



Semantic segmentation of cerebrospinal fluid and brain volume with a convolutional neural network in pediatric hydrocephalus—transfer learning from existing algorithms

Florian Grimm¹ · Florian Edl¹ · Susanne R. Kerscher^{1,2} · Kay Nieselt³ · Isabel Gugel¹ · Martin U. Schuhmann^{1,2}

Received: 30 March 2020 / Accepted: 1 June 2020 / Published online: 25 June 2020
© The Author(s) 2020

Abstract

Background For the segmentation of medical imaging data, a multitude of precise but very specific algorithms exist. In previous studies, we investigated the possibility of segmenting MRI data to determine cerebrospinal fluid and brain volume using a classical machine learning algorithm. It demonstrated good clinical usability and a very accurate correlation of the volumes to the single area determination in a reproducible axial layer. This study aims to investigate whether these established segmentation algorithms can be transferred to new, more generalizable deep learning algorithms employing an extended transfer learning procedure and whether medically meaningful segmentation is possible.

Methods Ninety-five routinely performed true FISP MRI sequences were retrospectively analyzed in 43 patients with pediatric hydrocephalus. Using a freely available and clinically established segmentation algorithm based on a hidden Markov random field model, four classes of segmentation (brain, cerebrospinal fluid (CSF), background, and tissue) were generated. Fifty-nine randomly selected data sets (10,432 slices) were used as a training data set. Images were augmented for contrast, brightness, and random left/right and X/Y translation. A convolutional neural network (CNN) for semantic image segmentation composed of an encoder and corresponding decoder subnetwork was set up. The network was pre-initialized with layers and weights from a pre-trained VGG 16 model. Following the network was trained with the labeled image data set. A validation data set of 18 scans (3289 slices) was used to monitor the performance as the deep CNN trained. The classification results were tested on 18 randomly allocated labeled data sets (3319 slices) and on a T2-weighted BrainWeb data set with known ground truth.

Results The segmentation of clinical test data provided reliable results (global accuracy 0.90, Dice coefficient 0.86), while the CNN segmentation of data from the BrainWeb data set showed comparable results (global accuracy 0.89, Dice coefficient 0.84). The segmentation of the BrainWeb data set with the classical FAST algorithm produced consistent findings (global accuracy 0.90, Dice coefficient 0.87). Likewise, the area development of brain and CSF in the long-term clinical course of three patients was presented.

Conclusion Using the presented methods, we showed that conventional segmentation algorithms can be transferred to new advances in deep learning with comparable accuracy, generating a large number of training data sets with relatively little effort. A clinically meaningful segmentation possibility was demonstrated.

Keywords Deep learning · Image segmentation · CSF volume · Brain volume · Pediatric hydrocephalus · Area determination

This article is part of the Topical Collection on *Pediatric Neurosurgery*

✉ Florian Grimm
florian.grimm@med.uni-tuebingen.de

¹ Department of Neurosurgery, University Hospital Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

² Division of Pediatric Neurosurgery, University Hospital Tübingen, Tübingen, Germany

³ Integrative Transcriptomics, Interfaculty Institute for Biomedical Informatics, University of Tübingen, Tübingen, Germany

Abbreviations

CSF	Cerebrospinal fluid
VP shunt	Ventriculoperitoneal shunt
ETV	Endoscopic third ventriculostomy
MRI	Magnetic resonance imaging
ABrain	Planar area of brain in cm ²
ACSF	Planar area of CSF in cm ²
FSL	Functional Magnetic Resonance Imaging of the Brain Software Library
BET	Brain extraction tool

FAST	FMRIB's Automated Segmentation Tool
true FISP	True fast imaging with steady state precession
SD	Standard deviation
TP	True positive
FN	False negative
FP	False positive
BFScore	Boundary contour matching score
IoU	Intersection-over-union
CNN	Convolutional neural network

Background

The analysis of medical image data sets with the help of deep learning algorithms can be of great benefit for extended patient care and specialized diagnostics. Today, these algorithms can provide a solid foundation for segmenting and categorizing image data of any modalities [14, 22, 31].

In principle, deep-learning procedures are inspired by the anatomy of biologically occurring neural networks. The networks are composed of artificial neurons that are organized in multiple layers and that are connected between consecutive layers. Each neuron receives, weighs, and processes incoming signals and passes them on to further neurons in the next layer. The transmission of stimuli between artificial neurons is modeled mathematically. As in the biological model, the transmission of stimuli is characterized by the input signals (excitation and inhibition) and by the connection strength (weights) to the neurons of the deeper layers [21, 29].

The signal processing starts with the input layer, which receives external signals and which is passed on to an output layer via several hidden layers. In the output layer, the final classification results. The data processing within the network takes place in an increasing abstraction of the input signal. In the case of MRI recognition and classification, the input signal could be one MRI slide. In the first layers, for example, the recognition of surfaces and edges takes place; in the following layers, the compositions of shapes are recognized; and finally, possibly using further (deeper) layers, the input image is classified based on the recognition of individual features [3, 21].

The detailed structure and architecture of artificial neural networks vary greatly, and different procedures have been established depending on the area of application [14]. Today, convolutional neural networks (CNNs) are the most common and established standard for processing image data. These particular subspecies are inspired by the principle of the receptive field in biological signal processing of visual signals. Similar to the biological model, individual neighboring artificial neurons react to overlapping parts of the upstream visual field [9]. CNNs are very memory-efficient and provide comparably robust image recognition.

Compared with biological models, the learning mechanism of artificial neural networks is highly variable and comparatively ineffective. Unlike the biological model, standard networks do not learn during use but have to be trained with separate data in the first step to perform a classification task. Once trained, the network then remains static in further applications but solves classification tasks extremely fast. In most cases, supervised learning takes place when known data is transferred to the neural network. For learning, a mathematical method, backpropagation, is usually used: many pre-segmented or classified images are needed to train a neural network. First, the known input images are sequentially fed into the untrained network and classification is calculated. In comparison with the known basic truth, the classification error is calculated from this. In the following, the weights of the neuron connections are then slowly adjusted based on the error rate to approach the desired classification. In this way, a lot of known training data is required and the data is repeated cyclically and a respective computing time or powerful hardware is needed. The advantage is that the weights are adjusted independently of user interaction until sufficient classification accuracy is achieved [21, 29].

However, the data set must be large enough so that the network does not adapt too much to the training data set. If there is not enough data present during training, the training data set is well recognized, but the sensitivity decreases during the classification of unknown data in later use (overfitting) [34]. Overfitting can also occur if the desired modality, sequence, or organ system do not have a sufficient number of data sets. To increase the data to a certain amount, augmentation is the method of choice, for example by changing the size, rotation, or position of the training image data randomly.

Providing that sufficient amount of data for training is a considerable challenge. There is a huge amount of available medical imaging data in archives; however, this is not classified and therefore not available for training. It is particularly advantageous for segmentation if large amounts of available data are labeled with known ground truth. Each pixel or voxel of the data set indicates which class of organ or tissue it represents [14]. As this classification is typically done manually, it is an extremely time-consuming and expensive method. Therefore, already publicly available, pre-segmented training data sets are often used, which are only available in limited numbers and densities.

To overcome this issue and to generate large densities of labeled data, classical established algorithms could be part of the solution. Classical algorithms developed prior to the era of deep learning provide valid segmentation by filtering algorithms or individually adapting machine learning algorithms to address very specific questions, such as the segmentation of the human brain [25]. While these algorithms offer remarkable results for specific issues, they are often not clinically established due to their high technical complexity and the

specificity of the problem. For general brain segmentation, several different algorithms already exist, but all of which fail to have a significant clinical implementation. The possibilities consist of semi-automatic segmentation [27], atlas-based [5] to extended algorithms with classical machine learning procedures [35, 39, 41]. Nevertheless, the implementation of volumetric analysis of the cerebral compartments brain and CSF also characterize an extended technical challenge as most of the algorithms are neither readily available nor free to use.

In our previous studies [10, 11], we addressed the segmentation of MRI data from children with hydrocephalus as a specific issue. Thin-layer true FISP data sets were used here, which in a T2-weighted approach provided reliable information on the anatomy and the amount of CSF [30]. Especially in the case of early childhood hydrocephalus, it is important to assess both the amount of cerebrospinal fluid and the brain volume as this crucial time of brain development determines the outcome [23, 24]. The CSF quantity yields information as to whether or not a therapy such as the implantation of a ventriculoperitoneal (VP) shunt or an endoscopic third ventriculostomy (ETV) may be successful. The development process of the brain volume, on the other hand, is more important for cognitive development and, thus, the outcome of the patient. Therapy of childhood hydrocephalus and its assessment should, therefore, be directed towards influencing the best neurocognitive outcome and thus should not only be assessed on ventricular size or CSF volume but also with information on brain volume.

For this reason, we implemented a well-known and widely used algorithm (FAST [42], FSL FMRIB Software Library [15]) for segmenting 3D data sets of hydrocephalic patients. FAST uses a robust segmentation algorithm based on a hidden Markov random field model, taking spatial orientation into account. The algorithm has already been implemented and adjusted into clinical practice and offers a reliable segmentation [25]. We could demonstrate that changes under therapy in brain volume and CSF can be reliably estimated automatically with this algorithm [10]. For this approach, however, the existence of the complete data set is required. For this reason, we investigated in the next step whether the area of CSF and brain on a representative 2D axial slice in the middle of the brain including the Foramen Monro would be sufficient. A very good correlation of volume and area was found so that an estimation of the clinical course is possible based solely on a single 2D layer [11].

Through this previous work [10, 12], a large amount of pre-segmented data was generated using the FAST algorithm. This study investigates whether a pre-initialized CNN (VGG16) can be trained in an extended transfer learning process with the generated segmented data and reliably deliver segmentation results. Furthermore, this paper examines whether the algorithm is capable of producing suitable segmentation results with known data. To do this, the BrainWeb

data set [20] containing an artificially generated T2 data set with known ground truth is used. In the final step, it is evaluated whether the course of therapy can be assessed in clinical examples, as in the preliminary work on the segmentation of a single layer.

Methods

Study cohort

Ninety-five routinely performed true FISP MRI sequences (1 mm isovoxel) were retrospectively analyzed in 47 patients with pediatric hydrocephalus (male $n = 24$, mean 5.8 ± 5.4 years, posthemorrhagic hydrocephalus $n = 14$, obstructive hydrocephalus $n = 30$, postmeningitic hydrocephalus $n = 1$, external hydrocephalus $n = 2$). Postoperative imaging was included of $n = 20$ patients following a ventriculoperitoneal shunt (VP shunt) and $n = 12$ patients after endoscopic third ventriculostomy (ETV).

Ground image segmentation

As performed in the previous studies [10, 12], a total of 95 routinely performed MRIs were evaluated using the freely available FMRIB Software Library (FSL). After preprocessing, the 3D data sets were fed into an automated script-based processing pipeline, consisting of the following steps: The first step was the masking of the inner skull compartments with the Brain Extraction Tool (BET) [33]. Subsequently, a 2-class segmentation into brain matter and CSF was carried out with FAST [42] with the result of 3-dimensional masks for the individual compartments. The segmentation of the remaining classes for tissue and background was performed using a threshold value (initially 30 units). If necessary, the threshold value was adjusted manually.

Each data set was visually inspected after segmentation to ensure a proper segmentation.

Training and test data sets

Data analysis was performed with Matlab Deep Learning Toolbox (MATLAB, (2019), version 9.5.0 (R2019b), Natick, Massachusetts: The MathWorks Inc.)

Axial, anterior-posterior oriented sections were generated from the 3D MRT files and stored as geometry corrected image files with a 1-mm voxel resolution (256×180). With the previous segmentation, appropriately labeled image data were created with four classes (CSF, brain, tissue, background). Since some of the structural data sets also included neck tissue, basal, or head incisions without relevant tissue apically, only axial layers that showed brain tissue in the segmentation were selected. The complete data set was randomly

split case wise so that 60% of the images (10,432 slices) were used as training data, 20% as validation data (3319 slices), and 20% test data (3289 slices). Since contrast and brightness values for MRIs differ significantly, image augmentation for these values was performed. Additionally, to the native training image, four images with the combinations of enhanced (+100%) or decreased (−50%) contrast level and enhanced (+30%) or decreased (−30%) brightness level were created. Additionally, random left/right translation and random X/Y translation of ±10 pixels were used for data augmentation.

As the data set is pre-segmented data by the FAST algorithm with residual uncertainty, an additional test data set with known ground truth was created. For this purpose, the T2-weighted data set of the BrainWeb data set [20] (slice thickness 1 mm, noise 3%, intensity non-uniformity 20%) was used. The labels were taken from the available Anatomical Model of Normal Brain according to the above requirements and a corresponding labeled data set was generated. This resulted in an additional 111 labeled test images with known ground truth.

Training of the segmentation network

A convolutional neural network (CNN) for semantic image segmentation composed of an encoder and corresponding decoder subnetwork was set up [2]. The network was pre-initialized with layers and weights from a pre-trained VGG 16 model [32].

The network used a pixel classification layer to predict the categorical label for every pixel in the input images. Class frequency of CSF (8.6%), brain (22.1%), tissue (14.3%), and background (55.0%) was obtained. Since the class “CSF” was underrepresented in the training data, a class weighting was carried out to balance classes.

A stochastic gradient descent with momentum (0.9) optimizer was used and a regularization term for the weights to the loss function was added with a weight decay of 0.0005. Cross-entropy was used as a loss function for optimizing the classification model. The initial learning rate was set to 0.001. Furthermore, the learning rate was reduced by a factor of 0.3 every 10 epochs. The network was tested against the validation data set every epoch to stop training when the validation accuracy converged. This prevented the network from overfitting on the training data set. The training was conducted on a single GPU (NVIDIA GeForce GTX 1060). The validation accuracy converged after 6000 repetitions.

Validation of segmentation

The accuracy of the segmentation of the neural network was evaluated by segmenting the deferred test data and the BrainWeb data set. Accuracy scores and exemplary segmentation results are illustrated for both groups.

For comparability of the segmentation results, the following scores were calculated:

Accuracy as the ratio of correctly classified pixels to the total number of pixels.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{TP—true positive, FN—false negative}$$

The boundary contour matching score (Mean BFScore) indicates how well the predicted boundary of each class matches the true boundary, defined as the harmonic mean of precision (Pc) and recall (Rc).

$$\text{Mean BFScore} = \frac{2 \text{Pc Rc}}{\text{Pc} + \text{Rc}}$$

The Intersection over Union (IoU, Jaccard similarity coefficient) indicating the amount of overlap per class.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad \text{FP—false positive}$$

And the commonly used Sørensen-Dice similarity coefficient.

$$\text{Dice coefficient} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

Additionally, confusion matrices were computed to illustrate the true and predicted classes for both data sets.

Results

Classification results test data

The classification results of the trained CNN were validated in 5046 randomly allocated test images (30% of the total data set). Table 1 and Fig. 1 give a detailed overview of the classification results as well as for the four classes’ brain, CSF, tissue, and background. Figure 2 shows exemplary segmentation results of individual patients with FAST and CNN.

Table 1 Classification results for global classification and each class separately of the clinical test data set. Accuracy, IoU (intersection-over-union), mean BFScore (boundary contour matching score), and Dice coefficient are reported for each class

	Accuracy	IoU	Mean BFScore	Dice coefficient
Global	0.90	0.74	0.83	0.86
CSF	0.86	0.68	0.70	0.81
Brain	0.85	0.79	0.79	0.88
Tissue	0.81	0.59	0.74	0.74
Background	0.95	0.94	0.84	0.97

Fig. 1 Confusion matrix of the segmentation result of the clinical test data. The columns represent the predicted class and the rows represent the true class. Data presented in % of classified pixels



CNN segmentation results of the BrainWeb data set

Additional validation of the segmentation was performed using the BrainWeb data set with known ground truth. This data set is a simulated T2-weighted data set. The segmentation results are shown in Table 2 and the confusion matrix in Fig. 3. Single layers of the BrainWeb data set with the segmentation results are shown in Fig. 5.

FAST segmentation results of the BrainWeb data set

For better comparability, the segmentation results of the FAST algorithm are also shown using the BrainWeb data. Table 3 and Fig. 4 display the validation data, while the segmentation of the individual layers is shown in Fig. 5.

Therapy course examples

To demonstrate the clinical viability of the segmentation results, ACSF and ABrain are presented in selected childhood hydrocephalus patients with varying clinical courses. Figure 6 exhibits the underlying MRI slides at the level of the foramen of Monro as well as the area segmentation results of FAST and CNN.

Discussion

In this work, we described how a classical, established machine learning algorithm can be conveyed to the modern method of segmentation using a pre-trained CNN in a double transfer learning algorithm. With existing specialized classical algorithms, it is relatively easy to create adequate training data

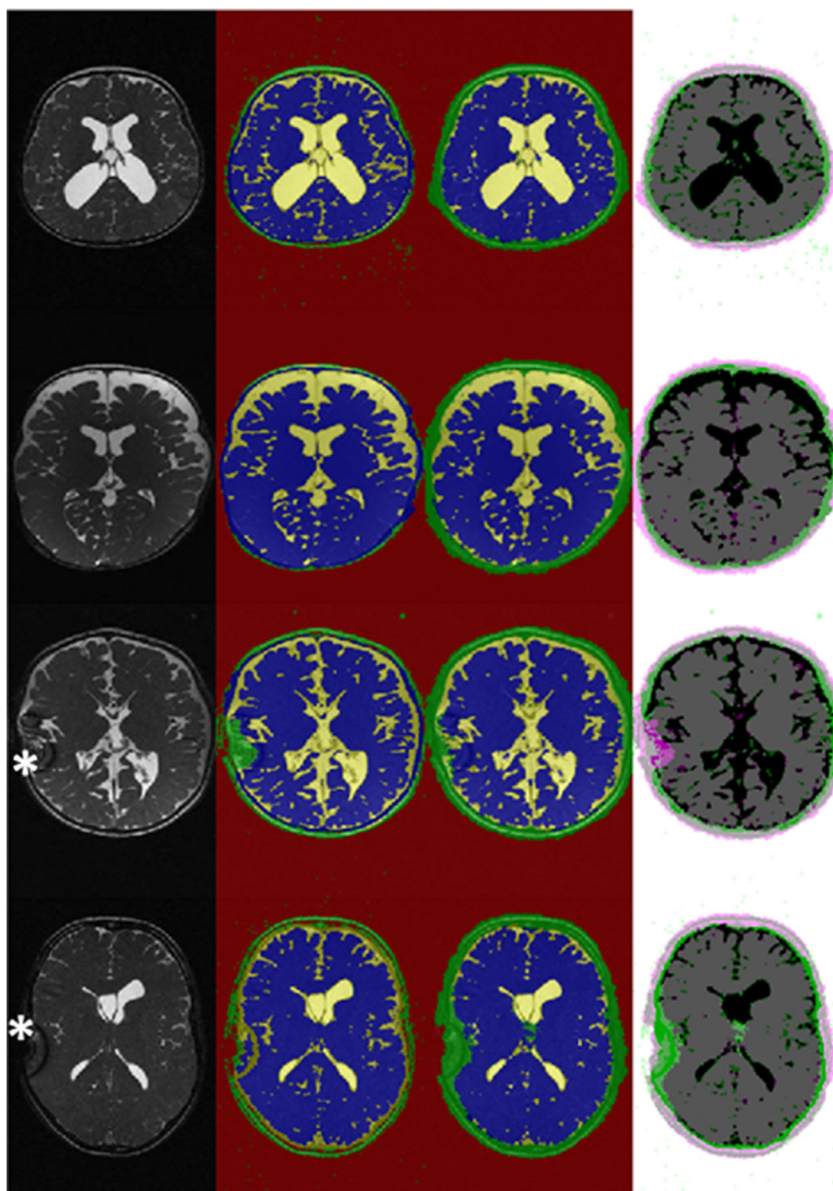
sets with sufficiently accurate labeling. Moreover, it is possible to train a CNN with this large data set. The segmentation by the CNN provides valid results and is comparable with the segmentation results of the original algorithm. This can ensure valid segmentation with clinical data that is consistent with the results of the classical algorithm.

Ground truth image segmentation to generate training data

The segmentation performance of the established FAST algorithm was transferred to our network by using its previously generated training data. Considering this ground truth in terms of reliability, the used FAST algorithm within the applied FSL-Toolbox [15] provided accurate segmentation results. FAST uses a robust segmentation algorithm based on a hidden Markov random field model, taking into account the spatial orientation of every voxel. The advantage of this algorithm is that segmentation can take place unaffected by the large anatomical variability often found in childhood hydrocephalus. Methods using atlas-based segmentation reach their limits in the case of severe anatomical changes, as the sometimes grotesquely altered ventricular cavities are often located far away from the conventional probability space. The FAST algorithm was evaluated in our previous studies in pediatric hydrocephalus. Each segmentation performed was reviewed and verified by medical professionals. An accurate segmentation for CSF and brain matter was found, and changes under therapy could be reliably assessed [10, 12].

In the MRBrainS Challenge [25], an external validation of the FAST algorithm took place. Three freely available

Fig. 2 Segmentation examples of clinical test data set. From left to right: original T2-weighted true FISP images, ground truth segmentation (FAST, CSF yellow, brain blue, tissue green, background red), segmentation result of CNN, differences of segmentation (deviant classes in green and pink, concordant classes greyscale). From top to bottom: 1-year-old toddler with posthemorrhagic hydrocephalus, preoperative imaging; 9-month-old toddler with occlusive hydrocephalus, preoperative imaging; 3-year-old girl with occlusive hydrocephalus, control imaging 18 months postimplantation of a gravity compensated VP shunt; 12-year-old boy with occlusive hydrocephalus, 12 months after post implantation of a gravity compensated VP shunt. Particularly noteworthy are the susceptibility artifacts of the shunt valve in the patients of the last two rows (*) resulting in a false classification of FAST and CNN in this area



segmentation toolboxes were evaluated (SPM 12 [25], Freesurfer [6], and FAST). FAST reached comparable good Dice coefficients of 93% for brain and 70% for CSF. The segmentation results of the MRBrainS Challenge are not

Table 2 Classification results of the CNN for global classification and each class separately of the BrainWeb data set

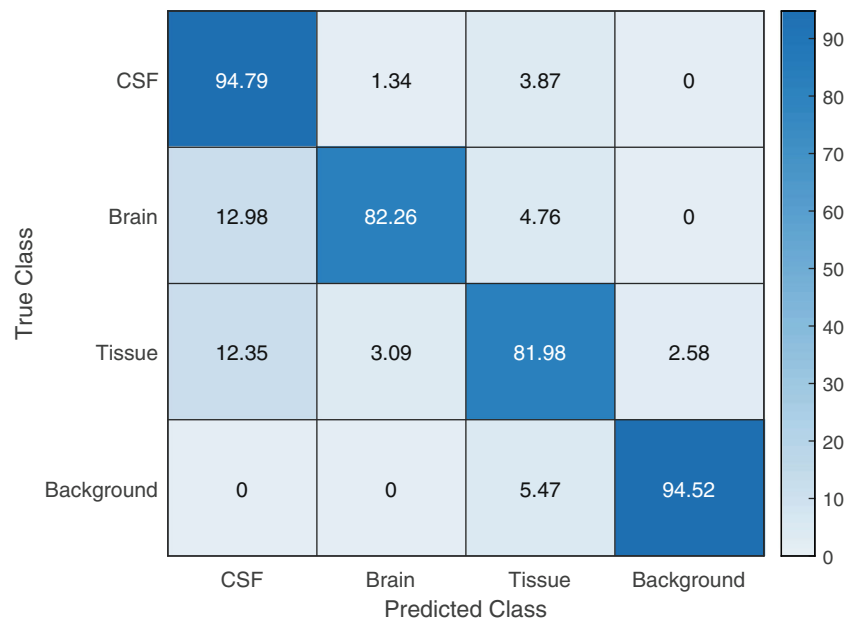
	Accuracy	IoU	Mean BFScore	Dice coefficient
Global	0.89	0.72	0.76	0.84
CSF	0.95	0.44	0.79	0.61
Brain	0.82	0.80	0.77	0.89
Tissue	0.82	0.71	0.57	0.83
Background	0.95	0.93	0.91	0.96

directly commensurate, as the data sets used were T1, IR, and FLAIR sequences.

Due to the higher CSF contrast in the used modeled T2-sequence of the BrainWeb data set for internal validation, the FAST algorithm even performed better in our study with 80% Dice coefficient for CSF and comparing 92% for brain.

It appears therefore advantageous and appropriate to use existing algorithms to generate the labeled training data sets. The accuracy of the segmentation is clinically significant, and large amounts of data can be generated in a relatively short time. However, if the data must first be segmented manually, the costs and time involved are very considerable and less training data can be generated, leading to poorer segmentation results. Nevertheless, one can assume a very accurate ground truth by manual segmentation.

Fig. 3 Confusion matrix of the CNN segmentation result of the BrainWeb data set. Columns represent the predicted class and rows the true class. Data presented in % of classified pixels



Segmentation results of the CNN

As a comparable study in literature, Han et al. used manual segmentation of 600 data sets to determine CSF without brain in early childhood hydrocephalus [13]. The algorithm in this study achieved a Dice coefficient for CSF of 88%. In another recent study by Klimont et al., the segmentation was performed with a U-net using 85 CT data sets in pediatric hydrocephalus. The segmentation was likewise performed manually. Here, CSF without brain volume was considered. In this study, a Dice coefficient CSF of 95% could be achieved [18].

Overall, our segmentation results of the CNN on the BrainWeb data set were very accurate; however, the Dice coefficient cut off was marginally inferior due to slightly higher false negative values in the tissue area, particularly in the lower layers. As these areas are not included in the classification of the FAST algorithm due to the previous extraction of the brain by BET, the previous masking of the inner skull spaces appears to be advantageous here. Furthermore, in the classical T2 sequences, the tissue fat signal was overrepresented when compared with the underlying training data with true

FISP sequences. The used true FISP sequences in our study are comparatively fast and offer high spatial resolution and excellent CSF contrast, making them therefore optimal for imaging pathologies of the CSF system [30].

The validation of the deferred clinical data sets shows optimal results with an overall accuracy of 90% and Dice coefficient for brain of 88% and CSF of 81%. The Mandell group achieved comparable values for brain (94%) and weaker results for CSF (57–67%) in the segmentation of pediatric hydrocephalus patients [24]. Han et al. achieved a Dice coefficient for CSF of 88%; a brain segmentation was not performed here [13]. With our method, clinically relevant areas or changes in volume can be detected. Figure 6 illustrates by the patient examples that the calculated areas can map an accurate course of CSF and brain mass making it possible to visualize clinically relevant changes.

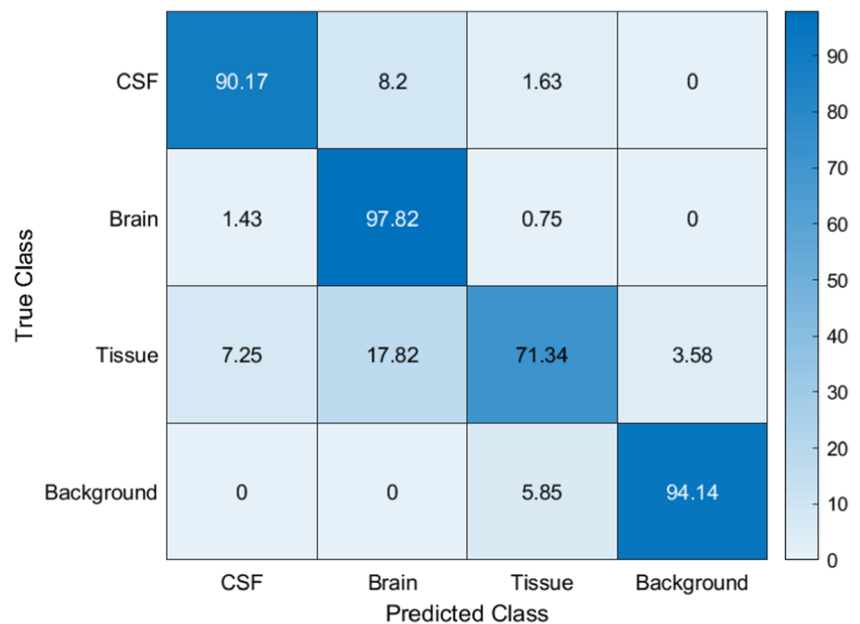
Network structure and transfer learning

In regard to the network structure and training data set in our study, we followed a multiple transfer learning approach. Transfer learning describes the ability of a system to use the knowledge acquired in a previous task for a new classification [4]. The study used a pre-initialized CNN designed for visual recognition (VGG 16), upon which the original weightings of the network were adopted and then fine-tuned with the training data. The closer the classification tasks of the networks are to each other, the better the classification accuracy can be achieved through transfer learning. Even if the original tasks are further apart, it could be shown that maintaining the weightings was more effective than randomizing the initialization of the network [34, 40]. This is similar to the studies with prenatal images by Wu et al. [37], who used the original

Table 3 Classification results of the FAST algorithm for global classification and each class separately of the BrainWeb data set

	Accuracy	IoU	Mean BFScore	Dice coefficient
Global	0.90	0.77	0.73	0.87
CSF	0.90	0.67	0.88	0.80
Brain	0.98	0.85	0.73	0.92
Tissue	0.71	0.65	0.55	0.79
Background	0.94	0.92	0.76	0.96

Fig. 4 Confusion matrix of the FAST algorithm segmentation result of the BrainWeb data set. Columns represent the predicted class and rows the true class. Data presented in % of classified pixels



weights from a general network (VGG 16) and then fine-tuned on the training data.

3D volumetry and 2D image segmentation

With the CNN segmentation algorithm described, both a two-dimensional (area determination) and a three-dimensional (volumetry) can be performed. As demonstrated in our previous work [10] and by other authors [23, 26, 36], 3D volumetry is an advanced method for quantitative and precise monitoring of the course of therapy in childhood hydrocephalus. Furthermore, only quantitative data on changes in brain and CSF provide an accurate foundation for comparing different treatment modalities. Previous volumetry studies have shown that the neurocognitive outcome depends primarily on the development of brain mass. In particular, the CSF volume may be less significant than the course of the brain volume when examining cognitive, motor, and speech development. It is postulated that these neurocognitive changes are caused by white matter lesions due to increased brain pressure and ventricular dilatation [7]. The volume effect of the brain mass would be a rather indirect sign to estimate these alterations. Therefore, the approach of determining the three-dimensional CSF and brain volume is preferable and a more objective basis for a comprehensive treatment evaluation [23, 24].

Howbeit, the volumetry method requires a complete 3D data set and significant computing time and power. The computing effort can be reduced through automation as previously shown [10]. Furthermore, patients with congenital hydrocephalus often require several MRI examinations to evaluate the course of their disease, especially concerning any necessary interventions or to check the functionality of an ETV or a VP shunt. Although MRI scans have no radiation exposure and

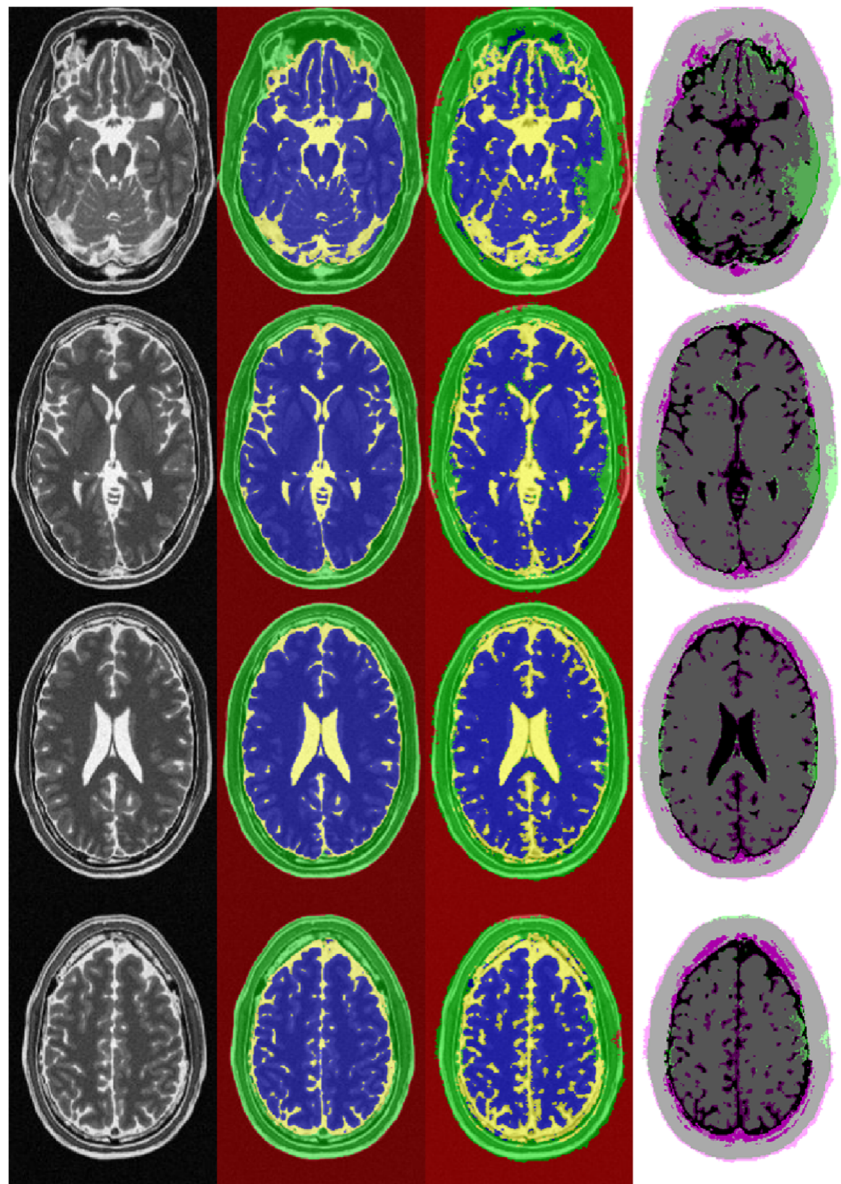
are therefore preferred to CT scans for follow-up checks, significantly longer examination time is required. This poses a problem for younger children in particular as they often cannot lie still for the entire duration of the examination and the quality of the MRI images is therefore significantly limited by movement artifacts. For this reason, sedation during the examination is almost always necessary in children [28].

The proposed method of planar area determination eliminates some of these disadvantages. There is no need for a complete thin-sliced 3D data set as an artifact-free slice in the plane of the foramen of Monro would be sufficient. The area result correlates excellently with the volumes of the compartments, and faster non-3D sequences can be used as a basis for an area-based evaluation. For a follow-up examination, it would be conceivable to perform only a short sequence including the proposed reference plane of high quality. Due to the excellent correlation, a reliable prediction of CSF and brain volume seems to be possible from the area data [11].

Limitations of the study

The clinically reliable calculation of the volume of an area determination in 2D resulted in the use of a 2D CNN as the basic network structure. Due to the excellent initial data situation with the available high-resolution 3D data sets and the already existing exact segmentation, a 3D CNN for the segmentation would also have been conceivable. Based on the higher spatial information in 3D, a greater accuracy of the segmentation could have been expected. While the benefits of 3D CNNs have been demonstrated in preliminary studies [8, 17], using a 3D model would have significantly influenced the clinical applicability of the method. As with the FAST algorithm, the complete 3D data set would have had to be

Fig. 5 CNN and FAST segmentation examples of BrainWeb data set. From left to right: original T2-weighted images, ground truth segmentation (FAST, CSF yellow, brain blue, tissue green, background red), segmentation result of CNN, differences of segmentation (deviant classes in green and pink, concordant classes grey). From top to bottom: ascending slices of the same data set



available artifact-free and in digital form and the processing of the data would have required increased technical and time expenditure. The pre-trained 2D CNN, which was developed within the scope of this study, can be easily transferred to e.g. a smartphone app, which could be used for volume estimation via screenshot in the clinical routine.

A general limitation of the comparability of study results is that only a certain number of validated test data sets are available for different modalities. For example, there are pre-segmented test data sets of the MRBrainS Challenge [25] in the modalities FLAIR (fluid-attenuated inversion recovery), IR (inversion recovery), and T1, but not for T2-weighted images. Pre-segmented true FISP data sets are not available at all to our knowledge. The used BrainWeb data set consists of artificially simulated MRI data and was originally

designed to validate various segmentation algorithms as a known basic truth [20]. It exhibits similarity of image morphological structures to in vivo acquired MRI data [20, 38] and is used frequently in studies for verification purposes [1, 16, 19].

With the proposed method, misclassifications were found in the area of susceptibility artifacts of the shunt implant since there was often complete signal annihilation. One disadvantage of our method is that systematic errors made by the underlying segmentation algorithm are learned by the CNN. An example of this can be seen in Fig. 2. Here, FAST and later CNN consistently incorrectly evaluate the susceptibility artifacts of the shunt system as tissue area. Of course, extinction artifacts pose a challenge to all algorithms, and possibilities of recognizing these artifacts and extrapolating them must be investigated in the future.

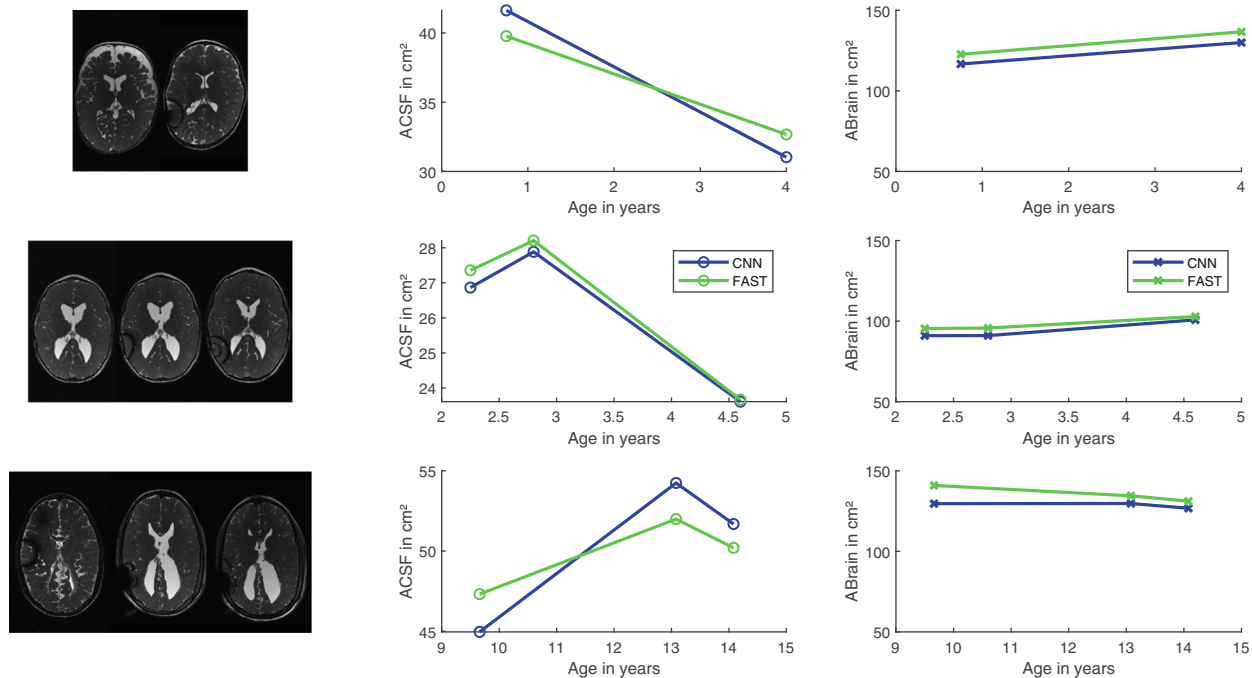


Fig. 6 Exemplary clinical follow-up data. From left to right: original MRI-data on the level of the foramen of Monro, diagrams for ACSF and ABrain determined by CNN and FAST, diagrams for the estimated areas by CNN and FAST on the left. Each data point is represented by the MRI-images on the left. Top: implantation of a gravity-compensated VP shunt in an infant with occlusive hydrocephalus at the age of 9 months. The regular decrease in CSF volume and an increase in brain volume are reflected in area estimates in control imaging at the age of 4 years. Middle: primary VP shunt implantation in a 2-year-old boy with post-hemorrhagic hydrocephalus, left is the pre-op scan. A slight increase of ventricular size was observed in the further course (second MRI from left). The adjustable valve unit was readjusted. In the subsequent MRI

controls, a further decrease of CSF and an increase in brain mass occurred. Bottom: postoperative course imaging of a girl with occlusive hydrocephalus after early childhood implantation of a VP shunt. After an initial complication-free course with regular CSF drainage (initial picture at the age of 9 years on the left), there was an increasing underdrainage symptomatology with fatigue and headache. The subsequent control MRI showed a strong increase in ventricular space. Afterward, at the age of 13 years, operative shunt revision was performed with the immediate postoperative improvement of the symptoms. The control MRT examination showed a comparative decrease in the CSF spaces after revision

Conclusion

Transfer learning from established classical segmentation algorithms via deep learning techniques is a promising method to establish a reliable segmentation. Large data sets can be created rather quickly, which are needed to achieve reliable segmentation accuracy through network training. With the created segmentation algorithm, MRI data can be segmented with reliable accuracy and the clinical course of brain and CSF development can be traced.

The presented area determination on a single layer allows an exact estimation of the volume of brain and CSF. These deep learning algorithms can be integrated into other applications e.g. via tensor flow. It would be logical to integrate the algorithms into a mobile phone app to allow broad access to this method. Thus, quantitative therapy monitoring in pediatric hydrocephalus therapy could be performed in daily practice and serve as a precise basis for future analysis and comparison of treatment options.

For a further generalization of the method, additional training of the network using other common imaging modalities is required. To expand this method to adult hydrocephalus, segmentation of CT data is necessary. For future research, it must be shown that an even more precise segmentation outside the intracranial spaces can be achieved, e.g. by additional probabilistic masks. Furthermore, it must be demonstrated how susceptibility artifacts can be extrapolated e.g. by mirroring the opposite side or by further deep-learning procedures.

Acknowledgments We thank Heather Smith for her help with proofreading.

Funding Information Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (University Hospital Tübingen,

Germany) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study, formal consent was not required.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arce-Santana ER, Mejia-Rodriguez AR, Martinez-Pena E, Alba A, Mendez M, Scalco E, Mastropietro A, Rizzo G (2019) A new Probabilistic Active Contour region-based method for multiclass medical image segmentation. *Med Biol Eng Comput* 57:565–576. <https://doi.org/10.1007/s11517-018-1896-y>
- Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39:2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35:1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- Chuen-Kai S, Chung-Hisang C, Chun-Nan C, Meng-Hsi W, Chang EY (2015) Transfer representation learning for medical image analysis. *Conf Proc IEEE Eng Med Biol Soc* 2015:711–714. <https://doi.org/10.1109/EMBC.2015.7318461>
- Ciofalo C, Barillot C (2009) Atlas-based segmentation of 3D cerebral structures with competitive level sets and fuzzy control. *Med Image Anal* 13:456–470. <https://doi.org/10.1016/j.media.2009.02.008>
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I Segmentation and surface reconstruction. *Neuroimage* 9:179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Del Bigio MR (2010) Neuropathology and structural changes in hydrocephalus. *Dev Disabil Res Rev* 16:16–22. <https://doi.org/10.1002/ddr.94>
- Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, Heng PA (2017) 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal* 41:40–54. <https://doi.org/10.1016/j.media.2017.05.001>
- Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36:193–202. <https://doi.org/10.1007/BF00344251>
- Grimm F, Edl F, Gugel I, Kerscher SR, Bender B, Schuhmann MU (2019) Automatic volumetry of cerebrospinal fluid and brain volume in severe paediatric hydrocephalus, implementation and clinical course after intervention. *Acta Neurochir*. <https://doi.org/10.1007/s00701-019-04143-5>
- Grimm F, Edl F, Gugel I, Kerscher SR, Schuhmann MU (2019) Planar single plane area determination is a viable substitute for total volumetry of CSF and brain in childhood hydrocephalus. *Acta Neurochir (Wien)* accepted
- Grimm F, Edl F, Gugel I, Kerscher SR, Schuhmann MU (2019) Planar single plane area determination is a viable substitute for total volumetry of CSF and brain in childhood hydrocephalus. *Acta Neurochir*. <https://doi.org/10.1007/s00701-019-04160-4>
- Han M, Quon J, Kim L, Shpanskaya K, Lee E, Kestle J, Lober R, Taylor M, Ramaswamy V, Edwards M, Yeom K (2019) One hundred years of innovation: automatic detection of brain ventricular volume using deep learning in a large-scale multi-institutional study (P5.6-022). *Neurology* 92:P5.6–P022
- Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging* 32:582–596. <https://doi.org/10.1007/s10278-019-00227-x>
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) Fsl. *Neuroimage* 62:782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Ji Z, Huang Y, Sun Q, Cao G, Zheng Y (2017) A rough set bounded spatially constrained asymmetric Gaussian mixture model for image segmentation. *PLoS One* 12:e0168449. <https://doi.org/10.1371/journal.pone.0168449>
- Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 36:61–78. <https://doi.org/10.1016/j.media.2016.10.004>
- Klimont M, Flieger M, Rzeszutek J, Stachera J, Zakrzewska A, Jonczyk-Potoczna K (2019) Automated ventricular system segmentation in paediatric patients treated for hydrocephalus using deep learning methods. *Biomed Res Int* 2019:3059170. <https://doi.org/10.1155/2019/3059170>
- Kong Y, Chen X, Wu J, Zhang P, Chen Y, Shu H (2018) Automatic brain tissue segmentation based on graph filter. *BMC Med Imaging* 18:9. <https://doi.org/10.1186/s12880-018-0252-x>
- Kwan RK, Evans AC, Pike GB (1999) MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans Med Imaging* 18:1085–1097. <https://doi.org/10.1109/42.816072>
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sanchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Mandell JG, Kulkarni AV, Warf BC, Schiff SJ (2015) Volumetric brain analysis in neurosurgery: part 2. Brain and CSF volumes discriminate neurocognitive outcomes in hydrocephalus. *J Neurosurg Pediatr* 15:125–132. <https://doi.org/10.3171/2014.9.PEDS12427>
- Mandell JG, Langelaan JW, Webb AG, Schiff SJ (2015) Volumetric brain analysis in neurosurgery: part 1. Particle filter segmentation of brain and cerebrospinal fluid growth dynamics from MRI and CT images. *J Neurosurg Pediatr* 15:113–124. <https://doi.org/10.3171/2014.9.PEDS12426>
- Mendrik AM, Vincken KL, Kuijff HJ, Breeuwer M, Bouvy WH, de Bresser J, Alansary A, de Bruijne M, Carass A, El-Baz A, Jog A, Katyal R, Khan AR, van der Lijn F, Mahmood Q, Mukherjee R, van Opbroek A, Paneri S, Pereira S, Persson M, Rajchl M, Sarikaya D, Smedby O, Silva CA, Vrooman HA, Vyas S, Wang C, Zhao L, Biessels GJ, Viergever MA (2015) MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Comput Intell Neurosci* 2015:813696. <https://doi.org/10.1155/2015/813696>
- Moeskops P, Benders MJ, Chit SM, Kersbergen KJ, Groenendaal F, de Vries LS, Viergever MA, Isgum I (2015) Automatic segmentation of MR brain images of preterm infants using supervised classification. *Neuroimage* 118:628–641. <https://doi.org/10.1016/j.neuroimage.2015.06.007>

27. Moore DW, Kovanlikaya I, Heier LA, Raj A, Huang C, Chu KW, Relkin NR (2012) A pilot study of quantitative MRI measurements of ventricular volume and cortical atrophy for the differential diagnosis of normal pressure hydrocephalus. *Neurol Res Int* 2012: 718150. <https://doi.org/10.1155/2012/718150>
28. O'Neill BR, Pruthi S, Bains H, Robison R, Weir K, Ojemann J, Ellenbogen R, Avellino A, Browd SR (2013) Rapid sequence magnetic resonance imaging in the assessment of children with hydrocephalus. *World Neurosurg* 80:e307–e312. <https://doi.org/10.1016/j.wneu.2012.10.066>
29. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
30. Schmitz B, Hagen T, Reith W (2003) Three-dimensional true FISP for high-resolution imaging of the whole brain. *Eur Radiol* 13: 1577–1582. <https://doi.org/10.1007/s00330-003-1846-3>
31. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. *Annu Rev Biomed Eng* 19:221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
32. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556
33. Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp* 17:143–155. <https://doi.org/10.1002/hbm.10062>
34. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35:1299–1312. <https://doi.org/10.1109/TMI.2016.2535302>
35. Vrooman HA, Cocosco CA, van der Lijn F, Stokking R, Ikram MA, Vernooij MW, Breteler MM, Niessen WJ (2007) Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. *Neuroimage* 37:71–81. <https://doi.org/10.1016/j.neuroimage.2007.05.018>
36. Warf B, Ondoma S, Kulkarni A, Donnelly R, Ampeire M, Akona J, Kabachelor CR, Mulondo R, Nsubuga BK (2009) Neurocognitive outcome and ventricular volume in children with myelomeningocele treated for hydrocephalus in Uganda. *J Neurosurg Pediatr* 4:564–570. <https://doi.org/10.3171/2009.7.PEDS09136>
37. Wu L, Xin Y, Li S, Wang T, Heng P, Ni D Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 18–21 April 2017 2017. pp 663–666. <https://doi.org/10.1109/ISBI.2017.7950607>
38. Xu Y, Raj A, Victor JD (2019) Systematic differences between perceptually relevant image statistics of brain MRI and natural images. *Front Neuroinform* 13:46. <https://doi.org/10.3389/fninf.2019.00046>
39. Yepes-Calderon F, Nelson MD, McComb JG (2018) Automatically measuring brain ventricular volume within PACS using artificial intelligence. *PLoS One* 13:e0193152. <https://doi.org/10.1371/journal.pone.0193152>
40. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? arXiv e-prints
41. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31:1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>
42. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57. <https://doi.org/10.1109/42.906424>

Comments

In this study, the authors build on their experience of classical machine learning algorithms and MRI segmentation in paediatric hydrocephalus. They transfer these algorithms to less specific and more generalisable deep learning ones. They evaluate whether this is successful, and, more importantly, clinically meaningful. The study is based on 95 MRI scans, in 47 patients, including a number of post- VP shunt and ETV patients. 60% of these scans are used to train a convolutional neural network, which is then validated and tested against the remaining 40%. The authors conclude that this method allows accurate segmentation of images into brain, CSF, tissue and background.

They present specific examples and demonstrate that these interpretations are clinically meaningful.

They note problems with artefacts such as shunts, which require further development.

As machine learning becomes more relevant to medical, and particularly radiological, practice, this study, as applied to paediatric hydrocephalus, represents a useful addition to the current literature.

This methodology can easily provide accurate volumes of brain and CSF, allowing not only monitoring of clinical trajectory in these patients, who often require multiple scans over short periods of time, but also research into the relationship of brain volume and cognition.

Kristian Aquilina
London, UK

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