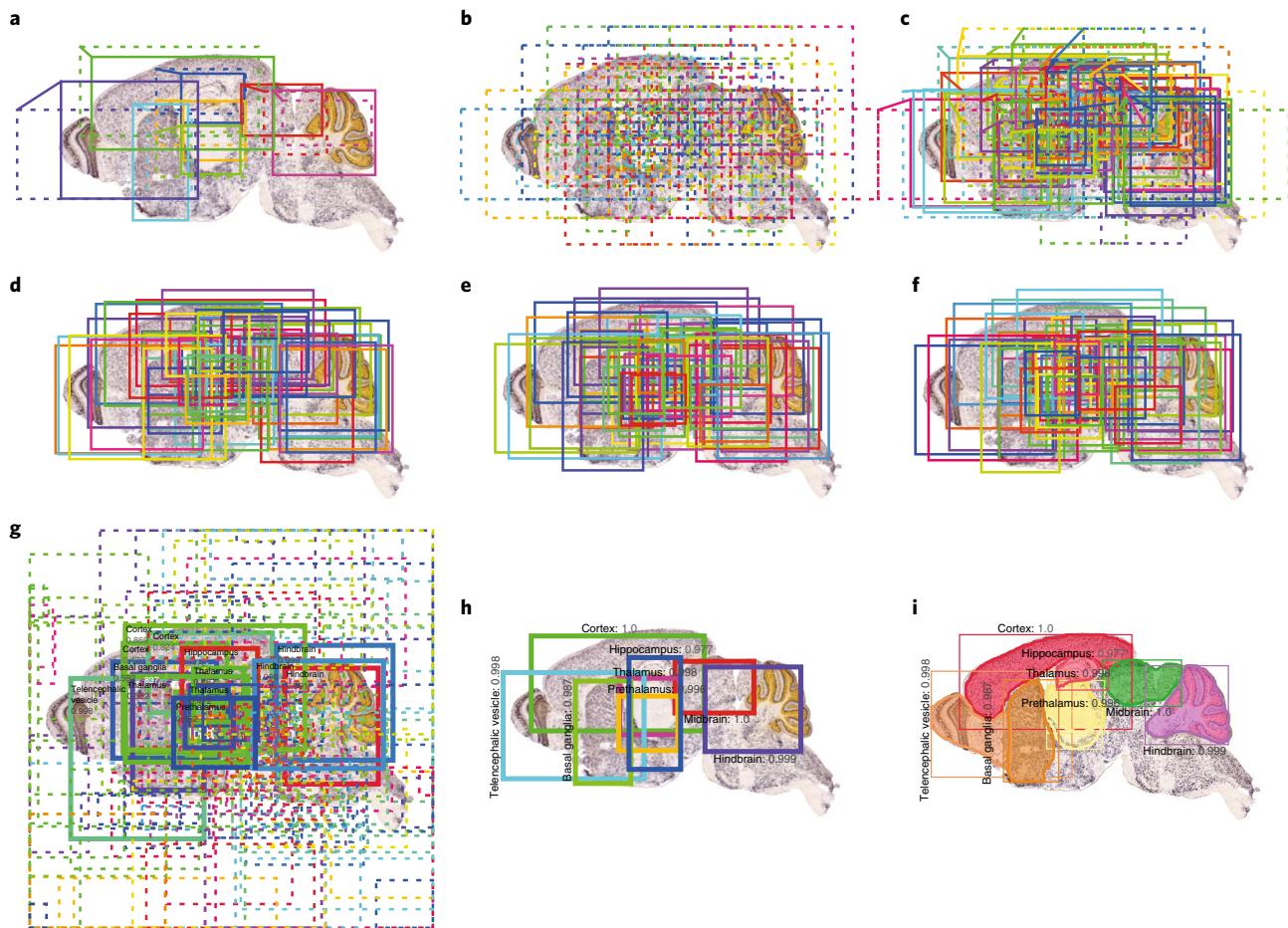


Fig. 2 | SeBRe multistage image processing pipeline. **a–f**, A step-by-step flow of brain region localization (**a–f**), classification (**g,h**) and segmentation (**i**). Thick boxes are labelled with the predicted class and confidence score; dashed boxes represent proposals classified as background. **a**, The bounding boxes of ground-truth targets generated for region proposal network (RPN) training. **b**, The boxes for proposed regions (anchors) that are predicted by the RPN. **c**, The RPN-predicted anchors after refinement. **d**, The RPN-predicted anchors after clipping to the image boundaries. **e**, The RPN-predicted anchors after applying non-maximum suppression. **f**, The final RPN-predicted anchors after coordinate normalization. **g**, The predictions of the feature pyramid network (FPN) classifier heads on RPN-predicted anchors before refinement. **h**, FPN-predicted classes with their classification scores after anchor refinement and non-maximum suppression. **i**, The final segmented masks of brain regions with their classification scores overlaid on the input brain section.

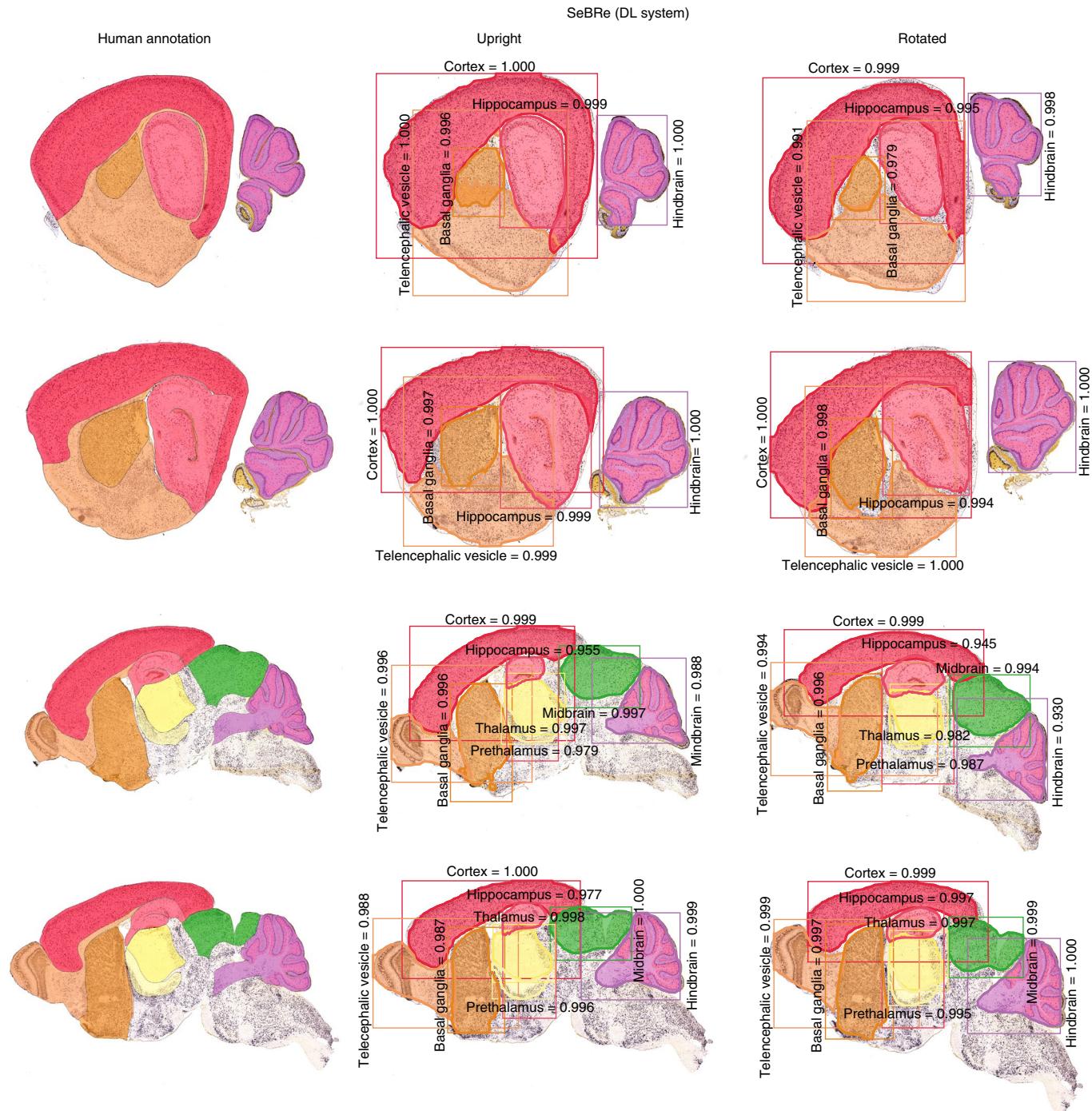


Fig. 3 | The performance of SeBRe in segmenting brain regions. A qualitative performance comparison of SeBRe on the lateral (rows 1 and 2) and medial (rows 3 and 4) sections of P14 mouse brains with human-annotated masks. SeBRe performs optimally at predicting masks of brain regions, for both upright (middle column) and rotated (right column) versions of the input brain sections (left column).

on images of two ISH brains of an intermediate mouse age (such as post-natal time point 14, or P14) to check the scalability of our method in segmenting brain regions of earlier (P4) and later (P28, P56) time points, whereas the second involved training and testing the DNN on an extended mouse dataset of nine ISH brains, which covers three post-natal mouse ages (P4, P14, P56) from different genetically modified animals. We used the trained model from the second experiment to demonstrate the generalizability of our method to segmenting regions in developing mouse brain images that are captured through different imaging modalities, such as fluorescent *in situ* hybridization (FISH).

The output of SeBRe for various sections of P14 mouse brains is demonstrated in Fig. 3, where the first column shows the human-expert-annotated masks for eight regions of interest in randomly selected mouse brain sections that are labelled with the neuronal markers glutamate decarboxylase 1 (GAD1) and vesicular GABA transporter (VGAT, where GABA is gamma-aminobutyric acid). The second column shows the performance of SeBRe in segmenting the eight regions in these brain sections. The third column shows the performance of the network on rotated $[-20^\circ, 20^\circ]$ versions of the same brain images. The network performs well in segmenting brain regions in upright as well as rotated versions of the brain images.

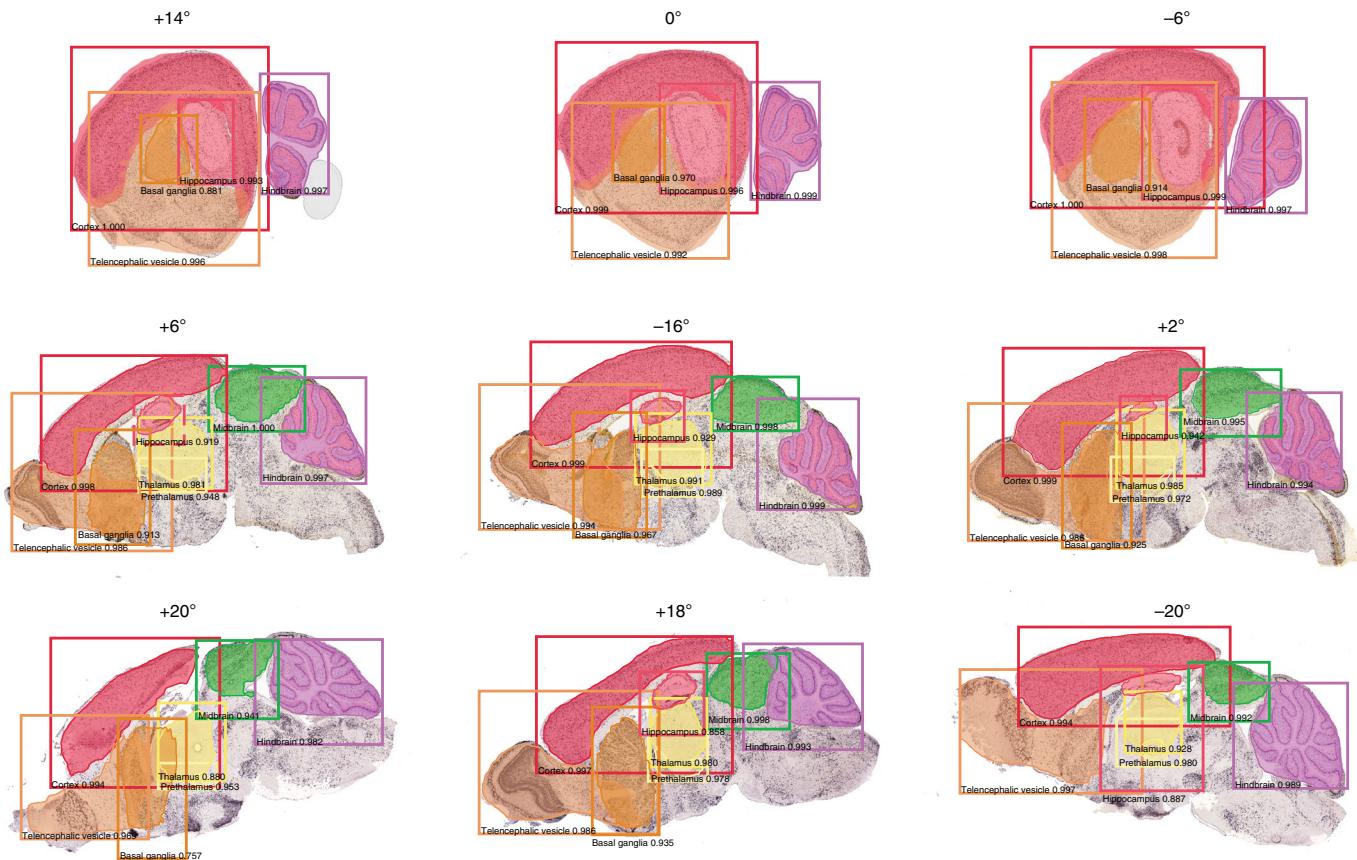


Fig. 4 | The rotationally invariant performance of SeBRe on the extended mouse dataset. The performance of SeBRe on a set of sample rotated sections from the mouse brain testing dataset is shown. Each rotation is given above the brain section, with reference to 0° (upright). The performance of SeBRe is unaffected by the rotation of a brain section in segmenting brain regions of interest, even in cases when the tissue is broken (for example, the cortex and telencephalic vesicle in the bottom-left section).

We further test the performance of SeBRe on rotated brain sections (Fig. 4) drawn from the extended mouse dataset to demonstrate that the method is unaffected by rotation. This approach is useful for registering brain sections in real time under the microscope, where the brain tissues could be placed at various angles for convenience.

The mean average precision scores achieved for brain sections in the original (P14) and extended (P4, P14, P56) test datasets are 0.84 and 0.87, respectively. The performance of the network is optimal in most brain regions, including the isocortex, hippocampus, basal ganglia, telencephalic vesicle and midbrain. The performance of the network for segmenting the prethalamus is limited due to large variation in the structure and size of this region as we move from the lateral to medial images across the sagittal plane.

The performance of SeBRe on mouse brain images labelled with previously ‘unseen’ markers and ages. After training and testing the performance of our network on P14 brain sections, we further assessed the scalability of our method by testing it on various mouse brain images from different developmental time points (P4, P14, P28 and P56) for commonly used neuronal markers that the network had not been trained on (CaMKIIa, Nissl, GAD1 and VGAT)¹¹. The largest population of neurons in the brain (excitatory (glutamatergic) neurons) was labelled with Ca²⁺/calmodulin-dependent protein kinase IIa (CaMKIIa), the second largest neural population (inhibitory (GABAergic) neurons) was labelled with GAD1 and VGAT, and all of the cells in the brain were labelled with Nissl (the results of the segmentation of various mouse brain section images across different ages and neuronal markers are shown in Supplementary Fig. 4, whereas the performance of SeBRe on ran-

domly selected sections from a P4 mouse brain labelled with GAD1 and VGAT is shown in Supplementary Fig. 4a,b). The network seems to perform optimally in segmenting the isocortex, thalamus and telencephalic vesicle, whereas detection of the midbrain and hindbrain regions proves challenging (the performance of SeBRe on randomly selected brain sections from Nissl and CaMKIIa brains at P4 is shown in Supplementary Fig. 4c,d). Although the network has been trained on GAD1 and VGAT tissue sections of only a single age (P14), it performs equally well in segmenting brain regions for various neuronal markers at different developmental ages (the segmentation results for P28+ adult mouse brain regions in GAD1, VGAT, CaMKIIa and Nissl brains are shown in Supplementary Fig. 4g–j). Similar to the P14 testing dataset, the network seems to occasionally omit a few regions from the medial brain sections of older mice (such as the midbrain and prethalamus), but the segmentation of the isocortex and other regions remains good.

The performance of SeBRe on mouse brain images acquired through a previously ‘unseen’ imaging modality. We tested the performance of SeBRe, trained on the extended mouse dataset, on a variety of fluorescently labelled brain images from various transgenic mouse lines that label different brain areas and cell types, and thus cover a diverse range of image intensity profiles¹¹. These lines are genetically modified using a Cre-LoxP system, with a variety of Cre driver lines recombining a red reporter allele and hence labelling different cell populations (represented in probe/driver-Cre-reporter notation). Many of them also display a second colour (green) after an ISH against GAD1, Slc17a6 or Rorb. We tested the performance of SeBRe in segmenting brain regions of the mouse brain

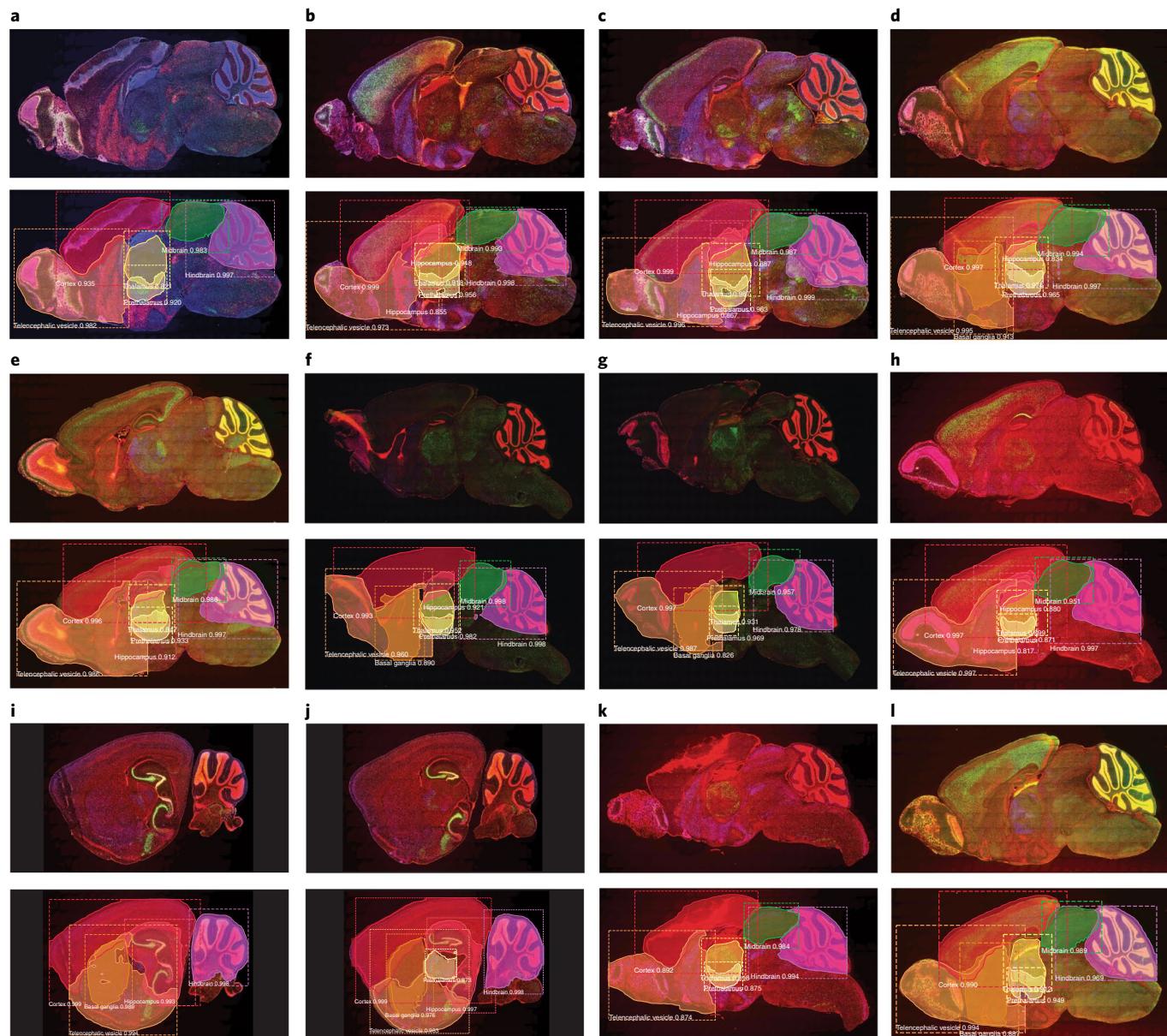


Fig. 5 | The generalized performance of SeBRe on unseen FISH brain sections. **a–j**, SeBRe is trained on the extended mouse brain dataset of ISH brain sections and tested on FISH brain sections. The input brain sections are shown at the top of each panel, with the output of the network displayed below. SeBRe is able to detect the brain regions captured through a different imaging modality without any additional preprocessing. The performance is shown for a variety of FISH brain sections collected randomly from a list of genetic labelling markers available on the Allen Brain Institute, such as GAD1/Cux2-CreERT2-Ai14(tdTomato) (**a–c**), GAD1/Grik4-Cre-Ai14(tdTomato) (**i,j**), Rorb/Scnn1a-Cre-Ai14(tdTomato) (**d–l**), Slc17a6/Slc32a1-Cre-Ai14(tdTomato) (**f,g**) and GAD1/Gpr26-CreKO250-Ai14(tdTomato) (**h,k**).

images from the following transgenic mouse brains: GAD1/Cux2-CreERT2-Ai14(tdTomato), GAD1/Grik4-Cre-Ai14(tdTomato), Rorb/Scnn1a-Cre-Ai14(tdTomato), Slc17a6/Slc32a1-Cre-Ai14(tdTomato) and GAD1/Gpr26-CreKO250-Ai14(tdTomato). Although only trained on GAD1 or VGAT ISH images, SeBRe performs reasonably well in segmenting mouse brain regions in FISH brains imaged under a fluorescent microscope (Fig. 5), even in cases when the brain region boundaries are not clearly defined (Fig. 5f,g) or are broken (Fig. 5a,k). It is interesting to observe that SeBRe performs well (independent of the brain imaging modality) because the DNN is able to detect features of interest without requiring any further training on the unfamiliar image set. This emphasizes the power of an artificial intelligence-based segmentation method over traditional registration techniques.

The performance of SeBRe in fine-scale segmentation of a mouse hippocampus. To evaluate whether our method can also be extended to handle the more complex task of fine-scale subregion segmentation within a brain tissue of interest, we tested the performance of SeBRe in isolating the four major subregions of the mouse hippocampus (CA1, CA2, CA3 and the dentate gyrus (DG)) in a Nissl-labelled ISH brain from the Allen Brain OPR. We trained the network on three-quarters of the dataset (selected randomly), and subsequently tested the performance on the remaining one-quarter; the results are shown in Fig. 6 for three sample mouse brain sections from different coronal planes. We further tested the performance of SeBRe on rotated samples of coronal brain sections with different orientations of two randomly selected rostral (Supplementary Fig. 5a) and caudal (Supplementary Fig. 5b) brain sections. The

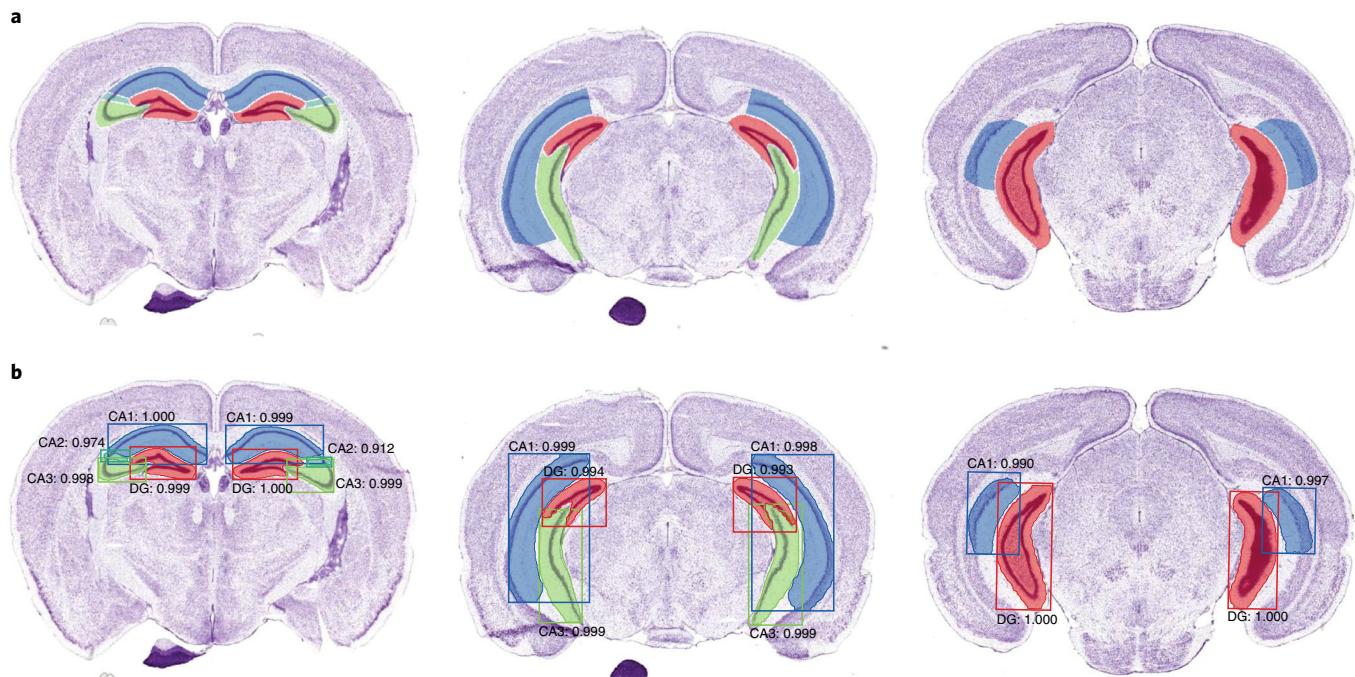


Fig. 6 | The performance of SeBRe in segmenting subregions of the hippocampus. **a**, Three mouse brain sections taken randomly from different coronal planes are shown, with overlaid human-annotated (ground-truth) masks for hippocampal subregions CA1, CA2, CA3 and the DG. **b**, The performance of SeBRe is demonstrated in segmenting hippocampal subregions in the corresponding upright brain sections in **a**.

subregion masks predicted by the network adequately match the human-annotated ground-truth segmentations for all four subregion classes in rostral as well as caudal brain sections that contain the hippocampus structure. The network has a high mean average precision score of 0.9 for detecting and distinguishing subregions of the mouse hippocampus. The network is also able to identify and segment the major subregions of the hippocampus in ISH brain sections that are labelled with a range of other neuronal markers spanning diverse spatial expression patterns (for example, the performance of SeBRe on a coronal section labelled with neurogranin (Nrgn) is shown in Supplementary Fig. 6b). We further tested whether our method is also generalizable to brain images captured under a previously unseen imaging modality. SeBRe (which was trained on ISH brain images) performed reasonably on a FISH transgenic mouse brain that simultaneously expressed a pan-neuronal red reporter and a green fluorescent marker for cell bodies and neurites (NF-160) (as illustrated in Supplementary Fig. 6a), despite significant differences from the bright-field image intensity range on which the network was originally trained.

The performance of SeBRe on human brain images. To evaluate the generalizability of our deep learning-based registration method to human data as well as other imaging modalities, we further tested the performance of SeBRe on an open-source, manually segmented, human T1-weighted magnetic resonance-imaging (MRI) brain dataset from the Internet Brain Segmentation Repository (IBSR)¹². We trained the network on two-thirds of the dataset (randomly selected), and subsequently tested the performance on the remaining one-third. Randomly selected horizontal sections drawn at different dorsoventral planes from the scanned brain volumes are shown in Fig. 7a, the human-annotated ground-truth masks for eight different subcortical structures overlaid onto the brain sections are shown in Fig. 7b and the segmentation performance of SeBRe on these input brain sections is shown in Fig. 7c. The network output closely matches the human-annotated ground-truth

segmentations for all eight subcortical regions, in both dorsal as well as ventral sections, giving a high mean average precision score of 0.95.

A comparison of SeBRe with image-processing-based brain registration techniques. To demonstrate the power of our registration-through-segmentation approach, the performance of SeBRe was compared with two other commonly used image registration methods: elastix, a toolbox for rigid and non-rigid registration of medical images¹³, and the neurodata registration module (ndreg), which uses affine and non-affine transformations to align a mouse brain reference atlas image to brain section images¹⁴. For a fair comparison, mouse brain reference atlas images (which only comprise the eight regions of interest on which SeBRe was trained) were registered onto the corresponding mouse brain sections using ndreg and elastix. Figure 8a visually demonstrates the registration performance of SeBRe, elastix and ndreg on the lateral and medial sagittal sections from GAD1 and VGAT brains. The mean squared error (MSE) score for the ‘registered’ masks returned by SeBRe was notably lower than the MSE scores for both elastix and ndreg (Fig. 8b,c), which indicates that our method achieved higher registration accuracy. It is interesting to observe that these two commonly used registration techniques seem to underperform on the same type of images, as indicated by the increased MSE score in registering rotated samples, in contrast to our machine intelligence-based method (Supplementary Fig. 7).

A comparison of SeBRe with other deep learning-based brain segmentation techniques. To demonstrate the robustness of our segmentation approach, the performance of SeBRe was also compared with previously built deep learning-based segmentation methods—including DeepLab¹⁵, BrainSegNet¹⁶ and QuickNAT¹⁷—across the eight subcortical regions of the IBSR dataset. The region-wise average Dice coefficient score is reported for all methods in Table 1 (columns 6–9): SeBRe gives the highest segmentation accuracy

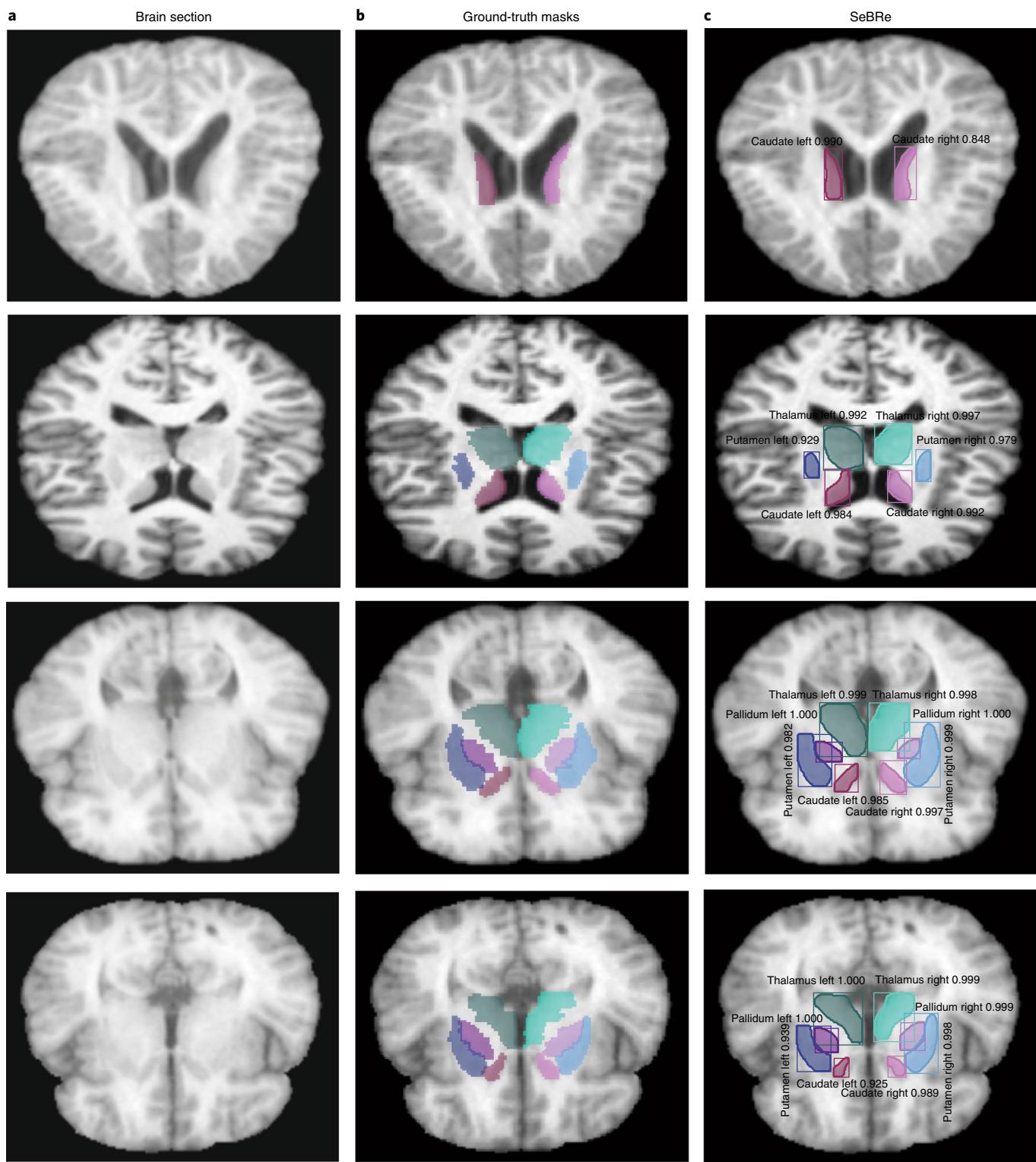


Fig. 7 | Performance of SeBRe on human brain sections. **a**, Randomly selected human brain MR scans are shown. **b**, Ground-truth masks of the corresponding MR scans in **a**. **c**, The performance of SeBRe is demonstrated in detecting eight brain regions: the caudate (left and right), thalamus (left and right), putamen (left and right) and pallidum (left and right), in the human MR scans.

among the evaluated methods in five out of eight classes (subcortical regions), and comparable accuracy for the remaining classes. The Hausdorff distance and contour mean distance scores for each subcortical region are further compared between SeBRe and DeepLab in Table 1 (columns 2–5), where it is shown that SeBRe provides consistently lower distance scores for segmentation of all of the regions.

In conclusion, we introduce a DNN-based method called SeBRe to classify and segment different regions in mouse and human brain images. To test the performance of our method, we used the open-source Allen Brain ISH mouse brain images and generated a human-annotated dataset of nine brains from three postnatal ages: P4, P14 and P56. We demonstrate proper and accurate segmentation of eight

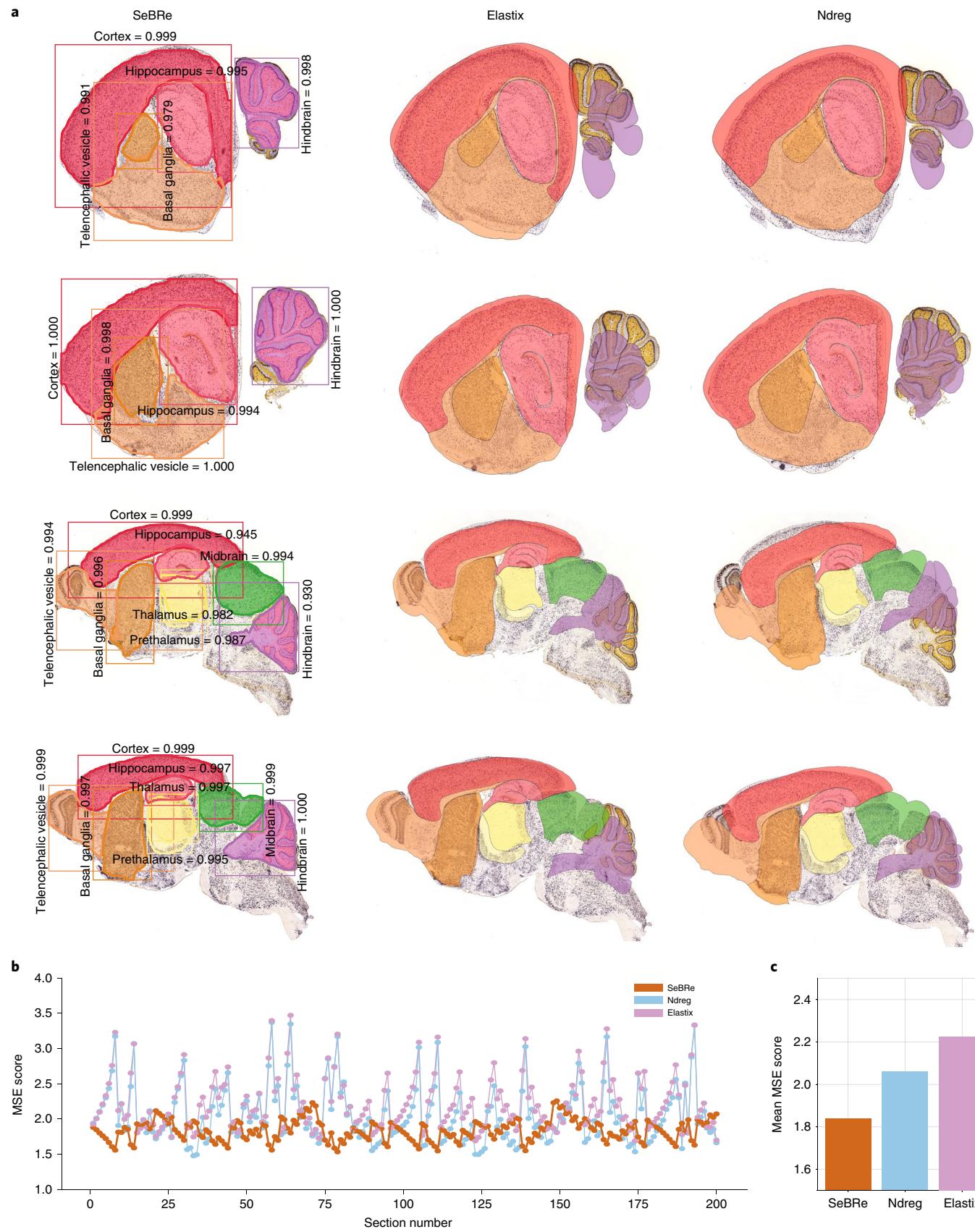


Fig. 8 | A comparison of SeBRe with commonly used brain registration methods. **a**, The performance of SeBRe on randomly selected lateral and medial brain sections is demonstrated in comparison with two commonly used brain registration methods: ndreg and elastix. **b**, A plot of MSE scores for all brain sections in the test dataset for SeBRe, ndreg and elastix. **c**, The mean MSE scores for SeBRe, ndreg and elastix on the complete test dataset.

Table 1 | A comparison of the performance of SeBRe with other deep learning-based brain region segmentation techniques

Brain region	SeBRe Hausdorff distance (mean)	F-CNN Hausdorff distance (mean)	SeBRe CMD (mean)	F-CNN CMD (mean)	SeBRe Dice (mean)	F-CNN Dice (mean)	BrainSegNet Dice (mean)	QuickNAT Dice (mean)
Caudate (left)	1.874 ^a	6.5	0.374 ^a	0.75	0.886 ^a	0.785	0.86	0.875
Pallidum (left)	1.971 ^a	4.5	0.477 ^a	0.77	0.849 ^a	0.77	0.81	0.81
Putamen (left)	1.828 ^a	5.0	0.412 ^a	0.72	0.889	0.84	0.91 ^a	0.89
Thalamus (left)	2.44 ^a	5.0	0.557 ^a	0.77	0.883 ^a	0.88	0.88	0.88
Caudate (right)	2.044 ^a	6.5	0.386 ^a	0.75	0.884 ^a	0.8	0.88	0.87
Pallidum (right)	2.124 ^a	4.5	0.502 ^a	0.80	0.844 ^a	0.75	0.83	0.825
Putamen (right)	1.875 ^a	7.0	0.449 ^a	0.75	0.872	0.85	0.91 ^a	0.89
Thalamus (right)	2.206 ^a	4.5	0.518 ^a	0.75	0.891	0.885	0.9 ^a	0.878

^aThe best performance of a DNN on the given brain region. CMD, contour mean distance.

regions in mouse brain images from the lateral and medial sagittal planes. Notably, our method is scale invariant; that is, it performs equally well in segmenting brain images that vary substantially in the relative sizes of the brain regions. This is crucial not only for accurately segmenting regions at different stages of development, but also in tissue sections in which the proper geometry of the brain has been compromised due to any methodological issues (for example, during brain tissue slicing) or even pathological deformities (such as neurodegeneration in Alzheimer's disease). This is demonstrated by the fact that although SeBRe was trained only on a P14 image dataset, the network performs equally well on other developmental ages, including P4, P28 and P56. We further demonstrate the performance of our method on brain sections obtained through a different imaging modality such as FISH, which points towards the generalizability of SeBRe in segmenting brain regions in a variety of bioimaging modalities.

Moreover, we demonstrate the performance of SeBRe in fine segmentation of the subregions of mouse hippocampus across different neuronal markers, as well as different imaging modalities (ISH and FISH). We show that the performance of SeBRe is extendable to other species by testing it on a publicly available human MRI dataset. We therefore believe that SeBRe could become a valuable tool for many bioimaging applications, such as segmenting brain regions in microscopy images for brain-wide analysis in animal models, or even human MRI and post-mortem tissue. Finally, although we demonstrate the application of our method in generating a reference atlas for mouse and human brain regions and subregions, it is worth noting that, in essence, a similar approach could also be adapted for any other primate or non-primate model organisms, such as monkey or drosophila.

Methods

Generating the ground-truth dataset for a mouse brain. We generated the ground-truth dataset by first obtaining the open-source ISH mouse brain images for two different genetic markers (GAD1 and VGAT) from the Allen Brain OPR¹¹. These two markers capture the population of GABAergic neurons in developing mouse brains. We chose P14 as the intermediate age because it captures enough variance between a young animal pup (P4) and an adult mouse (P28+) in terms of brain region development. These 20-μm-thick sections were cut in sagittal planes spaced at 200 μm intervals, covering one hemisphere of a whole brain from the lateral to medial (Supplementary Fig. 1 illustrates the process of ground-truth data generation on a sample GAD1 mouse brain section, and Supplementary Fig. 2 shows the sample human-annotated sagittal sections of a complete GAD1 brain at P14). The brain sections were overlaid by the Allen developing mouse brain reference atlas, with each region being assigned a unique colour code. Human experts manually registered the brain sections with scalar vector graphics files of the reference atlas by using the Boxy scalar vector graphics editor. The scalar vector graphics files of a developing mouse brain atlas were imported from the Allen Brain OPR. The GAD1 and VGAT brains at P14 comprise 36 brain sections (17 and 19, respectively). Six sections were removed from these two brains, as these did not meet the required quality criteria due to broken/damaged tissue.

To increase the variability of the sample images for each brain section, we introduced a synthetic variance of a 2° rotation (in the range of [−20, 20]), which resulted in 20 rotated versions per brain section. We applied the same procedure to all 30 brain sections, with the resulting augmented dataset comprising 600 brain sections. We applied a downsampling ratio of 25% to each image in the dataset to reduce the computational cost. Out of these 30 sections, we randomly chose two-thirds of the brain sections with their rotated versions (400 images) for training, and the remaining one-third with their rotated versions (200 images) for testing.

Generating masks for brain regions of a mouse. To demonstrate the performance of SeBRe, we chose eight major regions in the developing mouse brain for training and testing; namely, the isocortex, hippocampus, basal ganglia, thalamus, prethalamus, midbrain, telencephalic vesicle (olfactory bulb and partial forebrain) and hindbrain (cerebellum). Sample binary masks of five example regions, along with their corresponding brain sections, are shown in Supplementary Fig. 3. These masks were generated from the human expert-annotated ground-truth dataset, where each brain region is masked with a unique colour code (see Supplementary Fig. 2).

Generating the extended ground-truth dataset for a mouse brain across different ages. To scale our brain registration approach to different mouse developmental ages, we generated an extended ground-truth dataset comprising ISH mouse brains from the Allen Brain OPR that were labelled with different neuronal markers (GAD1, VGAT and Nissl), across three distinct developmental time-points (P4, P14 and P56; a total of nine brains). Each human-annotated brain section was overlaid with the manually registered Allen developing mouse brain reference atlas for the corresponding age. Ground-truth binary masks for the eight brain regions of interest were generated using unique colour codes for each region in the reference atlas, as explained above.

To increase training data variability, we applied image augmentation by introducing a synthetic variance of a 2° rotation with a range of [−20°, 20°] for each brain section after downsampling by a factor of 16 (6.25%). To demonstrate the generalizability of our registration framework across different ages, as well as different neuronal markers, we trained the network on the complete Nissl and VGAT brains (of P4, P14 and P56 mice) and tested the network on completely different brains that were labelled for GAD1 (across P4, P14 and P56 ages).

Generating the ground-truth dataset for a mouse hippocampus. To demonstrate the performance of the network in segmenting subregions of a brain region on an even finer scale, we generated a ground-truth dataset comprising an ISH adult mouse brain from the Allen Brain OPR that was labelled with Nissl. The coronal brain sections were evenly spaced 100-μm apart, covering the complete brain from the rostral to caudal planes. Only 27 coronal brain sections that contained the hippocampus structure were included in the dataset. Each coronal brain section was overlaid by the matching reference section from the Allen adult mouse brain coronal reference atlas, which was registered onto the right hemisphere of the Nissl brain. This atlas was mirrored and manually registered by the human experts onto the left hemisphere also. Ground-truth binary masks for the four major subregions of the mouse hippocampus (CA1, CA2, CA3 and the DG) are generated using the unique colour codes for each subregion in the reference atlas, as explained above.

To enhance training data variability, we applied image augmentation through synthetic variation: each coronal brain section was horizontally flipped, followed by rotation with a step of 2° in the range [−20°, 20°], thus creating 20 rotated versions of each brain section (as well as the flipped counterpart). Each brain section image in the dataset was downsampled by a factor of 16 (6.25%). To ensure that training and testing datasets were non-overlapping, we trained the network on 800 brain sections and tested it on 280 brain sections, where both sets were mutually exclusive.

Generating the ground-truth dataset for human brain. The ground-truth dataset was generated from the publicly available 18 T1-weighted MRI brain scans provided by the IBSR¹². The scanned brains belonged to 7- to 17-yr-old human subjects, therefore capturing adequate age variability. We used a ‘cropped’ version (158 × 123 × 145) of the original IBSR three-dimensional volumes as adapted in ref. ¹⁵, creating a two-dimensional dataset by drawing brain sections in the horizontal plane. Manually segmented labels were provided for 32 regions in the original dataset, of which we chose a subset of eight prominent subcortical structures: the caudate, pallidum, putamen and thalamus (in both hemispheres). These eight regions were selected to enable direct comparison with other deep learning-based segmentation methods, including DeepLab. Ground-truth binary masks for each region of interest generated from the human-annotated brain volume corresponding to each MR-scanned brain, where each brain region is masked with a unique colour code. A subset of brain sections with empty masks was removed from the dataset. From the complete dataset of 557 brain sections, two-thirds (372 images) were used for training the network, and the remaining one-third (185 sections) were used for testing.

Network architecture. SeBRe was designed by optimization of Mask R-CNN architecture, and constructed by using a convolutional backbone that comprises the first five stages of the very deep ResNet101¹⁸ and FPN¹⁹ architectures (our network architecture is shown in Supplementary Fig. 8). The feature map is processed by an RPN, which applies a convolutional neural network over the feature map in a sliding-window fashion. The RPN segregates and forwards the predicted n potential regions of interest (ROI) from each window to the Mask R-CNN ‘heads’ on the basis of the FPN. The ROI feature maps undergo a critical feature pooling operation by a pyramidal ROIAlign layer, which preserves a pixel-wise correspondence to the original image. Each level of the pyramidal ROIAlign layer is assigned an ROI feature map from the different levels of the FPN backbone (depending on the feature map area), which returns n pooled feature maps, $F_n [7 \times 7]$. Three arms of the FPN perform the core operations of brain region segmentation. The ‘classifier’ and ‘regressor’ heads— inherited from the Faster R-CNN⁵—detect and identify distinct brain regions and compute region-specific bounding boxes. The classifier output layer returns a discrete probability distribution, $[n, 9]$, for nine different object classes (eight brain regions plus the background). The regressor output layer gives the four (x coordinate, y coordinate, width, height) bounding-box regression offsets to be applied for each class, per ROI $[n, (4 \times 8)]$. Figure 2 illustrates the step-by-step procedure of brain region localization and classification performed by the FPN classifier and regressor heads.

A fully convolutional network forms the more recent mask-prediction arm, which returns a binary mask spanning each segmented brain region. The mask arm applies a mask of resolution of $m \times m$ for each class, for each ROI $[n, (8 \times m^2)]$. The output of the backbone architecture and the mask head for a single brain section is shown in Supplementary Fig. 9. The network is trained using a stochastic gradient descent algorithm that minimizes a multi-task loss (L) corresponding to each labelled ROI:

$$L = L_{\text{cls}} + L_{\text{reg}} + L_{\text{mask}}$$

where L_{cls} , L_{reg} and L_{mask} are the region classification, bounding box regression and predicted masks’ loss, respectively, as defined below:

$$L(p_i, \mathbf{q}_i) = \frac{1}{n_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \mu \frac{1}{n_{\text{reg}}} \sum_i L_{\text{reg}}(\mathbf{q}_i, \mathbf{q}_i^*)$$

where p_i is the probability of the i th proposed ROI, or anchor, enclosing an object (anchors with ≥ 0.5 intersection over union overlap with a ground-truth bounding box are considered positive, whereas anchors with < 0.5 intersection over union overlap are considered negative). The term p_i^* denotes whether the anchor is positive ($p_i^* = 1$) or negative ($p_i^* = 0$). The vector \mathbf{q}_i represents the four coordinates that characterize the predicted anchor bounding box, whereas vector \mathbf{q}_i^* represents the coordinates for the ground-truth box that correspond to a positive anchor. The L_{cls} for each anchor is calculated as logarithmic loss for two class labels (object versus non-object). The L_{reg} term is a regression loss function robust to outliers, as defined in ref. ²⁰. The parameters n_{cls} and n_{reg} are the normalizations for classification and regression losses, respectively, weighted by a balancing parameter, μ (ref. ⁵). The L_{mask} term is computed as average cross-entropy loss for per-pixel binary classification, applied to each ROI⁶.

Implementation of SeBRe. The implementation of SeBRe generally follows the original work in ref. ⁶, with limited hyperparameter optimization for the brain section dataset. Training on the brain section dataset is initialized with pretrained weights for the Microsoft COCO dataset¹⁰. Each batch slice consists of a single brain section image per graphics-processing unit; batch normalization layers are inactivated to optimize training for the small effective batch size. Training is performed using an NVIDIA GeForce GTX 970 graphics-processing unit. The training regime comprises two stages. First, the network heads are trained for 6,000 iterations at a learning rate of 0.001 and learning momentum of 0.9. Second, all of the layers are fine tuned for 9,000 iterations at a reduced learning rate of 0.0001. During inference, diverging from the original model, the mask branch is applied to the highest scoring eight detection boxes, proposed by the RPN. The maximum

number of ground-truth instances detected per image is also limited to eight (to avoid erroneous duplicate instances of region-specific masks). Adopting a more stringent approach, the minimum probability threshold for instance detection is raised to 0.9 to improve the accuracy of instance segmentation.

Data availability

The data that support the findings of this study are available from the corresponding author on reasonable request. The publicly available datasets that are used in this study are available at brain-map.org/api/index.html and https://www.nitrc.org/frs/shownotes.php?release_id=2316. The annotated datasets that are used in this study are available at <https://github.com/itsasimiqbal/SeBRe> and <https://bitbucket.org/theolab/>.

Code availability

We provide the code for the SeBRe toolbox at <https://github.com/itsasimiqbal/SeBRe> and <https://bitbucket.org/theolab/>.

Received: 9 July 2018; Accepted: 9 May 2019;

Published online: 10 June 2019

References

- Lein, Ed. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168 (2007).
- Fürth, D. et al. An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* **21**, 139 (2018).
- Niedworok, C. J. et al. aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data. *Nat. Commun.* **7**, 11879 (2016).
- Jarrett, K., Kavukcuoglu, K. & LeCun, Y. What is the best multi-stage architecture for object recognition? In *IEEE 12th International Conference on Computer Vision* 2146–2153 (IEEE, 2009).
- Ren, S., He, K., Girshick, R. & Sun, J. Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems* 91–99 (NIPS, 2015).
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. In *Proc. IEEE International Conference on Computer Vision* 2961–2969 (IEEE, 2017).
- de Vos, B. D., Berendsen, F. F., Viergever, M. A., Staring, M. & Isgum, I. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 204–212 (Springer, 2017).
- Li, H. & Fan, Y. Non-rigid image registration using fully convolutional networks with deep self-supervision. Preprint at <https://arxiv.org/abs/1709.00799> (2017).
- Milletari, F. et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput. Vis. Image Underst.* **164**, 92–102 (2017).
- Lin, T.-Y. et al. Microsoft COCO: common objects in context. In *European Conference on Computer Vision* (Springer, 2014).
- Allen Brain Atlas API (Allen Institute for Brain Science, 2015); <https://brain-map.org/api/index.html>
- Rohlfing, T. Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Trans. Med. Imaging* **31**, 153–163 (2012).
- Klein, S. et al. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).
- Kutten, K. S. et al. A large deformation diffeomorphic approach to registration of CLARITY images via mutual information. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (Springer, 2017).
- Shakeri, Mahsa, et al. Sub-cortical brain structure segmentation using F-CNN’s. In *IEEE 13th International Symposium on Biomedical Imaging* (IEEE, 2016).
- Mehta, R., Majumdar, A. & Sivaswamy, J. BrainSegNet: a convolutional neural network architecture for automated segmentation of human brain structures. *J. Med. Imaging* **4**, 024003 (2017).
- Roy, A. G., Conjeti, S., Navab, N. & Wachinger, C. & Alzheimer’s Disease Neuroimaging Initiative. QuickNAT: a fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage* **186**, 713–727 (2019).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
- Lin, T.-Y. et al. Feature pyramid networks for object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017).
- Girshick, R. Fast R-CNN. In *Proc. IEEE International Conference on Computer Vision* 1440–1448 (IEEE, 2015).

Acknowledgements

This work was supported by a grant from the European Research Council (ERC, 679175, T.K.).

Author contributions

A.I., R.K. and T.K. conceptualized the study and wrote the paper. A.I and R.K. developed the SeBRe method and performed the quantitative comparison with other registration and segmentation methods.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42256-019-0058-8>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019