Poselets [1] have been used in a variety of computer vision tasks, such as detection, segmentation, action classification, pose estimation and action recognition, often achieving state-of-the-art performance. Poselets are part classifiers trained to detect part of a human pose under a given viewpoint. Examples of poselet classifiers are a frontal face, a part of a face and left shoulder, or a hand next to a hip in a side view. Poselet evaluation, however, is computationally intensive as it involves running thousands of scanning window classifiers to detect hundreds of poselet types. We present an algorithm for training a hierarchical cascade of part-based detectors and apply it to speed up poselet evaluation. Our cascade hierarchy leverages common components shared across poselets. We generate a family of cascade hierarchies, including trees that grow logarithmically on the number of poselet classifiers.

Example of our cascade and evaluation algorithm is shown on Figure 1. At each node we train a classifier designed to distinguish between a subset of the parts and the background class. We compute two values at the node: the detection rate (the fraction of positive examples that the node classifier passes) and the retention rate (the fraction of examples the node classifier passes). The detection rate of the cascade is the product of detection rates of the chain of nodes from the root to the leaves, and the retention rate as a function of the number of poselet types. The number of poselet types classified by each node is indicated. Left corner: A zoom on part of the tree. At each node we show the average mask over all classifiers captured by the node, along with the horizontal stripe that was used to classify the node.

We use a dynamic programming approach to find the optimal cascade structure in this restricted state space. An example of the classification cascade is shown on Figure 2. We test our system on the PASCAL dataset [2] and show an order of magnitude speedup at less than 1% loss in AP (Figure 3-left). We also show that our algorithm evaluation cost scales logarithmically with the number of poselet classifiers (Figure 3-right).

Figure 3: Left: Average precision of our classifier (MaxK=4) on the PASCAL 2007 set for the Person category as a function of evaluation speed. We compare against the AP of Downsampling: standard poselet detector with coarser sampling in space and scale; Cascades: independent cascades for each individual poselet; Randcluster: cascade hierarchy with random partition and MaxK=k; cascade hierarchy where each node can have at most k children. Right: Computation time for the same detection rate as a function of the number of poselets.

Figure 1: Cascade hierarchy. Each node is a classifier trained to let through examples of a set of parts (represented as shapes within the node) while filtering out the background class. The set of parts at the node is partitioned and each child is responsible for a subset of them. When an example passes a node classifier, it is evaluated by all of its children.