

# Probabilistic Dipole Inversion for Adaptive Quantitative Susceptibility Mapping

Jinwei Zhang

Department of Biomedical Engineering, Cornell University, Ithaca, NY, USA  
Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

jz853@cornell.edu

Hang Zhang

Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA  
Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

hz459@cornell.edu

Mert Sabuncu

Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA  
Department of Biomedical Engineering, Cornell University, Ithaca, NY, USA  
Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

msabuncu@cornell.edu

Pascal Spincemaille

Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

pas2018@med.cornell.edu

Thanh Nguyen

Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

tdn2001@med.cornell.edu

Yi Wang

Department of Biomedical Engineering, Cornell University, Ithaca, NY, USA  
Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA

yiwang@med.cornell.edu

## Abstract

A learning-based posterior distribution estimation method, Probabilistic Dipole Inversion (PDI), is proposed to solve the quantitative susceptibility mapping (QSM) inverse problem in MRI with uncertainty estimation. In PDI, a deep convolutional neural network (CNN) is used to represent the multivariate Gaussian distribution as the approximate posterior distribution of susceptibility given the input measured field. Such CNN is first trained on healthy subjects via posterior density estimation, where the training dataset contains samples from the true posterior distribution. Domain adaptations are then deployed on patient datasets with new pathologies not included in pre-training, where PDI updates the pre-trained CNN's weights in an unsupervised fashion by minimizing the *Kullback-Leibler* divergence between the approximate posterior distribution represented by CNN and the true posterior distribution from the likelihood distribution of a known physical model and pre-defined prior distribution. Based on our experiments, PDI provides additional uncertainty estimation compared to the conventional MAP approach, while addressing the potential issue of the pre-trained CNN when test data deviates from training. Our code is available at [https://github.com/Jinwei1209/Bayesian\\_QSM](https://github.com/Jinwei1209/Bayesian_QSM).

**Keywords:** Variational Inference, Uncertainty Estimation, Convolutional Neural Network, Quantitative Susceptibility Mapping

## 1. Introduction

Quantitative susceptibility mapping (QSM) (de Rochefort et al., 2010) is an image contrast in magnetic resonance imaging (MRI) to measure the underlying tissue apparent magnetic susceptibility, which enables quantification of specific biomarkers such as iron, calcium and

gadolinium (Wang and Liu, 2015). The forward model of QSM in three dimensional image space is:

$$b = d * \chi + n \quad (1)$$

where  $\chi$  is the tissue susceptibility,  $b$  is the measured local magnetic field,  $d$  is the spatial dipole convolution kernel, and  $n$  is the measurement noise. Dipole convolution can also be defined in k-space (Fourier space) as follows:

$$b = F^H DF\chi + n \quad (2)$$

where  $F$  is the Fourier transform operator and  $D$  is the point-wise multiplication operator with the dipole kernel in k-space. The k-space formulation is more computationally efficient because of the fast Fourier transform, so Eq. 2 is often implemented in practice. The standard deviation (SD) of the Gaussian noise  $n$  is obtained by computing the variance of the least squares fit of the magnetic field  $b$  from the acquired multi-echo data (Liu et al., 2013). The problem is to recover  $\chi$  from  $b$  due to the ill-posedness of the dipole kernel in QSM (Wang and Liu, 2015; Kee et al., 2017).

Two representative methods have been proposed to solve the QSM inverse problem. The first one is called COSMOS (calculation of susceptibility through multiple orientation sampling) (Liu et al., 2009). COSMOS relies on multiple orientation scans to calculate the susceptibility map with high accuracy. As a result, it has been used as the gold standard reference when developing new QSM algorithms. However, the drawback of COSMOS is that it requires at least three orientation scans, which is infeasible for clinical use. Another method called MEDI (morphology enabled dipole inversion) (Liu et al., 2011b) was proposed to solve the QSM problem with a single orientation scan. MEDI uses a morphology-related regularization term and solves the following optimization problem:

$$\hat{\chi} = \arg \min_{\chi} \|W(F^H DF\chi - b)\|_2^2 + \lambda \|M\nabla\chi\|_1 \quad (3)$$

where  $W$  is derived from the observation noise covariance matrix,  $\lambda$  is the tunable parameter of weighted total variation (TV) regularization (Rudin et al., 1992) with binary weighting matrix  $M$  of susceptibility's spatial gradients which only penalizes regions outside the brain tissue edges in order to suppress image-space artifacts introduced by dipole inversion (Liu et al., 2011b). Numerical optimization algorithms for solving Eq. 3 are reviewed by Kee et al. (2017). With efficient numerical solvers, MEDI generates reasonable susceptibility maps compared to COSMOS as a reference (Liu et al., 2011b) and requires only single orientation scan. As a result, MEDI has been used for clinical applications in the last decade (Wang et al., 2017; Chen et al., 2014).

From the Bayesian point of view, Eq. 3 belongs to the maximum a posteriori probability (MAP) estimation problem with the likelihood distribution defined as a multivariate Gaussian:

$$p(b|\chi) = \mathcal{N}(b|F^H DF\chi, \Sigma_{b|\chi}) \quad (4)$$

where  $n \sim \mathcal{N}(0, \Sigma_{b|\chi})$  with  $\Sigma_{b|\chi}$  diagonal, and the prior distribution defined as the Laplace of the spatial gradient:

$$p(\chi) \propto e^{-\lambda \|M\nabla\chi\|_1}. \quad (5)$$

Based on Bayes’s rule, the full posterior distribution  $p(\chi|b)$  given the measured local field  $b$  can also be estimated in principle, which will quantify the uncertainty in the solutions delivered and may have some clinical implications. In this paper, motivated by the posterior distribution estimation problem in QSM and advances in deep learning based density estimation techniques, we introduce a set of neural network parameterized distributions to learn an approximate posterior distribution of susceptibility  $\chi$  for any given  $b$  with an adaptive training strategy. We validate our method on both healthy subjects and patients and show good performance of the proposed method. This paper is extended from previously published work (Zhang et al., 2020b) at MIDL 2020. The additions include a detailed methodology section, comparisons to PDI-VI0 as another baseline in Figures 2-4 and Table 1, an experiment on multiple sclerosis patients in Figure 3, amortized versus subject-specific variational inference in Figure 5 and 6, uncertainty estimation evaluation in Figure 7, and the discussion section.

## 2. Related Work

In recent years, posterior distribution estimation in imaging inverse problems has become a new topic in medical imaging (Repetti et al., 2019; Chappell et al., 2009), in which variance of a random variable is provided from posterior distribution to measure the uncertainty of the solution. However, posterior distribution estimation requires a complicated or even intractable integral from Bayes formula. Therefore, approximate inference methods are used to reduce the computational cost and intractability of the problem. Markov chain Monte Carlo (MCMC) (Andrieu et al., 2003) and variational inference (VI) (Bishop, 2006) are two classes of approximate inference approaches to the Bayesian estimation problem. In MCMC, Markov chain based sampling methods are used to generate random samples from the true posterior distribution in order to represent an empirical distribution which resembles the true distribution. MCMC is general in that it is able to achieve the exact inference given infinite computational time. However, in imaging inverse problems, the computational cost of MCMC for Bayesian estimation is often several magnitudes higher than that of the optimization method of MAP estimation, because of the curse of dimensionality (Pereyra, 2017). In addition, convergence of the Markov chain is hard to diagnose, raising concerns on the quality of the samples.

An alternative approach is to use VI, in which an approximate distribution is proposed with tractable function form and unknown parameters, and an optimization algorithm is used (for example, expectation-maximization (EM) algorithm (Blei et al., 2017)) to learn these parameters by minimizing the divergence between the true and approximate posterior distributions. After convergence, the approximate posterior distribution is expected to represent the true posterior distribution. Compared to MCMC, VI is fairly efficient as the inference problem is reduced to the optimization problem with respect to the distribution parameters. However, VI may make the model less expressive and thus lead to suboptimal performance. Although more complicated approximate function has a better representation ability in some cases, it introduces higher computational cost. Such accuracy-computation trade-off cannot be achieved easily as the inference performance depends on the tricky design of the approximate distribution form.

Due to advances in deep learning in the past few years, using deep neural network as the approximate function has become a new trend in VI. This is especially true for generative models such as variational auto-encoder (VAE) (Rezende et al., 2014; Kingma and Welling, 2013), in which an encoder network is built to approximate the latent variable distribution conditioned on the observed data and a decoder network is built to represent the observed data distribution conditioned on the latent variable. In addition, because of the generalization ability of a deep neural network with millions of trainable weights, amortized formulation with regularization is applied on the training dataset to learn the network weights for faster inference on the test dataset than classic VI per data, but at the expense of lower precision (Cremer et al., 2018). As a result, this leads to a new trade-off between inference speed and amortization accuracy.

Another topic related to posterior distribution estimation with deep learning are the deep generative models trained with maximum likelihood, such as autoregressive (Oord et al., 2016) and flow models (Dinh et al., 2016). In these models, neural network parameterized models are built to deploy tractable maximum likelihood training and generate new samples after training. If the parameterized model family is highly expressive with enough training samples, maximum likelihood training is expected to learn parameters which fit to the true data density well and generate new data with high fidelity. Autoregressive and flow models differ from VAE in that exact likelihood is evaluated in the former while approximate evaluation is applied for the latter. Such tractable inference makes training simpler but models less expressive, except for flow models which provide a combination of tractability and high expressiveness.

In this work, we propose to solve the posterior distribution estimation problem in QSM using a neural network parameterized distribution family by combining posterior density estimation from samples with posterior distribution approximation via VI for domain adaptation. Assuming multivariate Gaussian represented by a CNN as the posterior distribution of susceptibility given the input local field, a COSMOS (Liu et al., 2009) dataset of field susceptibility pairs were used as samples from the true posterior distribution to train such CNN with an MAP loss function. Applying the likelihood in Eq. 4 and prior in Eq. 5, the pre-trained CNN was adapted using VI posterior distribution approximation on different patient datasets which only contained input measured fields. Compared to MAP estimation MEDI (Liu et al., 2011b) in Eq. 3 and other deep learning QSM methods, QSMnet (Yoon et al., 2018) and FINE (Zhang et al., 2020a), the proposed method estimated the full posterior distribution of susceptibility with uncertainty quantification, while achieving domain adaptations on various datasets.

### 3. Methodology

Based on the assumption that the pattern from field  $b$  to  $p(\chi|b)$  is recoverable, a general distribution  $p_{data}(\chi|b)$  for any given  $b$  can be approximated with a learning-based approach. To accomplish that, a set of parameterized distributions  $q_\psi(\chi|b)$  using a neural network with parameters  $\psi$  are introduced and learned on a cohort of datasets including healthy subjects and patients. In this work, we assume a multivariate Gaussian distribution with diagonal covariance matrix as the approximate posterior distribution, i.e.,  $q_\psi(\chi|b) = \mathcal{N}(\mu_{\chi|b}, \Sigma_{\chi|b})$ , and use a dual-decoder network architecture (Figure 1) extended from 3D U-Net (Ron-

neberger et al., 2015; Çiçek et al., 2016) to represent  $q_\psi(\chi|b)$ , with dual decoders’ outputs representing mean  $\mu_{\chi|b}$  and variance  $\Sigma_{\chi|b}$  maps.

### 3.1 Posterior Density Estimation

The modeling process consists of two steps. The first step employs the COSMOS dataset. Since COSMOS provides gold standard QSM images based on multiple orientation scans, we can treat COSMOS field-susceptibility data pairs as the samples from the true posterior data distribution. Given the large amount of samples, they define an empirical distribution as follows:

$$\hat{p}_{data}(\chi|b) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}[\chi = \chi_i|b = b_i] \quad (6)$$

where  $(b_i, \chi_i)$  is the  $i$ -th susceptibility-field data pair sampled from  $p_{data}(\chi|b)$  with total  $N$  samples, and  $\mathbf{1}[\cdot]$  is the indicator function. We use *Kullback–Leibler* (KL) divergence as the loss function to measure the distance between the empirical distribution defined by the COSMOS samples and the parameterized approximate distribution defined by the network, i.e.,  $KL[\hat{p}_{data}(\chi|b)||q_\psi(\chi|b)]$ , which is equivalent to the following loss function:

$$KL[\hat{p}_{data}(\chi|b)||q_\psi(\chi|b)] = \frac{1}{N} \sum_{i=1}^N -\log q_\psi(\chi_i|b_i) + H(\hat{p}_{data}) \quad (7)$$

where the first term computes the expectation of negative log posterior density with respect to the empirical distribution and the second term is the entropy of the empirical distribution. Since the second term does not include the learnable parameters  $\psi$ , only the first term is used during parameter learning. Notice that training using this loss function is equivalent to maximizing the parametrized approximate posterior distribution by fitting to the dataset. Inserting  $q_\psi(\chi|b) = \mathcal{N}(\mu_{\chi|b}, \Sigma_{\chi|b})$  into the first term of Eq. 7 and removing the second term of entropy, we get the loss function of posterior density estimation with the COSMOS dataset:

$$\frac{1}{N} \sum_{i=1}^N -\log q_\psi(\chi_i|b_i) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} (\chi_i - \mu_{\chi|b_i})^T \Sigma_{\chi|b_i}^{-1} (\chi_i - \mu_{\chi|b_i}) + \frac{1}{2} \ln |\Sigma_{\chi|b_i}|. \quad (8)$$

We refer to  $q_\psi(\chi|b)$  trained with the COSMOS dataset as Probabilistic Dipole Inversion (PDI).

### 3.2 VI Domain Adaptation

After training with the COSMOS dataset using Eq. 8 and obtaining the learned parameters  $\psi^o$ , we can simply estimate  $p(\chi|b)$  as  $q_{\psi^o}(\chi|b)$  given a test local field  $b$ . However, for a new test dataset that deviate from the COSMOS training dataset such as containing a new pathology, inferior outputs may be produced. To address this issue,  $q_{\psi^o}(\chi|b)$  can be adapted by deploying VI on a subset of the new test dataset with only local field data needed in the loss function. Specifically, the pre-trained approximation network  $q_\psi(\chi|b)$  with initial weights  $\psi^o$  can be fine-tuned by minimizing the KL divergence between  $p(\chi|b)$  and  $q_\psi(\chi|b)$ :

$$\begin{aligned}
 & \text{KL}[q_\psi(\chi|b)||p(\chi|b)] \\
 &= \mathbb{E}_q[\log q_\psi(\chi|b) - \log p(\chi|b)] \\
 &= \mathbb{E}_q[\log q_\psi(\chi|b) - \log p(\chi, b)] + \log p(b) \\
 &= \text{KL}[q_\psi(\chi|b)||p(\chi)] - \mathbb{E}_q[\log p(b|\chi)]
 \end{aligned} \tag{9}$$

where the first term in the last equation imposes the approximate posterior to be similar to the prior, which works as the regularization term for training, and the second term encourages data consistency in the likelihood with the QSM forward model. Constant term  $\log p(b)$  is omitted in the last equation. Inserting the prior distribution in Eq. 5 and likelihood distribution in Eq. 4, the KL divergence in Eq. 9 becomes:

$$\begin{aligned}
 & \text{KL}[q_\psi(\chi|b)||p(\chi|b)] \\
 &= -\frac{1}{2}\ln|\Sigma_{\chi|b}| + \frac{1}{2K} \sum_{k=1}^K \lambda \|M\nabla\chi_k\|_1 + \frac{1}{2K} \sum_{k=1}^K (F^H DF\chi_k - b)^T \Sigma_{b|\chi}^{-1} (F^H DF\chi_k - b)
 \end{aligned} \tag{10}$$

where  $-\frac{1}{2}\ln|\Sigma_{\chi|b}|$  is derived from the entropy of  $q_\psi(\chi|b)$  in  $\text{KL}[q_\psi(\chi|b)||p(\chi)]$ ,  $-\mathbb{E}_q[\ln p(\chi)]$  and  $-\mathbb{E}_q[\log p(b|\chi)]$  are approximated through Monte Carlo (MC) sampling with  $K$  samples  $\chi'_k$ s from  $q_\psi(\chi|b)$ . The reparameterization strategy can be used to implement back-propagation (Kingma and Welling, 2013), where samples from the standard Normal distribution were used to generate samples from the predicted susceptibility distribution by scaling and translating operations, in order to make the predicted susceptibility mean and variance map learnable through back-propagation. In VI domain adaptation, Eq. 10 is minimized across the new subjects. Once trained, the adapted  $q_\psi(\chi|b)$  can be used to predict  $\mu_{\chi|b}$  and  $\Sigma_{\chi|b}$  for new test subject directly, which is the so-called amortized VI. We refer to the fine-tuned approximate distribution with Eq. 10 as PDI-VI. Amortized VI can also be deployed without any COSMOS pre-training, in which only the target dataset with single orientation local field maps are needed to learn the probabilistic dipole inversion network using Eq. 10. We refer to amortized VI without COSMOS pre-training as PDI-VI0.

The amortized formulation of VI in Eq. 10 achieves fast inference during test time compared to the classic VI per case, but potentially at the expense of suboptimal performance (Cremer et al., 2018). This inference suboptimality can be explained as the *inference gap*, which can be decomposed as follows:

$$\underbrace{\text{KL}[q_{\psi^*}(\chi|b)||p(\chi|b)]}_{\text{Approximation gap}} + \underbrace{\text{KL}[q_\psi(\chi|b)||p(\chi|b)] - \text{KL}[q_{\psi^*}(\chi|b)||p(\chi|b)]}_{\text{Amortization gap}} \tag{11}$$

where  $\psi$  and  $\psi^*$  are obtained by amortized and subject-specific VIs of Eq. 10. As a result,  $\text{KL}[q_\psi(\chi|b)||p(\chi|b)]$  is decomposed into the two terms above: the approximation gap and the amortization gap. The approximation gap is determined by the capacity of the parameterized model family  $q_\psi(\chi|b)$  to approximate the true posterior distribution. The amortization gap is determined by the ability of the learned variational parameters  $\psi$  to generalize to a new test case. Initialized with the pre-trained PDI from Eq. 8, we deployed and compared both amortized and subject-specific VI for QSM posterior distribution estimation.

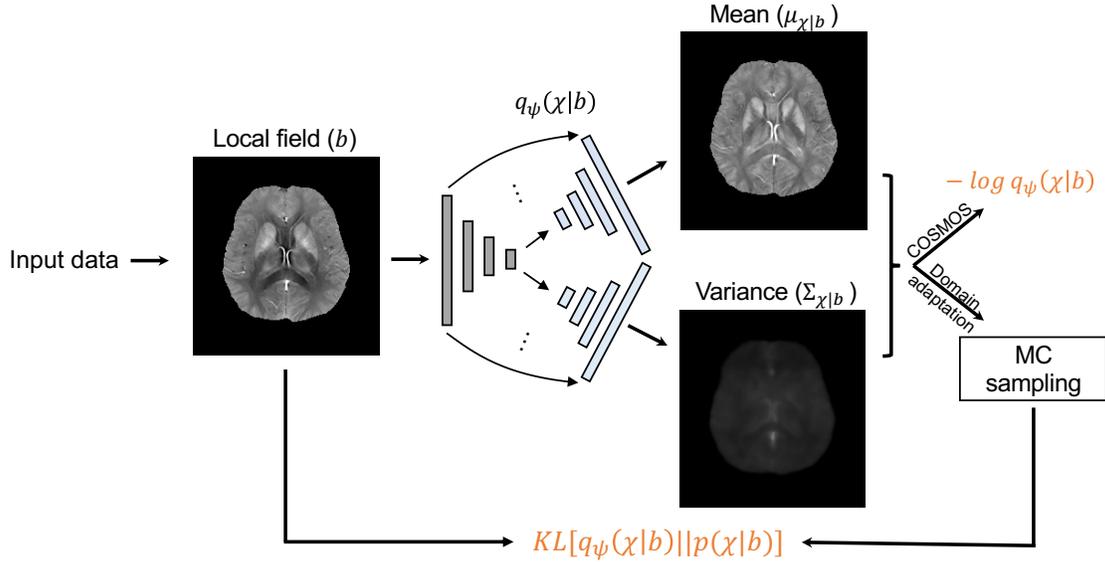


Figure 1: The network architecture of the proposed method. Two upsampling paths’ outputs represent mean and variance maps of susceptibility. The COSMOS dataset was used to perform posterior density estimation in Eq. 9. Domain adaptation VI with MC sampling in Eq. 10 were applied on other datasets.

### 3.3 Relation to VAE

The proposed VI domain adaptation strategy in Eq. 9 resembles the unsupervised variational auto-encoder (Kingma and Welling, 2013). In VAE, the auto-encoder architecture is used to learn both the approximate inference network as the encoder for the latent space variable  $z$  conditioned on the input data  $x$ , and the generative network as the decoder of data  $x$  given samples of  $z$ .  $x$  is expected to be reconstructed from  $z$ . Evidence lower bound (ELBO) is used to approximate the log density of data  $x$  by training the encoder and decoder simultaneously, where the optimal encoder of ELBO is the true posterior distribution of  $z$  given  $x$ , at which point the ELBO is tight and equals the log density of data  $x$ .

In the proposed PDI-VI strategy for QSM, the approximate posterior distribution is also a neural network ”encoder” from the input field  $b$  to the ”latent” susceptibility  $\chi$ , whereas the ”decoder” is no longer a neural network and does not need to be trained. Instead, this ”decoder” is the likelihood distribution from the forward dipole convolution model with additive Gaussian noise in Eq. 4. In addition, the prior distribution of the ”latent” variable  $\chi$  in Eq. 5 also comes from the domain knowledge of solving the QSM inverse problem. From physics-based likelihood and prior distributions, the same ELBO loss function in Eq. 9 is applied. Therefore, the proposed PDI-VI combines the modeling principle of distribution approximation and learning in VAE with the domain knowledge from medical physics in QSM.

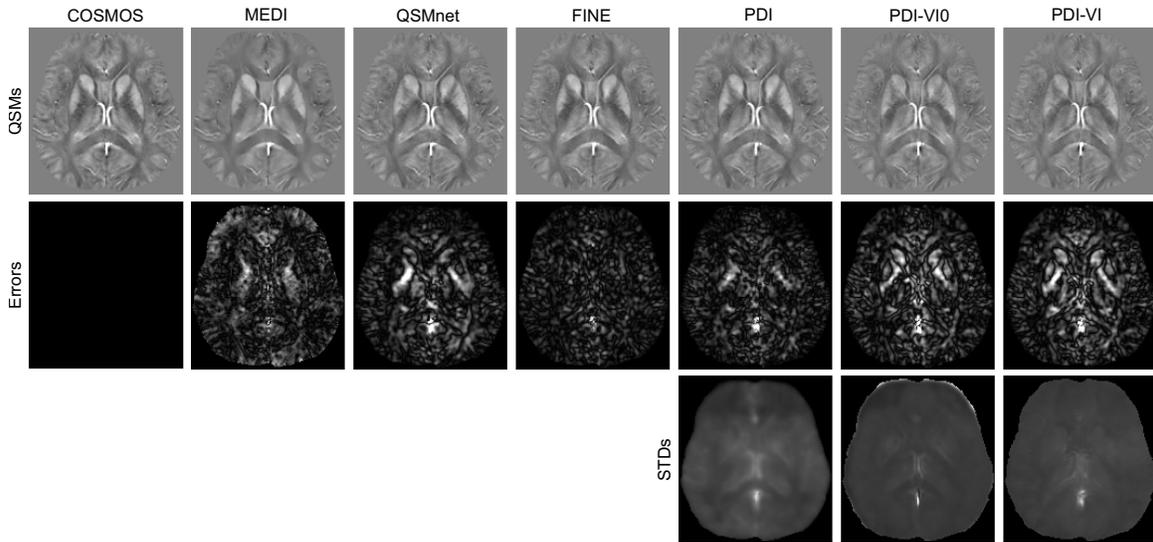


Figure 2: Reconstructions (first row,  $[-0.15, 0.15]$  ppm) and absolute error maps (second row,  $[0, 0.05]$  ppm) of one COSMOS test subject in one orientation, with COSMOS as the gold standard. FINE achieved the lowest reconstruction error, while the other methods had comparable results. SD maps of PDI, PDI-VI0 and PDI-VI (third row,  $[0, 0.05]$  ppm) showed high uncertainties at the sagittal sinus and globus pallidus, which was consistent with their error maps.

Table 1: Average quantitative metrics of 10 test COSMOS brains reconstructed by different methods. FINE gave the best reconstruction at the expense of significantly increased computational time. The other methods had comparable results.

	pSNR ( $\uparrow$ )	RMSE ( $\downarrow$ )	SSIM ( $\uparrow$ )	HFEN ( $\downarrow$ )	GPU time (s)
MEDI (Liu et al., 2012)	46.39	41.16	0.9569	31.30	17.54
QSMnet (Yoon et al., 2018)	46.35	41.29	0.9705	43.31	0.60
FINE (Zhang et al., 2020a)	48.12	33.66	0.9789	31.97	65.42
PDI (Eq. 8)	47.77	35.08	0.9772	35.17	0.61
PDI-VI0 (Eq. 10)	46.05	42.74	0.9704	42.27	0.61
PDI-VI (Eq. 8, then Eq. 10)	46.31	41.51	0.9707	40.58	0.61

### 3.4 Network Architecture

The proposed network architecture of  $q_\psi(\chi|b)$  is shown in Figure 1. This network is inspired by the widely used U-Net architecture (Ronneberger et al., 2015; Çiçek et al., 2016) for image-to-image mapping tasks in the biomedical deep learning field. The extension of the proposed architecture is to have one downsampling and two upsampling paths, where each upsampling path generates the mean or variance map from the same compressed feature

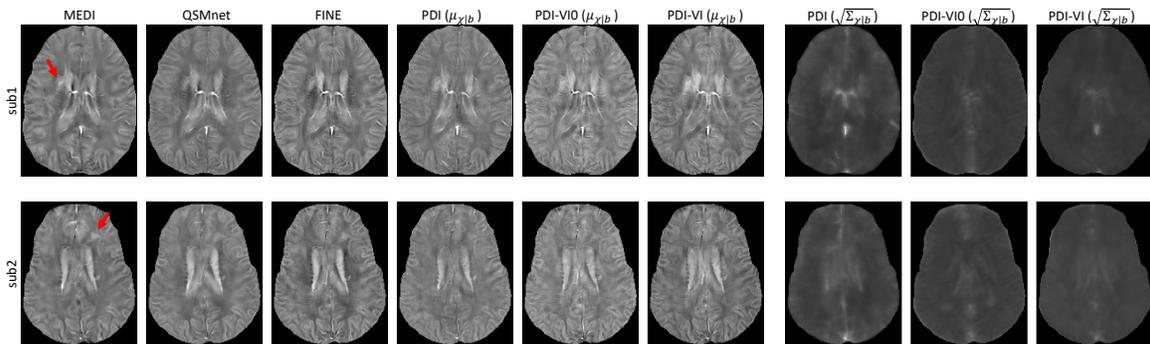


Figure 3: Two MS patient reconstructions (first six columns,  $[-0.15, 0.15]$  ppm) and SD maps (last three columns,  $[0, 0.05]$  ppm). Lesions indicated by the red arrows near the ventricle had lower susceptibility values in QSMnet and PDI, but were recovered in FINE and PDI-VI. Compared to PDI-VI, lesions reconstructed by PDI-VI0 also had lower susceptibility.

maps. Skip concatenations between downsampling and upsampling are applied for spatial information sharing and better gradient back-propagation. Loss functions in Eqs. 8 and 10 are used for training on COSMOS and other datasets. For the loss function in Eq. 10, Monte Carlo sampling with reparameterization strategy is applied to stochastically optimize  $q_\psi(\chi|b)$ . The 3D convolutional kernel size is  $3 \times 3 \times 3$ . The numbers of filters from the highest feature level to the lowest are 32, 64, 128, 256 and 512. Batch normalization (Ioffe and Szegedy, 2015) after each convolutional layer, and max pooling operation for downsampling and deconvolutional operation for upsampling are used.

## 4. Experiments

### 4.1 Data Acquisition and Preprocessing

MRI was performed on 7 healthy subjects with 5 brain orientations using a 3T scanner (GE, Waukesha, WI) equipped with a multi-echo 3D gradient echo (GRE) sequence. The acquisition matrix was  $256 \times 256 \times 48$  and voxel size was  $1 \times 1 \times 3 \text{ mm}^3$ . The input local tissue field data  $b$  was generated by sequentially deploying non-linear fitting across multi-echo phase data (Kressler et al., 2009), graph-cut based phase unwrapping (Dong et al., 2014) and background field removal (Liu et al., 2011a). A reference QSM reconstruction was obtained using COSMOS (Liu et al., 2011b). Two other datasets were obtained by performing single orientation GRE MRI on 9 patients with multiple sclerosis (MS) and 7 patients with intracerebral hemorrhage (ICH), which were acquired using the same scanning parameters and image processing procedures as above, except for the COSMOS reconstruction step. Data were acquired following an IRB approved protocol.

For the COSMOS dataset, data from 4/1 subjects (20/5 brain volumes) were used as the training/validation dataset, with augmentation by in-plane rotation of  $\pm 15^\circ$ . The brain volume data in the training and validation dataset was divided into 3D patches with patch size  $64 \times 64 \times 32$  and extraction step  $21 \times 21 \times 11$ , generating 9659/2874 patches

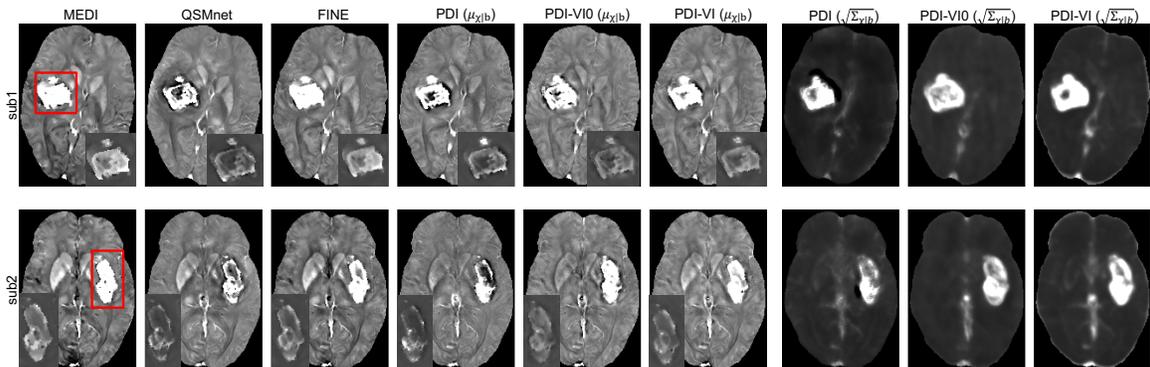


Figure 4: Two ICH patient Reconstructions (first six columns,  $[-0.15, 0.15]$  ppm) with the insets ( $[-0.6, 1.5]$  ppm) and SD maps (last three columns,  $[0, 0.05]$  ppm). Hemorrhage susceptibility was lower on QSMnet and PDI as compared to MEDI. This issue was reduced in FINE and PDI-VI. PDI-VI0 gave comparable hemorrhage reconstructions to PDI-VI. High variance inside the hemorrhage was consistent with high measured noise in the same region.

for training/validation. Data from the remaining 2 subjects (10 brain volumes in total) were used for testing. For the MS patient dataset, data from 6/1 subjects were used as the training/validation dataset and data from the remaining 2 subjects were used for testing. For the ICH patient dataset, data from 4/1 subjects were used as the training/validation dataset and data from the remaining 2 subjects were used for testing.

## 4.2 Implementation Details

The loss function in Eq. 8 was applied for posterior density estimation on the COSMOS dataset with Adam optimizer (Kingma and Ba, 2014) (learning rate:  $10^{-3}$ , Number of epochs: 60), yielding a trained network  $q_{\psi^o}(\chi|b)$ , denoted as PDI. Initialized with  $\psi^o$ , VI domain adaptations using the loss function in Eq. 10 were deployed on both MS and ICH datasets with Adam optimizer (learning rate:  $10^{-3}$ , Number of epochs: 100), denoted as PDI-VI. VIs using Eq. 10 and without  $\psi^o$  initialization were also performed and compared for all datasets (Adam learning rate:  $10^{-3}$ , Number of epochs: 100), denoted as PDI-VI0. MC sampling size  $K$  in VI was chosen as 5 due to limited GPU memory. The hyperparameter  $\lambda$  in Eq. 10 was chosen as 20 to balance the streaking artifact suppression and over-smoothing effect of TV regularization. While training and validation were implemented using 3D patches, whole brain volumes were fed into the network during COSMOS testing and all VI experiments. We implemented the proposed method using PyTorch (Python 3.6) on an RTX 2080Ti GPU.

## 4.3 COSMOS Dataset

For the COSMOS test dataset, we compared PDI (Eq. 8), PDI-VI0 (Eq. 10 without PDI pre-training) and PDI-VI (Eq. 10 with PDI pre-training) to MAP estimation MEDI (Liu

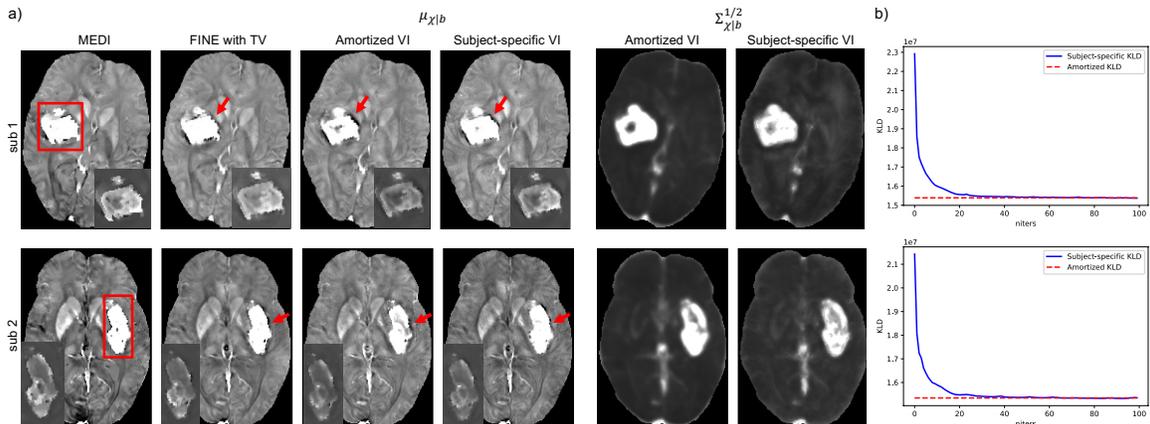


Figure 5: (a) Reconstructions ( $[-0.15, 0.15]$  ppm) with the insets ( $[-0.6, 1.5]$  ppm) and SD maps ( $[0, 0.05]$  ppm) and (b) KL divergence values of two ICH test patients using amortized and subject-specific VIs. MEDI and FINE with TV were used for comparison. Although an almost zero amortization gap (Eq. 11) was achieved by amortized VI (b) for both cases, reconstruction quality at the hemorrhage center and surrounding hemorrhage was still marginally better for subject-specific VI. FINE with TV and subject-specific VI achieve comparably image quality.

et al., 2012) and two deep learning reconstructions QSMnet (Yoon et al., 2018) and FINE (Zhang et al., 2020a). Reconstruction maps of one orientation from one test subject are shown in Figure 2. Quantitative metrics of each reconstruction method averaged among 10 test brains are shown in Table 1. FINE gave the best overall quantitative results with the expense of significantly increased computational time. The other methods had comparable results. All deep learning methods achieved fast inference time on GPU except FINE. In Figure 2, error maps of PDI, PDI-VI0 and PDI-VI’s mean outputs  $\mu_{\chi|b}$  matched their SD outputs  $\sqrt{\Sigma_{\chi|b}}$ , with high uncertainty/error located at the sagittal sinus and globus pallidus. The SD output of PDI-VI0 and PDI-VI were sharper than PDI with lower white-grey matter variation.

#### 4.4 Patient Datasets

The reconstruction maps of two MS patients in the test dataset are shown in Figure 3. Lesions indicated by the red arrows had susceptibility values lower in QSMnet and PDI than in MEDI, but were recovered in FINE and PDI-VI. Compared to PDI-VI, lesions reconstructed by PDI-VI0 also had lower susceptibility, which qualitatively indicated the advantage of the COSMOS dataset pre-training for PDI-VI.

The QSMs for two ICH patients in the test dataset are shown in Figure 4. Compared to MEDI and FINE which had hyperintensity inside the hemorrhage, both QSMnet and PDI had lower susceptibility inside this region, which might result from the fact that such pathology was not encountered during training. After amortized VI domain adaptation, susceptibility value inside the hemorrhage was recovered in PDI-VI. Shadow artifacts surround-

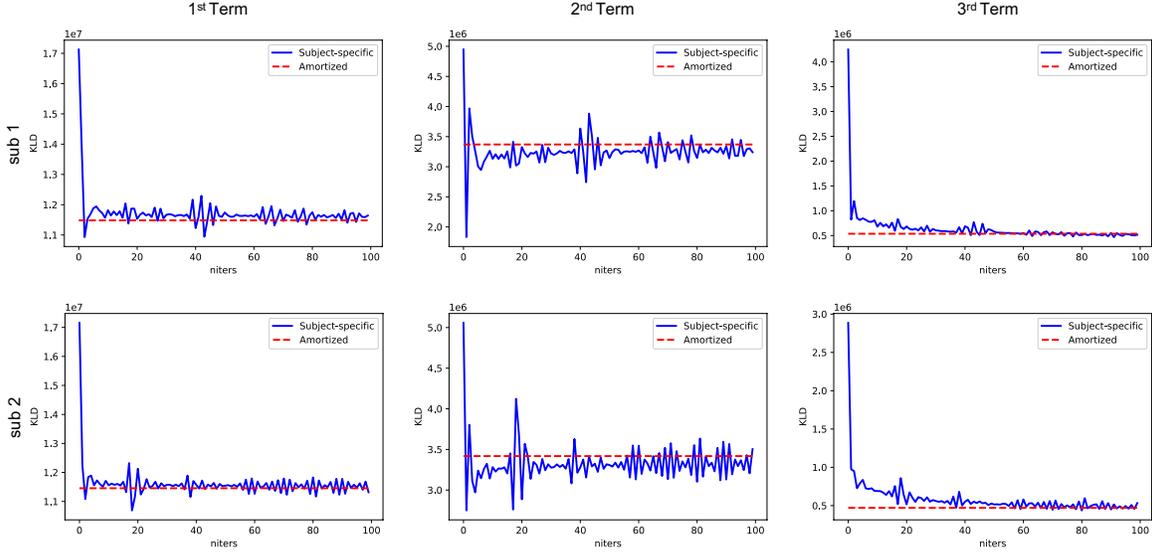


Figure 6: Value changes of three individual terms in Eq. 10 of subject-specific VI during iterations, with the value of amortized VI as a reference. The second term of TV regularization was slightly lower in subject-specific VI after convergence, while the other two terms were similar between amortized and subject-specific VIs.

ing the hemorrhage were also reduced in PDI-VI from PDI. PDI-VI0 yielded hemorrhage reconstructions that were comparable to PDI-VI. High SD map inside the reconstructed hemorrhage as shown in the last three columns of Figure 4 implied high reconstruction uncertainty of this region.

#### 4.5 Amortized vs Subject-specific VI

The inference gap in Eq. 11 was investigated on two ICH test cases shown in Figure 5, where subject-specific VI using Eq. 10 initialized from the weights of PDI was deployed with 100 iterations for convergence. MAP estimations in Eq. 3 of iterative reconstruction MEDI and network parametrized reconstruction FINE with TV ( $\lambda = 20, 100$  iterations) were also deployed for comparison. As demonstrated in Figure 5a, both amortized and subject-specific VIs recovered the susceptibility value inside the hemorrhage from PDI in Figure 4. Compared to amortized VI, the susceptibility values at the center of hemorrhage (insets in Figure 5a) were further recovered and shadow artifacts surrounding the hemorrhage (red arrows in Figure 5) were reduced in subject-specific VI. In addition, subject-specific VI had similar reconstructions to MEDI and FINE with TV for both test cases, which confirmed that the mean susceptibility map by subject-specific VI equals the MAP susceptibility maps by MEDI and FINE with TV. Figure 5b shows that KL divergence of Eq. 10 during subject-specific VIs converged to the value of amortized VIs with almost zero amortization gap (Eq. 11). Figure 6 shows the value changes of three individual terms in Eq. 10 during subject-specific VI iterations, where the second term ( $\frac{1}{2K} \sum_{k=1}^K \lambda \|M \nabla \chi_k\|_1$ ) was slightly lower on

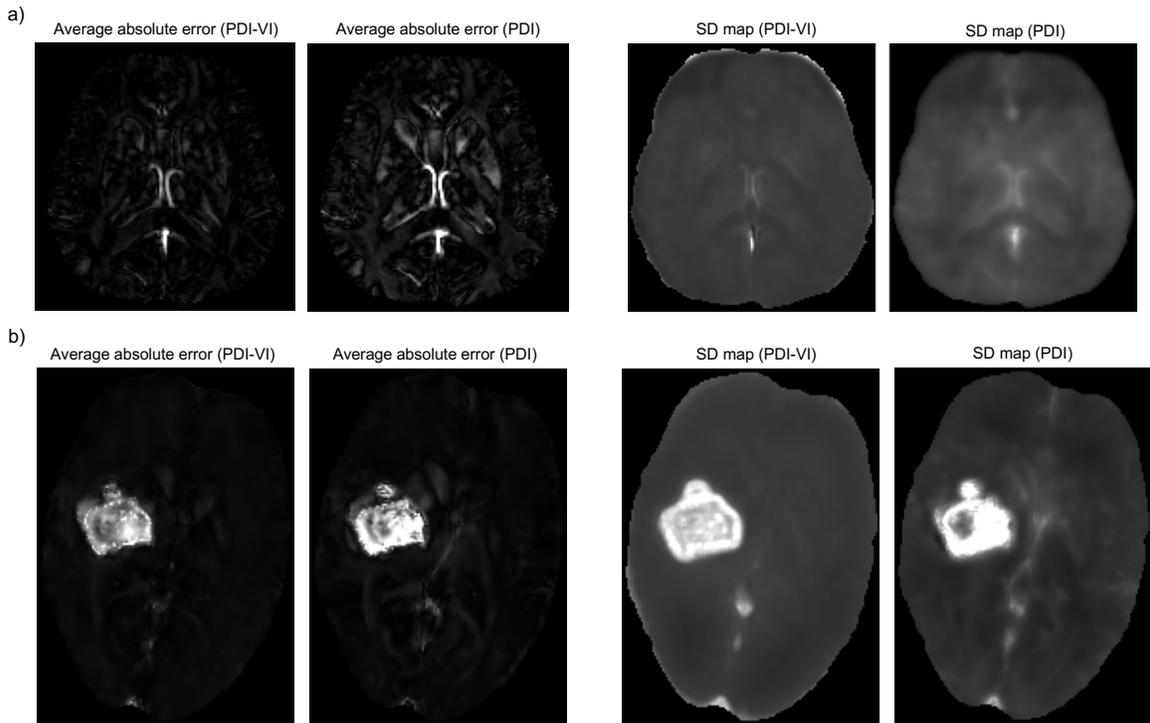


Figure 7: PDI and PDI-VI’s average absolute error maps (first two columns,  $[0, 0.05]$  ppm) through simulations and predicted SD maps (last two columns,  $[0, 0.05]$  ppm) of (a) healthy and (b) hemorrhagic brains. The SD maps resembled the error maps in both cases for PDI and PDI-VI.

average than the one of amortized VI for both test cases, which might contribute to the improvement of shadow artifact reduction.

#### 4.6 Uncertainty Map Evaluation

To evaluate uncertainty estimation performance of the predicted SD map, absolute error maps of PDI and PDI-VI’s mean predictions to the ground truth susceptibilities were computed via simulation, then correlation between susceptibility SD and error maps was examined. Local field inputs were simulated from (a) COSMOS test data in Figure 2 and (b) FINE reconstruction of the ICH patient in Figure 4a through multi-echo data synthesis with additive noise, nonlinear field fitting and phase unwrapping. Details of the simulation steps are shown in Appendix A. Such simulation was repeated 100 times to generate 100 local fields as inputs to PDI and PDI-VI. 100 mean maps of PDI and PDI-VI were predicted accordingly to compute the average absolute errors. Figure 7 shows the average absolute error maps and predicted SD maps of PDI and PDI-VI. In Figure 7a, large errors in the cerebral veins and sagittal sinus were reflected in the predicted SD maps for both PDI and PDI-VI, while in Figure 7b, large errors in the hemorrhage were also predicted in

PDI and PDI-VI’s SD maps, which demonstrates good correlation between the error map and the predicted SD map of the proposed method for uncertainty estimation.

## 5. Discussion

The adaptive learning strategy proposed in this paper tackles the domain adaptation challenge in medical imaging with deep learning from a probabilistic distribution refinement point of view. Since the high quality COSMOS samples are acquired only from healthy subjects, posterior density estimation with COSMOS samples may not generalize well to the patients with pathology not covered by the COSMOS dataset. As a result, even though the COSMOS pre-trained PDI performs well on COSMOS test dataset from the same distribution (Figure 2), inferior mapping happens evidenced by lower susceptibility values for lesions when applying PDI to the patients directly (Figures 3 and 4). Based on the distribution approximation principle, the pre-trained density estimation network PDI needs fine-tuning in order to fit to the patient data distribution as well. VI with KL divergence as a measure of similarity between two distributions is used for approximate distribution refinement, which helps reduce the generalization error of PDI (Figure 3 and 4). However, in terms of other domain generalizations such as different imaging resolutions, PDI-VI with KL divergence loss function for weight adjustment has not been tested and may suffer in accuracy, which will be explored in the future work. Another domain adaptation method FINE works better than PDI-VI (Figure 4) to reduce generalization error of the pre-trained network, since FINE fits to every test case by minimizing the fidelity loss, which has the major drawback of significantly increased computational time (Table 1).

The relationship between PDI-VI (Eq. 9) and VAE (Kingma and Welling, 2013) is described in the methods section. The key point is that the generative network (the decoder) from latent variable to data in VAE is replaced by a physics-based likelihood model (Eq. 4) in PDI-VI. This implies a general way of learning the posterior distribution of image data conditioned on the measured signal for any imaging modality, where a specific forward imaging model is used to form the ”decoder” and only the ”encoder” is learned with input measured signals in an unsupervised fashion like VAE. The training strategy of PDI-VI utilizes both widely available measured signals in clinic and well-defined imaging physical models to improve the reconstruction fidelity of the trained model. When gold standard reconstructions are available for training, as in the COSMOS dataset, combining direct conditional density estimation using high quality images with VI domain adaptation on measured input signals could improve the performance of VI trained on the measured signals alone (Figure 3).

PDI defines a set of parameterized distributions using a neural network and learns these parameters from samples to approximate the true distribution, where the expressiveness of the distribution family affects their approximation ability. The network architecture (Figure 1) is inspired by 3D U-Net (Ronneberger et al., 2015; Çiçek et al., 2016), which was originally proposed for medical image segmentation tasks and has also been successfully used in deep QSM reconstructions (Yoon et al., 2018; Zhang et al., 2020a; Bollmann et al., 2019), therefore such architecture should be expressive enough for field-to-susceptibility mapping. The COSMOS experiment indicates satisfactory image-to-image mapping ability of the proposed architecture (Figure 2 and Table 1). The simulation experiment verifies

correlation between the predicted SD map and the error map, indicating reasonable uncertainty quantification of PDI and PDI-VI. Despite such merits, the choice of variational posterior form in this work is simply a Gaussian distribution with diagonal covariance matrix, which is known as the mean field approximation for modeling and calculation simplicity in classic VI. This factorized Gaussian does not consider correlation between voxels in the reconstructed susceptibility map, but in view of the forward convolution operation (Eq. 1) which aggregates the global susceptibility into the measured field at each location, taking into account the dependency between local voxels in the susceptibility map may make the variational posterior more expressive. Possible options could be improving the Gaussian posterior with a non-diagonal covariance matrix and using an autoregressive (Oord et al., 2016) or flow-based (Dinh et al., 2016) model to capture the dependency.

The prior distribution of susceptibility (Eq. 5) used in PDI-VI comes from MEDI (Liu et al., 2011b), where weighted TV regularization was used to suppress streaking artifacts appeared on QSM dipole inversion (Kee et al., 2017). In general, the prior distribution  $p(x)$  captures the density of data  $x$  from a prior knowledge, where higher quality data  $x$  has a higher density value. In this sense, estimating the density from sufficient data may build a more comprehensive prior distribution and therefore become more efficient to regularize the inverse problem solution. In fact, learning the prior density for MAP estimation of the imaging inverse problem has been explored by Tezcan et al. (2018) and Luo et al. (2019), where VAE and PixelCNN++ (Salimans et al., 2017) were deployed to learn the explicit prior distribution of MR images. Lønning et al. (2019) proposed the recurrent inference machine by implicitly encoding the prior distribution for MAP estimation. These deep prior approaches inspire us to extend our work in the future by learning and evaluating a prior density from data and inserting them into Eqs. 9 and 10 for VI.

The inference gap (Eq. 11) summarizes two types of errors when applying the amortized inference strategy. Amortized VI has the advantage of fast inference during test time. However, it has slightly worse visual quality inside and surrounding the hemorrhage than subject-specific VI (Figure 5a). Even though an almost zero amortization gap was achieved (Figure 5b), the regularization term of KL divergence (Eq. 10) was still better imposed in subject-specific VI, which may contribute to its better reconstruction of the hemorrhage. However, such advantage comes at a cost of extra inference time. To accelerate the inference speed of subject-specific VI, optimizing the initialization of variational parameters is useful to reduce the number of VI optimization steps. Meta-learning (Naik and Mammone, 1992; Hochreiter et al., 2001) may be applied to optimize the optimization process of VI per data, where a learner can be designed during pre-training to learn an inference algorithm that generalizes well to the data of interest.

## 6. Conclusion

In conclusion, we demonstrate a neural network parametrized distribution which yields an approximate posterior distribution of susceptibility given an input local field map for the Bayesian QSM inverse problem. The network was pre-trained on a COSMOS dataset by fitting to the empirical distribution and adapted to different domains using amortized VI. The proposed method computes adaptive reconstructions of susceptibility together with an uncertainty estimation. Future work will include building a more expressive posterior

distribution family, learning a deep prior density for regularization and optimizing the subject-specific VI algorithm using meta-learning.

## Acknowledgments

The authors would like to thank Adrian Dalca for useful feedback and discussions. This research was supported in part by National Institute of Health (R01NS090464, R01NS095562, R01NS105144, R01DK116126, R01CA181566, S10OD021782, 1R21AG050122, R01LM012719 and R01AG053949), National Science Foundation (1748377 and 1707312) and National Multiple Sclerosis Society (RR-1602-07671).

## Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript. Data were acquired following an IRB approved protocol.

## Conflicts of Interest

We declare we don't have conflicts of interest.

## Appendix A.

In this appendix we show the steps of simulation in section 4.6:

- Synthesize local magnetic field data  $b_{syn}$  from COSMOS gold standard susceptibility  $\chi_{COSMOS}$  using dipole convolution model:

$$b_{syn} = \chi_{COSMOS} * d$$

- Synthesize multi-echo MR images  $S_j$  from  $b_{syn}$  (above),  $R2^*$  ( $T2^*$  decay rate) and  $M_0$  (water) using forward physical model:

$$S_j = M_0 e^{-R2^* \cdot t_j} e^{i(\phi_0 + b_{syn} \cdot t_j)} + n_j,$$

where  $t_j$  is the echo time of the  $j$ -th echo,  $n_j$  is the i.i.d. Gaussian noise on real and imag parts for all voxels and initial phase  $\phi_0 = 0$  is assumed for all voxels.

- Deploy nonlinear field fitting and spatial phase unwrapping to estimate the noisy local field data.

## References

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

- Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.
- Steffen Bollmann, Kasper Gade Bøtker Rasmussen, Mads Kristensen, Rasmus Guldhammer Blendal, Lasse Riis Østergaard, Maciej Plochanski, Kieran O’Brien, Christian Langkammer, Andrew Janke, and Markus Barth. Deepqsm-using deep learning to solve the dipole inversion for quantitative susceptibility mapping. *Neuroimage*, 195:373–383, 2019.
- Michael A Chappell, Adrian R Groves, Brandon Whitcher, and Mark W Woolrich. Variational bayesian inference for a nonlinear forward model. *IEEE Transactions on Signal Processing*, 57(1):223–236, 2009.
- Weiwei Chen, Wenzhen Zhu, Ilhami Kovanlikaya, Arzu Kovanlikaya, Tian Liu, Shuai Wang, Carlo Salustri, and Yi Wang. Intracranial calcifications and hemorrhages: characterization with quantitative susceptibility mapping. *Radiology*, 270(2):496–505, 2014.
- Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- Chris Cremer, Xuechen Li, and David Duvenaud. Inference suboptimality in variational autoencoders. *arXiv preprint arXiv:1801.03558*, 2018.
- Ludovic de Rochefort, Tian Liu, Bryan Kressler, Jing Liu, Pascal Spincemaille, Vincent Lebon, Jianlin Wu, and Yi Wang. Quantitative susceptibility map reconstruction from mr phase data using bayesian regularization: validation and application to brain imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 63(1):194–206, 2010.
- Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.
- Jianwu Dong, Tian Liu, Feng Chen, Dong Zhou, Alexey Dimov, Ashish Raj, Qiang Cheng, Pascal Spincemaille, and Yi Wang. Simultaneous phase unwrapping and removal of chemical shift (spurs) using graph cuts: application in quantitative susceptibility mapping. *IEEE transactions on medical imaging*, 34(2):531–540, 2014.
- Sepp Hochreiter, A Steven Younger, and Peter R Conwell. Learning to learn using gradient descent. In *International Conference on Artificial Neural Networks*, pages 87–94. Springer, 2001.
- Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- Youngwook Kee, Zhe Liu, Liangdong Zhou, Alexey Dimov, Junghun Cho, Ludovic De Rochefort, Jin Keun Seo, and Yi Wang. Quantitative susceptibility mapping (qsm)

- algorithms: mathematical rationale and computational implementations. *IEEE Transactions on Biomedical Engineering*, 64(11):2531–2545, 2017.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Bryan Kressler, Ludovic De Rochefort, Tian Liu, Pascal Spincemaille, Quan Jiang, and Yi Wang. Nonlinear regularization for per voxel estimation of magnetic susceptibility distributions from mri field maps. *IEEE transactions on medical imaging*, 29(2):273–281, 2009.
- Jing Liu, Tian Liu, Ludovic de Rochefort, James Ledoux, Ildar Khalidov, Weiwei Chen, A John Tsiouris, Cynthia Wisnieff, Pascal Spincemaille, Martin R Prince, et al. Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. *Neuroimage*, 59(3):2560–2568, 2012.
- Tian Liu, Pascal Spincemaille, Ludovic De Rochefort, Bryan Kressler, and Yi Wang. Calculation of susceptibility through multiple orientation sampling (cosmos): a method for conditioning the inverse problem from measured magnetic field map to susceptibility source image in mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 61(1):196–204, 2009.
- Tian Liu, Ildar Khalidov, Ludovic de Rochefort, Pascal Spincemaille, Jing Liu, A John Tsiouris, and Yi Wang. A novel background field removal method for mri using projection onto dipole fields. *NMR in Biomedicine*, 24(9):1129–1136, 2011a.
- Tian Liu, Jing Liu, Ludovic De Rochefort, Pascal Spincemaille, Ildar Khalidov, James Robert Ledoux, and Yi Wang. Morphology enabled dipole inversion (medi) from a single-angle acquisition: comparison with cosmos in human brain imaging. *Magnetic resonance in medicine*, 66(3):777–783, 2011b.
- Tian Liu, Cynthia Wisnieff, Min Lou, Weiwei Chen, Pascal Spincemaille, and Yi Wang. Nonlinear formulation of the magnetic field to source relationship for robust quantitative susceptibility mapping. *Magnetic resonance in medicine*, 69(2):467–476, 2013.
- Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan WA Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous mri data. *Medical image analysis*, 53:64–78, 2019.
- GuanXiong Luo, Na Zhao, Wenhao Jiang, and Peng Cao. Mri reconstruction using deep bayesian inference. *arXiv preprint arXiv:1909.01127*, 2019.
- Devang K Naik and Richard J Mammone. Meta-neural networks that learn by learning. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 1, pages 437–442. IEEE, 1992.

- Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*, 2016.
- Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017.
- Audrey Repetti, Marcelo Pereyra, and Yves Wiaux. Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM Journal on Imaging Sciences*, 12(1):87–118, 2019.
- Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.
- Olaf Ronneberger, Philipp Fischer, and Thomas 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.
- Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.
- Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017.
- Kerem C Tezcan, Christian F Baumgartner, Roger Luechinger, Klaas P Pruessmann, and Ender Konukoglu. Mr image reconstruction using deep density priors. *IEEE transactions on medical imaging*, 2018.
- Yi Wang and Tian Liu. Quantitative susceptibility mapping (qsm): decoding mri data for a tissue magnetic biomarker. *Magnetic resonance in medicine*, 73(1):82–101, 2015.
- Yi Wang, Pascal Spincemaille, Zhe Liu, Alexey Dimov, Kofi Deh, Jianqi Li, Yan Zhang, Yihao Yao, Kelly M Gillen, Alan H Wilman, et al. Clinical quantitative susceptibility mapping (qsm): biometal imaging and its emerging roles in patient care. *Journal of Magnetic Resonance Imaging*, 46(4):951–971, 2017.
- Jaeyeon Yoon, Enhao Gong, Itthi Chatnuntawech, Berkin Bilgic, Jingu Lee, Woojin Jung, Jingyu Ko, Hosan Jung, Kawin Setsompop, Greg Zaharchuk, et al. Quantitative susceptibility mapping using deep neural network: Qsmnet. *Neuroimage*, 179:199–206, 2018.
- Jinwei Zhang, Zhe Liu, Shun Zhang, Hang Zhang, Pascal Spincemaille, Thanh D Nguyen, Mert R Sabuncu, and Yi Wang. Fidelity imposed network edit (fine) for solving ill-posed image reconstruction. *NeuroImage*, page 116579, 2020a.
- Jinwei Zhang, Hang Zhang, Mert Sabuncu, Pascal Spincemaille, Thanh Nguyen, and Yi Wang. Bayesian learning of probabilistic dipole inversion for quantitative susceptibility mapping. In *Medical Imaging with Deep Learning*, 2020b.