Training object class detectors with click supervision

Dim P. Papadopoulos, Jasper R. R. Uijlings, Frank Keller and Vittorio Ferrari

Train object detectors

Full supervision: draw bounding boxes

time consuming (35s per box):
ImageNet protocol [Su AAAI’12]

Crowd-sourcing clicks

Weak supervision: image labels

very cheap, but low quality detectors
[Bilen CVPR’16, Cinbis CVPR’14, Deselaers ECCV’10, Siva ICCV’11]

Ours: center click supervision

• reduce annotation time by 9x-18x
• high quality detectors without ever drawing any bounding-boxes

Incorporating clicks into WSOL

Multiple Instance Learning (MIL)

A. Re-training

object detectors

B. Re-localizing

objects

- initialization: full images
  (Cinbis CVPR’14, Pandey ICCV’11)
- re-localization: pick proposal with the highest appearance score
  $S_\text{det}(p, c, \sigma_\text{det})$

One-click supervision

- initialization: largest proposal centered on click
- re-localization: use both appearance and center click

Two-click supervision

- estimate object center even more accurately
- estimate object area from distance between 2 clicks

Learn hyper-parameters

• CorLoc

Quantitative results

- PASCAL VOC 2007
- Fast R-CNN, AlexNet, EdgeBoxes proposals

Simulated results on COCO

Create a realistic scenario

Click supervision: human verification

Training:

Two-click supervision

One-click supervision

Human Verification

Drawing

MIL

Results

Weak supervision vs 1-click vs 2-click

- substantial better than WSOL at a modest cost
- two-click supervision performs even better
- reduces annotation time by 9x-18x,
  almost as good as fully supervised
- even better trade-off than human verification
  [Papadopoulos CVPR’16]

Annotations images

• Amazon Mechanical Turk
  • PASCAL VOC 2007 trainval set
  • 19612 clicks for 5011 images
• Annotation time: 1.9s per click
• Average error: 19.5 pixels
• Just 75s in total

Clicks available online:

ImageNet

Just human verification

90% mAP of full supervision,
9x less human annotation time

Success

Failure

Estimate object area from distance between 2 clicks

Learn \( \sigma_\text{det} \)

Learn \( \sigma_\text{area error} \)

Simulated results on COCO

MIL

Human Verification

Click supervision

Training set (80 classes – 82,783 images)

Test set: validation set (40,137 images)

Center clicks clearly outperform human verification (3.5x cheaper)