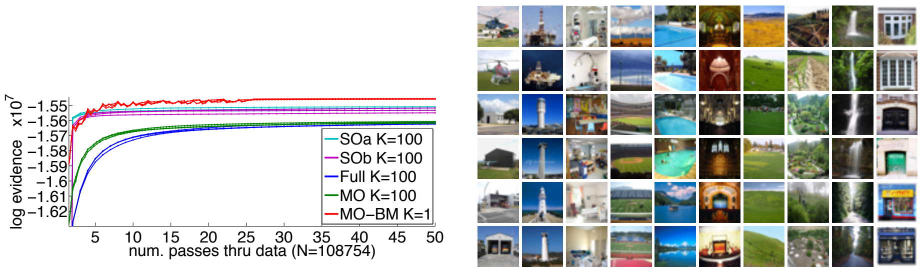

Figure 4: SUN-397 tiny images. *Left:* ELBO during training. *Right:* Visualization of 10 of 28 learned clusters for best MO-BM run. Each column shows two images from the top 3 categories aligned to one cluster.

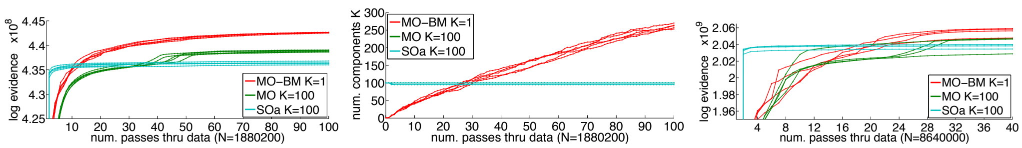

Figure 5: 8×8 image patches. *Left:* ELBO during training, $N = 1.88$ million. *Center:* Effective truncation-level K during training, $N = 1.88$ million. *Right:* ELBO during training, $N = 8.64$ million.

4.3 Tiny image clustering

We next learn a full-mean DP-GMM for tiny, 32×32 images from the SUN-397 scene categories dataset [18]. We preprocess all 108754 color images via PCA, projecting each example down to $D = 50$ dimensions. We start MO-BM at $K = 1$, while other methods have fixed $K = 100$. Fig. 4 plots the training ELBO as more data is seen. Our MO-BM runs surpass all other algorithms.

To verify quality, Fig. 4 shows images from the 3 most-related scene categories for each of several clusters found by MO-BM. For each learned cluster k, we rank all 397 categories to find those with the largest fraction of members assigned to k via $\hat{r}_{\cdot k}$. The result is quite sensible, with clusters for tall free-standing objects, swimming pools and lakes, doorways, and waterfalls.

4.4 Image patch modeling

Our last experiment applies a zero-mean, full-covariance DP-GMM to learn the covariance structures of natural image patches, inspired by [19, 20]. We compare online algorithms on $N = 1.88$ million 8×8 patches, a dense subsampling of all patches from 200 images of the Berkeley Segmentation dataset. Fig. 5 shows that our birth-merge memoized algorithm started at $K = 1$ can consistently add useful components and reach better solutions than alternatives. We also examined a much bigger dataset of $N = 8.64$ million patches, and still see advantages for our MO-BM.

Finally, we perform denoising on 30 heldout images, using code from [19]. Our best MO-BM run on the 1.88 million patch dataset achieves PSNR of 28.537 dB, within 0.05 dB of the PSNR achieved by [19]'s publicly-released GMM with $K = 200$ trained on a similar corpus. This performance is visually indistinguishable, highlighting the practical value of our new algorithm.

5 Conclusions

Our novel memoized online variational algorithm avoids noisiness and sensitivity inherent in stochastic methods. Our birth and merge moves successfully escape local optima. These innovations are applicable to common nonparametric models beyond the Dirichlet process.

Acknowledgments This research supported in part by ONR Award No. N00014-13-1-0644. M. Hughes supported in part by an NSF Graduate Research Fellowship under Grant No. DGE0228243.