Automated Analysis of Corpora Callosa

Mikkel B. Stegmann[1], Rhodri H. Davies[2]

[1]Informatics and Mathematical Modelling – IMM, Technical University of Denmark,
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark
Corresponding author: `mbs@imm.dtu.dk`, `http://www.imm.dtu.dk/∼mbs/`

[2]Division of Imaging Science and Biomedical Engineering – ISBE,
University of Manchester, Manchester, M13 9PT, England

Abstract. This report describes and evaluates the steps needed to perform modern model-based interpretation of the corpus callosum in MRI. The process is discussed from the initial landmark-free contours to full-fledged statistical models based on the Active Appearance Models framework. Topics treated include landmark placement, background modelling and multi-resolution analysis. Preliminary quantitative and qualitative validation in a cross-sectional study show that fully automated analysis and segmentation of the corpus callosum are feasible.

1 Introduction

Many neurological studies indicate that the size and shape of the corpus callosum are related to various human characteristics, dysfunctions et cetera [9]. In short, corpus callosum is the nervous tissue that connects the two cerebral hemispheres of the human brain. The gold standard for such morphometry studies is magnetic resonance imaging, which allows acquisition of accurate images of the anatomy (and function) of the brain. However, doing manual tracings of the corpus callosum is both time-consuming, error-prone and operator dependent. Instead, medical image analysis should aim at replacing this task with automated and efficient methods eliminating all subjectivity. In this report we demonstrate that the generic Active Appearance Models (AAMs) [7, 2] can be adapted to this problem and proceed by performing a quantitative and qualitative assessment of the method. Earlier quantitative segmentation studies of the corpus callosum include the work of Lundervold et al. [9], Brejl and Sonka [1] and van Ginneken et al. [8].

In addition to being a truly generic approach to corpus callosum segmentation, AAMs provide a reference coordinate system for every new example. This makes propagation of reference areas such as the rostrum, genu, truncus, isthmus and splenium a trivial matter. Further, regression analyses against gender, age, motor abilities, lifestyle et cetera can be carried out directly since shape variation and size are both sensibly and compactly encoded.

To supplement to this report, a continuation of our work is presented in [11] using a slightly different version of the data set and evolved versions of the AAM and MDL implementation described below.

2   Data Material

The data material comprises 17 cross-sectional, mid-sagittal magnetic resonance images (MRI) of the brain. The corpus callosum was manually annotated by an expert drawing one closed, landmark-free contour in each image. Figure 1 shows all shapes in Procrustes aligned form. The corresponding MR images are shown in Appendix A.

Fig. 1: 17 corpus callosum annotations (row-wise) and the mean shape (lower right).

3   Methods

3.1   Active Appearance Models

Active Appearance Models are generative models capable of synthesising images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform image interpretation. The modelled object properties are shape and pixel intensities (called *texture*). Training objects are defined by marking up each example image with points of correspondence. Variability is modelled by means of principal component analyses (PCA). Prior to PCA modelling shapes are Procrustes aligned and textures are warped into a shape-free reference frame and sampled. Let there be given P training examples, and let each example be represented by a set of N landmark points and M texture samples. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding means. New instances can be generated by adjusting the PC scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s$ and $\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t$ where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are eigenvectors of the shape and texture dispersions estimated from the training set. To recover any correlation between shape and texture and obtain a combined parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined in a third PCA,

$$\begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}}(\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}}(\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c} = \mathbf{\Phi}_c \mathbf{c}. \tag{1}$$

Here, $\mathbf{W}_s$ is a diagonal matrix weighting pixel distances against intensities.

Synthetic examples, parameterised by $\mathbf{c}$, are generated by $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c}$ and $\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}$ and rendered into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$. Using an iterative updating scheme the model can be fitted to an unseen image. See [2,3] for the details.

3.2   Landmark Placement

Establishing points of correspondence, the so-called *landmarks*, on training examples by hand can not only be time-consuming, tedious and error-prone, but in some cases directly impossible when no good anatomical points are available.

We have tried to manually identify two landmarks on the corpus callosum; one at the rostrum (left "tip") and one at the splenium (right "tip"). Further, 78 semi-landmarks were interpolated along the contour using a uniform arc-length parameterisation.

To supplement this, a method for automatic landmarking in 2D and 3D was also applied. This was the MDL-framework proposed by Davies et al. [5, 6] that – in the spirit of Occam's razor – seeks to minimise the total description length of the resulting PCA shape model. As a starting point, the method requires the shapes to be Procrustes aligned, which was obtained from our manually landmarked shapes. Then landmarks with a minimal description length were calculated for all shapes and the alignment was inverted so that the MDL-based landmarks were given in image coordinates.

3.3 Background Awareness

By modelling the full appearance of objects, AAMs achieve an impressive degree of robustness, e.g. when compared to its predecessor the Active Shape Models (ASMs) [4]. However, one widely recognised advantage of the ASMs is that they also model the appearance of the background variation. In fact, the typical ASM implementation has equal weights on appearance of the object and background. Obviously, this is good when the background is fairly constant or, more importantly, when the object has a rather homogeneous appearance. On the other hand, modelling object appearance only is preferred when the background is highly varying. However, this requires that the object modelled have internal features that can constrain the outer border, e.g. mouth, nose and eyes in face images.

In medical images, organs and structures typically have a very homogenous appearance. Since AAMs only model the object appearance the cost function can easily have perfectly fine minimum inside the object, where both a plausible shape and good texture fit can be obtained. Corpora callosa are no exception to this. One typical fix is to model the convex hull of the shape points. However, this only helps at the concave parts of the modelled contour. Therefore, we propose a simple scheme for modelling background variation by adding normals outwards from the shape, which we call *whiskers*, at all contour points. These are implicitly added during texture sampling and scaled according to the current shape size. Texture samples obtained from whiskers can now be weighted against the normal AAM samples. As in Active Shape Models (ASMs) [4] this unfortunately introduces the need for one more parameter setting, namely the length of the whiskers.

3.4 Multi-Resolution Analysis

To improve both convergence and robustness AAM analysis can be performed at different resolutions. A widely used scheme is to recursively decimate an input image thereby creating a scale pyramid representation. This can be carried out on the training set, thus producing one model for each level of the pyramid.

In our formulation, model optimisations are initialised at the smallest level and propagated toward the full image resolution. Let $\mathbf{c}_i$ denote the converged model configuration at level i. The corresponding shape, $\mathbf{s}_i$ is trivially propagated to level $i + 1 = j$ by $\mathbf{s}_j = 2\mathbf{s}_i$ and the texture is sampled into $\mathbf{t}_j$. Assuming that the optimisation at level i produced a match up to the accuracy given by image resolution, $\mathbf{c}_j$ can now be estimated by projecting $\mathbf{s}_j$ and $\mathbf{t}_j$ into the corresponding combined eigenspace

$$\hat{\mathbf{c}}_j = \boldsymbol{\Phi}_{c,j}^{\mathsf{T}} \begin{bmatrix} \mathbf{W}_{s,j} \boldsymbol{\Phi}_{s,j}^{\mathsf{T}} (\mathbf{s}_j - \bar{\mathbf{s}}_j) \\ \boldsymbol{\Phi}_{t,j}^{\mathsf{T}} (\mathbf{t}_j - \bar{\mathbf{t}}_j) \end{bmatrix} \tag{2}$$

and model optimisation can continue at level j. If $\boldsymbol{\Phi}_{s,i} = \boldsymbol{\Phi}_{s,j} \; \forall \; i, j$ then $\mathbf{b}_{s,i}$ can be used to ease the calculation of $\hat{\mathbf{c}}_j$, but more importantly; $\mathbf{s}_j$ and $\mathbf{t}_j$ would lie in (or at least be very close to) the space spanned by $\boldsymbol{\Phi}_{c,j}$. This means that $\mathbf{t}_j$ can be used directly to calculate the first update of $\hat{\mathbf{c}}_j$.

4 Experimental Results

To assess the applicability of AAMs a set of test scenarios were set up. We started by inspecting the quality of the two different shape models provided by the manual and the MDL annotation. Figure 2 shows the three first principal modes of shape variation for both shape sets. As these are ordered in terms of decreasing variance it might surprise that the first mode in the manual case seems to deliver less shape change than the second. The cause is clearly illustrated in Figure 3, which shows that variation in the manual case almost solely stems from landmark movements *along* the contour. However, in the MDL case deformations are, by design, much closer to being orthogonal to the tangent of the contour.
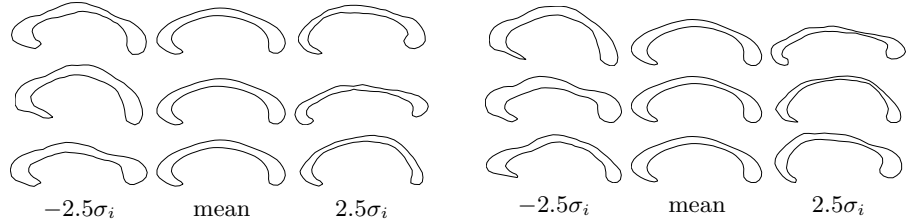

Fig. 2: Shape deformation using the three largest principal modes (top-down). Left: Manual annotation. Right: MDL annotation.
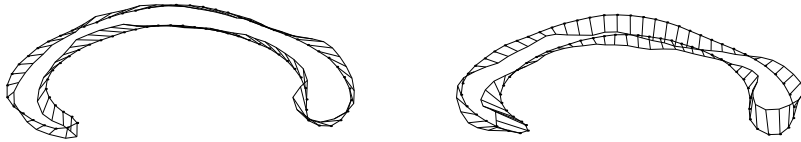
Fig. 3: The first principal shape mode shown as displacements from the mean shape. Left: Manual annotation. Right: MDL annotation.

Table 1: Segmentation Results in Units of Pixels

	Whiskers added	MDL shapes	Scale pyramid	Pt.pt. Mean	Pt.pt. Median	Pt.crv. Mean	Pt.crv. Median	Lower bound (pt.pt.) Leave-all-in	Lower bound (pt.pt.) Shape model
A	-	-	-	3.64	2.70	0.90	0.88	1.61	1.06
B	-	-	Yes	2.54	2.36	0.91	0.93	n/a	n/a
C	-	Yes	-	6.98	2.29	3.85	0.91	1.74	0.93
D	-	Yes	Yes	2.24	1.97	0.94	0.89	n/a	n/a
E	Yes	-	-	2.92	2.33	0.80	0.76	0.72	1.06
F	Yes	-	Yes	2.40	2.12	0.89	0.81	n/a	n/a
G	Yes	Yes	-	3.01	1.93	0.85	0.77	1.51	0.93
H	Yes	Yes	Yes	1.85	1.80	0.86	0.77	n/a	n/a
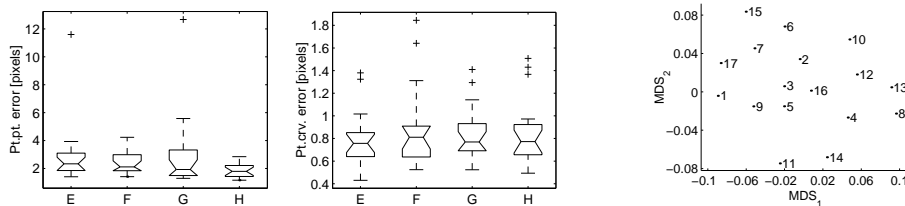
Fig. 4: Left: Boxplots of the segmentation accuracy for models built with whiskers. Right: Multidimensional scaling (MDS) of the MDL-based principal shape scores.

But how much does this affect the segmentation accuracy? Trying to answer this, we also tested the effects of the multi-resolution analysis and the addition of whiskers. Each test was carried out using cross-validation in leave-one-out scheme. Used performance measures were the average distance between i) model and ground truth landmarks (pt.pt.), and ii) model and ground truth contour (pt.crv.). Further, two lower bounds on the segmentation error were calculated. The first is the error obtained from a leave-all-in segmentation test. This means that the optimisation process had knowledge of the image being optimised on and the shape, texture and combined models should – apart from the truncation – span the space of the test example. The second lower bound is obtained from discarding any error introduced from the lack of span from the combined model and the error introduced by an erroneous optimisation. Fitting the independent leave-one-out shape model directly to the ground truth carried this out. All results are shown in Table 1 and a subset is shown as boxplots in Figure 4 (left). All models were automatically initialised using a method we have described in [10]. Texture samples from whiskers influenced the texture model by one third. The convex hull was retained and the whisker length was three times the distance between landmark 1 and 2 on the mean shape sized to mean size.

A qualitative impression of the average segmentation result is given in Figure 5 (middle) and an example of failure is shown in Figure 5 (right). In this case the pt.pt. error was 11.6 pixels but the pt.crv. error was only 1.3 pixels. This stems from the non-symmetry of the latter measure. Hence, the vanilla pt.crv. formulation is clearly misleading in cases with search failures. A simple patch to this is to use $E_{\mathrm{sym.pt.crv.}}(a, b) = \max(\ E_{\mathrm{pt.crv.}}(a, b) + E_{\mathrm{pt.crv.}}(b, a)\ )$ instead.

Fig. 5: Left: Corpus callosum annotation using 80 landmarks. Middle: Average pt.pt. segmentation result (case H). Right: Segmentation failure (case E).

Finally, a multidimensional scaling (MDS) map of the principal shape scores is shown in Figure 4 (right), which gives a planar approximation to the Euclidean distances between shapes in the high-dimensional space of $\mathbf{b}_s$.

5   Discussion and Conclusion

We observe that the addition of the multi-resolution representation, the whiskers and the MDL-based landmarking consistently lead to better models. This was achieved by a combination of eliminating initialisation and search failures and improving the general accuracy. Further, the validity of Occam's razor is confirmed by the last column of Table 1.

To the best of our knowledge, this is the first work documenting the feasibility of applying Active Appearance Models for fully automated shape analysis and segmentation of the corpus callosum. We have evaluated three different improvements, which all lead to better models. The three combined lead to a landmark error median of 1.80 pixels pt.pt. and 0.77 pixels pt.crv. We consider this acceptable considering the rather small training set.

We conclude that AAMs are feasible as a tool for automated analysis of the corpus callosum in MRI images. Future work should aim at a clinical validation performed on a larger data set.

Acknowledgements

# References

1. M. Brejl and M. Sonka. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. *Medical Imaging, IEEE Transactions on*, 19(10):973–985, 2000.
2. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498. Springer, 1998.
3. T. F. Cootes and C. J. Taylor. *Statistical Models of Appearance for Computer Vision*. Tech. Report. Feb 2000, University of Manchester, 2000.
4. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
5. R. H. Davies. *Learning Shape: Optimal Models for Analysing Natural Variability*. PhD thesis, Department of Imaging Science and Biomedical Engineering, University of Manchester, Stopford building, University of Manchester, 2002.
6. R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, 21(5):525–537, 2002.
7. G. J. Edwards, C. J. Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–5. IEEE Comput. Soc, 1998.
8. B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. Ter Haar Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.
9. A. Lundervold, N. Duta, T. Taxt, and A. K. Jain. Model-guided segmentation of corpus callosum in MR images. *Computer Vision and Pattern Recognition*, 1999.
10. M. B. Stegmann. Object tracking using active appearance models. In *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark*, volume 1, pages 54–60. DIKU, 2001.
11. M. B. Stegmann and R. H. Davies. Corpus callosum analysis using MDL-based sequential models of shape and appearance. In *preparation*, 2003.

A   Corpus Callosum MRIs
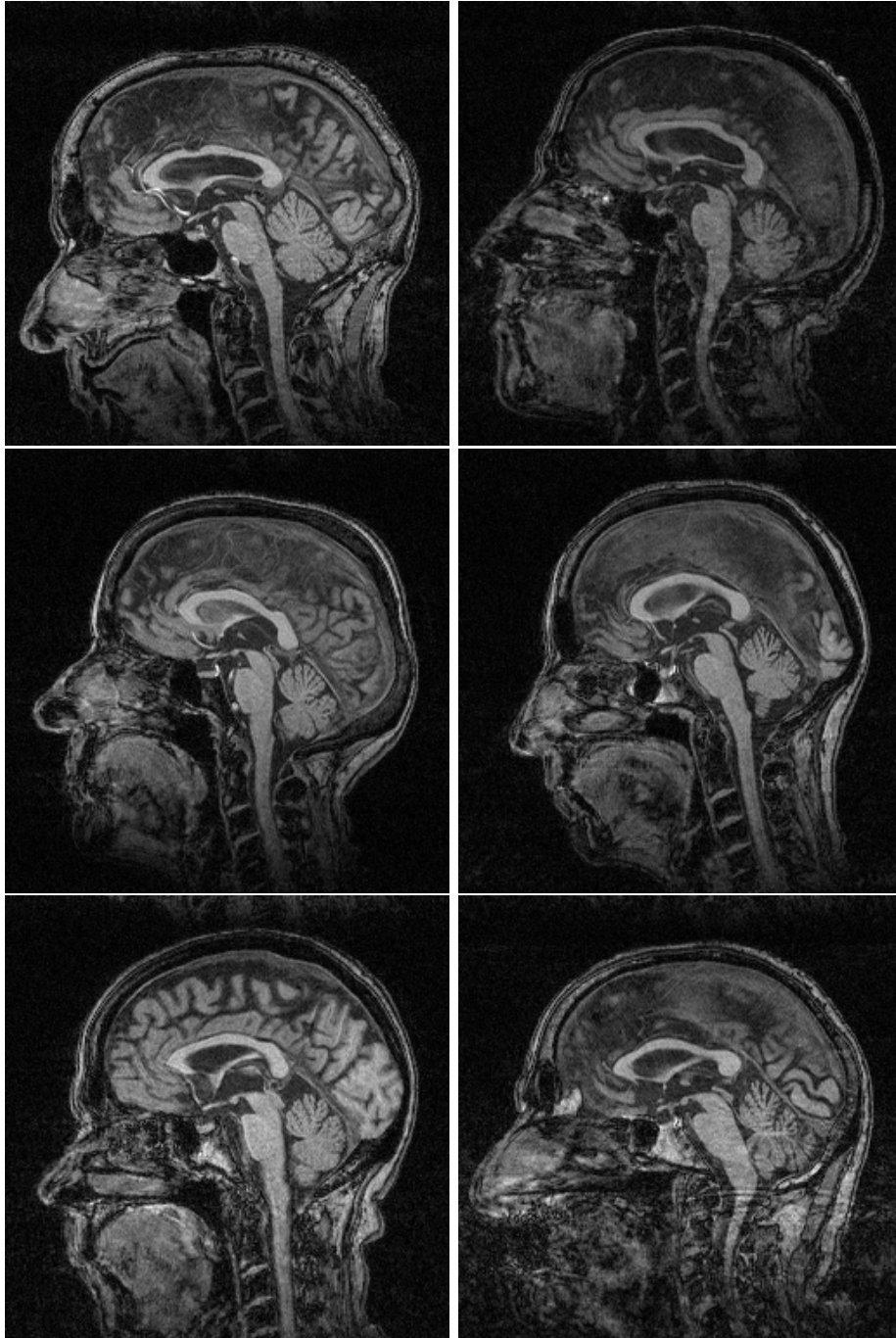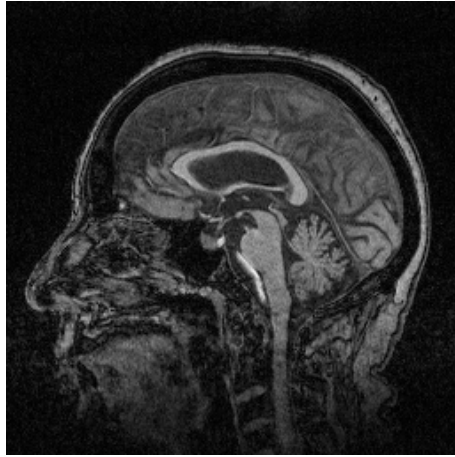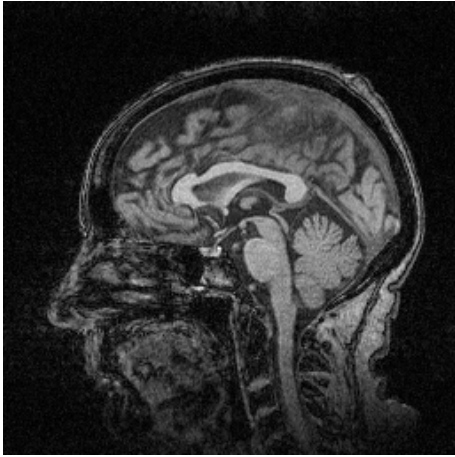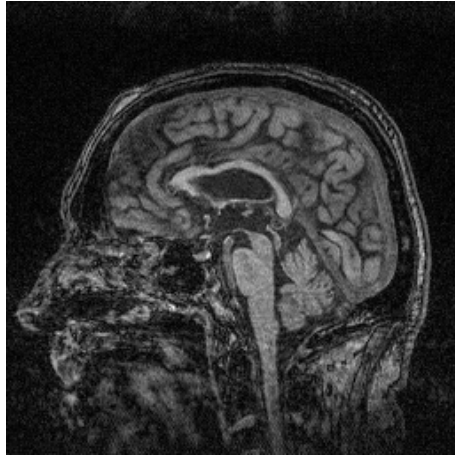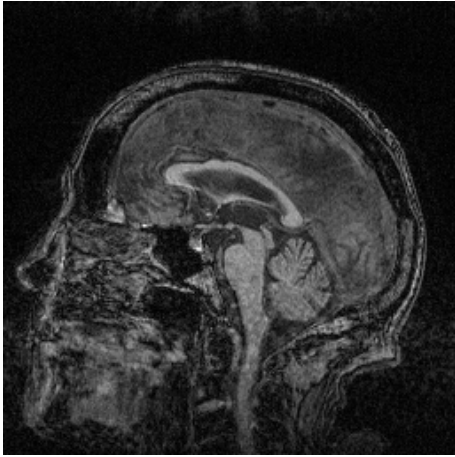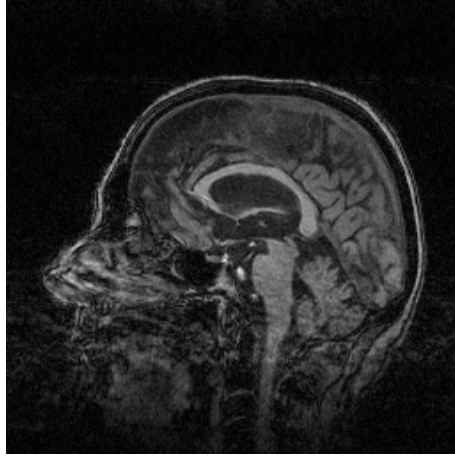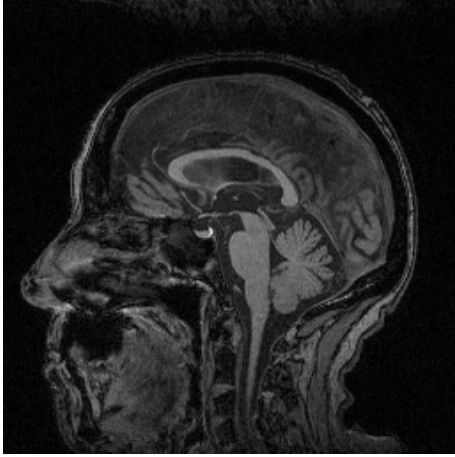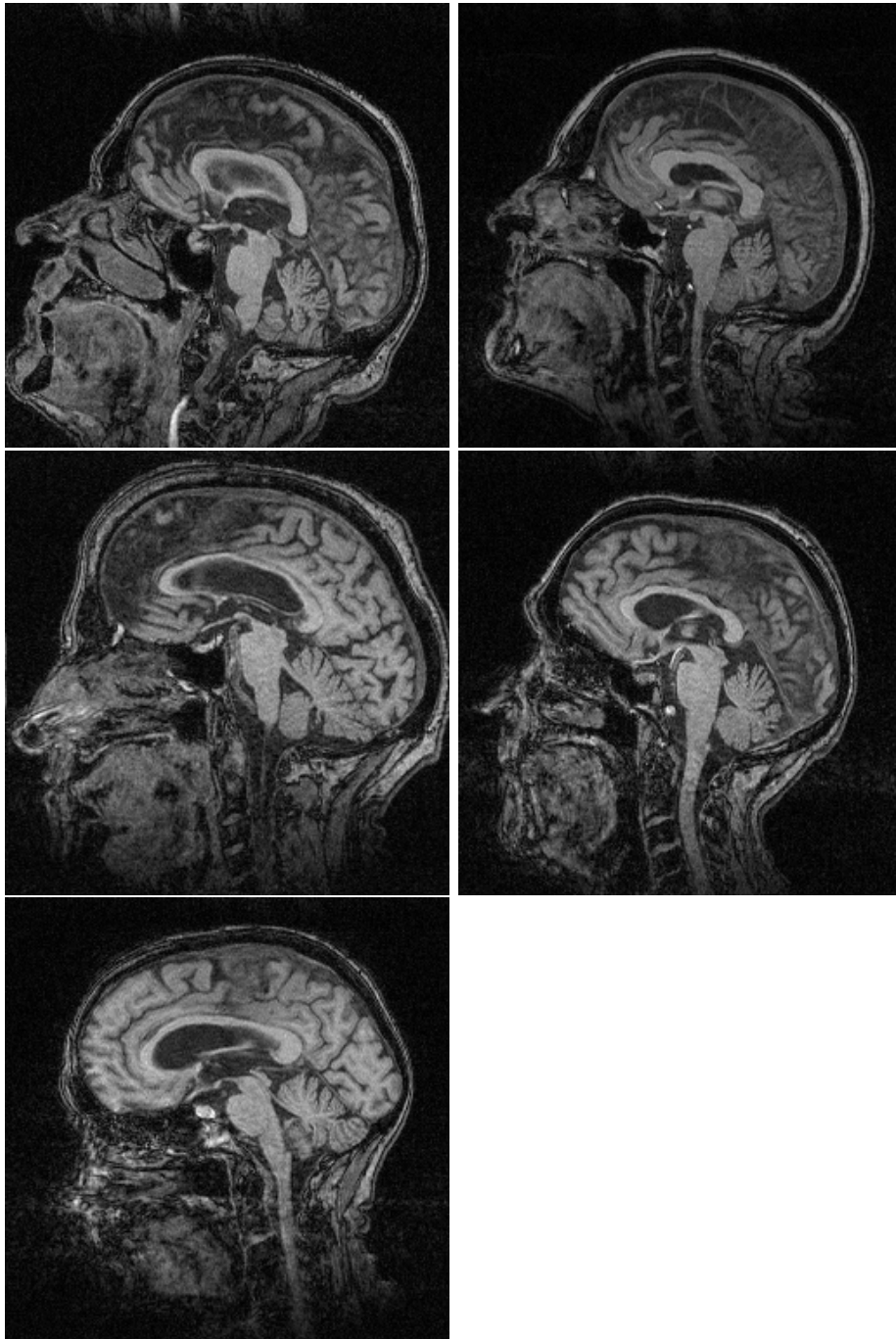
Images are presented row-wise in same ordering as the corresponding annotations in Figure 1.