A novel method for automatic generation of pseudo ground truth data (less than 4.1% error rate) from images with weak annotations.

- Exploiting existing text metadata as weak annotations, i.e., product databases.
- The method works with an arbitrary detector/recognizer.
- The accuracy of the best-performing open-source recognition model is improved by 3.7% on average.
- A single model surpasses/matches the performance of different second-best methods on multiple benchmarks.

Images with text metadata are common - our method is publicly available. The field heavily relies on synthetic images.

The core idea is to associate the output of the detector and the recognizer with weak annotations. The resulting set of text detections with weak label classifications is referred to as pseudo-ground truth (PGT). The PGT is added to the training set to finetune the recognizer, whose output in turn updates the PGT.

The PGT generation algorithm:

1. **Input**: D - fully-annotated dataset, F - weakly annotated dataset, D - dataset with PGT
2. **Output**: PGT
3. **Algorithm**
   - foreach (det, trg) in D do
     - if (det, trg) is in FindOptimalBox(det, trg) then
       - PGT = add(det, trg)
   - return PGT

The AssignWeak method is implemented as a mutually nearest weak label assignment in a bipartite graph, where detections (and their text transcriptions) and weak labels become pgt candidates if they are mutually nearest, using the edit distance metric.

Experiments: Benchmarks

Single model fine-tuned with PGT from both ABC and UT, 3.7% average accuracy improvement.

On the right: Without punctuation for comparison with other methods.