Perfect MCMC Sampling in Bayesian MRFs for Uncertainty Estimation in Segmentation

Saurabh Garg, Suyash P. Awate

Computer Science and Engineering (CSE) Department, Indian Institute of Technology (IIT) Bombay, Mumbai, India.

Introduction and Motivation

- Accounting for uncertainty in automated segmentation results can improve risk analysis in clinical procedures and reliability in clinical diagnosis and studies.
- Typical segmentation methods, e.g., using graph cuts or using expectation maximization (EM) and hidden Markov random fields (MRFs), typically produce a single optimal solution, and don't provide information about (i) object-boundary uncertainty or (ii) alternate close-to-optimal solutions.
- To estimate uncertainty, some methods intend to sample segmentations from label-image posterior models using Markov chain Monte Carlo (MCMC) sampling or perturbation models. However, they cannot guarantee sampling from the true posterior, deviating significantly in practice.
- We propose methods that guarantee exact sampling, in finite time, from generic Bayesian MRFs to estimate uncertainty.

Our Approach for Uncertainty Estimation

- We introduce a new framework for uncertainty estimation in segmentation by relying on perfect MCMC sampling, in finite time, from generic Bayesian MRF models.
- Perfect MCMC tracks *parallel coupled* chains (one chain started at each point in state space, each chain using the same random number generator) and then checks *coalescence*.
- We propose perfect sampling of label images in two ways:
  (i) by combining coupling-from-the-past (CFTP) [Propp-Wilson 1996 Rand. Struct. Algo.] with the bounding-chain (BC) [Huber 2004 Ann. Appl. Prob.] scheme, which we call CFTP-BC.
  (ii) by theoretically extending Fill's algorithm (FA) [Fill 1998 Ann. Appl. Prob.] using the BC scheme, which we call FA-BC.

CFTP-BC Algorithm for Perfect Sampling

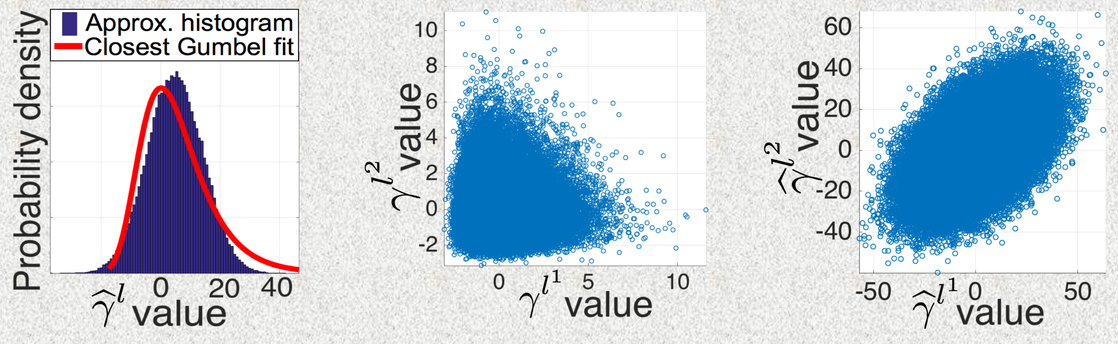
- Consider Markov chain $\mathring{\mathcal{M}}$ with state space $(2^{\mathcal{L}})^V$, where $2^{\mathcal{L}}$ is the set of subsets of label set $\mathcal{L} := \{1, \ldots, L\}$
- For $\mathring{\mathcal{M}}$, each state, say, $\mathring{X}$, contains a set of states $X \in \mathcal{L}^V$
- Initialize $\mathring{X}_v$ to the label set $\mathcal{L}$, for all voxels *v*.
- At voxel *v*, let $P^{\min}(X_v = l|x_{-v})$ and $P^{\max}(X_v = l|x_{-v})$ be the max and min conditional probabilities over all possible chains.
- At each voxel *v*, do the following:
  (1) In the bounding chain $\mathring{\mathcal{M}}$, initialize the set of labels $\mathring{X}_v = \phi$
  (2) Draw l uniformly from the label set $\mathcal{L}$. Draw $u \sim U(0,1)$
  (3) If $u > P^{\max}(X_v = l|x_{-v})$, then do nothing.
  (4) If $u \in [P^{\min}(X_v = l|x_{-v}), P^{\max}(X_v = l|x_{-v})]$, insert l into $\mathring{X}_v$
  (5) If $u < P^{\min}(X_v = l|x_{-v})$, then insert l into set $\mathring{X}_v$ and exit.
  (6) Repeat from Step 2.
- When $\forall v$, $\mathring{X}_v$ is a singleton, say $\{\hat{x}_v\}$, then all Markov chains $\mathcal{M}$ have coalesced to the label image $\hat{x}$

Our FA-BC Algorithm for Perfect Sampling

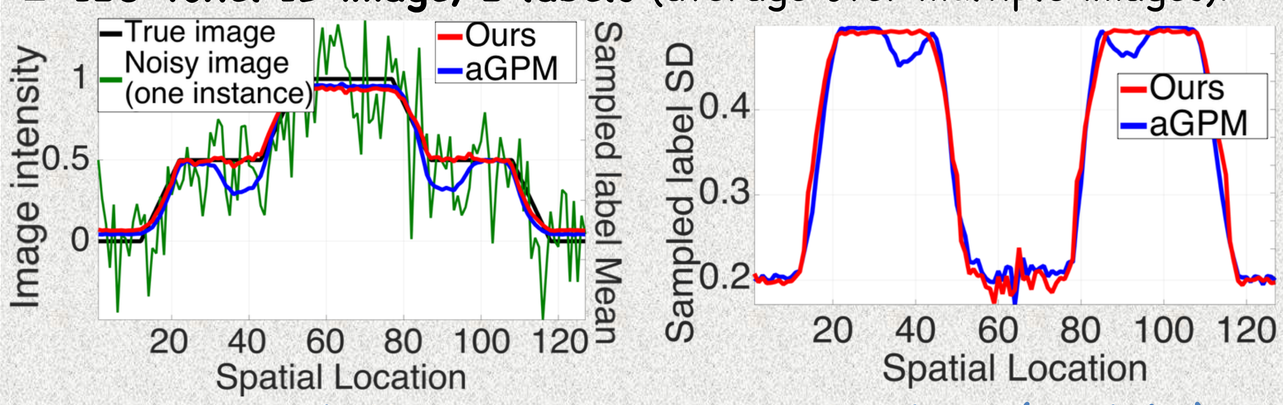
- Based on acceptance-rejection sampling. Generate proposal: pick randomly, a state *z* and integer *T* > *0*; run (reverse) Markov chain for T steps to take *z* -> *x*. Accept proposal x by simulating parallel coupled chains, constrained so that *x* -> *z*, and checking for coalescence after *T* steps.
- Let l^* be label at voxel v for time $t+1$ along Markov chain path $x \to z$ and let $P^*(X_v^t = l^*|x_{-v}^t)$ be conditional probability conditioned on the neighbor-pixels' labels for the path $x \to z$
- Initialize $t := 0, x^0 := x$
  1. At time t, do the following at each voxel v:
     a) In the bounding chain $\mathring{\mathcal{M}}$, initialize set of labels $\mathring{X}_v = \phi$
     b) Draw l uniformly from label set $\mathcal{L}$
     c) If $l \neq l^*$, draw $u \sim U(P^*(X_v^t = l^*|x_{-v}^t), 1)$; otherwise draw $u \sim U(0,1)$
     d) If $u > P^{\max}(X_v^t = l|x_{-v}^t)$, then do nothing.
     e) If $u \in [P^{\min}(X_v^t = l|x_{-v}^t), P^{\max}(X_v^t = l|x_{-v}^t)]$, then insert label l into the set $\mathring{X}_v$
     f) If $u < P^{\min}(X_v^t = l|x_{-v}^t)$, then insert label l into $\mathring{X}_v$. Exit
  2. Increment t by 1. If $t < T$, repeat step 1. If $t = T$ and coalescence has occurred then accept x as a draw.

Validation on Simulated Data

- Mean and standard deviation (SD) per voxel (for multi-category case, we generalize SD by square-root of unalikeability)
- Difference between ideal Gumbel (γ) and its tractable approximation aGPM [Alberts et. al. 2016 ISBI] ($\hat{\gamma}$):

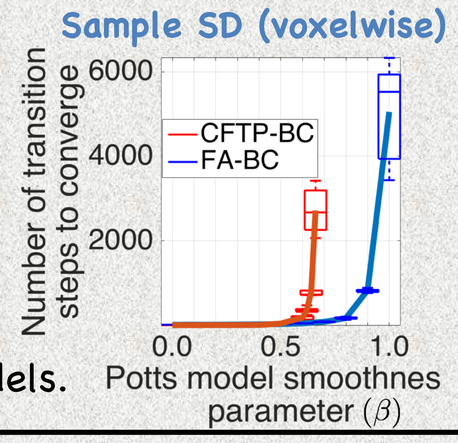


- **128-voxel 1D image, 2 labels** (average over multiple images):



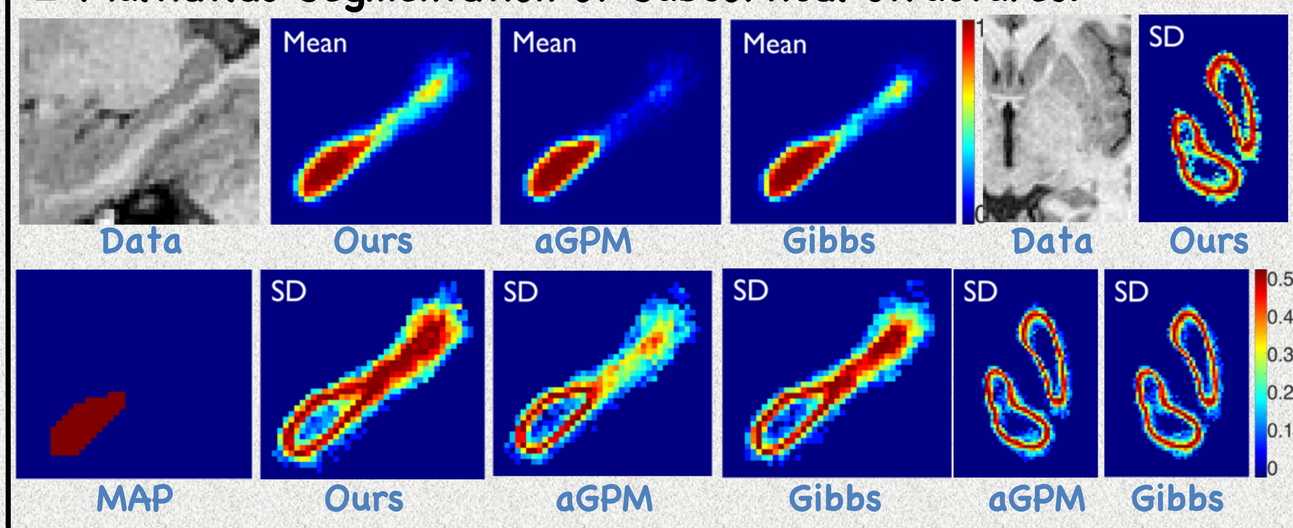
Sample Mean | Sample SD (voxelwise)

- aGPM can be lead to a strong bias in the empirical mean estimate near edges. Our empirical mean estimate is much closer to ground truth.
- Our FA-BC converges faster than our CFTB-BC for many MRF models.

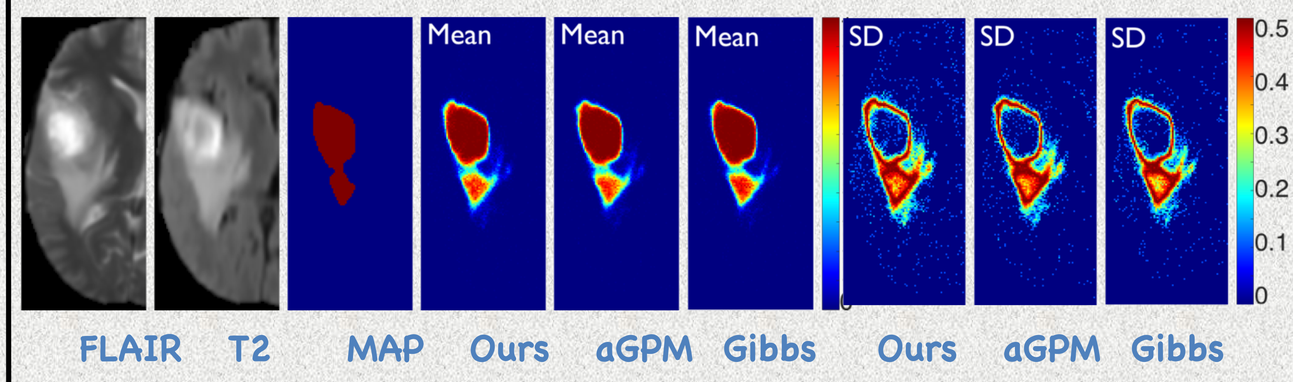


Results on Classic Segmentation Problems

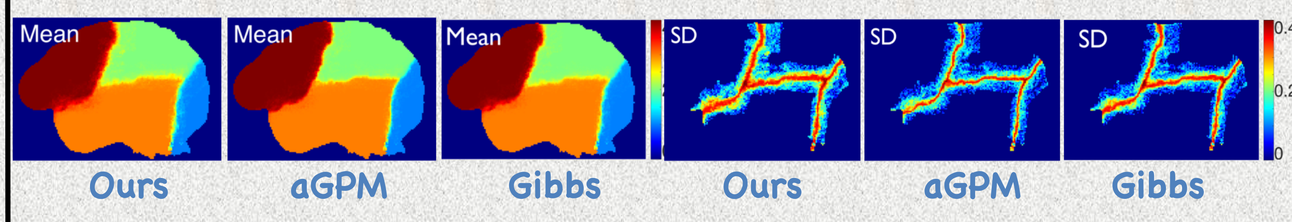
- **Multiatlas Segmentation of Subcortical structures:**



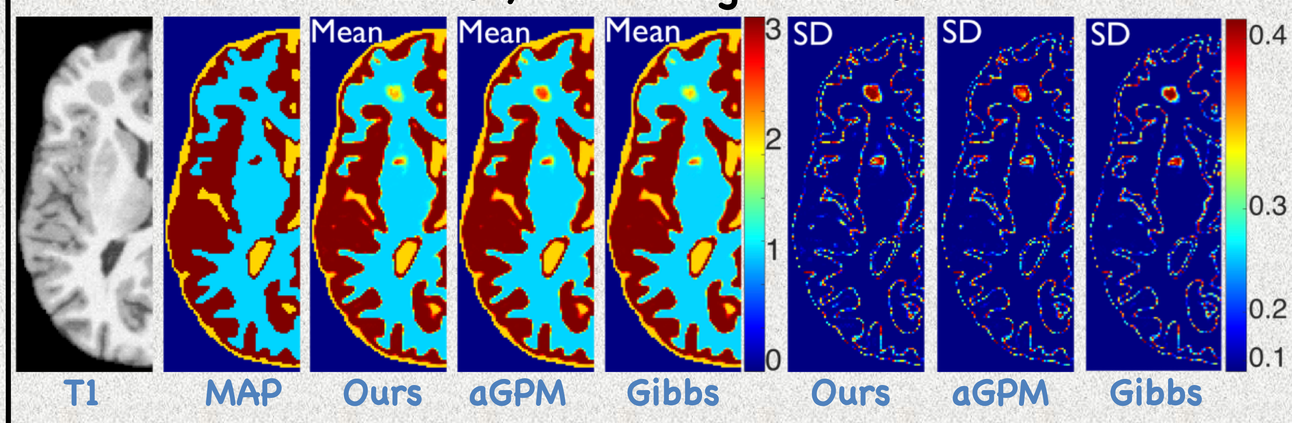
- **Multimodal Brain MRI, Tumor Segmentation:**



- **Multiatlas Segmentation of Lobes:**



- **Simulated Mild Lesion, Tissue Segmentation:**



- Gibbs's convergence time varies severely with the MRF model and the data, making it very difficult to predict burn-in.
- For a safe-side Gibbs burn-in of 5000, FA-BC is 10-20x faster.

Thanks to support from IIT Bombay Seed Grant 14IRCCSG010