Auto-annotation and self-assessment in ImageNet

ILSVRC workshop
December 2013

Vittorio Ferrari
University of Edinburgh

Thanks to M. Guillaumin, D. Kuettel, A. Veznevetohts

We want annotations!

Object location annotations are necessary for many applications

- **Object class detection**
  [Felzenszwalb et al., PAMI 10; Dalal and Triggs CVPR 05; Harzallah ICCV 09, Vedaldi ICCV 09, Wang ICCV 13]

- **Foreground segmentation**
  [Kumar ICCV 11; Kuettel CVPR 12; Tu IJCV 05]

- **Tracking objects in video**
  [Leibe ICCV 07; Breitenstein ICCV 09; Sivic CIVR 05]

- **Shape modeling**
  [Cootes CVIU 95; Shotton ICCV 05; Sebastian PAMI 04; Eslami CVPR 2012]

- **Pose estimation (objects or humans)**
  [Gu ECCV 10; Jiang ICCV 09]
Enter ImageNet ...

- Semantic hierarchy
- 20k classes; 15M images
- 7% with bounding-boxes
- 0 with segmentations

Deng et al. CVPR 2009
Goal

- Goal: automatically add many more annotations

- Bounding-boxes: transfer knowledge from source classes to related target classes
Goal

- Goal: automatically add many more annotations

- Bounding-boxes: transfer knowledge from source classes to related target classes
Goal

- **Goal:** automatically add many more annotations

- **Bounding-boxes:** transfer knowledge from **source classes** to related **target classes**

- **Segmentations:** transfer also from external seed dataset (PASCAL VOC)
Part 1

Bounding-boxes
Monolithic knowledge transfer

- each source-type induces a distribution over windows \(\rightarrow\) reduce location uncertainty
- combine them all to maximize location accuracy of most probable window
- select most probable window – train detector – iterate
Some good outputs ... but which?

IoU for 62k images with ground-truth (out of 0.5M auto-annotated)
Self-assessment

Any auto-annotation machine will make mistakes

• Can we automatically find the good outputs?

• How do we quantify prediction’s uncertainty?

• New scheme:
  localization with self-assessment

Paper available at
Probabilistic self-assessment score

\[ \eta(w, \lambda) \equiv \arg \max_\xi P(Y(w) = \xi | w) \geq \lambda \]

- Largest overlap \( Y(w) \) that \( w \) is predicted to have with at least \( \lambda \) probability
- Problem: probabilistic inference in high dimensional space is hard (e.g. GPs)

Rasmussen and Williams, 06
Overview

1. Objectness windows
2. Low-dimensional Associative Embedding
3. Knowledge transfer with Gaussian Processes
4. Windows scoring with self-assessment

$\eta(w, \lambda)$

Alexe CVPR 2010
Associative window representation

How does window look?

What does window look like?

HOG descriptor
Exemplar SVMs

Source

Malisiewicz et al. ICCV’11; Aghazadeh et al. ECCV’12; Dong et al. CVPR’12; Juneja CVPR’13
Associative Embedding

Source

Target
Associative Embedding

- $x_i \theta^T$
- E-SVMs $\{\theta\}$
- Latent semantic analysis
- window = vector of E-SVM outputs rich, but also highly correlated
- Just 3 dimensions from initial SURF BoW or HOG features
Associative Embedding

\[
\min_{U, V} \| A - UV^T \|_F^2 \quad \text{s.t. dim}(u) = \text{dim}(v) = d
\]

In practice: perform truncated SVD
\rightarrow just 3 dimensions enough for low reconstruction error

Hofmann “Probabilistic latent semantic indexing”, SIGIR’99
Embedding Scottish deerhound

- dimensions tend to correspond to aspects of appearance
- can be built on top of any feature representation
Overview

1. Objectness windows
2. Low-dimensional Associative Embedding
3. Knowledge transfer with Gaussian Processes
4. Windows scoring with self-assessment

Alexe CVPR 2010
Probabilistic inference using GP

Source windows in AE space

GP transfers IoU from source windows to target ones

Final output: self-assessed scoring of all target windows
Probabilistic inference using GP

\[ \eta(w, \lambda) \equiv \arg \max_{\xi} P(Y(w) = \xi | w) \geq \lambda \]

- Probabilistic inference enables self-assessment
- Compute \( \eta \) as mean - stdv
- Non-parametric model, highly non-linear

Rasmussen and Williams, 06; Li et al. CVPR’10
Experiments: ImageNet

- 219 target classes with some bounding-boxes for themselves and their siblings

- Total 500k images, 92k with bounding-boxes (source = 60k, target = 32k)

- evaluate PASCAL VOC intersection-over-union, averaged over images and classes
Experiments: source sets

Target set

Source = Family

Marsupial

Koala

Kangaroo

Cuscus
Experiments: ImageNet

Distribution of ground-truth boxes size w.r.t. image size.

Highly varied dataset, with many difficult images
(see also http://www.image-net.org/challenges/LSVRC/2012/analysis/)
Automatically selected localizations
\( \eta(w,0.5) > 0.7 \), window area < 0.5*image area, 1 per class
Automatically selected localizations

$\eta(w,0.5) > 0.7$, window area $< 0.5 \times$ image area, 1 per class

schooner
screen
seaplane
ski-plane
skunk
soccer ball
stallion
streetlight
tree squirrel
van
warthog
water buffalo
weasel
wild boar
wirehair
wombat
Automatically selected localizations
\( \eta(w,0.5) > 0.7 \), window area < 0.25*image area, 1 per class

- airliner
- airship
- asian wild ox
- australopithecine
- balloon
- baseball
- basketball
- bomber
- bowhead
- buckle
- button
- candle
- chimpanzee
- domestic sheep
Automatically selected localizations
η(w,0.5)>0.7, window area < 0.25*image area, 1 per class

propeller plane
rabbit
rorqual
sailboat
schipperke
schooner
screen
seaplane
soccer ball
streetlight
television
warthog
water buffalo
Automatically selected errors!
\[ \eta(w,0.5) < 0.4, \text{ 1 per class} \]

Armadillo

Caribou

Cow

Elk

Polecat

Rorqual
Baselines and competitors

• Guillaumin and Ferrari CVPR 2012

• MKL-SVM
  On Source, train a single MKL-SVM with same features as ours:
  – HOG (linear)
  – SURF BoW ($\chi^2$, linearized with explicit feature maps)
  – Location and scale
  – Objectness score
  – SVM output = self-assessment score

Vedaldi CVPR 2010 & PAMI 2012
- AE+GP with weakest source (siblings) > competitors with any source
- cusps in competitor curves: poor self-assessment
- mixed supervision (self->family) confuses MKL-SVM, but not AE+GP
- difference greater earlier in the curve: validates our self-assessment
Self-assessment curves

- Net effect for user: querying for boxes with predicted overlap > 70% with probability 0.5 returns 30% with mean IoU 73% \(\rightarrow 150k\) images!

- Useful? MKL +6% IoU on targets when adding this as extra training data
Conclusion for Part 1

• 500k bounding-boxes produced so far
• Overall PASCAL-level detection rate ±70%
• Automatically return 30% of data with high localization accuracy
• Aim to process all ImageNet and produce **5M accurate BBs**

(hopefully ;)

All new bounding-boxes will be online soon

http://groups.inf.ed.ac.uk/calvin
Objectness measure v2.2

- class-generic proposals to speedup class-specific detectors (CVPR 10)

- Objectness probability crucial to support other applications (e.g. weakly supervised learning / auto-annotation, tracking, content-aware resizing, assessing image quality or difficulty, saliency measures)

- Objectness good at low number of windows
Part 2

Quick update

Segmentations

Kuettel, Guillaumin, Ferrari, *Segmentation Propagation in ImageNet*, ECCV 2012,

→

Guillaumin, Kuettel, Ferrari,

*ImageNet auto-annotation with segmentation propagation*, submitted to IJCV
Exploit all available information

- Segmented images help segment images with similar objects
- Bounding-boxes constrain segmentations
- Semantically related object classes can share appearance
Segmentation propagation in a hierarchy

- segmented images (source) help segmenting new ones (target): segmentation transfer
- Proceed recursively: propagation
Segmentation propagation in a hierarchy

- Start from the easiest images

Source

Target

transportation

wheeled vehicle

car

aircraft

VOC10
Segmentation propagation in a hierarchy

Source

Segmentation transfer

VOC10

Target

transportation

aircraft

wheeled vehicle

car
Segmentation propagation in a hierarchy

Joint segmentation

VOC10

transportation

aircraft

wheeled vehicle

car
Segmentation propagation in a hierarchy

VOC10

Source

Segmentation transfer

Target

aircraft

transportation

wheeled vehicle
car
Segmentation propagation in a hierarchy

Joint segmentation

VOC10

transportation

aircraft

wheeled vehicle

car
Segmentation propagation in a hierarchy

VOC10

Semantic relation

aircraft

wheeled vehicle

car

transportation
Segmentation propagation in a hierarchy
Segmentation propagation in a hierarchy
Segmentation propagation in a hierarchy

VOC10

transportation

aircraft

wheeled vehicle

car
Segmentation propagation in a hierarchy

VOC10

transportation

aircraft

wheeled vehicle

car
• Related to earlier annotation transfer works
  [Russel NIPS07, Liu CVPR09, Guillaumin ICCV09, Rosenfeld ICCV11, Kuettel CVPR12]
1. Sample windows on objects

Segmentation transfer

Source

Target image

Objectness sampling

[Alexe CVPR10]
1. Sample windows on objects
2. Find visually similar windows

Segmentation transfer

HOG + compact binary code for efficient retrieval

[Dalal CVPR05, Torralba CVPR08, Gong CVPR11]
1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations
1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations

Segmentation transfer

Target image

[Kuettel CVPR12]
1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations
4. Initialize and run GrabCut
Segmentation transfer

- Window-level vs image-level
  + less variability, easier to match
  + compositionality
- Fast thanks to binary codes (<2 sec/image)
- Code available: [http://groups.inf.ed.ac.uk/calvin/software.html](http://groups.inf.ed.ac.uk/calvin/software.html)

[Kuettel CVPR12; Gong CVPR11]
Joint segmentation with shared appearance

• Extend GrabCut to multiple images [Rother SIGGRAPH04]
• Additional unary potentials for transfer mask, class-wide appearance model, and appearance model from related classes segmented before
• *Linear* in the number of images $\rightarrow$ very fast
Experiments on ImageNet
Experiments on ImageNet
Experiments on ImageNet

Setup
• 0.5M images, 577 classes
• 10 images X 446 classes annotated with Mechanical Turk

Results (intersection-over-union)
• 24.0  GrabCut initialized from image center
• 52.7  Transfer from fixed source pool VOC10
• 57.3  Full propagation

+ segmentation transfer much better than baseline
+ propagation helps
+ IoU better than accuracy for measuring segmentation performance
Experiments on ImageNet

Breakout over stages

+ ground-truth bounding-boxes help stage 1 a lot
+ propagation helps at all stages (+2%)
+ >>ECCV12: corrected segmentation available on our website
+ useful? +3% accuracy on dog/horse segmentation when adding this
Conclusions for Part 2

- Segmentation Propagation: an efficient scheme to recursively segment images in ImageNet
- Produced 500k segmentations with average IoU 57.3%

Updated segmentations (better than last year’s ;)
http://www.vision.ee.ethz.ch/~mguillau/imagenet.html

Segmentation transfer code:
http://biwinas03.ee.ethz.ch/duettel/cvpr12/