

# Data augmentation using learned transforms for one-shot medical image segmentation

Amy Zhao  
MIT

xamyzhao@mit.edu

Guha Balakrishnan  
MIT

balakg@mit.edu

Frédo Durand  
MIT

fredo@mit.edu

John V. Guttag  
MIT  
gutttag@mit.edu

Adrian V. Dalca  
MIT, MGH  
adalca@mit.edu

## Abstract

Biomedical image segmentation is an important task in many medical applications. Segmentation methods based on convolutional neural networks attain state-of-the-art accuracy; however, they typically rely on supervised training with large labeled datasets. Labeling datasets of medical images requires significant expertise and time, and is infeasible at large scales. To tackle the lack of labeled data, researchers use techniques such as hand-engineered preprocessing steps, hand-tuned architectures, and data augmentation. However, these techniques involve costly engineering efforts, and are typically dataset-specific.

We present an automated data augmentation method for medical images. We demonstrate our method on the task of segmenting magnetic resonance imaging (MRI) brain scans, focusing on the one-shot segmentation scenario – a practical challenge in many medical applications. Our method requires only a single segmented scan, and leverages other unlabeled scans in a semi-supervised approach. We learn a model of transforms from the images, and use the model along with the labeled example to synthesize additional labeled training examples for supervised segmentation. Each transform is comprised of a spatial deformation field and an intensity change, enabling the synthesis of complex effects such as variations in anatomy and image acquisition procedures. Augmenting the training of a supervised segmenter with these new examples provides significant improvements over state-of-the-art methods for one-shot biomedical image segmentation. Our code is available at <https://github.com/xamyzhao/brainstorm>.

## 1. Introduction

Semantic image segmentation is crucial to many biomedical imaging applications, such as performing pop-

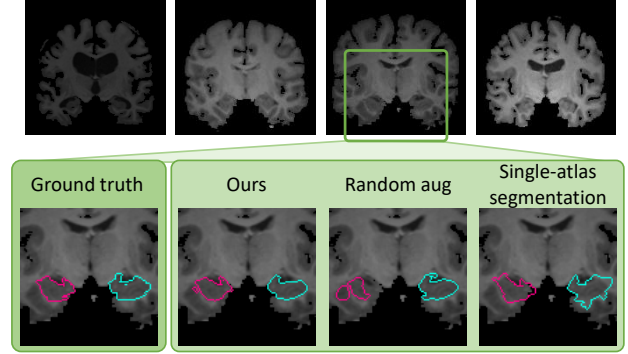


Figure 1: Biomedical images often vary widely in anatomy, contrast and texture (top row). Our method enables more accurate segmentation of anatomical structures compared to other one-shot segmentation methods (bottom row).

ulation analyses, diagnosing disease, and planning treatments. When enough labeled data is available, supervised deep learning-based segmentation methods produce state-of-the-art results. However, in the medical domain, obtaining manual segmentation labels for medical images requires considerable expertise and time. In most clinical image datasets, there are very few manually labeled images. The problem of limited labeled data is exacerbated by differences in image acquisition procedures across machines and institutions, which can produce wide variations in resolution, image noise, and tissue appearance [43].

To overcome these challenges, many supervised biomedical segmentation methods focus on hand-engineered preprocessing steps and architectures [50, 54]. It is also common to use hand-tuned data augmentation to increase the number of training examples [2, 52, 54, 60, 62]. Data augmentation functions such as random image rotations or random nonlinear deformations are easy to implement, and

have been shown to be effective at improving segmentation accuracy in some settings [52, 54, 60, 62]. However, these functions have limited ability to emulate diverse and realistic examples [25], and can be highly sensitive to the choice of parameters [24].

We propose to address the challenges of limited labeled data by learning to synthesize diverse and realistic labeled examples. Our novel, automated approach to data augmentation leverages unlabeled images. Using learning-based registration methods, we model the set of spatial and appearance transforms between images in the dataset. These models capture the anatomical and imaging diversity present in the unlabeled volumes. We then synthesize new examples by sampling transforms and applying them to a single labeled example.

We demonstrate our method on the task of one-shot segmentation of brain magnetic resonance imaging (MRI) scans. We use our method to synthesize new labeled training examples, enabling the training of a supervised segmentation network. This strategy outperforms state-of-the-art one-shot biomedical segmentation approaches, including single-atlas segmentation and supervised segmentation with hand-tuned data augmentation.

## 2. Related work

### 2.1. Medical image segmentation

We focus on the segmentation of brain MR images, which is challenging for several reasons. Firstly, human brains exhibit substantial anatomical variations [27, 56, 73]. Secondly, MR image intensity can vary as a result of subject-specific noise, scanner protocol and quality, and other imaging parameters [43]. This means that a tissue class can appear with different intensities across images – even images of the same MRI modality. Segmenting such scans based on appearance alone is a difficult task.

Many existing segmentation methods rely on scan pre-processing to mitigate these intensity-related challenges. Pre-processing methods can be costly to run, and developing techniques for realistic datasets is an active area of research [14, 70]. Our augmentation method tackles these intensity-related challenges from another angle: rather than removing intensity variations, it enables a segmentation method to be robust to the natural variations in MRI scans.

A large body of classical segmentation methods use *atlas-based* or *atlas-guided* segmentation, in which a labeled reference volume, or *atlas*, is aligned to a target volume using a deformation model, and the labels are propagated using the same deformation [6, 13, 21, 31]. When multiple atlases are available, they are each aligned to a target volume, and the warped atlas labels are fused [35, 40, 65, 75]. In atlas-based approaches, anatomical variations between subjects are captured by a deformation

model, and the challenges of intensity variations are mitigated using pre-processed scans, or intensity-robust metrics such as normalized cross-correlation. However, ambiguities in tissue appearances (*e.g.*, indistinct tissue boundaries, image noise) can still lead to inaccurate registration and segmentations. We aim to address this limitation by training a segmentation model on diverse realistic examples, making the segmenter more robust to such ambiguities. We focus on having a single atlas, and demonstrate that our strategy outperforms atlas-based segmentation. If more than one segmented example is available, our method can leverage them.

Supervised learning approaches to biomedical segmentation have gained popularity in recent years. To mitigate the need for large labeled training datasets, these methods often use data augmentation in conjunction with hand-engineered pre-processing steps and architectures [2, 39, 50, 54, 60, 62, 78]. For instance, multi-resolution image patches and convolutional weight-sharing are used in [50] for few-shot segmentation.

Semi-supervised and unsupervised approaches have also been proposed to combat the challenges of small training datasets. These methods do not require paired image and segmentation data. Rather, they leverage collections of *segmentation* data to build anatomical priors [20], to train an adversarial network [38], or to train a novel semantic constraint [28]. In practice, collections of *images* are more readily available than segmentations. Rather than rely on segmentations, our method leverages a set of unlabeled images.

### 2.2. Spatial and appearance transform models

Models of shape and appearance have been used in a variety of image analyses. In medical image registration, a spatial deformation model is used to establish semantic correspondences between images. This mature field spans optimization-based methods [4, 7, 64, 67], and recent learning-based methods [8, 9, 19, 41, 59, 69, 76]. We leverage VoxelMorph [8, 9], a recent unsupervised learning-based method, to learn spatial transforms.

Many registration methods focus on intensity-normalized images or intensity-independent objective functions, and do not explicitly account for variations in image intensity. For unnormalized images, spatial and appearance transform models have been used together to register objects that differ in texture or appearance, as well as shape. Many works build upon the framework of Morphable Models [37] or Active Appearance Models (AAMs) [15, 16], in which statistical models of shape and texture are constructed. In the medical domain, AAMs have been used to localize anatomical landmarks [17, 55] and perform segmentation [49, 53, 74]. We build upon these concepts by using convolutional neural networks to learn

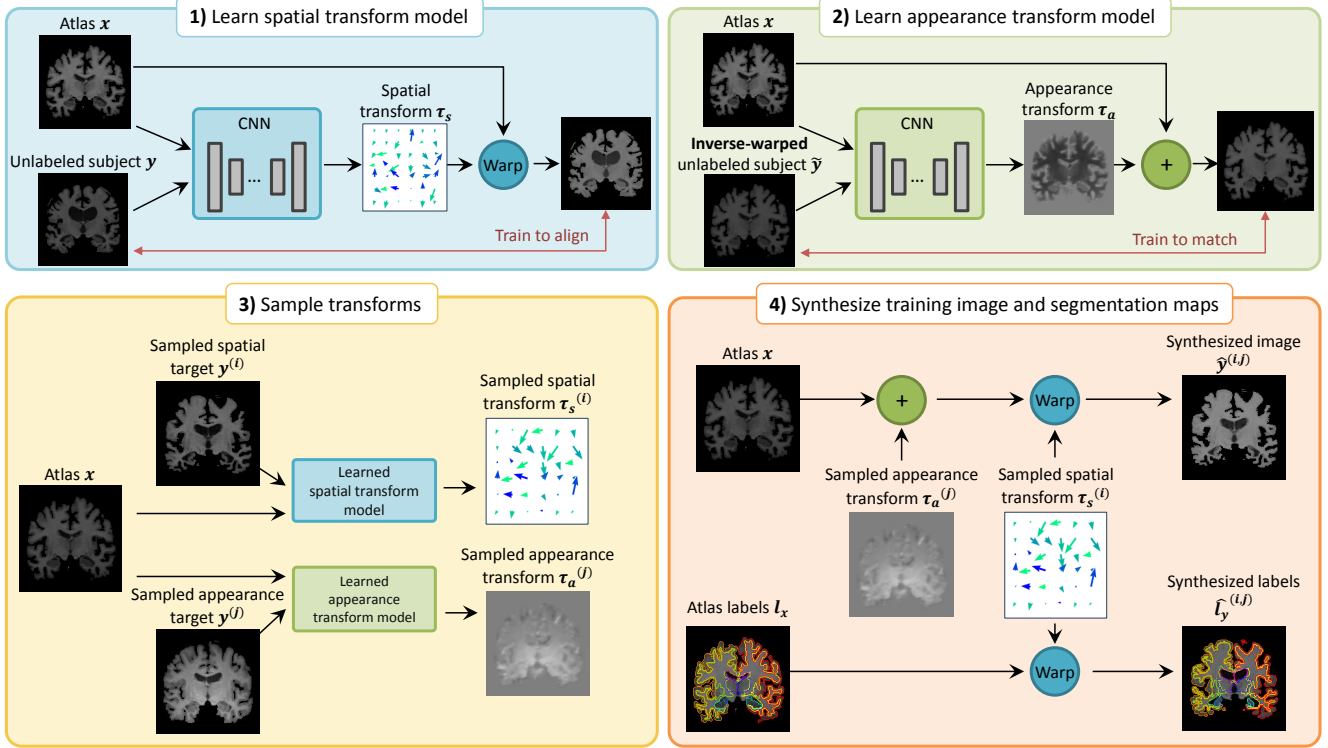


Figure 2: An overview of the proposed method. We learn independent spatial and appearance transform models to capture the variations in our image dataset. We then use these models to synthesize a dataset of labeled examples. This synthesized dataset is used to train a supervised segmentation network.

models of unconstrained spatial and intensity transform fields. Rather than learning transform models for the end goal of registration or segmentation, we sample from these models to synthesize new training examples. As we show in our experiments, augmenting a segmenter’s training set in this way can produce more robust segmentations than performing segmentation using the transform models directly.

### 2.3. Few-shot segmentation of natural images

Few-shot segmentation is a challenging task in semantic segmentation, video object segmentation and interactive segmentation. Existing approaches focus mainly on natural images. Methods for few-shot semantic segmentation incorporate information from prototypical examples of the classes to be segmented [23, 66]. Few-shot video segmentation is frequently implemented by aligning objects in each frame to a labeled reference frame [36, 72]. Other approaches leverage large labeled datasets of supplementary information such as object appearances [11]. Guided networks have been used to incorporate additional information (e.g., human-in-the-loop) to perform few-shot segmentation in various settings [57]. Medical images present a different set of challenges from natural images; for instance, the vi-

sual differences between tissue classes are very subtle compared to the differences between objects in natural images.

### 2.4. Data augmentation

In image-based supervised learning tasks, data augmentation is commonly performed using simple parameterized transforms such as rotation and scaling. In the medical imaging domain, random smooth flow fields are often used to simulate anatomical variations [48, 60, 61]. These parameterized transforms can reduce overfitting and improve test performance [33, 42, 48, 60, 61]. However, the performance gains imparted by these transforms vary with the selection of transformation functions and parameter settings [24].

Recent works have proposed learning data augmentation transformations from data. Hauberg *et al.* [30] focus on data augmentation for classifying MNIST digits. They learn digit-specific spatial transformations, and sample training images and transformations to create new examples aimed at improving classification performance. We learn an appearance model in addition to a spatial model, and we focus on the problem of MRI segmentation.

Ratner *et al.* [58] present a semi-automated approach to learning spatial and color transformations for data aug-

mentation. They rely on user input to create compositions of simple parameterized transformation functions (*e.g.*, rotation and contrast enhancement). They learn to generate new compositions of transform functions using a generative adversarial network. In contrast, our approach is fully automated.

### 3. Method

We propose to improve one-shot biomedical image segmentation by synthesizing realistic training examples in a semi-supervised learning framework.

Let  $\{y^{(i)}\}$  be a set of biomedical image volumes, and let the pair  $(x, l_x)$  represent a labeled reference volume, or *atlas*, and its corresponding segmentation map. In our problem of brain MRI segmentation, each  $x$  and  $y$  is a grayscale 3D volume. We focus on the challenging case where only one labeled atlas is available, since it is often difficult in practice to obtain many segmented volumes. If more segmented volumes become available, our method can be easily extended to leverage them.

To perform data augmentation, we apply transforms  $\tau^{(k)}$  to the labeled atlas  $x$ . We first learn separate spatial and appearance transform models to capture the distribution of anatomical and appearance differences between the labeled atlas and each unlabeled volume. Using the two learned models, we synthesize labeled volumes  $\{(\hat{y}^{(k)}, \hat{l}_y^{(k)})\}$  by applying a spatial transform and an appearance transform to the atlas volume, and by warping the atlas label maps using the spatial transform. Compared to single-atlas segmentation, which suffers from uncertainty or errors in the spatial transform model, we use the same spatial transform to synthesize the volume *and* label map, ensuring that the newly synthesized volume is correctly labeled. These synthetic examples form a labeled dataset that characterizes the anatomical and appearance variations in the unlabeled dataset. Along with the atlas, this new training set enables us to train a supervised segmentation network. This process is outlined in Fig. 2.

#### 3.1. Spatial and appearance transform models

MR images can exhibit substantial inter-scan variation in anatomy and appearance. We describe the differences between scans using a combination of spatial and intensity transforms. Specifically, we define a transform  $\tau$  from one volume to another as a composition of a spatial transform  $\tau_s(\cdot)$  and an intensity or *appearance* transform  $\tau_a(\cdot)$ , *i.e.*,  $\tau(\cdot) = \tau_s(\tau_a(\cdot))$ .

We assume a spatial transform takes the form of a smooth voxel-wise displacement field  $u$ . Following the medical registration literature, we define the *deformation* function  $\phi = \text{id} + u$ , where  $\text{id}$  is the identity function. We use  $x \circ \phi$  to denote the application of the deformation  $\phi$  to  $x$ . To model the distribution of spatial transforms in our

dataset, we compute the deformation that warps atlas  $x$  to each volume  $y^{(i)}$  using  $\phi^{(i)} = g_{\theta_s}(x, y^{(i)})$ , where  $g_{\theta_s}(\cdot, \cdot)$  is a parametric function that we describe later. We write approximate inverse deformation of  $y^{(i)}$  to  $x$  as  $\phi^{-1(i)} = g_{\theta_s}(y^{(i)}, x)$ .

We model the appearance transform  $\tau_a(\cdot)$  as a per-voxel addition in the spatial frame of the atlas. We compute this per-voxel volume using the function  $\psi^{(i)} = h_{\theta_a}(x, y^{(i)} \circ \phi^{-1(i)})$ , where  $y^{(i)} \circ \phi^{-1(i)}$  is a volume that has been registered to the atlas space using our learned spatial model. In summary, our spatial and appearance transforms are:

$$\tau_s^{(i)}(x) = x \circ \phi^{(i)}, \quad \phi = g_{\theta_s}(x, y^{(i)}) \quad (1)$$

$$\tau_a^{(i)}(x) = x + \psi^{(i)}, \quad \psi^{(i)} = h_{\theta_a}(x, y^{(i)} \circ \phi^{-1(i)}). \quad (2)$$

#### 3.2. Learning

We aim to capture the distributions of the transforms  $\tau_s$  and  $\tau_a$  between the atlas and the unlabeled volumes. We estimate the transform functions  $g_{\theta_s}(\cdot, \cdot)$  and  $h_{\theta_a}(\cdot, \cdot)$  defined in Eqs. (1) and (2) using separate convolutional neural networks, with each network using the general architecture outlined in Fig. 3. Drawing on insights from Morphable Models [37] and Active Appearance Models [16, 17], we optimize the spatial and appearance models independently.

For our spatial model, we leverage VoxelMorph [8, 9, 19], a recent unsupervised learning-based approach with an open-source implementation. VoxelMorph learns to output a smooth displacement vector field that registers one image to another by jointly optimizing an image similarity loss and a displacement field smoothness term. We use a variant of VoxelMorph with normalized cross-correlation as the image similarity loss, enabling the estimation of  $g_{\theta_s}(\cdot, \cdot)$  with unnormalized input volumes.

We use a similar approach to learn the appearance model. Naively, one might define  $h_{\theta_a}(\cdot, \cdot)$  from Eq. (2) as a simple per-voxel subtraction of the volumes in the atlas space. However, this leads  $h_{\theta_a}(\cdot, \cdot)$  to produce extraneous details when the registration function  $\phi$  is imperfect, resulting in image details that do not match the correspondingly warped anatomical labels. We instead design  $h_{\theta_a}(\cdot, \cdot)$  as a neural network that produces a per-voxel intensity change in an anatomically consistent manner. Specifically, we use an image similarity loss as well as a semantically-aware smoothness regularization. Given the network output  $\psi^{(i)} = h_{\theta_a}(x, y^{(i)} \circ \phi^{-1(i)})$ , we define a smoothness regularization function based on the atlas segmentation map:

$$\mathcal{L}_{smooth}(c_x, \psi) = (1 - c_x) \nabla \psi, \quad (3)$$

where  $c_x$  is a binary image of anatomical boundaries computed from the atlas segmentation labels  $l_x$ , and  $\nabla$  is the spatial gradient operator. Intuitively, this term discourages

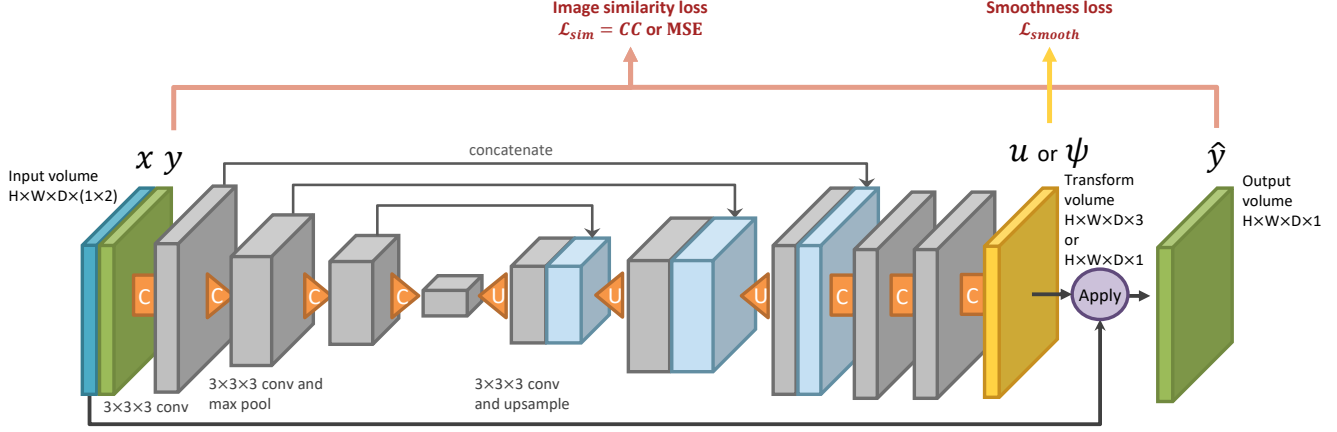


Figure 3: We use a convolutional neural network based on the U-Net architecture [60] to learn each transform model. The application of the transform is a spatial warp for the spatial model, and a voxel-wise addition for the appearance model. Each convolution uses  $3 \times 3 \times 3$  kernels, and is followed by a LeakyReLU activation layer. The encoder uses max pooling layers to reduce spatial resolution, while the decoder uses upsampling layers.

dramatic intensity changes within the same anatomical region.

In the total appearance transform model loss  $\mathcal{L}_a$ , we use mean squared error for the image similarity loss  $\mathcal{L}_{sim}(\hat{y}, y) = \|\hat{y} - y\|^2$ . In our experiments, we found that computing the image similarity loss in the spatial frame of the subject was helpful. We balance the similarity loss with the regularization term  $\mathcal{L}_{smooth}$ :

$$\begin{aligned} \mathcal{L}_a(x, y^{(i)}, \phi^{(i)}, \phi^{-1(i)}, \psi^{(i)}, c_x) \\ = \mathcal{L}_{sim}((x + \psi^{(i)}) \circ \phi^{(i)}, y^{(i)}) + \lambda_a \mathcal{L}_{smooth}(c_x, \psi^{(i)}), \end{aligned}$$

where  $\lambda_a$  is a hyperparameter.

### 3.3. Synthesizing new examples

The models described in Eqs. (1) and (2) enable us to sample spatial and appearance transforms  $\tau_s^{(i)}, \tau_a^{(j)}$  by sampling target volumes  $y^{(i)}, y^{(j)}$  from an unlabeled dataset. Since the spatial and appearance targets can be different subjects, our method can combine the spatial variations of one subject with the intensities of another into a single synthetic volume  $\hat{y}$ . We create a labeled synthetic example by applying the transforms computed from the target volumes to the labeled atlas:

$$\begin{aligned} \hat{y}^{(i,j)} &= \tau_s^{(i)}(\tau_a^{(j)}(x)), \\ \hat{l}_y^{(i,j)} &= \tau_s^{(i)}(l_x). \end{aligned}$$

This process is visualized in steps 3 and 4 in Fig. 2. These new labeled training examples are then included in the labeled training set for a supervised segmentation network.

### 3.4. Segmentation network

The newly synthesized examples are useful for improving the performance of a supervised segmentation network. We demonstrate this using a network based on the state-of-the-art architecture described in [63]. To account for GPU memory constraints, the network is designed to segment one slice at a time. We train the network on random slices from the augmented training set. We select the number of training epochs using early stopping on a validation set. We emphasize that the exact segmentation network architecture is not the focus of this work, since our method can be used in conjunction with any supervised segmentation network.

### 3.5. Implementation

We implemented all models using Keras [12] and Tensorflow [1]. The application of a spatial transform to an image is implemented using the differentiable 3D spatial transformer layer as described in [8], and a similar layer that uses nearest neighbor interpolation is used to transform segmentation maps. For simplicity, we capture the forward and inverse spatial transforms described in Section 3.1 using two identical neural networks. For the appearance transform model, we use the hyperparameter setting  $\lambda_a = 0.02$ . We train our transform models with a single pair of volumes in each batch, and train the segmentation model with a batch size of 16 slices. All models are trained with a learning rate of  $5e^{-4}$ . Our code is available at <https://github.com/xamyzhao/brainstorm>.



## 4. Experiments

We demonstrate how our automatic augmentation method can be used to improve brain MRI segmentation. We focus on one-shot segmentation of unnormalized scans – a challenging but practical scenario. Intensity normalization methods such as bias field correction [26, 68, 71] can work poorly in realistic situations (*e.g.*, clinical-quality scans, or scans with stroke [70] or traumatic brain injury).

### 4.1. Data

We use the publicly available dataset of T1-weighted MRI brain scans described in [8]. The dataset is compiled from eight databases: ADNI [51], OASIS [44], ABIDE [46], ADHD200 [47], MCIC [29], PPMI [45], HABS [18], and Harvard GSP [32]. As in [8], we resample the brains to  $256 \times 256 \times 256$  with 1mm isotropic voxels, and affinely align and crop the images to  $160 \times 192 \times 224$ . We do not perform any intensity corrections. We obtain anatomical segmentation maps for all scans using FreeSurfer [26], and perform skull-stripping by zeroing out voxels with no anatomical label. For evaluation, we use segmentation maps of the 30 anatomical labels described in [8].

We focus on the task of segmentation using a single labeled example. We randomly select 101 brain scans to be available at training time. In practice, the atlas is usually selected to be close to the anatomical average of the population. We select our atlas from the training set by finding the most similar scan to the anatomical average computed in [8]. This atlas is the single labeled example that is used in the training process of our method; the labels of the other 100 training brains are not used. We use an additional 50 scans as a validation set, and an additional 100 scans as a held-out test set.

### 4.2. Segmentation baselines

We compare our method to the following baselines:

**Single-atlas segmentation (SAS):** We train the state-of-the-art registration algorithm described in [8] to register the labeled atlas to each training volume. At test time, we use the trained spatial transform model in a single-atlas segmentation framework: we register the atlas to each test volume and warp the atlas labels using the computed deformation field [6, 13, 21, 31, 40]. That is, for each test image  $y^{(i)}$ , we compute  $\phi^{(i)} = g_{\theta_s}(x, y^{(i)})$  and predict labels  $\hat{l}_y^{(i)} = l_x \circ \phi^{(i)}$ .

**Data augmentation using single-atlas segmentation (SAS-aug):** We use SAS results as labels for the unannotated training brains, which we then include as training examples for supervised segmentation. This adds 100 new training examples to the segmenter training set. Even though SAS can produce imperfect labels because of mistakes or

ambiguity in registration, training on multiple coarse labels can result in improved segmentation performance [79].

### Hand-tuned random data augmentation (*rand-aug*):

Random smooth deformations have been shown to be useful for data augmentation [48, 60, 61], and are particularly relevant in biomedical applications since they can simulate anatomical variations in tissues [60]. Similarly to [48, 60, 61], we create a random smooth deformation field by first sampling random displacement vectors on a sparse grid, and then applying bilinear interpolation and spatial blurring. We evaluated several settings for the amplitude and smoothness of the deformation field, including the ones described in [60]. We use the settings that result in the best segmentation performance on a validation set.

We synthesize variations in tissue imaging intensity using a random global intensity multiplicative factor, similar to [34, 39]. We sample this factor uniformly from the range [0.5, 1.5], which we determined by inspection to match the intensities in the dataset. This is representative of how augmentation parameters are tuned in practice. This augmentation method synthesizes a new randomly transformed brain in each training iteration.

**Supervised:** We train a fully-supervised segmentation network that uses ground truth labels for all 101 examples in our training dataset. Apart from the atlas labels, these labels are not available for any of the other methods. This method serves as an upper bound.

### 4.3. Variants of our method

**Independent sampling (*ours-indep*):** As described in Section 3.3, we sample spatial and appearance target images independently to compute  $\tau_s^{(i)}, \tau_a^{(j)}$ . With 100 unlabeled targets, we obtain 100 spatial transforms and 100 appearance transforms, enabling the synthesis of 10000 different labeled examples. Due to memory constraints, we synthesize a random labeled example in each training iteration, rather than adding all 10000 new examples to the training set.

**(Ablation study) coupled sampling (*ours-coupled*):** To highlight the efficacy of our independent transform models, we compare *ours-indep* to a variant of our method where we sample each of the spatial *and* appearance transforms from the same target image. This results in 100 possible synthetic examples. As in *ours-indep*, we synthesize a random example in each training iteration.

***Ours-indep* + *rand-aug*:** When training the segmenter, we alternate between training on examples synthesized using *ours-indep*, and examples synthesized using *rand-aug*. The addition of hand-tuned augmentation to our synthetic augmentation could introduce additional variance that is unseen even in the unlabeled set, improving the robustness of the segmenter.

Table 1: Segmentation performance in terms of Dice score [22], evaluated on a held-out test set of 100 brain scans. We report the mean Dice score (and standard deviation in parentheses) across all 30 anatomical labels and all 100 test subjects. We also report the mean pairwise improvement of each method over the SAS baseline.

Method	Dice score	Pairwise Dice improvement
SAS	0.759 (0.137)	-
SAS-aug	0.775 (0.147)	0.016 (0.041)
Rand-aug	0.765 (0.143)	0.006 (0.088)
Ours-coupled	0.795 (0.133)	0.036 (0.036)
Ours-indep	<b>0.804</b> (0.130)	<b>0.045</b> (0.038)
Ours-indep + rand-aug	<b>0.815</b> (0.123)	<b>0.056</b> (0.044)
Supervised (upper bound)	0.849 (0.092)	0.089 (0.072)

#### 4.4. Evaluation metrics

We evaluate the accuracy of each segmentation method in terms of Dice score [22], which quantifies the overlap between two anatomical regions. A Dice score of 1 indicates perfectly overlapping regions, while 0 indicates no overlap. The predicted segmentation labels are evaluated relative to anatomical labels generated using FreeSurfer [26].

#### 4.5. Results

##### 4.5.1 Segmentation performance

Table 1 shows the segmentation accuracy attained by each method. Our methods outperform all baselines in mean Dice score across all 30 evaluation labels, showing significant improvements over the next best baselines *rand-aug* ( $p < 1e-15$  using a paired t-test) and *SAS-aug* ( $p < 1e-20$ ).

In Figs. 4 and 5, we compare each method to the single-atlas segmentation baseline. Fig. 4 shows that our methods attain the most improvement on average, and are more consistent than hand-tuned random augmentation. Fig. 5 shows that *ours-indep* + *rand-aug* is consistently better than each baseline on every test subject. *Ours-indep* alone is always better than *SAS-aug* and *SAS*, and is better than *rand-aug* on 95 of the 100 test scans.

Fig. 6 shows that *rand-aug* improves Dice over *SAS* on large structures, but is detrimental for smaller anatomical structures. In contrast, our methods produce consistent improvements over *SAS* and *SAS-aug* across all brain structures. We show several examples of segmented hippocampi in Fig. 7.

##### 4.5.2 Synthesized images

Our independent spatial and appearance models enable the synthesis of a wide variety of brain appearances. Fig. 8 shows several examples where combining transforms from our models produces realistic results with accurate labels.

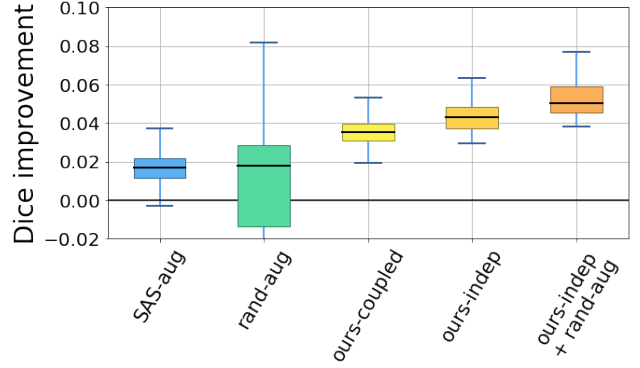


Figure 4: Pairwise improvement in mean Dice score (with the mean computed across all 30 anatomical labels) compared to the SAS baseline, shown across all test subjects.

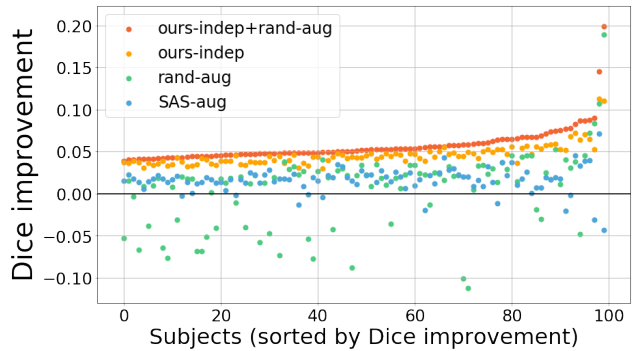


Figure 5: Pairwise improvement in mean Dice score (with the mean computed across all 30 anatomical labels) compared to the SAS baseline, shown for each test subject. Subjects are sorted by the Dice improvement of our method (*ours-indep*+*rand-aug*).

#### 5. Discussion

**Why do we outperform single-atlas segmentation?** Our methods rely on the same spatial registration model that is used for *SAS* and *SAS-aug*. Both *ours-coupled* and *SAS-aug* augment the segmenter training set with 100 new images.

To understand why our method produces better segmentations, we examine the augmented images. Our method warps the image in the same way as the labels, ensuring that the warped labels match the transformed image. On the other hand, *SAS-aug* applies the warped labels to the original image, so any errors or noise in the registration results in a mis-labeled new training example for the segmenter. Fig. 9 highlights examples where our method synthesizes image texture within the hippocampus label that is more consistent with the texture of the ground truth hippocampus, resulting in a more useful synthetic training example.

**Extensions** Our framework lends itself to several plausible future extensions. In Section 3.1, we discussed the use

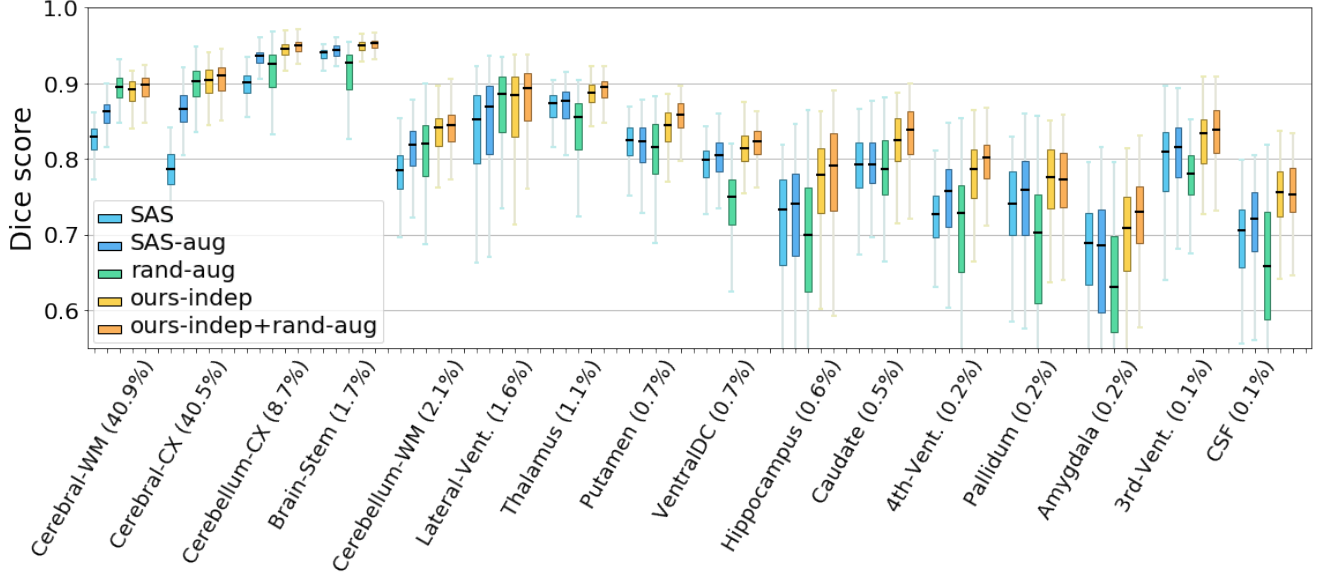


Figure 6: Segmentation accuracy of each method across various brain structures. The percentage of the brain occupied by each label in the atlas is shown in parentheses. Labels are sorted by the volume of each structure in the atlas, and labels consisting of left and right structures (e.g. Hippocampus) are combined. We abbreviate the labels: white matter (WM), cortex (CX), ventricle (vent), and cerebrospinal fluid (CSF).

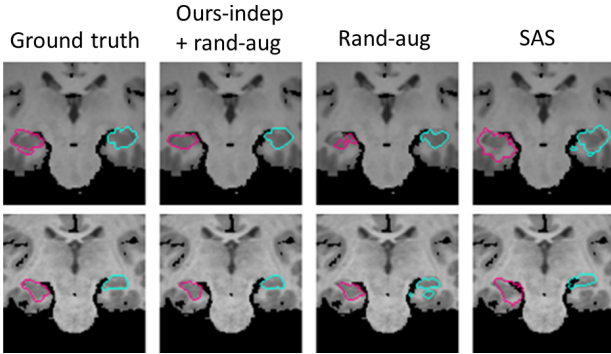


Figure 7: Hippocampus segmentation predictions for two test subjects (rows). Our method (column 2) produces more accurate predictions than the baselines (columns 3 and 4).

of an approximate inverse deformation function for learning the appearance transform in the reference frame of the atlas. Rather than learning a separate inverse spatial transform model, in the future we will leverage existing work in diffeomorphic registration [3, 5, 10, 19, 77].

We sample transforms from a discrete set of spatial and appearance transforms. This could be extended to span the space of transforms more richly, e.g., through interpolation between transforms, or using compositions of transforms.

We demonstrated our approach on brain MRIs. Since the method uses no brain- or MRI-specific information, it is feasible to extend it to other anatomy or imaging modalities, such as CT.

## 6. Conclusion

We presented a learning-based method for data augmentation, and demonstrated it on one-shot medical image segmentation.

We start with one labeled scan and a set of unlabeled examples. Using learning-based registration methods, we model the set of spatial and appearance transforms between the labeled and unlabeled examples. These transforms capture effects such as non-linear deformations and variations in imaging intensity. We synthesize new labeled examples by sampling transforms and applying them to the labeled example, producing a wide variety of realistic new scans.

We use these synthesized examples to train a supervised segmentation model. The segmenter out-performs existing one-shot segmentation methods on every example in our test set, approaching the performance of a fully supervised model. This framework enables segmentation in many applications, such as the clinical setting where time constraints permit the manual annotation of only a few scans.

In summary, this work shows that:

- learning independent models of spatial and appearance transforms from unlabeled images enables the synthesis of diverse and realistic labeled examples, and
- these synthesized examples can be used to train a segmentation model that out-performs existing methods in a one-shot scenario.



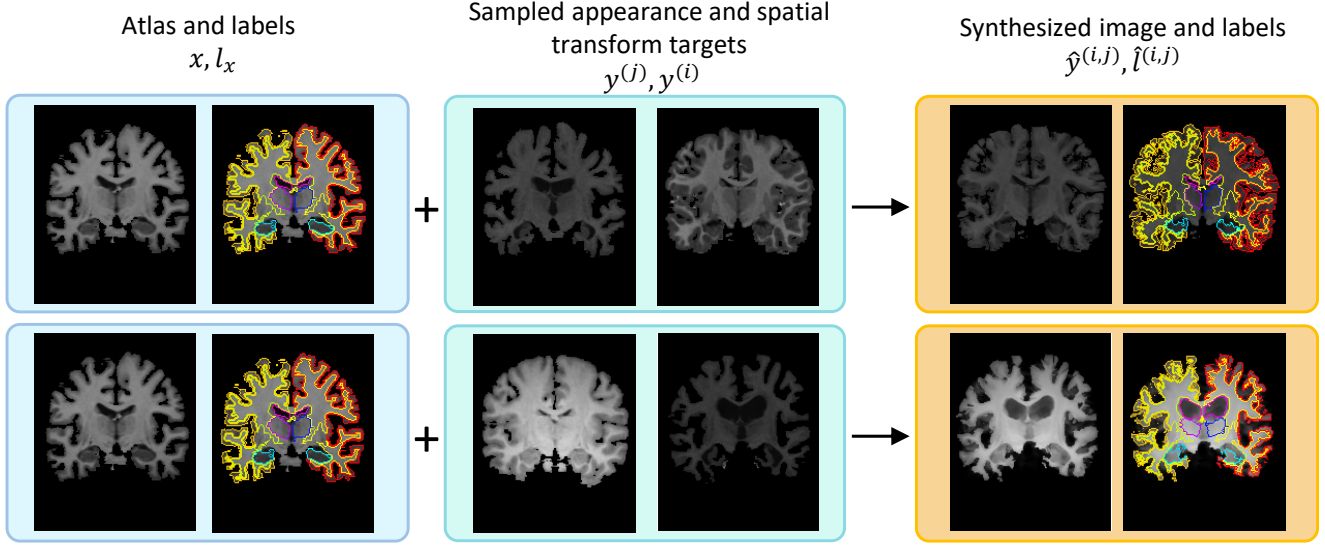


Figure 8: Since we model spatial and appearance transforms independently, we are able to synthesize a wide variety of combined effects. The top row shows a synthetic image where the appearance transform target produced a darkening effect, and the spatial transform shrunk the ventricles and widened the whole brain. In the second row, the atlas is brightened and the ventricles are enlarged.

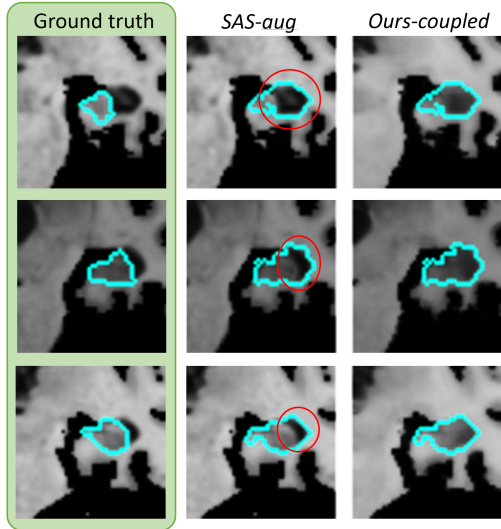


Figure 9: Several examples of synthetic training examples produced by *SAS-aug* (column 2) and *ours-coupled* (column 3). When the spatial registration model (used by both methods) produces imperfect warped labels, our method still synthesizes a useful training example by matching the synthesized image texture to the label. *SAS-aug*, on the other hand, pairs the imperfect warped label with incorrect image textures.

## References

- [1] M. Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016. 5
- [2] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017. 1, 2
- [3] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. 8
- [4] J. Ashburner and K. Friston. Voxel-based morphometry-the methods. *Neuroimage*, 11:805–821, 2000. 2
- [5] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008. 8
- [6] C. Baillard, P. Hellier, and C. Barillot. Segmentation of brain 3d mr images using level sets and dense registration. *Medical image analysis*, 5(3):185–194, 2001. 2, 6
- [7] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46:1–21, 1989. 2
- [8] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9252–9260, 2018. 2, 4, 5, 6
- [9] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 2019. 2, 4
- [10] M. F. Beg, M. I. Miller, A. Trounev, and L. Younes. Computing large deformation metric mappings via geodesic flows of

- diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005. 8
- [11] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool. One-shot video object segmentation. In *CVPR 2017*. IEEE, 2017. 3
- [12] F. Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015. 5
- [13] C. Ciofolo and C. Barillot. Atlas-based segmentation of 3d cerebral structures with competitive level sets and fuzzy control. *Medical image analysis*, 13(3):456–470, 2009. 2, 6
- [14] D. Coelho de Castro and B. Glocker. Nonparametric density flows for mri intensity normalisation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 206–214, 09 2018. 2
- [15] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor. A unified framework for atlas matching using active appearance models. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 322–333. Springer, 1999. 2
- [16] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685, 2001. 2, 4
- [17] T. F. Cootes and C. J. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *Medical Imaging 2001: Image Processing*, volume 4322, pages 236–249. International Society for Optics and Photonics, 2001. 2, 4
- [18] A. Dagley, M. LaPoint, W. Huijbers, T. Hedden, D. G. McLaren, J. P. Chatwal, K. V. Papp, R. E. Amariglio, D. Blacker, D. M. Rentz, et al. Harvard aging brain study: dataset and accessibility. *NeuroImage*, 144:255–258, 2017. 6
- [19] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018. 2, 4, 8
- [20] A. V. Dalca, J. Guttag, and M. R. Sabuncu. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9290–9299, 2018. 2
- [21] B. M. Dawant, S. L. Hartmann, J.-P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel. Automatic 3-d segmentation of internal structures of the head in mr images using a combination of similarity and free-form transformations. i. methodology and validation on normal subjects. *IEEE transactions on medical imaging*, 18(10):909–916, 1999. 2, 6
- [22] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. 7
- [23] N. Dong and E. P. Xing. Few-shot semantic segmentation with prototype learning. In *BMVC*, volume 3, page 4, 2018. 3
- [24] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox. Discriminative unsupervised feature learning with exemplar convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1734–1747, 2016. 2, 3
- [25] Z. Eaton-Rosen, F. Bragman, S. Ourselin, and M. J. Cardoso. Improving data augmentation for medical image segmentation. In *International Conference on Medical Imaging with Deep Learning*, 2018. 2
- [26] B. Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012. 6, 7
- [27] M. A. Frost and R. Goebel. Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage*, 59(2):1369–1381, 2012. 2
- [28] P.-A. Ganaye, M. Sdika, and H. Benoit-Cattin. Semi-supervised learning for segmentation under semantic constraint. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 595–602. Springer, 2018. 2
- [29] R. L. Gollub et al. The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3):367–388, 2013. 6
- [30] S. Hauberg, O. Freifeld, A. B. L. Larsen, J. Fisher, and L. Hansen. Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In *Artificial Intelligence and Statistics*, pages 342–350, 2016. 3
- [31] P. Hellier and C. Barillot. A hierarchical parametric algorithm for deformable multimodal image registration. *Computer Methods and Programs in Biomedicine*, 75(2):107–115, 2004. 2, 6
- [32] A. J. Holmes et al. Brain genomics superstruct project initial data release with structural, functional, and behavioral measures. *Scientific data*, 2, 2015. 6
- [33] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016. 3
- [34] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin. Differential data augmentation techniques for medical imaging classification tasks. In *AMIA Annual Symposium Proceedings*, volume 2017, page 979. American Medical Informatics Association, 2017. 6
- [35] J. E. Iglesias and M. R. Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015. 2
- [36] S. D. Jain, B. Xiong, and K. Grauman. Fusionseg: Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos. In *Proc. CVPR*, volume 1, 2017. 3
- [37] M. J. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. *International Journal of Computer Vision*, 29(2):107–131, 1998. 2, 4
- [38] T. Joyce, A. Chatsias, and S. A. Tsafaris. Deep multi-class segmentation without ground-truth labels. 2018. 2
- [39] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017. 2, 6

- [40] A. Klein and J. Hirsch. Mindboggle: a scatterbrained approach to automate brain labeling. *NeuroImage*, 24(2):261–280, 2005. 2, 6
- [41] J. Krebs et al. Robust non-rigid registration through agent-based action learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 344–352. Springer, 2017. 2
- [42] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 3
- [43] K. K. Leung, M. J. Clarkson, J. W. Bartlett, S. Clegg, C. R. Jack Jr, M. W. Weiner, N. C. Fox, S. Ourselin, A. D. N. Initiative, et al. Robust atrophy rate measurement in alzheimer’s disease using multi-site serial mri: tissue-specific intensity normalization and parameter selection. *Neuroimage*, 50(2):516–523, 2010. 1, 2
- [44] D. S. Marcus et al. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, non-demented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007. 6
- [45] K. Marek et al. The parkinson progression marker initiative. *Progress in neurobiology*, 95(4):629–635, 2011. 6
- [46] A. D. Martino et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014. 6
- [47] M. P. Milham et al. The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6:62, 2012. 6
- [48] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016. 3, 6
- [49] S. C. Mitchell, J. G. Bosch, B. P. Lelieveldt, R. J. Van der Geest, J. H. Reiber, and M. Sonka. 3-d active appearance models: segmentation of cardiac mr and ultrasound images. *IEEE transactions on medical imaging*, 21(9):1167–1178, 2002. 2
- [50] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Išgum. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1252–1261, 2016. 1, 2
- [51] S. G. Mueller et al. Ways toward an early diagnosis in alzheimer’s disease: the alzheimer’s disease neuroimaging initiative (adni). *Alzheimer’s & Dementia*, 1(1):55–66, 2005. 6
- [52] A. Oliveira, S. Pereira, and C. A. Silva. Augmenting data when training a cnn for retinal vessel segmentation: How to warp? In *Bioengineering (ENBENG), 2017 IEEE 5th Portuguese Meeting on*, pages 1–4. IEEE, 2017. 1, 2
- [53] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, 2011. 2
- [54] S. Pereira, A. Pinto, V. Alves, and C. A. Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251, 2016. 1, 2
- [55] V. Potesil, T. Kadir, G. Platsch, and M. Brady. Personalized graphical models for anatomical landmark localization in whole-body medical images. *International Journal of Computer Vision*, 111(1):29–49, 2015. 2
- [56] J. Rademacher, U. Bürgel, S. Geyer, T. Schormann, A. Schleicher, H.-J. Freund, and K. Zilles. Variability and asymmetry in the human precentral motor system: a cytoarchitectonic and myeloarchitectonic brain mapping study. *Brain*, 124(11):2232–2258, 2001. 2
- [57] K. Rakelly, E. Shelhamer, T. Darrell, A. A. Efros, and S. Levine. Few-shot segmentation propagation with guided networks. *arXiv preprint arXiv:1806.07373*, 2018. 3
- [58] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré. Learning to compose domain-specific transformations for data augmentation. *arXiv preprint arXiv:1709.01643*, 2017. 3
- [59] M.-M. Rohé et al. Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 266–274. Springer, 2017. 2
- [60] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 2, 3, 5, 6
- [61] H. R. Roth, C. T. Lee, H.-C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers. Anatomy-specific classification of medical images using deep convolutional nets. *arXiv preprint arXiv:1504.04003*, 2015. 3, 6
- [62] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 556–564. Springer, 2015. 1, 2
- [63] A. G. Roy, S. Conjeti, D. Sheet, A. Katouzian, N. Navab, and C. Wachinger. Error corrective boosting for learning fully convolutional networks with limited data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 231–239. Springer, 2017. 5
- [64] D. Rueckert et al. Nonrigid registration using free-form deformation: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999. 2
- [65] M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE transactions on medical imaging*, 29(10):1714–1729, 2010. 2
- [66] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots. One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410*, 2017. 3
- [67] D. Shen and C. Davatzikos. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002. 2

- [68] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998. 6
- [69] H. Sokooti et al. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 232–239. Springer, 2017. 2
- [70] R. Sridharan, A. V. Dalca, K. M. Fitzpatrick, L. Cloonan, A. Kanakis, O. Wu, K. L. Furie, J. Rosand, N. S. Rost, and P. Golland. Quantification and analysis of large multimodal clinical image studies: Application to stroke. In *International Workshop on Multimodal Brain Image Analysis*, pages 18–30. Springer, 2013. 2, 6
- [71] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig. Parametric estimate of intensity inhomogeneities applied to mri. *IEEE Trans. Med. Imaging*, 19(3):153–165, 2000. 6
- [72] Y.-H. Tsai, M.-H. Yang, and M. J. Black. Video segmentation via object flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3899–3908, 2016. 3
- [73] D. C. Van Essen and D. L. Dierker. Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron*, 56(2):209–225, 2007. 2
- [74] G. Vincent, G. Guillard, and M. Bowes. Fully automatic segmentation of the prostate using active appearance models. *MICCAI Grand Challenge: Prostate MR Image Segmentation*, 2012, 2012. 2
- [75] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):611–623, 2013. 2
- [76] X. Yang et al. Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage*, 158:378–396, 2017. 2
- [77] M. Zhang, R. Liao, A. V. Dalca, E. A. Turk, J. Luo, P. E. Grant, and P. Golland. Frequency diffeomorphisms for efficient image registration. In *International conference on information processing in medical imaging*, pages 559–570. Springer, 2017. 8
- [78] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015. 2
- [79] A. Zlateski, R. Jaroensri, P. Sharma, and F. Durand. On the importance of label quality for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1479–1487, 2018. 6