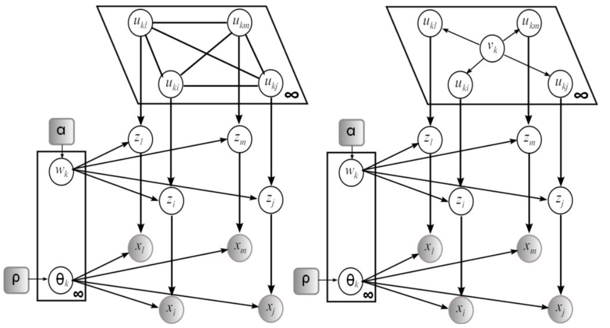

Figure 1. **Generative models of image partitions.** *Left*. Spatially dependent PY model, *(right)* low rank model. Shaded nodes represent observed random variables. $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$ is a low dimensional Gaussian random variable and $\mathbf{u}_k$ is the corresponding N dimensional layer. $w_k \sim \text{Beta}(1-\alpha_a, \alpha_b + k\alpha_a)$ controls expected layer size and are governed by Pitman-Yor hyper-parameters $\alpha = (\alpha_a, \alpha_b)$. The Dirichlet hyper-parameters $\rho = (\rho^t, \rho^c)$ parametrize appearance distributions. Finally, the color and texture histograms describing super-pixel n are represented as $x_n = (x_n^t, x_n^c)$

serves this PY construction, while adding spatial dependence among super-pixels by associating a layer (real valued function) drawn from a zero mean *Gaussian process* (GP) $\mathbf{u}_k \sim GP(\mathbf{0}, \Sigma)$ with each segment k. Σ captures the spatial correlation amongst super-pixels, and without loss of generality we assume that it has a unit diagonal. Each super-pixel can now be associated with a layer following the procedure described in the previous paragraph, n.e.,

$$z_n = \min\left\{k \mid u_{kn} < \Phi^{-1}(w_k)\right\},\ u_{kn} \sim \mathcal{N}(0, \Sigma_{nn} = 1) \quad (3)$$

Here, $u_{kn} \perp u_{\ell n}$ for $k \neq \ell$ and $\Phi(u)$ is the standard normal *cumulative distribution function* (CDF). Let $\delta_k = \Phi^{-1}(w_k)$ denote a threshold for layer k. Since $\Phi(u_{kn})$ is uniformly distributed on $[0, 1]$, we have

$$\begin{aligned} \mathbb{P}[z_n = 1] &= \mathbb{P}[u_{1n} < \delta_1] = \mathbb{P}[\Phi(u_{1n}) < w_1] = w_1 = \pi_1 \\ \mathbb{P}[z_n = 2] &= \mathbb{P}[u_{1n} > \delta_1]\, \mathbb{P}[u_{2n} < \delta_2] = (1 - w_1)w_2 = \pi_2 \end{aligned} \quad (4)$$

and so on. The extent of each layer is determined via the region on which a real-valued function lies below the threshold δ_{layer}, akin to level set methods. If $\Sigma = \mathbf{I}$, we recover the BOF model. More general covariances can be used to encode the prior probability that each feature pair occupies the same segment; developing methods for learning these probabilities is a major contribution of this paper.

The power law prior on segment sizes is retained by transforming priors on stick proportions $w_k \sim \text{Beta}(1 - \alpha_a, \alpha_b + k\alpha_a)$ into corresponding randomly distributed thresholds $\delta_k = \Phi^{-1}(w_k)$:

$$p(\delta_k \mid \alpha) = \mathcal{N}(\delta_k \mid 0, 1) \cdot \text{Beta}(\Phi(\delta_k) \mid 1 - \alpha_a, \alpha_b + k\alpha_a) \quad (5)$$

Figure 1 displays corresponding graphical model. Image features are generated as in the BOF model.

2.4. Low-Rank Representation

In the preceding generative model, the layer support functions $\mathbf{u}_k \sim \mathcal{N}(0, \Sigma)$ are samples from a Gaussian distribution over N super-pixels. Inference involving GPs involve inverting Σ which is in general a $O(N^3)$ operation and thus scales poorly with increasing image sizes. To cope, we employ a low-rank representation based on $D \leq N$ dimensions, analogous to factor analysis models. We proceed by defining a Gaussian distributed D dimensional latent variable $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$, we then set $\mathbf{u}_k = A\mathbf{v}_k + \epsilon_k$, where A is a N-by-D dimensional factor loading matrix and $\epsilon_k \sim \mathcal{N}(0, \Psi)$, with Ψ being a diagonal matrix. Observe that marginalizing over $\mathbf{v}_k$ results in a model equivalent to the full rank model of the preceding section with $\Sigma = AA^T + \Psi$. The low rank model replaces the $O(N^3)$ operation with an $O(ND^2)$ operation, thus scaling linearly with N[1]. Figure 1 displays the corresponding graphical model.

3. Inference

This section describes a novel, robust to local optima, inference algorithm which is an example of a Maximization Expectation (ME) [29] technique. In contrast to the popular Expectation Maximization algorithms, ME algorithms marginalize model parameters and directly maximize over the latent variables. In our model, the latent variables correspond to segment assignments of super-pixels (z_n). Any configuration of these variables defines a partition of the image. Our strategy is to explore the space of these image partitions by climbing the posterior $p\,(\mathbf{z} \mid \mathbf{x}, \eta)$ surface, where $\eta = \{\alpha, \rho, A, \Psi\}$. It is worth noting that since different partitions will have different numbers of segments, we are in fact searching over models of varying complexities akin to traditional model selection techniques.

The algorithm proceeds by first evaluating the posterior for an initial image partition $\mathbf{z}$. It then modifies the partition in an interesting fashion to generate a new partition $\mathbf{z}'$ which is accepted if $p(\mathbf{z}' \mid \mathbf{x}, \eta) \geq p(\mathbf{z} \mid \mathbf{x}, \eta)$. This process is repeated until convergence. By caching the various mutated partitions, we approximate the posterior distribution over partitions (Figure 5). In what follows, we first describe the innovations required for evaluating the posterior marginal and then the procedure for mutating a partition.

3.1. Posterior Evaluation

In our model (*Figure 1*), the posterior $p\,(\mathbf{z} \mid \mathbf{x}, \eta)$ factorizes as $p\,(\mathbf{z} \mid \mathbf{x}, \eta) \propto p\,(\mathbf{x} \mid \mathbf{z}, \rho) p\,(\mathbf{z} \mid \alpha, A, \Psi)$. The likelihood:

$$p\,(\mathbf{x} \mid \mathbf{z}, \rho) = \int_{\Theta} p\,(\mathbf{x} \mid \mathbf{z}, \Theta) p\,(\Theta \mid \rho) d\Theta \quad (6)$$

[1] A complete time complexity analysis is available in the supplement.