# Active Learning for Structured Probabilistic Models with Histogram Approximation

Qing Sun[1], Ankit Laddha[2], Dhruv Batra[1]
[1]Virginia Tech. [2]Carnegie Mellon University.

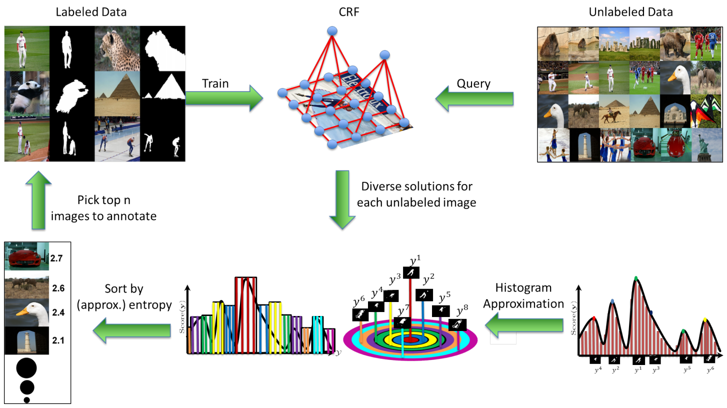



Figure 1: Overview of our approach. We begin with a structured probabilistic model (CRF) trained on a small set of labeled images; then search the large unlabeled pool for a set of informative images to annotate where our current model is most uncertain, *i.e.* has highest entropy. Since computing the exact entropy is NP-hard for loopy models, we approximate the Gibbs distribution with a *coarsened histogram* over $M$ bins. The bins we use are 'circular rings' of varying hamming-ball radii around the highest scoring solution. This leads to a novel variational approximation of entropy in structured models, and an efficient active learning algorithm.

A number of problems in Computer Vision – image segmentation, geometric labeling, human body pose estimation – can be written as a mapping from an input image $\mathbf{x} \in \mathcal{X}$ to an exponentially large space $\mathcal{Y}$ of *structured outputs*. For instance, in semantic segmentation, $\mathcal{Y}$ is the space of all possible (super-)pixel labelings, $|\mathcal{Y}| = L^n$, where $n$ is the number of (super-)pixels and $L$ is the number of object labels that each (super-)pixel can take.

As a number of empirical studies have found [4, 8, 13], the amount of training data is one of the most significant factors influencing the performance of a vision system. Unfortunately, unlike *unstructured* prediction problems – binary or multi-class classification – data annotation is a particularly expensive activity for structured prediction. For instance, in image segmentation annotations, we must label every (super-)pixel in every training image, which may easily run into millions. In pose estimation annotations, we must label 2D/3D locations of all body parts and keypoints of interest in thousands of images. As a result, modern dataset collection efforts such as PASCAL VOC [3], ImageNet [2], and MS COCO [6] typically involve spending thousands of human-hours and dollars on crowdsourcing websites such as Amazon Mechanical Turk.

**Active learning** [10] is a natural candidate for reducing annotation efforts by seeking labels only on the most *informative* images, rather than the annotator passively labeling all images, many of which may be uninformative. Unfortunately, active learning for structured-output models is challenging. Even the simplest definition of "informative" involves computing the entropy of the learnt model over the output-space:

$$H(\mathrm{P}) = -\mathbb{E}_{\mathrm{P}(\mathbf{y}|\mathbf{x})}[\log(\mathrm{P}(\mathbf{y}|\mathbf{x}))] \tag{1a}$$

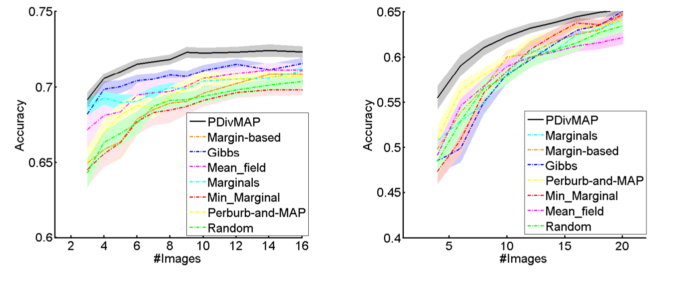
$$= -\sum_{\mathbf{y}\in\mathcal{Y}} \mathrm{P}(\mathbf{y}|\mathbf{x}) \log \mathrm{P}(\mathbf{y}|\mathbf{x}), \tag{1b}$$

which is intractable due to the summation over an exponentially-large output space $\mathcal{Y}$.

**Overview and Contributions.** In this paper, we study active learning for probabilistic models such as Conditional Random Fields (CRFs) that encode probability distributions over an exponentially-large structured output space.

Our main technical contribution is a variational approach [12] for approximate entropy computation in such models. Specifically, we present a crude yet surprisingly effective *histogram approximation* to the Gibbs distribution, which replaces the exponentially-large support with a *coarsened* distribution that may be viewed a histogram over $M$ bins. As illustrated in Fig. 1, each bin in the histogram corresponds to a subset of solutions – for instance, all segmentations where size of foreground (number of ON pixels) is in a specific range $[L\ U]$. Computing the entropy of this coarse distribution is simple since $M$ is a small constant ($\sim$10). Importantly, we prove that the *optimal histogram*, *i.e.* one that minimizes the KL-divergence to the Gibbs distribution, is composed of the mass of the Gibbs distribution in each bin, *i.e.* $\sum_{\mathbf{y}\in \text{bin}} P(\mathbf{y}|\mathbf{x})$. Unfortunately, the problem of estimating sums of the Gibbs distribution under general hamming-ball constraints continues to be #P-complete [11]. Thus, we upper bound the mass of the distribution in a bin with the maximum entry in a bin multiplied by the size of the bin. Fortunately, finding the most probable configuration in a hamming ball has been recently studied in the graphical models literature [1, 7, 9], and efficient algorithms have been developed, which we use in this work.

We perform experiments on figure-ground image segmentation and coarse 3D geometric labeling [5]. As shown in Fig. 2, our proposed algorithm significantly outperforms a large number of baselines and can help save hours of human annotation effort.




(a) Binary Segmentation (b) Geometric Labeling

Figure 2: Accuracy vs the number of images annotated (shaded regions indicate confidence intervals, achieved from 20 and 30 runs respectively). We can see that our approach `Active-PDivMAP` outperforms all baselines and is very quickly able to reach the same performance as annotating the entire dataset.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Dhruv Batra, Payman Yadollahpour, Abner Guzman-Rivera, and Greg Shakhnarovich. Diverse M-Best Solutions in Markov Random Fields. In *ECCV*, 2012.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009.

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, June 2010.

[4] James Hays and Alexei A. Efros. im2gps: estimating geographic information from a single image. In *CVPR*, 2008.

[5] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Recovering surface layout from an image. *IJCV*, 75(1), 2007.

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context, 2014.

[7] Franziska Meier, Amir Globerson, and Fei Sha. The More the Merrier: Parameter Learning for Graphical Models with Multiple MAPs. In *ICML Workshop on Inferning: Interactions between Inference and Learning*, 2013.

[8] D. Parikh and C.L. Zitnick. The role of features, algorithms and data in visual recognition. In *CVPR*, 2010.

[9] Adarsh Prasad, Stefanie Jegelka, and Dhruv Batra. Submodular meets structured: Finding diverse subsets in exponentially-large structured item sets. 2014.

[10] B. Settles. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2012.

[11] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2), 1979.

[12] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[13] Xiangxin Zhu, Carl Vondrick, Deva Ramanan, and Charless Fowlkes. Do we need more training data or better models for object detection? In *BMVC*, 2012.