1. Motivation

- **Existing deep models**: treated pedestrian detection as a single binary classification task, which can not capture rich pedestrian variations.

- **Our idea**: jointly optimizes pedestrian detection with auxiliary semantic tasks, including pedestrian attributes (e.g., ‘backpack’, ‘gender’, and ‘views’) and scene attributes (e.g., ‘vehicle’, ‘tree’, and ‘vertical’).

2. Overall Pipeline

2.1 Scene Attributes

- Shared: enable the learning of shared representation across multiple datasets

- Unshared: Enhance the diversity of attributes

2.2 Pedestrian Attributes

3. Feature Visualization

- Objective: as additive information to bridge the visual gaps between the background datasets and the pedestrian dataset

- Computation:
  1. Organizing the positive and negative patches of pedestrian dataset into two tree structures by HOG feature clustering. Specifically, each tree partitions patches top-down, and have 3 layers and 50 leaf nodes
  2. SPV of each sample is obtained by concatenating the 3 layers and 50 leaf nodes

4. Experimental Results

4.1 Effectiveness of Auxiliary Tasks

4.2 Overall Results on Caltech

4.3 Attribute Prediction (mean accuracy 75+%)

5. Discussions

5.1 Improvement Potentials

- Pre-train models on ImageNet
- Explore more attribute configurations
- Learn sharing representations between multiple pedestrian datasets

5.2 Future Work

- Simultaneously detection and scene parsing