Fine-Grained Recognition of Thousands of Object Categories with Single-Example Training

Leonid Karlinsky, Joseph Shtok, Yochay Tzur, Asaf Tzadok. IBM Research

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 2017 HONOLULU

Problem: How to quickly detect and recognize thousands of object categories with training on one example per category

Examples:

1. Detect retail products “in the wild” by training on a single image per product
2. Detect brand logos by training on a single graphic per logo type
3. Detect 3D poses of objects inside 2D images, by training on a sparse subset of (partial) object views

Solution: We use a non-parametric probabilistic model for initial detection, CNN-based refinement, and temporal integration (where applicable)

Results: Achieving state-of-the-art performance in a variety of experiments on both existing benchmarks and our own

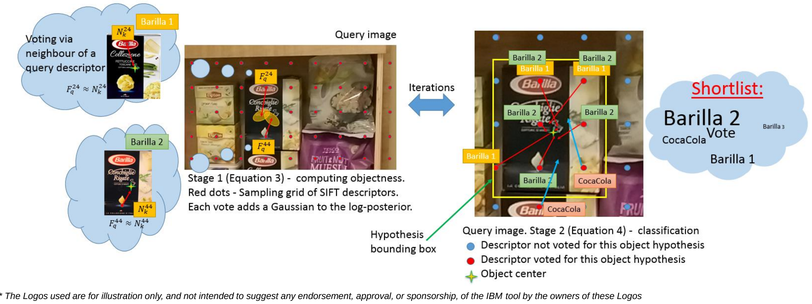
Algorithm Phases

Phase 1:

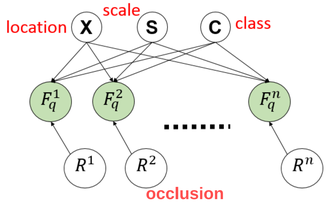
1. **Objectness** – Propose object regions regardless of class
2. **Recognition** – Classify the objects into a shortlist of possibilities
3. **Refinement** – Refine object regions for top-scoring classes

Phase 2: Re-classify objects using CNN

Phase 3: Temporal integration (where applicable)



Probabilistic Model



Observed: Descriptors F_q^i

Unobserved: Object of category C appearing at image location X with scale S (relative to nominal); occlusion event R^i for descriptor F_q^i

Inference: Max over X & S (**objectness**) → sample top C given X & S (**recognition**) → max over X & S given C (**refinement**) → CNN to refine C (**re-classification**) → combine and filter assuming temporal smoothness (**temporal integration**)

Dataset /Algorithm	[7]	[10]	FRCNN [26]	ours phase 1	ours phases 1+2	ours full 1+2+3	ours full top 5
(a) Grocery Products-3.2k [7]	23.49%			42.97%	44.72%		52.16%
(a) Grocery Products: 27 super-classes [7]			81.1%	86.47%			
(b) Grozi-120 [20]				43.22%		49.7%	49.8%
(b) Grozi-120 subset from [7]	13.21%			54.22%		62.64%	62.77%
(f) Flickr32 [27]		74.4%		78.5%	79.3%		
(c) GameStop			27.17%	81.3%	87.5%	89.1%	93.4%
(d) Retail 121			53.67%	84.6%	84.7%	91.3%	91.9%
(e) PCPE 3D pose dataset				93.5%			

[7] M. George and C. Floerkemeier. *Recognizing products. A per-exemplar multi-label image classification approach*. Computer Vision ECCV 2014
[10] F. N. Iandola, A. Shen, P. Gao, and K. Keutzer. *Deeplogo: Hitting logo recognition with the deep neural network hammer*. CoRR, abs/1510.02131, 2015
[20] M. Merler, C. Galleguillos, and S. Belongie. *Recognizing groceries in situ using in vitro training data*. In CVPR 2007.
[26] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN. *Towards real-time object detection with region proposal networks*. ANIPS 2015
[27] S. Romberg, L. G. Pueyo, R. Lienhart, and R. van Zwol. *Scalable logo recognition in real-world images*. ICMR 2011

Learning data augmentation for CNN

Synthesize multiple realistic appearances of an object image using learned photometric filters