AnchorNet: A Weakly Supervised Network to Learn Geometry-sensitive Features For Semantic Matching

**Introduction**

- **Objective**
  - Semantic matching with weak supervision (image-level labels)
  - Learn a deep representation that surpasses handcrafted features at this task

- **Semantic matching**
  - Given a pair of semantically related objects → estimate matches between corresponding parts

**Motivation**

1. Weak supervision → 
   - more target weak supervision because fully supervised approaches require expensive annotations / synthetic dataset (5.6)
   - standard weakly supervised approaches
     - Step 1: Extract pixel-wise descriptors → pooled deep features, HoG, ... → DSP [1], Proposal Flow [2], SIFT Flow [3], ...
     - Step 2. Use a matching algorithm > top-3, Visual Words (V), SIFT Flow (V)
   - Deep features do not improve because:
     - networks trained with a global classification loss → attention to the most discriminative regions
     - insensitivity to geometry of the objects

2. Pixel-wise descriptors → matching traditional features with deep alternatives does not improve semantic matching accuracy on Pascal VOC [5]

**Key idea**

- Anchoring principle: A set of discriminative and diverse features results in distinct keypoints sensitive to geometry

**Proposed approach - overview**

- Given a large dataset with object category image-level labels
- Learn distinct features of the object categories → object-specific parts
- Use the features within a matching algorithm

---

**Semantic matching**

- Given a pair of images of the same object category → estimate matches between corresponding parts

**Evaluation procedure**

- **Pascal Parts [7]**
  - Evaluation of segmentation transfer between images of meaningful classes
  - related object classes share part segmentations
  - cross-class semantic matching
  - estimate matches between corresponding parts

- **Animal Parts [8]**
  - Cross-class keypoint matching accuracy

**Experiments**

**References**

