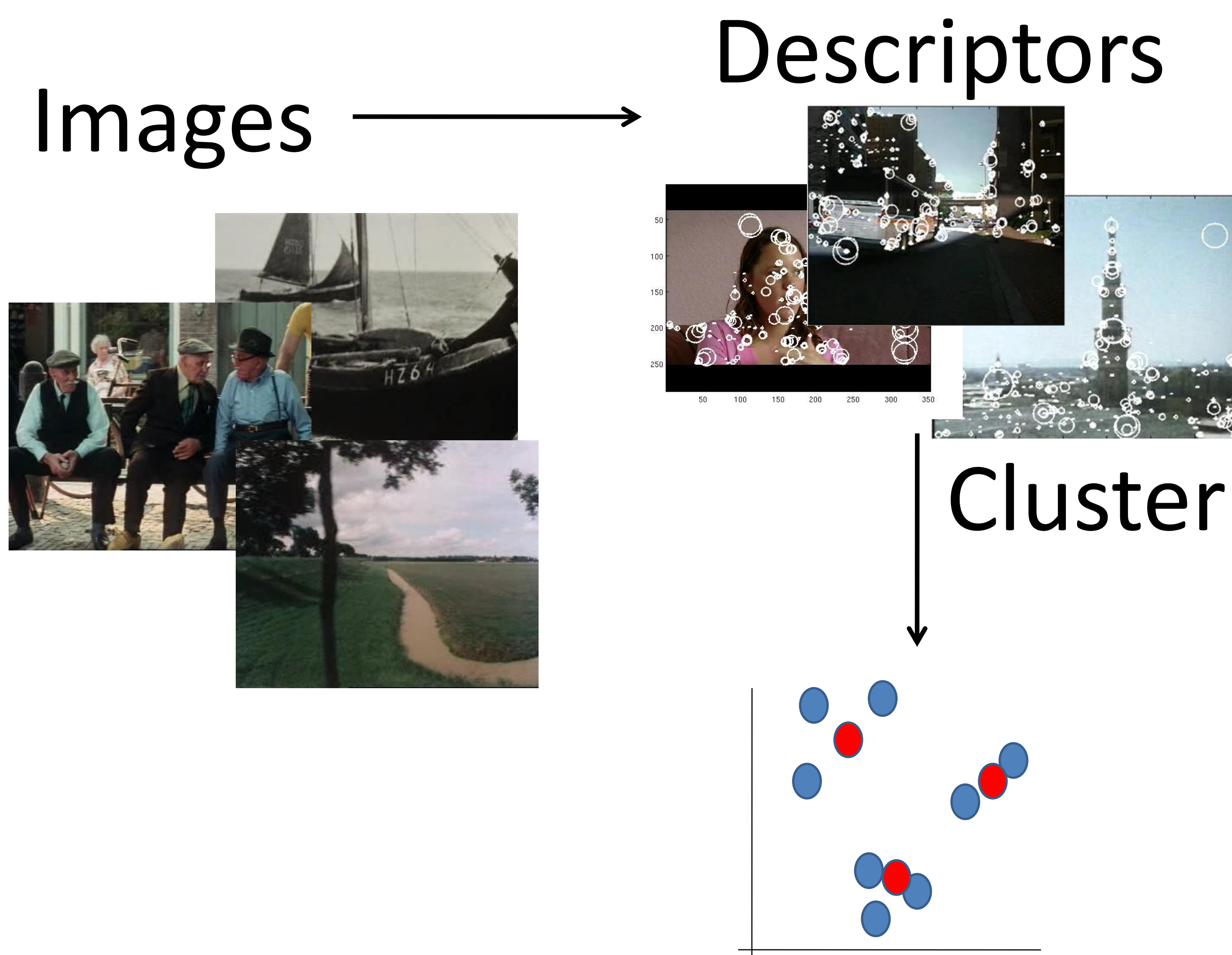


**Introduction:** TRECVID is an international competition in information video retrieval. Specifically, the team at UCF is participating in the high level feature extraction task. This task involves automatically extracting from over 250 hours of video specific shots containing certain concepts such as “Person playing soccer” and “Boat/Ship.”

TRECVID is a difficult data set, containing videos with lots of background clutter, occlusion, and variation, among other challenges. Additionally, the concepts to be detected have large interclass variation. In order to maximize its success, the UCF team first implemented a baseline “bag of words” system, then made incremental improvements, experimenting with bootstrapping, weighting of words, and the addition of new features.

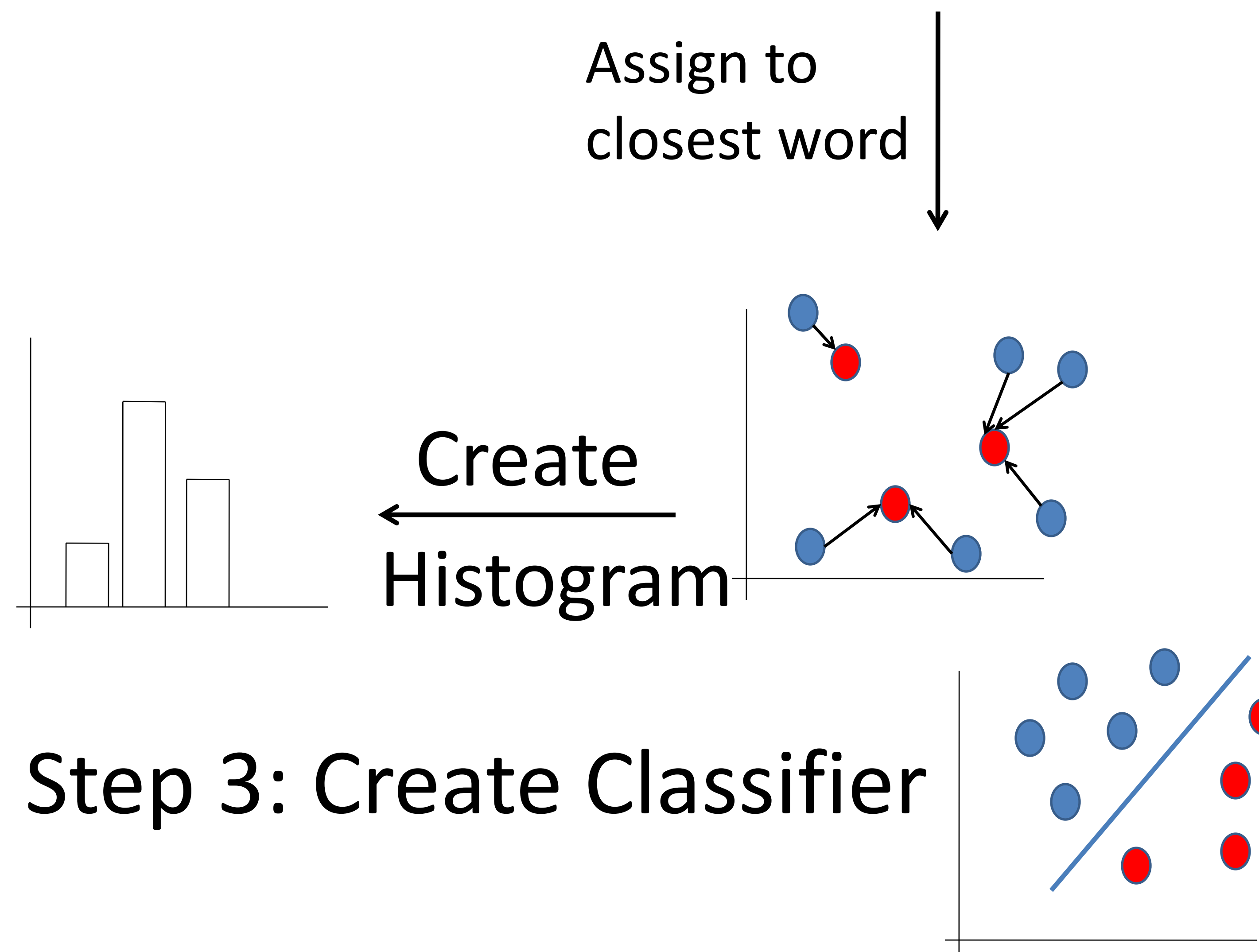
## Bag-of-Words Approach

### Step 1: Create “Code Book”



### Step 2: Create Word Histograms

#### Images → Descriptors



### Step 3: Create Classifier

#### Weighting Words

Soft-Assignment: Assign descriptors to multiple words

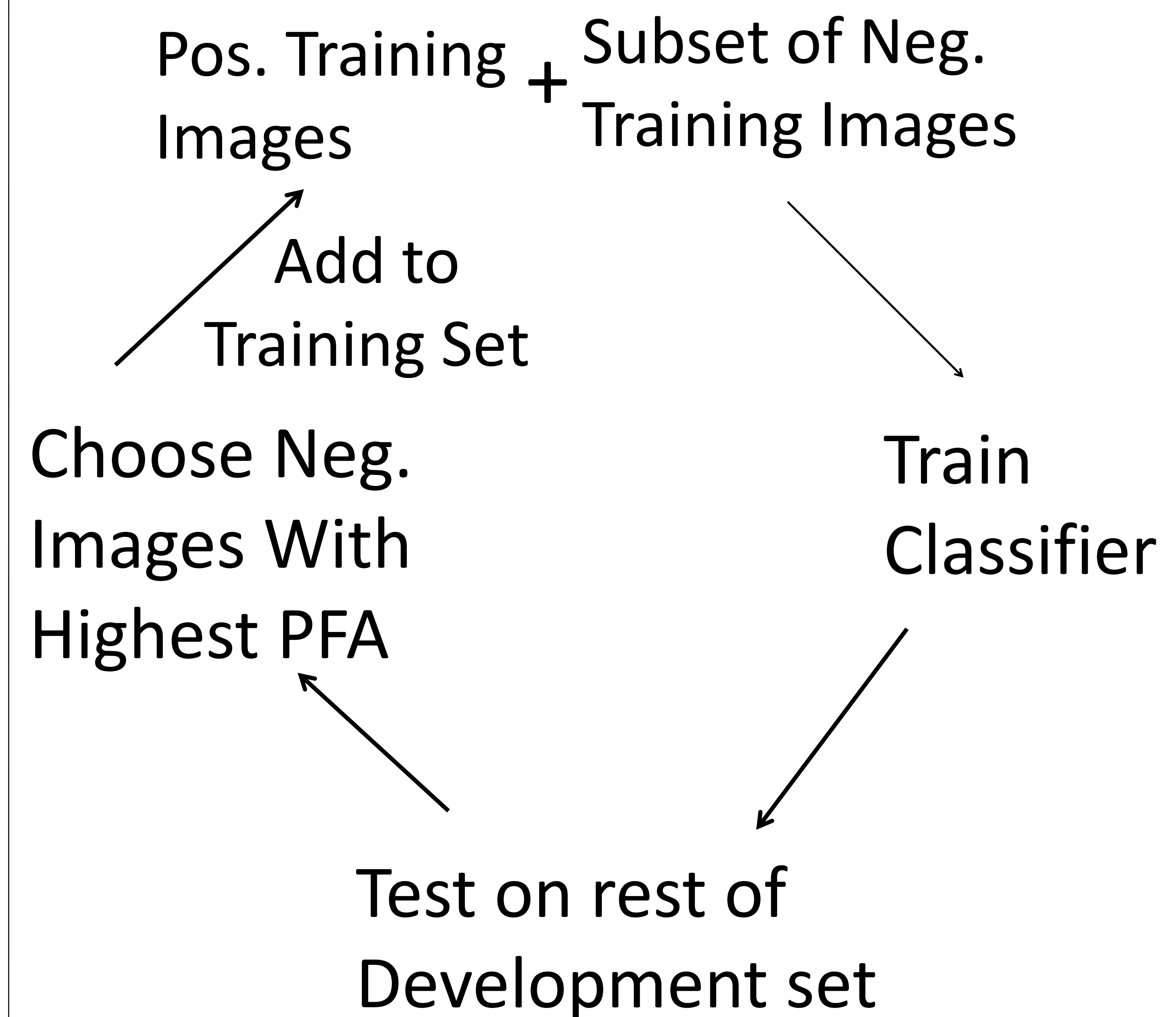
(Mean TF\*IDF)<sup>2</sup>: Give more weight to words that appear infrequently and words that when present in an image, appear multiple times

Base-Line		With Word Weighting	
PD	PFA	PD	PFA
73.5%	26.5%	77.1%	23.9%

## Bootstrapping

Problems:

- 1) Limited number of positive examples
  - 2) Large number of negative examples
- Solution: Find and train on the “hardest” negative examples



Feature	Original		After Bootstrapping	
	PD	PFA	PD	PFA
Chair	90%	31%	90%	8%
Classroom	60%	44%	70%	8%
Infant	70%	33%	80%	9%