Active Deformable Part Models Inference

Menglong Zhu   Nikolay Atanasov   George J. Pappas   Kostas Daniilidis

GRASP Laboratory, University of Pennsylvania
3330 Walnut Street, Philadelphia, PA 19104, USA

Abstract. This paper presents an active approach for part-based object detection, which optimizes the order of part filter evaluations and the time at which to stop and make a prediction. Statistics, describing the part responses, are learned from training data and are used to formalize the part scheduling problem as an *offline* optimization. Dynamic programming is applied to obtain a policy, which balances the number of part evaluations with the classification accuracy. During inference, the policy is used as a look-up table to choose the *part order* and the *stopping time* based on the observed filter responses. The method is faster than cascade detection with deformable part models (which does not optimize the part order) with negligible loss in accuracy when evaluated on the PASCAL VOC 2007 and 2010 datasets.

1 Introduction

The state-of-the-art performance in object detection nowadays is obtained by star-structured models such as deformable part models (DPM) [3]. DPM-based detectors achieve unrivaled accuracy on standard datasets but their computational demand is significant as it is proportional to the number of parts in the model and the number of locations at which to evaluate the part filters.

Our method is inspired by approaches for speeding-up the part-based models inference such as the cascade DPM [2], branch-and-bound (BB) [7,9,10,8], and multi-resolution schemes [4,11,14,12], which use the responses obtained from initial part-location evaluations to reduce the future computation. For example, BB has been used to prioritize the search over image locations by using an upper bound on the classification score. This paper introduces two novel ideas, which are missing in the state-of-the-art methods for speeding up DPM inference.

First, at each location in the image pyramid, a part-based detector has to make a decision: whether to evaluate more parts and in what order or to stop and predict a label. We formalize this as a planning problem and unlike existing approaches which rely on a predetermined sequence of parts we optimize the order in which to apply the part filters so that a minimal number of part evaluations provides maximal classification accuracy at each location. Second, instead of the threshold-based stopping criteria utilized by most other approaches, we minimize a decision loss which quantifies the trade-off between false positive and false negative mistakes. The novelty and the main advantage of our approach over the related active classification approaches [2,7,13,15,5,6] is that the part

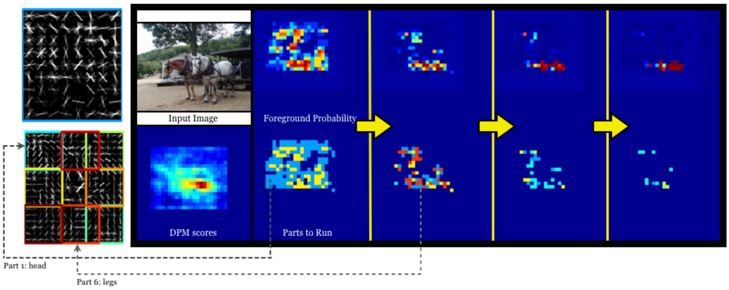

Fig. 1: **Active DPM Inference**: A deformable part model of horse is shown with colored root and parts in the first column. The second column contains an input image and the baseline original DPM scores. The rest of the columns illustrate the ADPM inference which proceeds in rounds. The foreground probability of a horse being present is maintained at each image location (top row) and is updated sequentially. A policy (learned off-line) is used to select the best sequence of parts to apply at different locations. The bottom row shows the part filters applied at consecutive rounds with colors corresponding to the parts on the left. The policy decides to stop the inference at each location based on the confidence of foreground.

order and the stopping criterion are optimized *jointly* and a globally optimal solution is obtained. These ideas have enabled us to propose a novel object detector, Active Deformable Part Models (ADPM), named so because of the active part selection. The detection procedure consists of two phases: an offline phase, which learns a part scheduling policy from training data and an online phase (inference), which uses the policy to optimize the detection task on test images (see Fig. 1). We make the following **contributions** to the state of the art in part-based object detection:
1. A part scheduling policy is obtained to optimize the order of the filter evaluations and to balance their total number with the classification accuracy.
2. The ADPM detector achieves a significant speed-up versus the cascade DPM without sacrificing accuracy.
3. The approach is independent of the representation. It can be generalized to any classification task with several stages (parts) and a linear additive score.

2 Technical approach

We seek an active inference procedure of the Deformable Part Models with n parts and one root. We start by considering the root filter equally as a part but with infinity deformation cost. Now, we can think of the model consisting of just $n+1$ parts and the final score is the sum of individual part scores, i.e. $score(x) = \sum_{k=0}^{n} m_k(x) + b$, where m_k is the k-th part score and b is the bias.

2.1 Score Likelihoods for the Parts

The part scores at a fixed location are random variables, which depend on the input image and the true label. We model the part responses by learning non-

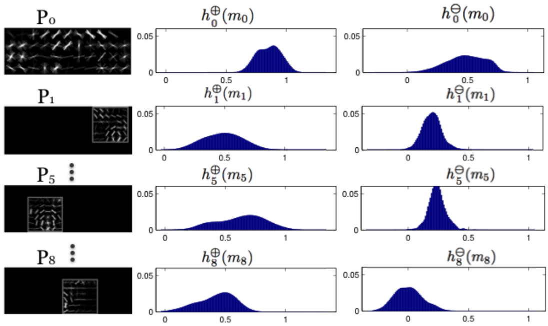

Fig. 2: Score likelihoods for several parts from a car DPM model. The root and three parts of the model are shown on the left. The corresponding positive and negative score likelihoods are shown on the right.

parametric representations of the joint probability density functions (pdf) of the part scores conditioned on the true labels. We assume the responses of the parts of a star-structured model with a given root location are independent conditioned on the the true label. The assumption enables us to learn the joint distribution as products of individual part score distributions and avoids over-fitting. The assumption is empirically verified on VOC 2007 object models. Individual part score distributions are learned by kernel density estimation on scores collected from training set. Let $\{h_k^{\oplus}, h_k^{\ominus}\}$ be the pdf of the response of part k conditioned on the true label being positive or negative. Fig. 2 shows several examples of the score likelihoods obtained from the part responses of a car model.

2.2 Active Part Selection

The active part selection problem is to select and ordered subset of the $n+1$ parts, which when applied at a given image location, has a small probability of mislabeling the location. More specifically, the goal is to learn a policy to non-myopically select which part to run next sequentially, depending on the part responses obtained in the past. The policy learning problem is formulated as minimizing the expected stopping time with the constraint of bounded probability of making false positive and false negative mistakes.

Problem (Active Part Selection)**.** *Given $\epsilon > 0$, choose an admissible part policy π with minimum expected stopping time and probability of error bounded by ϵ:*

$$\begin{aligned} \min_{\pi \in \Pi} \quad & \mathbb{E}[\tau(\pi)] \\ s.t. \quad & Pe(\pi) \leq \epsilon, \end{aligned} \tag{1}$$

where the expectation is over the label y and the part scores $M_{k(0)}, \ldots, M_{k(\tau-1)}$.

The constraint is subsequently incorporated in the objective function by introducing a Lagrangian multiplier. The resulting optimization problem can be solved efficiently with Dynamic Programming, yielding a global optimum.

2.3 Active DPM Inference

During inference, the learned policy is used to select a sequence of parts to apply at each image location or make a prediction. We take a Bayesian approach and maintain a probability of a positive label at each location conditioned on the part scores from the previous rounds. At each round, the policy is queried to obtain either the next part to run or a predicted label based on the score of the last part run and positive label probability. Note that querying the policy at each round is an O(1) operation since it is stored as a lookup table.

3 Results

We compare ADPM[1] versus two baselines, the DPM and the cascade DPM (Cascade) in terms of average precision (AP), relative number of part evaluations (RNPE), and relative wall-clock time speedup. Experiments were carried out on all 20 classes in the PASCAL VOC 2007 and 2010 datasets [1].

1) ADPM was compared to DPM in terms of AP and RNPE to demonstrate the ability of ADPM to reduce the number of part evaluations with minimal loss in accuracy irrespective of the features used. ADPM achieves a significant decrease (90 times on average) in the number of evaluated parts with negligible loss in accuracy. 2) The improvement in detection speed achieved by ADPM is demonstrated via a comparison to Cascade in terms of AP, RNPE, and wall-clock time. To make a fair comparison, we adopted a similar **two-stage** approach with both PCA filters and original filters. An additional policy is was learned using PCA score likelihoods and was used to schedule PCA filters during the first pass. On average, ADPM is 7 times faster than Cascade in RNPE and 3 times faster in seconds. The discrepancy between RNPE and wall-clock speedup is due to the fact that a single full-filter evaluation is significantly slower than the PCA-filter.

4 Conclusion

This paper presents an active part selection approach which substantially speeds up inference with pictorial structures without sacrificing accuracy. Statistics learned from training data are used to pose an optimization problem, which balances the number of part filter convolution with the classification accuracy. Unlike existing approaches, which use a pre-specified part order and hard stopping thresholds, the resulting part scheduling policy selects the part order and the stopping criterion adaptively based on the filter responses obtained during inference. Potential future extensions include consider different computational costs for different filters, optimizing the part selection across scales and image positions and detecting multiple classes simultaneously.

[1] ADPM source code is available at: http://cis.upenn.edu/~menglong/adpm.html

References

1. Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) Challenge. IJCV 88(2), 303–338 (2010)
2. Felzenszwalb, P.F., Girshick, R.B., McAllester, D.: Cascade object detection with deformable part models. In: CVPR. pp. 2241–2248. IEEE (2010)
3. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. PAMI 32(9), 1627–1645 (2010)
4. Fleuret, F., Geman, D.: Coarse-to-fine face detection. IJCV (2001)
5. Gao, T., Koller, D.: Active classification based on value of classifier. In: NIPS (2011)
6. Karayev, S., Fritz, M., Darrell, T.: Anytime recognition of objects and scenes. In: CVPR (2014)
7. Kokkinos, I.: Rapid deformable object detection using dual-tree branch-and-bound. In: NIPS (2011)
8. Lampert, C.H.: An efficient divide-and-conquer cascade for nonlinear object detection. In: CVPR. IEEE (2010)
9. Lampert, C.H., Blaschko, M.B., Hofmann, T.: Beyond sliding windows: Object localization by efficient subwindow search. In: CVPR. pp. 1–8. IEEE (2008)
10. Lehmann, A., Leibe, B., Van Gool, L.: Fast prism: Branch and bound hough transform for object class detection. IJCV 94(2), 175–197 (2011)
11. Pedersoli, M., Vedaldi, A., Gonzalez, J.: A coarse-to-fine approach for fast deformable object detection. In: CVPR. pp. 1353–1360. IEEE (2011)
12. Sapp, B., Toshev, A., Taskar, B.: Cascaded models for articulated pose estimation. In: ECCV. Springer (2010)
13. Sznitman, R., Becker, C., Fleuret, F., Fua, P.: Fast object detection with entropy-driven evaluation. In: CVPR (June 2013)
14. Weiss, D., Sapp, B., Taskar, B.: Structured prediction cascades. arXiv preprint arXiv:1208.3279 (2012)
15. Wu, T., Zhu, S.C.: Learning near-optimal cost-sensitive decision policy for object detection. In: ICCV. pp. 753–760. IEEE (2013)