$\mathcal{HC}$-Search for Structured Prediction in Computer Vision

Michael Lam[1], Janardhan Rao Doppa[2], Sinisa Todorovic[1], Thomas G. Dietterich[1]
[1]Oregon State University. [2]Washington State University.

This paper explores whether $\mathcal{HC}$-Search [3] can be effective in computer vision problems; it has already shown the state-of-the-art performance on structured prediction problems in natural language processing.

The mainstream approach to structured prediction problems in computer vision is to learn an energy function such that the solution minimizes that function. At prediction time, this approach must solve an often challenging optimization problem. Search-based methods provide an alternative that has the potential to achieve higher performance. These methods learn to control a search procedure that constructs and evaluates candidate solutions. The recently-developed $\mathcal{HC}$-Search method [3] has been shown to achieve state-of-the-art results in natural language processing, but mixed success when applied to vision problems. This paper studies whether $\mathcal{HC}$-Search can achieve similarly competitive performance on basic vision tasks such as object detection, scene labeling, and monocular depth estimation, where the leading paradigm is energy minimization.

Instead of training a single global energy function and then solving a global optimization problem, as is done in mainstream approaches, $\mathcal{HC}$-Search decomposes the problem into three steps: (Step 1) Find an initial complete solution, (Step 2) Explore a search tree of alternative candidate solutions rooted at the initial solution, and (Step 3) Score each of these candidates to select the best one (see Figure 1). Any existing method can perform Step 1. Step 2 is guided by a learned heuristic function $\mathcal{H}$, and Step 3 is performed by a learned cost (energy) function $\mathcal{C}$.

Previous work has tested $\mathcal{HC}$-Search on two vision problems [4]. Results on semantic scene labeling were promising, but initial experiments on object detection against significant background clutter were disappointing. In this paper, we claim that the shortcoming of previous work was in Step 2: the formulation of the search space and the learning algorithm employed to train $\mathcal{H}$. This is because the $\mathcal{HC}$-Search specification in previous work [4] used a naive search space defined over relatively small image patches, and thus required an immense branching factor and a very deep search depth (to find a good solution). Previous work developed ad hoc search spaces that sometimes worked and sometimes did not.

To this end, we introduced two improvements to $\mathcal{HC}$-Search: (1) a search operator suited to the vision domain and (2) a more robust learning algorithm by applying the DAGGER algorithm.

We call our search operator the randomized segmentation search space. The search space is defined by probabilistic sampling of a plausible image segmentation from the hierarchy of Berkeley segmentations of the image (UCM [1, 2]). This sampling is realized by randomly picking a threshold on the saliency of region boundaries present in the segmentation. Each search step then involves changing the label of the regions defined by intersecting the chosen segmentation with connected regions defined by the current candidate solution. By choosing different thresholds for UCM, we obtain segmentations at different scales, which when intersected with the candidate solution give us search steps at multiple scales.

This search space is a big improvement, but to successfully apply $\mathcal{HC}$-Search, we also complement this search space by training $\mathcal{H}$ using the advanced imitation learning algorithm DAGGER [6]. The deeper search depths of computer vision problems mean that training $\mathcal{H}$ using simple exact-imitation learning [5] is not sufficient, because exact-imitation learning does not "learn from its own mistakes." During learning, DAGGER blends the current heuristic $\mathcal{H}$ with an oracle heuristic, which is more effective at teaching $\mathcal{H}$ to recover from its errors.

Our evaluation shows that these improvements reduce the branching factor and search depth, and thus yield a significant performance boost. We show that $\mathcal{HC}$-Search with these improvements gives performance comparable to or better than the state of the art across three diverse vision tasks: semantic scene labeling, monocular depth estimation, and object detection in biological images.

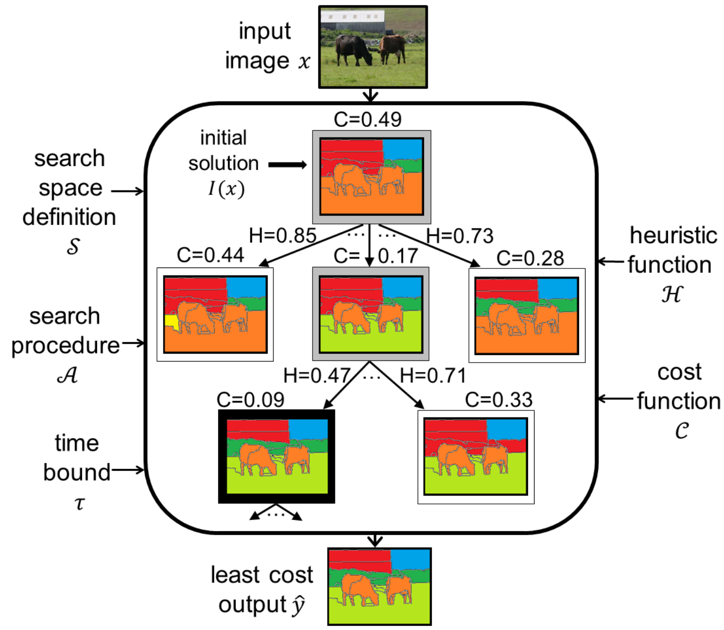

Figure 1: An overview of $\mathcal{HC}$-Search for the semantic scene labeling problem. Given an input x, an initial solution $s_0 = I(x)$ is computed by an (unspecified) procedure I. This forms the starting state in a search space $\mathcal{S}$. Each node in $\mathcal{S}$ consists of a candidate labeling y for x. Nodes are expanded according to search procedure $\mathcal{A}$, which is guided by a learned heuristic $\mathcal{H}$. The search continues until time bound τ. In this figure, $\mathcal{A}$ is greedy search, and the nodes with grey outlines are the nodes visited by $\mathcal{A}$. These nodes are then evaluated by learned cost function $\mathcal{C}$, and the node with the lowest cost is selected as the predicted output $\hat{y}$.

[1] Pablo Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *IEEE Workshop Perceptual Organization (POCV)*, page 182, 2006.

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.161.

[3] Janardhan Rao Doppa, Alan Fern, and Prasad Tadepalli. HC-search: A learning framework for search-based structured prediction. *J. Artif. Intell. Res. (JAIR)*, 50:369–407, 2014.

[4] Michael Lam, Janardhan Rao Doppa, Xu Hu, Sinisa Todorovic, Thomas Dietterich, Abigail Reft, and Marymegan Daly. Learning to detect basal tubules of nematocysts in sem images. In *ICCV 2013 Workshop on Computer Vision for Accelerated Biosciences*, 2013.

[5] Stéphane Ross and Drew Bagnell. Efficient reductions for imitation learning. *Journal of Machine Learning Research - Proceedings Track*, 9:661–668, 2010.

[6] Stéphane Ross, Geoffrey J. Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. *Journal of Machine Learning Research - Proceedings Track*, 15:627–635, 2011.

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