# Medical Image Imputation From Image Collections

Adrian V. Dalca<sup>®</sup>, Katherine L. Bouman<sup>®</sup>, William T. Freeman<sup>®</sup>, Natalia S. Rost, Mert R. Sabuncu<sup>®</sup>, and Polina Golland for the Alzheimer's Disease Neuroimaging Initiative

Abstract—We present an algorithm for creating highresolution anatomically plausible images consistent with acquired clinical brain MRI scans with large inter-slice spacing. Although large data sets of clinical images contain a wealth of information, time constraints during acquisition result in sparse scans that fail to capture much of the anatomy. These characteristics often render computational analysis impractical as many image analysis algorithms tend to fail when applied to such images. Highly specialized algorithms that explicitly handle sparse slice spacing do not generalize well across problem domains. In contrast, we aim to enable the application of existing algorithms that were originally developed for high-resolution research scans to significantly undersampled scans. We introduce a generative model that captures a fine-scale anatomical structure across subjects in clinical image collections and derives an

Manuscript received June 3, 2018; accepted July 8, 2018. Date of publication August 22, 2018; date of current version February 1, 2019. This work was supported in part by NIH NINDS under Grant R01NS086905, in part by NIH NICHD under Grant U01HD087211, in part by NIH NIBIB NAC under Grant P41EB015902, in part by NIH under Grant R41AG052246-01, Grant 1K25EB013649-01, and Grant 1R21AG050122-01, in part by NSF IIS under Grant 1447473, in part by Wistron Corporation, in part by SIP, in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) through the National Institutes of Health under Grant U01 AG024904, and in part by the Department of Defense ADNI under Award W81XWH-12-2-0012. The work at ADNI was supported in part by the National Institute on Aging, in part by the National Institute of Biomedical Imaging, in part by Bioengineering, and in part by the generous contributions from several agencies listed at http://adni.loni.usc.edu/about/. (Corresponding author: Adrian V. Dalca.)

A. V. Dalca is with the Computer Science and Artificial Intelligence Laboaratory, MIT, Cambridge, MA 02139 USA, and also with the Martinos Center for Biomedical Imaging, Massachusetts General Hospital, HMS, Boston, MA 02114 USA (e-mail: adalca@csail.mit.edu).

K. L. Bouman and P. Golland are with the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139 USA.

W. T. Freeman is with the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139 USA, and also with Google, Cambridge, MA 02142 USA.

N. S. Rost is with the Department of Neurology, Massachusetts General Hospital, HMS, Boston, MA 02114 USA.

M. R. Sabuncu is with the School of Electrical and Computer Engineering and the Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14850 USA.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how to apply/ ADNI Acknowledgement List.pdf

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2018.2866692

algorithm for filling in the missing data in scans with large inter-slice spacing. Our experimental results demonstrate that the resulting method outperforms the state-of-the-art upsampling super-resolution techniques, and promises to facilitate subsequent analysis not previously possible with scans of this quality. Our implementation is freely available at https://github.com/adalca/papago.

EMB WEEE

*Index Terms*—Imputation, super-resolution, clinical scans, MRI, sparse slices, thick slices, brain scans.

#### I. INTRODUCTION

**I** NCREASINGLY open image acquisition efforts in clinical practice are driving dramatic increases in the number and size of patient cohorts in clinical archives. Unfortunately, clinical scans are typically of dramatically lower resolution than the research scans that motivate most methodological development. Specifically, while slice *thickness* can vary depending on the clinical study or scan, inter-slice *spacing* is often significantly larger than the in-plane resolution of individual slices. This results in missing voxels that are typically filled via interpolation.

Our work is motivated by a study that includes brain MRI scans of thousands of stroke patients acquired within 48 hours of stroke onset. The study aims to quantify white matter disease burden [23], necessitating skull stripping and deformable registration into a common coordinate frame [27], [31], [32]. The volumes are severely undersampled (0.85mm × 0.85mm × 6mm) due to constraints of acute stroke care (Fig. 1). Such undersampling is typical of modalities, such as T2-FLAIR, that aim to characterize tissue properties, even in research studies like ADNI [15].

In undersampled scans, the image is no longer smooth, and the anatomical structure may change substantially between consecutive slices (Fig. 1). Since such clinically acquired scans violate underlying assumptions of many algorithms, even basic tasks such as skull stripping and deformable registration present significant challenges, yet are often necessary for downstream analysis [4], [7], [12], [15], [23], [30], [31].

We present a novel method for constructing high resolution anatomically plausible volumetric images consistent with the available slices in sparsely sampled clinical scans. Importantly, our method does not require any high resolution scans or expert annotations for training. It instead imputes the missing structure by learning solely from the available collection of sparsely sampled clinical scans. The restored images represent plausible anatomy. They promise to act as a medium for enabling computational analysis of clinical scans

0278-0062 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications\_standards/publications/rights/index.html for more information.



Fig. 1. An example scan from our clinical dataset. The three panels display axial, sagittal and coronal slices, respectively. While axial in-plane resolution can be similar to that of a research scan, slice spacing is significantly larger. We visualize the saggital and coronal views using nearest neighbor interpolation.

with existing techniques originally developed for high resolution, isotropic research scans. For example, although imputed data should not be used in clinical evaluation, the brain mask obtained through skull stripping of the restored scan can be applied to the original clinical scan to improve subsequent analyses.

## A. Prior Work

Many image restoration techniques depend on having enough information in a single image to synthesize data. Traditional interpolation methods, such as linear, cubic or spline [28], assume a functional representation of the image. They treat the low resolution voxels as samples, or observations, and estimate function parameters to infer missing voxel values. Patch-based super resolution algorithms use fine-scale redundancy within a single scan [10], [11], [19], [20], [22]. The key idea is to fill in the missing details by identifying similar image patches in the same image that might contain relevant detail [19], [22]. This approach depends on having enough repetitive detail in a scan to capture and re-synthesize high frequency information. Unfortunately, clinical images are often characterized by sampling that is too sparse to adequately fit functional representations or provide enough fine-scale information to recover the lost detail. For example, 6mm slice spacing, typical of many clinical scans including our motivating example, is far too high to accurately estimate approximating functions without prior knowledge. In such cases, a single image is unlikely to contain enough fine-scale information to provide anatomically plausible reconstructions in the direction of slice acquisition, as we demonstrate later in the paper.

Alternatively, one can use additional data to synthesize better images. Many superresolution algorithms use multiple scans of the same subject, such as multiple low resolution acquisitions with small shift differences to synthesize a single volume [2], [16], [22]. However, such acquisitions are not commonly available in the clinical setting.

Nonparametric and convolutional neural-network (CNN) based upsampling methods that tackle the problem of superresolution often rely on an external dataset of high resolution data or cannot handle extreme undersampling present in clinical scans. For example, some methods fill in missing data by matching a low resolution image patch from the input scan with a high resolution image patch from the training dataset [3], [13], [16], [17], [24], [25]. Similarly, CNN-based upsampling methods approximate completion functions, but require high resolution scans for training [8], [21]. A recent approach to improve resolution from a collection of scans with sparse slices jointly upsamples all images using non-local means [26]. However this method has only been demonstrated on slice spacing of roughly three times the in-plane resolution, and in our experience similar nonparametric methods fail to upsample clinical scans with more significant undersampling.

Our work relies on a low dimensional embedding of image patches with missing voxels. Parametric patch methods and low dimensional embeddings have been used to model the common structure of image patches from full resolution images, but are typically not designed to handle missing data. Specifically, priors [33] and Gaussian Mixture Models [35], [36] have been used in both medical and natural images for classification [1] and denoising [9], [36]. The procedures used for training of these models rely on having full resolution patches with no missing data in the training phase.

Unfortunately, high (full) resolution training datasets are not readily available for many image contrasts and scanners, and may not adequately represent pathology or other properties of clinical populations. Acquiring the appropriate high resolution training image data is often infeasible, and here we explicitly focus on the realistic clinical scenario where only sparsely sampled images are available.

#### B. Method Overview

We take advantage of the fact that local fine scale structure is shared in a population of medical images, and each scan with sparse slices captures some partial aspect of this structure. We borrow ideas from Gaussian Mixture Model (GMM) for image patch priors [36], low dimensional Gaussian embeddings [14], [34], and missing data models [14], [18] to develop a probabilistic generative model for sparse 3D image patches around a particular location using a low-dimensional GMM with partial observations. We derive the EM algorithm for maximum likelihood estimation of the model parameters and discuss related modeling choices. Given a new sparsely sampled scan, the maximum a posteriori estimate of the latent structure yields the imputed high resolution image. We evaluate our algorithm using scans from the ADNI cohort, and demonstrate its utility in the context of the motivating stroke study. We investigate the behaviour of our model under different parameter settings, and illustrate an example of potential improvements in the downstream analysis using an example task of skull stripping.

This paper extends the preliminary version of the method presented at the 2017 Conference on Information Processing in Medical Imaging [5]. Here, we improve model inference by removing parameter co-dependency between iterations and providing new parameter initialization. We provide detailed derivations and discuss an alternative related model. Finally, we provide an analysis of important model parameters, present results for more subjects, and illustrate more example reconstructions. The paper is organized as follows. Section II introduces the model and learning algorithm. Section III discusses implementation details. We present experiments and analysis of the algorithm's behavior in Section IV. We discuss



Fig. 2. Image imputation for a subvolume. (a) Full resolution images, shown for illustration only. These are unobserved by the algorithm. (b) Sparse planes acquired in clinical scans. (c) During learning, we train a GMM that captures the low dimensional nature of patch variability in a region around a particular location (white dot). (d) Given a sparsely sampled scan, we infer the most likely cluster for each 3D patch, and restore the missing data using the learned model and the observed voxels. We form the final volume from overlapping restored patches. 2D images are shown for illustration only, the algorithms operate fully in 3D.

important modeling aspects and related models in Section V. We include an Appendix and Supplementary Material with detailed derivations of the EM algorithm for the proposed models.

#### II. METHOD

In this section, we construct a generative model for sparse image patches, present the resulting learning algorithm, and describe our image restoration procedure.

Let  $\{Y_1, \ldots, Y_N\}$  be a collection of scans with large inter-slice spaces, *roughly* aligned into a common atlas space (we use affine transformations in our experiments). For each image  $Y_i$  in the collection, only a few slices are observed. We seek to restore an anatomically plausible high resolution volume by imputing the missing voxel values.

We capture local structure using image patches. We assume a constant patch shape, and in our experiments use a 3D 11 × 11 × 11 shape. We use  $y_i$  to denote a *D*-length vector that contains voxels of the image patch centered at a certain location in image  $Y_i$ . We perform inference at each location independently and stitch the results into the final image as described later in this section. Fig. 2 provides an overview of the method.

# A. Generative Model

We treat an image patch as a high dimensional manifestation of a low dimensional representation, with the intuition that the covariation within image patches has small intrinsic dimensionality relative to the number of voxels in the patch. To capture the anatomical variability across subjects, we employ a Gaussian Mixture Model (GMM) to represent local structure of 3D patches in the vicinity of a particular location across the entire collection. We then explicitly model the observed and missing information. Fig 3 presents the corresponding graphical model.



Fig. 3. **Graphical representation of our model.** Circles indicate random variables and rounded squares represent parameters. Shading represents observed quantities and the plate indicates replication. The observed patch voxels  $y_i^{\mathcal{O}_i}$  form a subset of patch  $y_i$  extracted by the mask  $\mathcal{O}_i$  and are generated from a multivariate Gaussian distribution conditioned on the latent cluster  $k_i$  and the latent patch representation  $x_i$ . Parameters  $\mu$  and W define the mean and the variance of the Gaussian components of the mixture, and  $\sigma^2$  is the image noise variance.

We model the latent low dimensional patch representation  $x_i$ of length d < D as a normal random variable

$$x_i \sim \mathcal{N}(0, I_{d \times d}),\tag{1}$$

where  $\mathcal{N}(\mu, \Sigma)$  denotes the multivariate Gaussian distribution with mean  $\mu$  and covariance  $\Sigma$ . We draw latent cluster assignment k from a categorical distribution defined by a length-K vector  $\pi$  of cluster probabilities, and treat image patch  $y_i$  as a high dimensional observation of  $x_i$  drawn from a K-component multivariate GMM. Specifically, conditioned on the drawn cluster k,

$$y_i = \mu_k + W_k x_i + \epsilon_i, \text{ where}$$
(2)  
$$\epsilon_i \sim \mathcal{N}(0, \sigma_k^2 I_{D \times D}), \text{ and } \epsilon_i \perp x_i.$$

Vector  $\mu_k$  is the patch mean of cluster k, matrix  $W_k$ shapes the covariance structure of  $y_i$ , and  $\sigma_k^2$  is the variance of image noise. This model implies  $\mathbb{E}[y_i|k] = \mu_k$  and  $C_k \stackrel{\Delta}{=} \mathbb{E}[(y_i - \mu_k)(y_j - \mu_k)^T|k] = W_k W_k^T + \sigma_k^2 I_{D \times D}$ . Defining  $\theta = \{\mu_k, W_k, \sigma_k^2, \pi_k\}_{k=1}^K$ , the likelihood of all patches  $\mathcal{Y} = \{y_i\}$  at this location under the mixture model is

$$p(\mathcal{Y};\theta) = \prod_{i} \sum_{k} \pi_{k} \mathcal{N}(y_{i};\mu_{k},C_{k}).$$
(3)

In our clinical images, only a few slices are known. To model sparse observations, we let  $O_i$  be the set of observed voxels in patch  $y_i$ , and  $y_i^{O_i}$  be the corresponding vector of their intensity values:

$$y_i^{\mathcal{O}_i} = \mu_k^{\mathcal{O}_i} + W_k^{\mathcal{O}_i} x_i + \epsilon_i^{\mathcal{O}_i}, \qquad (4)$$

where  $W_k^{\mathcal{O}_i}$  comprises rows of  $W_k$  that correspond to the observed voxel set  $\mathcal{O}_i$ . The likelihood of the *observed* data  $\mathcal{Y}^{\mathcal{O}} = \{y_i^{\mathcal{O}_i}\}$  is therefore

$$p(\mathcal{Y}^{\mathcal{O}};\theta) = \prod_{i} \sum_{k} \pi_{k} \mathcal{N}(y_{i}^{\mathcal{O}_{i}};\mu_{k}^{\mathcal{O}_{i}},C_{k}^{\mathcal{O}_{i}}\mathcal{O}_{i}), \qquad (5)$$

where matrix  $C_k^{\mathcal{O}_i \mathcal{O}_i}$  extracts the rows and columns of  $C_k$  that correspond to the observed voxel subset  $\mathcal{O}_i$ .

We do not explicitly model slice thickness, as in many clinical datasets this thickness is unknown or varies by site, scanner or acquisition. Instead, we simply treat the original data as high resolution thin planes and analyze the effects of varying slice thickness on the results in the experimental evaluation of the method.

We also investigated an alternative modeling choice where each missing voxel of patch  $y_i$  is modelled as a latent variable. This assumption can optionally be combined with the latent low-dimensional patch representation. We discuss this alternative choice in Section V, and provide parameter updates in the Supplementary Material, in the supplementary files/multimedia tab. Unfortunately, the resulting algorithm is prohibitively slow.

## B. Learning

Given a collection of observed patches  $\mathcal{Y}^{\mathcal{O}}$ , we seek the maximum likelihood estimates of the model parameters { $\mu_k$ ,  $W_k$ ,  $\sigma_k^2$ } and  $\pi$  under the likelihood (5). We derive the Expectation Maximization algorithm [6] in Appendix VI, and present the update equations and their interpretations below.

The expectation step updates the class memberships:

$$\gamma_{ik} \stackrel{\Delta}{=} p(k|y_i^{\mathcal{O}_i}; \theta) = \frac{\pi_k \mathcal{N}(y_i^{\mathcal{O}_i}; \mu_k^{\mathcal{O}_i}, C_k^{\mathcal{O}_i})}{\sum_{k'} \pi_{k'} \mathcal{N}(y_i^{\mathcal{O}_i}; \mu_{k'}^{\mathcal{O}_i}, C_{k'}^{\mathcal{O}_i})},$$
(6)

and the statistics of the low dimensional representation  $x_i$  for each image patch  $y_i^{O_i}$  as "explained" by cluster k:

$$\widehat{x}_{ik} \stackrel{\Delta}{=} \mathbb{E}[x_i|k] = \left( (W_k^{\mathcal{O}_i})^T (W_k^{\mathcal{O}_i}) + \sigma_k^2 I_{d \times d} \right)^{-1} (W_k^{\mathcal{O}_i})^T (y_i^{\mathcal{O}_i} - \mu_k^{\mathcal{O}_i}),$$
(7)

$$S_{ik} \stackrel{\Delta}{=} \mathbb{E}[x_i x_i^T | k] - \widehat{x}_{ik} \widehat{x}_{ik}^T$$
  
=  $\sigma_k^2 \left( (W_k^{\mathcal{O}_i})^T (W_k^{\mathcal{O}_i}) + \sigma_k^2 I_{d \times d} \right)^{-1}.$  (8)

We let  $\mathcal{P}_j$  be the set of patches in which voxel *j* is observed, and form the following normalized mean statistics:

$$\delta_{ik} = \frac{\gamma_{ik}}{\sum_{i' \in \mathcal{P}_j} \gamma_{i'k}} \tag{9}$$

$$b_{j} = \sum_{i \in \mathcal{P}_{i}} \delta_{ik} \mathbf{E}[x_{i}|k] = \sum_{i \in \mathcal{P}_{i}} \delta_{ik} \widehat{x}_{ik}$$
(10)

$$A_j = \sum_{i \in \mathcal{P}_j} \delta_{ik} \mathbf{E}[x_i x_i^T | k] = \sum_{i \in \mathcal{P}_j} \delta_{ik} (\widehat{x}_{ik} \widehat{x}_{ik}^T + S_{ik}).$$
(11)

The **maximization step** uses the observed voxels to update the model parameters. We let  $y_i^j$  be the  $j^{\text{th}}$  element of vector  $y_i$ , and update the cluster mean as a convex combination of observed voxels:

$$\mu_k^j \leftarrow \frac{\sum_{i \in \mathcal{P}_j} \gamma_{ik} (1 - \hat{x}_i^T A_j^{-1} b_j) y_i^j}{\sum_{i' \in \mathcal{P}_j} \gamma_{i'k} (1 - \hat{x}_{i'}^T A_j^{-1} b_j)}.$$
(12)

The covariance factors and image noise variance are updated based on the statistics of the low dimensional representation from (10) and (11):

$$W_k^j \leftarrow \sum_{i \in \mathcal{P}_j} \delta_{ik} (y_i^j - \mu_k^j) \widehat{x}_{ik}^T A_j^{-1},$$
(13)

$$\sigma_k^2 \leftarrow \frac{\sum_j \sum_{i \in \mathcal{P}_j} \gamma_{ik} \left[ (y_i^j - \mu_k^j - W_k^j \widehat{x}_{ik})^2 + W_k^j S_{ik} (W_k^j)^T \right]}{\sum_j \sum_{i' \in \mathcal{P}_j} \gamma_{i'k}}.$$
(14)

where  $W_k^j$  is the *j*<sup>th</sup> row of matrix  $W_k$ . Finally, we update the cluster proportions:

$$\pi_k = \frac{1}{N} \sum_i \gamma_{ik}.$$
 (15)

Intuitively, learning our model with sparse data is possible because each image patch provides a slightly different subset of voxel observations that contribute to the parameter estimation (Fig. 2). In our experiments, all subject scans have the same acquisition direction. Despite different affine transformations to the atlas space for each subject, some voxel *pairs* are still never observed in the same patch, resulting in missing entries of the covariance matrix. Using a low-rank approximation for the covariance matrix regularized the estimates.

Upon convergence of the EM updates, we compute the cluster covariance  $C_k = W_k W_k^T + \sigma_k^2 I_{D \times D}$  for each k.

#### C. Imputation

To restore an individual patch  $y_i$ , we compute the *maximum-a-posteriori* (MAP) estimate of the image patch:

$$\hat{y}_{i} = \arg \max_{y_{i}} p(y_{i}|y_{i}^{\mathcal{O}_{i}}; \theta)$$

$$= \arg \max_{y_{i}} \sum_{k} p(k|y_{i}^{\mathcal{O}_{i}}) \int_{x_{i}} p(y_{i}|x_{i}) p(x_{i}|k, y_{i}^{\mathcal{O}_{i}}) dx_{i}$$

$$= \arg \max_{y_{i}} \sum_{k} \gamma_{ik} \int_{x_{i}} p(y_{i}|x_{i}) p(x_{i}|k, y_{i}^{\mathcal{O}_{i}}) dx_{i}$$

$$= \arg \max_{y_{i}} \sum_{k} \gamma_{ik} \mathcal{N}(y_{i}; \mu_{k} + W_{k} \widehat{x}_{ik}, \Sigma_{ik}),$$

where  $\Sigma_{ik} = \sigma_k^2 I_{D \times D} + W_k (S_{ik} + \hat{x}_{ik} \hat{x}_{ik}^T) W_k^T$ . Due to the high-dimensional nature of the data, most cluster membership



Fig. 4. **Representative restorations in the ADNI dataset**. Reconstruction by NLM, linear interpolation, and our method, and the original high resolution images for two representative subjects in the study. Our method reconstructs more anatomically plausible substructures as can be especially seen in the close-up panels of the skull, ventricles, and temporal lobe. Additional examples are available in the Supplementary Materials (supplementary files / multimedia tab).

estimates are very close to 0 or 1. We therefore first estimate the most likely cluster  $\hat{k}$  for patch  $y_i$  by selecting the cluster with the highest membership  $\gamma_{ik}$ . We estimate the low dimensional representation  $\hat{x}_{i\hat{k}}$  given the observed voxels  $y_i^{\mathcal{O}_i}$  using (7), which yields the high resolution imputed patch:

$$\hat{y}_i = \mu_{\hat{k}} + W_{\hat{k}} \hat{x}_{i\hat{k}}.$$
(16)

By restoring the scans using this MAP solution, we perform *conditional mean imputation* (c.f. 17, Sec.4.2.2), and demonstrate the reconstructions in our experiments. In addition, our model enables imputation of each patch by sampling the posterior  $p(\cdot|y_i^{\mathcal{O}_i}; \theta) \approx \mathcal{N}(\cdot; \mu_{\hat{k}} + W_{\hat{k}}\hat{x}_{i\hat{k}}, \Sigma_{i\hat{k}})$ , providing a better estimation of the residual noise. Depending on the desired downstream application, sampling-based imputation may be desired.

We average overlapping restored patches using standard techniques [18] to form the restored volume.

#### **III. IMPLEMENTATION**

We work in the atlas space, and approximate voxels as either observed or missing in this space by thresholding interpolation weights. To limit interpolation effects due to affine alignment on the results, we set a higher threshold for regions with high image gradients than in regions with low gradients. Parameter estimation could be implemented to include transformation of the model parameters into the subject-specific space in order to optimally use the observed voxels, but this leads to computationally prohibitive updates. We stack together the affinely registered sparse images from the entire collection. We learn a single set of mixture model parameters within overlapping subvolumes of  $21 \times 21 \times$ 21 voxels in the isotropically sampled common atlas space. Subvolumes are centered 11 voxels apart in each direction. We use a cubic patch of size  $11 \times 11 \times 11$  voxels, and instead of selecting just one patch from each volume at a given location, we collect all overlapping patches within the subvolume centered at that location. This aggregation provides more data for each model, which is crucial when working with severely undersampled volumes. Moreover, including nearby voxels offers robustness in the face of image misalignment. Given the learned parameters at each location, we restore all overlapping patches within a subvolume.

While learning is performed in the common atlas space, we restore each volume in its original image space to limit the effects of interpolation. Specifically, we apply the inverse of the estimated subject-specific affine transformation to the cluster statistics prior to performing subject-specific inference.

Our implementation is freely available at https://github.com/ adalca/papago.

## **IV. EXPERIMENTS**

We demonstrate the proposed imputation algorithm on two datasets and evaluate the results both visually and quantitatively. We also include an example of how imputation can aid in a skull stripping task.



Fig. 5. **Reconstruction accuracy statistics**. Accuracy for different image restoration methods (top), and improvement over nearest neighbor interpolation using MSE (bottom). All statistics were computed over 50 scans randomly chosen from the ADNI dataset. Image intensities are scaled to a [0,1] range.

## A. Data: ADNI Dataset

We evaluate our algorithm using 826 T1-weighted brain MR images from ADNI [15].<sup>1</sup> We downsample the isotropic  $1 \text{mm}^3$  images to slice separation of 6mm ( $1 \text{mm} \times 1 \text{mm}$  in-plane) in the axial direction to be of comparable quality with the clinical dataset. We use these low resolution images as input. All downsampled scans are affinely registered to a T1 atlas. The original images serve as the ground truth for quantitative comparisons. After learning model parameters using the data set, we evaluate the quality of the resulting imputations.

#### B. Evaluation

We compare our algorithm to three upsampling methods: nearest neighbour (NN) interpolation, non-local means (NLM)



Fig. 6. Regions used for hyper-parameter analysis. Representative example of four subvolumes used in analyses, shown in saggital, coronal and axial views.

upsampling, and linear interpolation [19]. We compare the reconstructed images to the original isotropic volumes both visually and quantitatively. We use the mean squared error,

MSE 
$$(Z, Z_o) = \frac{1}{N} \sum ||Z - Z_o||^2$$
, (17)

of the reconstructed image Z relative to the original high resolution scan  $Z_o$ . We also compute the related peak signal to noise ratio,

$$PSNR = \log_{10} \frac{\max(Z_o)}{MSE(Z, Z_o)}.$$
 (18)

Both metrics are commonly used in measuring the quality of reconstruction of compressed or noisy signals.

#### C. Results

Fig. 4 illustrates representative restored images for subjects in the ADNI dataset. Our method produces more plausible structure. The method restores anatomical structures that are almost entirely missing in the other reconstructions, such as the dura or the sulci of the temporal lobe by learning about these structures from the image collection. We provide additional example results in the Supplementary Materials in the supplementary files / multimedia tab.

Fig. 5 reports the error statistics in the ADNI data. Due to high variability of MSE among subject scans, we report improvements of each method over the nearest neighbor interpolation baseline in the same scan. Our algorithm offers significant improvement compared to nearest neighbor, NLM, and linear interpolation ( $p \le 10^{-5}, 10^{-42}, 10^{-27}$ , respectively). Our method performs significantly better on all subjects. The improvement in MSE is observed in every single scan. Similarly, our method performs consistently better using the PSNR metric (not shown), with mean improvements of up to  $1.4 \pm 0.44$  compared to the next best restored scan.

## D. Parameter Setting

We analyze the performance of our algorithm while varying the values of the parameters, and the sparsity patterns of the observed voxels. For these experiments, we use four distinct

<sup>&</sup>lt;sup>1</sup>Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).



Fig. 7. **Mask Analysis**. Top: example masks for each simulation, shown in the saggital plane. The first experiment reflects the limited variability of axial-only acquisitions, whereas the second and third experiments represent increasingly more varied patterns of observed voxels. Bottom: imputation errors. More varied masks leads to improved reconstructions.

subvolumes that encompass diverse anatomy from ADNI data, as illustrated in Fig. 6. We start with isotropic data and use different observation masks as described in each experiment.

1) Hyper-Parameters: We evaluate the sensitivity of our method under different hyper parameters: the number of clusters,  $k \in [1, 2, 5, 10, 15]$  and the number of dimensions of the low dimensional embedding  $d \in [10, 20, 30, 40, 50]$ . While different regions give optimal results with different settings, overall our algorithm produces comparable results for the middle range of these parameters. We run all of our experiments with k = 5 and d = 30.

2) Sparsity Patterns: First, we evaluate how our algorithm performs under three different mask patterns, all of which allow for the same number of observed voxels. Specifically, we (i) use the true sparsely observed planes as in the first experiment; (ii) simulate random rotations of the observation planes mimicking acquisitions in different directions; and (iii) simulate random mask patterns. The latter setup is useful for denoising or similar tasks, and is instructive of the performance of our algorithm. Fig. 7 demonstrates that our algorithm performs better under acquisition with different directions, and similarly under truly random observations as more entries of the cluster covariance matrices are directly observed.



Fig. 8. Slice thickness simulation. Top: saggital close-up of a region where axial slices were blurred in the direction perpendicular to the acquisition direction; followed by respective imputed results. Bottom: performance of our algorithm under different slice thickness simulations are shown MSE (solid line) and standard deviation interval (shaded region).

This demonstrates a promising application of this model to other settings where different patterns of image voxels are observed.

3) Slice Thickness: We also investigate the effects of slice thickness on the results. The model treats the original data as high resolution planes. Here, we simulate varying slice thickness by blurring isotropic data in the direction perpendicular to the slice acquisition direction. We then use the sampling masks of the scans used in the main experiments to identify observed, albeit blurred, voxels. Fig. 8 shows that although the algorithm performs worse with larger slice thickness, it provides plausible imputation results. For example, results show minimal noticeable differences, even for a blur kernel of  $\sigma =$  1mm, simulating a slice with significant signal contribution from 4mm away. Our method, which treats observed slices as thin, is nevertheless robust to slice thicknesses variations.

## E. Skull Stripping

We also illustrate how imputed data might facilitate downstream image analysis. Specifically, the first step in many analysis pipelines is brain extraction – isolating the brain from the rest of the anatomy. Typical algorithms assume that the brain consists of a single connected component separated from



Fig. 9. Skull Stripping Example. Example of a skull stripping failure for linear and NLM interpolation. Skull stripping dramatically improves when applied to the imputed image for this example.

the skull and dura by cerebral spinal fluid [29]. Thus, they often fail on sparsely sampled scans that no longer include a clear contrast between these regions. Fig. 9 provides an example where the brain extraction fails on the original subject scan but succeeds on our reconstructed image.

# F. Clinical Dataset

We also demonstrate our algorithm on a clinical set of 766 T2-FLAIR brain MR scans in a stroke patient cohort. These scans are severely anisotropic  $(0.85 \times 0.85 \text{mm in-plane}, \text{slice separation of 6mm})$ . All subjects are affinely registered to an atlas and the intensity is normalized.

Fig. 10 illustrates representative restoration improvements in T2-FLAIR scans from a clinical population. Our method produces more plausible structure, as can be especially seen in the close-up panels focusing on anatomical details. We provide additional example results in the Supplementary Materials in the supplementary files/multimedia tab.

# V. DISCUSSION

# A. Modeling Choices

We explicitly model and estimate a latent low-dimensional embedding for each patch. The likelihood model (5) does not include the latent patch representation  $x_i$ , leading to observed patch covariance  $C_k^{\mathcal{O}_i,\mathcal{O}_i} = W_k^{\mathcal{O}_i}(W_k^{\mathcal{O}_i})^T + \sigma_k^2 I$ . Since the set of observed voxels  $\mathcal{O}_i$  varies across subjects, the resulting Expectation Maximization algorithm [6] becomes intractable if we marginalize the latent representation out before estimation. Introducing the latent structure simplifies the optimization problem.

We investigated an alternative modeling choice that instead treats each missing voxel as a latent variable. In particular we consider the missing values of patch  $y_i$  as latent variables, which can be optionally combined with the latent low-dimensional patch representation. These assumptions lead to an Expectation Conditional Maximization (ECM) [14], [18], a variant of the Generalized Expectation Maximization where parameter updates depend on the previous parameter estimates. The resulting algorithm estimates the expected missing voxel mean and covariance directly, and then updates the cluster parameters (see Supplementary Materials for a complete derivation). The most notable difference between this formulation and simpler algorithms that iteratively fill in missing

voxels and then estimate GMM model parameters is in the estimation of the expected data covariance, which captures the covariance of the missing and observed data ([18], Ch.8). We found that compared to the method presented in Section II, this variant often got stuck in local minima, had difficulty moving away from the initial missing voxel estimates, and was an order of magnitude slower than the presented method. We provide both implementations in our code.

## **B.** Restoration

Our restoration method assumes that the observed voxels are noisy manifestations of the low dimensional patch representation, and reconstructs the entire patch, including the observed voxels, leading to smoother images. This formulation assumes the original observed voxels are noisy observations of the true data. Depending on the downstream analyses, the original voxels could be kept in the reconstruction. In addition, we also investigated an alternative reconstruction method of filling in the missing voxels given the observed voxels as noiseless ground truth (not shown). This formulation leads to sharper but noisier results. The two restoration methods therefore yield images with different characteristics. This tradeoff is a function of the noise in the original acquisition: higher noise in the clinical acquisition leads to noisier reconstructions using the alternative method, whereas in the ADNI dataset the two methods perform similarly. In addition, imputation can be achieved by sampling the posterior distribution rather than using conditional mean estimation, enabling a better estimate of the residual noise for downstream analysis.

# C. Usability

Our model assumes that whether a voxel is observed is independent of the intensity of that voxel. Although the voxels missing in the sparsely sampled images clearly form a spatial pattern, we assume there is no correlation with the actual intensity of the voxels. The model can therefore be learned from data with varying sparseness patterns, including restoring data in all acquisition directions simultaneously.

The proposed method can be used for general image imputation using datasets of varying resolution. For example, although acquiring a large high resolution dataset for a clinical study is often infeasible, our algorithm will naturally make use of any additional image data available. Even a small number of



Fig. 10. Representative restorations in the clinical dataset. Reconstruction using NLM, linear interpolation and our method for two representative subjects. Our method reconstructs more plausible substructures, as can be especially seen in the close-up panels of the skull and the periventricular region. Additional examples are available in the Supplementary Materials in the supplementary files / multimedia tab.

acquisitions in different directions or higher resolution than the study scans promise to improve the accuracy of the resulting reconstruction.

The presented model depends on the image collection containing similar anatomical structures roughly aligned, such as affinely aligned brain or cardiac MR scans. Smaller datasets that contain vastly different scans, such as traumatic brain injuries or tumors, may not contain enough consistency to enable the model to learn meaningful covariation. However, a wide range of clinical datasets contain the anatomical consistency required, and can benefit from the proposed method.

# D. Initialization

We experimented with several initialization schemes, and provide them in our implementation. A natural initialization is to first learn a simple GMM from the linearly interpolated volumes, and use the resulting parameter values as initializations for our method. This leads to results that improve on the linear interpolation but still maintain somewhat blocky effects caused by interpolation. More agnostic initializations, such as random parameter values, lead to more realistic anatomy but noisier final estimates. Different methods perform well in different regions of the brain. The experimental results are initialized by first learning a simple GMM from the linearly interpolated volumes, and using the resulting means with diagonal covariances as an initial setting of the parameters. We start with a low dimensional representation to be of dimension 1, and grow it with every iteration up to the desired dimension. We found that this approach outperforms all other strategies.

# **VI. CONCLUSIONS**

We propose an image imputation method that employs a large collection of low-resolution images to infer fine-scale anatomy of a particular subject. We introduce a model that captures anatomical similarity across subjects in large clinical image collections, and imputes, or fills in, the missing data in low resolution scans. The method produces anatomically plausible volumetric images consistent with sparsely sampled input scans.

Our approach does not require high resolution scans or expert annotations for training. We demonstrate that our algorithm is robust to many data variations, including varying slice thickness. The resulting method enables the use of untapped clinical data for large scale scientific studies and promises to facilitate novel clinical analyses.

# APPENDIX EXPECTATION MAXIMIZATION UPDATES

Following (5), the complete likelihood of our model is:

$$p(\mathcal{Y}^{\mathcal{O}}, \mathcal{X}; \theta) = \prod_{i} \sum_{k} \pi_{k} \mathcal{N}(y_{i}^{\mathcal{O}_{i}}, x_{i}; \mu_{k}^{\mathcal{O}_{i}}, C_{k}^{\mathcal{O}_{i}\mathcal{O}_{i}}), \quad (19)$$

where  $\mathcal{X} = \{x_i\}$ . The expectation of this probability is then

$$Q(\theta|\theta) = \mathbf{E}_{\mathcal{X}|\mathcal{Y}^{\mathcal{O}},\theta} \left[ \log p(\mathcal{Y}^{\mathcal{O}}, \mathcal{X}; \theta) \right]$$
$$= \sum_{i,k} \mathbf{E}[k(-\frac{d}{2}\log 2\pi - \frac{1}{2}\log |\sigma_k^2 I| -\frac{1}{2\sigma_k^2}(y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}}x_i - \mu_k^{\mathcal{O}_i})^T(y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}}x_i - \mu_k^{\mathcal{O}_i}) -\frac{d}{2}\log 2\pi - \frac{1}{2}\log |I| - \frac{1}{2}x_i x_i^T)].$$
(20)

Computing this expectation requires evaluating  $\mathbb{E}[k]$ ,  $\mathbb{E}[x_i|k]$ , and  $\mathbb{E}[x_i x_i^T|k]$ , which is trivially done to obtain the expectation step updates (6)-(8).

For the maximization step, we optimize (20) with respect to the model parameters.

$$\frac{\partial Q}{\partial \mu_{k}} \propto \sum_{i \in \mathcal{P}_{j}} \frac{\partial}{\partial \mu_{k}} \mathbf{E} \Big[ k(y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}}x_{i} - \mu_{k}^{\mathcal{O}_{i}})^{T} \\ \times (y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}}x_{i} - \mu_{k}^{\mathcal{O}_{i}}) \Big] \\ \propto \sum_{i \in \mathcal{P}_{j}} \mathbf{E} \Big[ k(y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}}x_{i} - \mu_{k}^{\mathcal{O}_{i}}) \Big] \\ = \sum_{i \in \mathcal{P}_{j}} \gamma_{ik}(y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}}\widehat{x}_{ik} - \mu_{k}^{\mathcal{O}_{i}}) = 0 \\ \mu_{k}^{j} = \frac{1}{\sum_{i'} \gamma_{i'k}} \sum_{i \in \mathcal{P}_{j}} \gamma_{ik}(y_{i}^{j} - W_{k}^{j}\widehat{x}_{ik})$$
(21)

 $\frac{\partial Q}{\partial W_{L}}$ 

$$\propto \sum_{i \in \mathcal{P}_{j}} \frac{\partial \mathbf{E} \left[ k(y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}} x_{i} - \mu_{k}^{\mathcal{O}_{i}})^{T} (y_{i}^{\mathcal{O}_{i}} - W_{k}^{\mathcal{O}_{i}} x_{i} - \mu_{k}^{\mathcal{O}_{i}}) \right] }{\partial W_{k}}$$

$$\propto \sum_{i \in \mathcal{P}_{j}} \mathbf{E} \left[ k(W_{k}^{\mathcal{O}_{i}} x_{i} - (y_{i}^{\mathcal{O}_{i}} - \mu_{k}^{\mathcal{O}_{i}})) x_{i}^{T} \right]$$

$$= \sum_{i \in \mathcal{P}_{j}} \gamma_{ik} W^{\mathcal{O}_{i}} (\widehat{x}_{ik} \widehat{x}_{ik}^{T} + S_{ik}) - (y_{i}^{\mathcal{O}_{i}} - (\mu_{k}^{\mathcal{O}_{i}})) \widehat{x}_{ik}^{T} = 0$$

$$W_{k}^{j} = \left[ \sum_{i \in \mathcal{P}_{j}} \gamma_{ik} (\widehat{x}_{ik} \widehat{x}_{ik}^{T} + S_{ik}) \right]^{-1} \sum_{i \in \mathcal{P}_{j}} \gamma_{ik} (y_{i}^{j} - \mu_{k}^{j}) \widehat{x}_{ik}^{T}$$

$$= \sum_{i \in \mathcal{P}_{j}} \delta_{ik} (y_{i}^{j} - \mu_{k}^{j}) \widehat{x}_{ik}^{T} A_{j}^{-1}$$

$$(22)$$

where  $\delta$  and A are defined in (9) and (11), respectively. By combining (22) and (21), we obtain

$$\mu_k^j \sum_i \gamma_{ik} = \sum_{i \in \mathcal{P}_j} \gamma_{ik} y_i^j - W_k^j \sum_{i \in \mathcal{P}_j} \gamma_{ik} \widehat{x}_{ik}$$

$$\mu_{k}^{j} = \sum_{i \in \mathcal{P}_{j}} \delta_{ik} y_{i}^{j} - W_{k}^{j} b_{j}$$

$$\mu_{k}^{j} = \sum_{i \in \mathcal{P}_{j}} \delta_{ik} y_{i}^{j} - \sum_{i \in \mathcal{P}_{j}} \delta_{ik} (y_{i}^{j} - \mu_{k}^{j}) \widehat{x}_{ik}^{T} A_{j}^{-1} b_{j}$$

$$\mu_{k}^{j} (1 - \sum_{i \in \mathcal{P}_{j}} \delta_{ik} \widehat{x}_{ik}^{T} A_{j}^{-1} b_{j})$$

$$= \sum_{i \in \mathcal{P}_{j}} \delta_{ik} y_{i}^{j} (1 - \widehat{x}_{ik}^{T} A_{j}^{-1} b_{j})$$

$$\mu_{k}^{j} = \frac{\sum_{i \in \mathcal{P}_{j}} \delta_{ik} y_{i}^{j} (1 - \widehat{x}_{ik}^{T} A_{j}^{-1} b_{j})}{\sum_{i \in \mathcal{P}_{j}} \delta_{ik} (1 - \widehat{x}_{ik}^{T} A_{j}^{-1} b_{j})}.$$
(23)

We therefore update  $\mu_k^j$  via (23), followed by  $W_k^j$  using (22). Finally,

$$\frac{\partial Q}{\partial \sigma_k^2} \propto \sum_{i \in \mathcal{P}_j} -\frac{\partial}{\partial \sigma_k^2} \frac{1}{\sigma_k^2}$$

$$\mathbf{E} \Big[ k(y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i})^T (y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i}) \Big]$$

$$-\frac{\partial}{\partial \sigma_k^2} N \log \sigma_k^2$$

$$= \sum_{i \in \mathcal{P}_j} \frac{1}{\sigma_k^4} \mathbf{E} \Big[ k(y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i})^T$$

$$\times (y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i}) \Big] - \frac{N}{\sigma_k^2} = 0$$

$$\sigma_k^2 = \sum_{i \in \mathcal{P}_j} \mathbf{E} \Big[ k(y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i})^T$$

$$\times (y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i})^T$$

$$\times (y_i^{\mathcal{O}_i} - W_k^{\mathcal{O}_i} x_i - \mu_k^{\mathcal{O}_i}) \Big]$$

$$\sigma_k^2 = \frac{\sum_j \sum_{i \in \mathcal{P}_j} \gamma_{ik} \Big[ (y_i^j - \mu_k^j - W_k^j \hat{x}_{ik})^2 + W_k^j S_{ik} (W_k^j)^T \Big]}{\sum_j \sum_{i \in \mathcal{P}_j} \gamma_{ik}}.$$
(24)

#### REFERENCES

- K. K. Bhatia, A. Rao, A. N. Price, R. Wolz, J. V. Hajnal, and D. Rueckert, "Hierarchical manifold learning for regional image analysis," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 444–461, Feb. 2014.
- [2] E. Carmi, S. Liu, N. Alon, A. Fiat, and D. Fiat, "Resolution enhancement in MRI," *Magn. Reson. Imag.*, vol. 24, no. 2, pp. 133–154, 2006.
- [3] P. Coupé, J. V Manjón, V. Fonov, J. Pruessner, M. Robles, and L. D. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [4] A. V. Dalca, A. Bobu, N. S. Rost, and P. Golland, "Patch-based discrete registration of clinical brain images," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.* Cham, Switzerland: Springer, 2016, pp. 60–67.
- [5] A. V. Dalca, K. L. Bouman, W. T. Freeman, N. S. Rost, M. R. Sabuncu, and P. Golland, "Population based image imputation," in *Information Processing in Medical Imaging.* Cham, Switzerland: Springer, 2017.
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Roy. Stat. Soc. B, Methodol., vol. 39, no. 1, pp. 1–38, 1977.
- [7] A. D. Martino *et al.*, "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
- [8] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 184–199.

- IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 38, NO. 2, FEBRUARY 2019
- [9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [10] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based superresolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.
- [11] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 349–356.
- [12] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, no. 3, pp. R1–R45, 2001.
  [13] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker,
- [13] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, and B. Fischl, "Is synthesizing mri contrast useful for inter-modality analysis?" in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 8149. Berlin, Germany: Springer, 2013, pp. 631–638.
- [14] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 1957–2000, 2010.
- [15] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.
- [16] A. Jog, A. Carass, and J. L. Prince, "Improving magnetic resonance resolution with supervised learning," in *Proc. IEEE ISBI*, Apr./May 2014, pp. 987–990.
- [17] E. Konukoglu, A. van der Kouwe, M. R. Sabuncu, and B. Fischl, "Example-based restoration of high-resolution magnetic resonance image acquisitions," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 8149. Berlin, Germany: Springer, 2013, pp. 131–138.
- [18] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*. Hoboken, NJ, USA: Wiley, 2014.
- [19] J. V. Manjón, P. Coupé, A. Buades, V. Fonov, D. L. Collins, and M. Robles, "Non-local MRI upsampling," *Med. Image Anal.*, vol. 14, no. 6, pp. 784–792, 2010.
- [20] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity," *Med. Image Anal.*, vol. 16, no. 1, pp. 18–27, 2012.
- [21] O. Oktay et al., "Multi-input cardiac image super-resolution using convolutional neural networks," in Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent. Cham, Switzerland: Springer, 2016, pp. 246–254.
- [22] E. Plenge, D. H. J. Poot, W. J. Niessen, and E. Meijering, "Superresolution reconstruction using cross-scale self-similarity in multislice MRI," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 8151. Berlin, Germany: Springer, 2013, pp. 123–130.

- [23] N. S. Rost *et al.*, "White matter hyperintensity burden and susceptibility to cerebral ischemia," *Stroke*, vol. 41, no. 12, pp. 2807–2811, 2010.
- [24] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1852–1862, Oct. 2011.
- [25] F. Rousseau and Alzheimer's Disease Neuroimaging Initiative, "A non-local approach for image super-resolution using intermodality priors," *Med. Image Anal.*, vol. 14, no. 4, pp. 594–605, 2010.
- [26] F. Rousseau, K. Kim, and C. Studholme, "A groupwise super-resolution approach: application to brain MRI," in *Proc. IEEE ISBI*, Apr. 2010, pp. 860–863.
- [27] P. Schmidt *et al.*, "An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis," *NeuroImage*, vol. 59, no. 4, pp. 3774–3783, 2012.
- [28] I. J. Schoenberg, Cardinal Spline Interpolation. Philadelphia, PA, USA: SIAM, 1973.
- [29] F. Ségonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage*, vol. 22, no. 3, pp. 1060–1075, 2004.
- [30] W. Shi et al., "Cardiac image super-resolution with global correspondence using multi-atlas PatchMatch," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 8151. Berlin, Germany: Springer, 2013, pp. 9–16.
- [31] R. Sridharan *et al.*, "Quantification and analysis of large multimodal clinical image studies: Application to stroke," *Multimodal Brain Image Analysis*, (Lecture Notes in Computer Science), vol. 8159. Cham, Switzerland: Springer, 2013, pp. 18–30.
- [32] K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 677–688, Aug. 2001.
- [33] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [34] R. Zhang, C. A. Bouman, J.-B. Thibault, and K. D. Sauer, "Gaussian mixture Markov random field for image denoising and reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2013, pp. 1089–1092.
- [35] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 479–486.
  [36] D. Zoran and Y. Weiss, "Natural images, Gaussian mixtures and
- [36] D. Zoran and Y. Weiss, "Natural images, Gaussian mixtures and dead leaves," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1736–1744.