# ImageNet Auto-Annotation with Segmentation Propagation

**Matthieu Guillaumin · Daniel Küttel · Vittorio Ferrari**

Bryan Anenberg & Michela Meister

# Outline

**Goal & Motivation**

System Overview

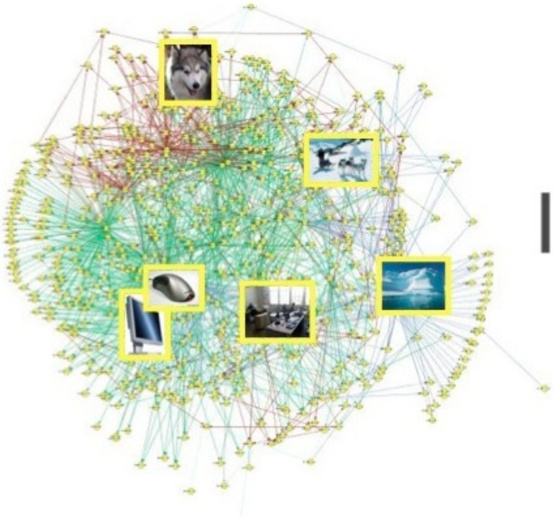
Segmentation Transfer

Joint Segmentation

Results

# Goal

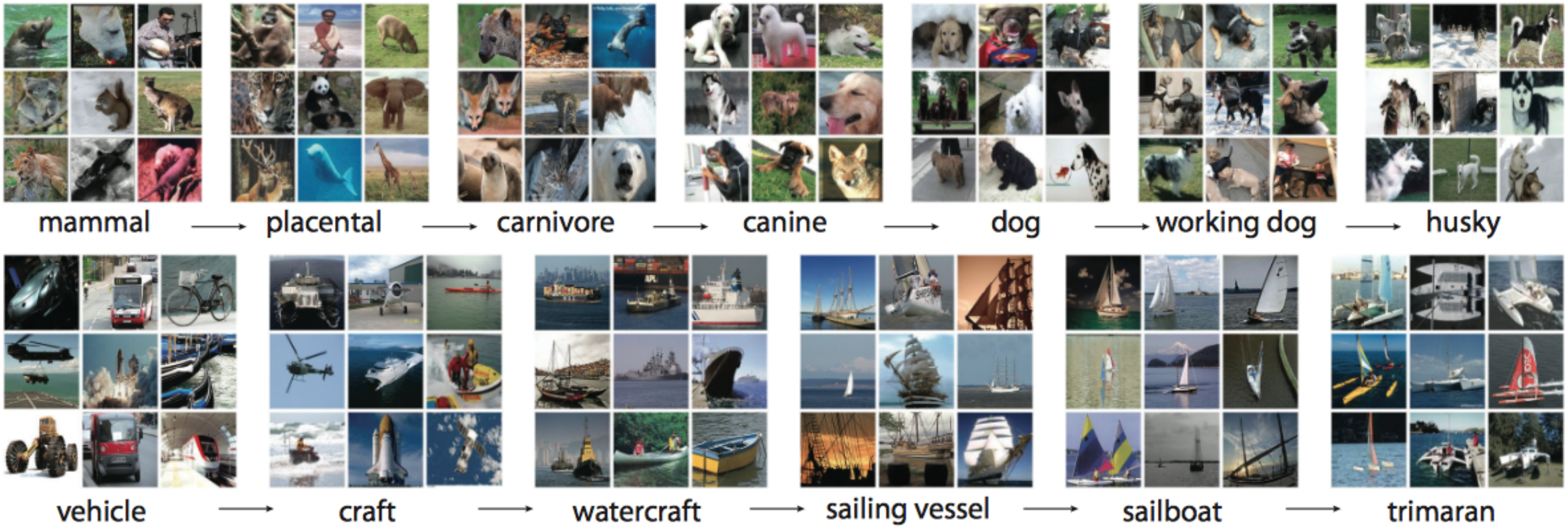
Automatic foreground pixel-level segmentation of ImageNet



IMAGENET

# ImageNet

- large-scale, hierarchical
- 15,000,000 images
- 22,000 classes



# Outline

Goal & Motivation

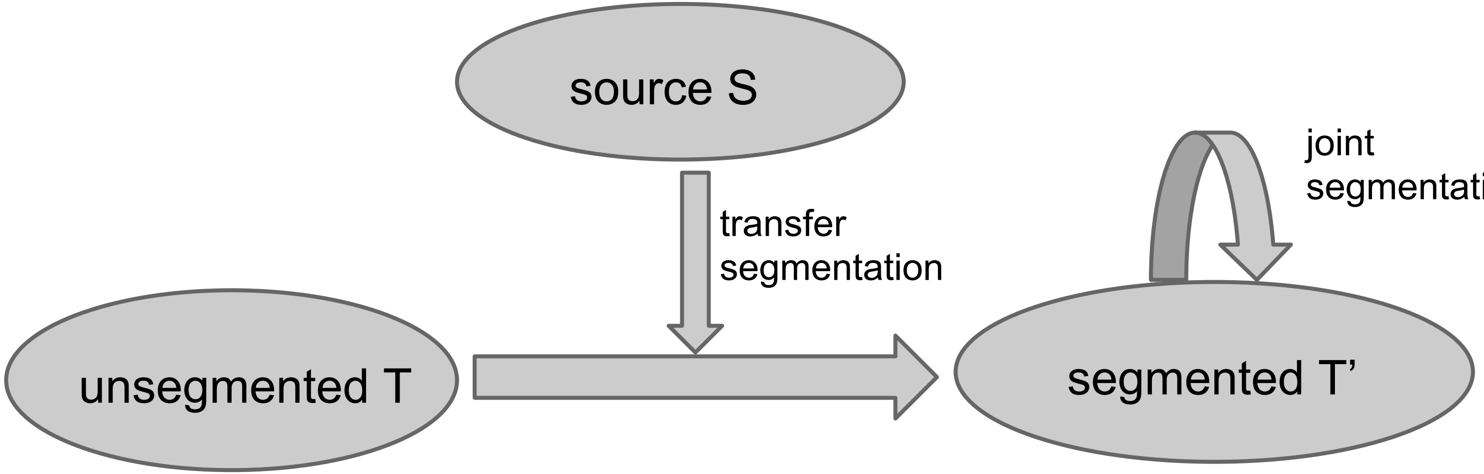
System Overview

Segmentation Transfer

Joint Segmentation

Results

# System Overview




new source = S U T’

[3] Guillamin, Kuettel, Ferrari

# Exploit all available information

External GT segmentations

*PASCAL VOC10*

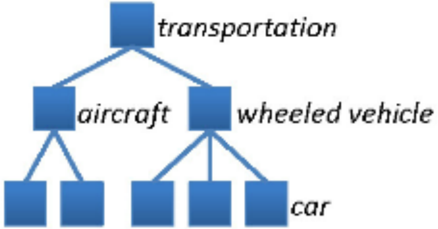
Bounding boxes

*aircraft*

Object class

*cat*

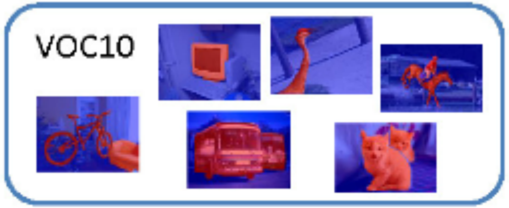
Semantic hierarchy (WordNet)




- Segmented images help segment images with similar objects
- Bounding boxes constrain segmentations
- Semantically related object classes can share appearance

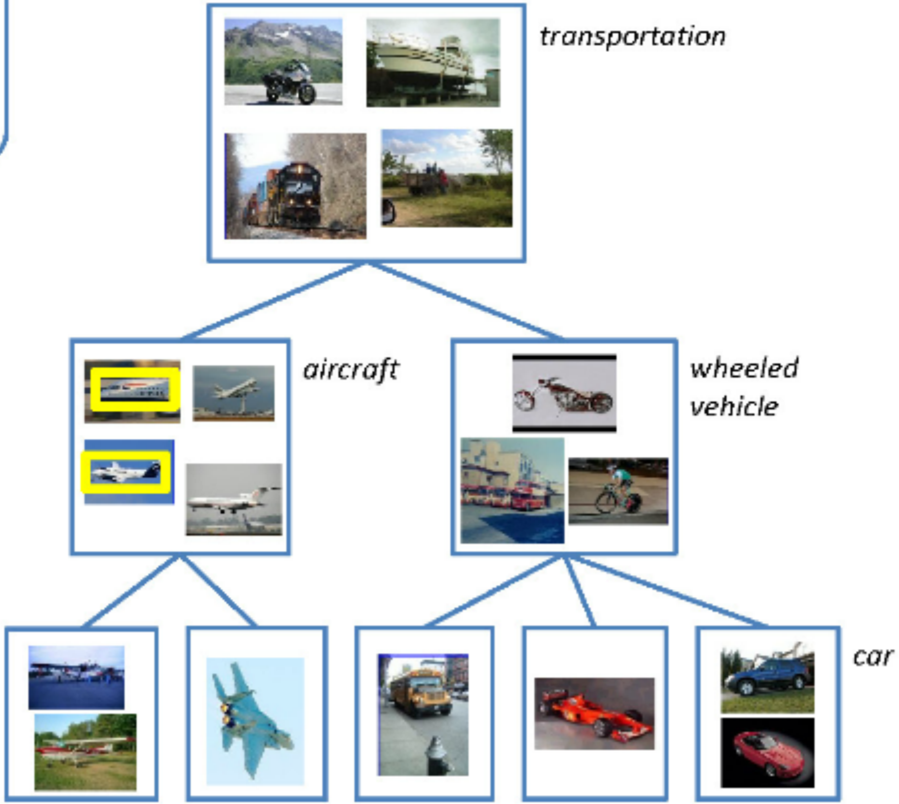
slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



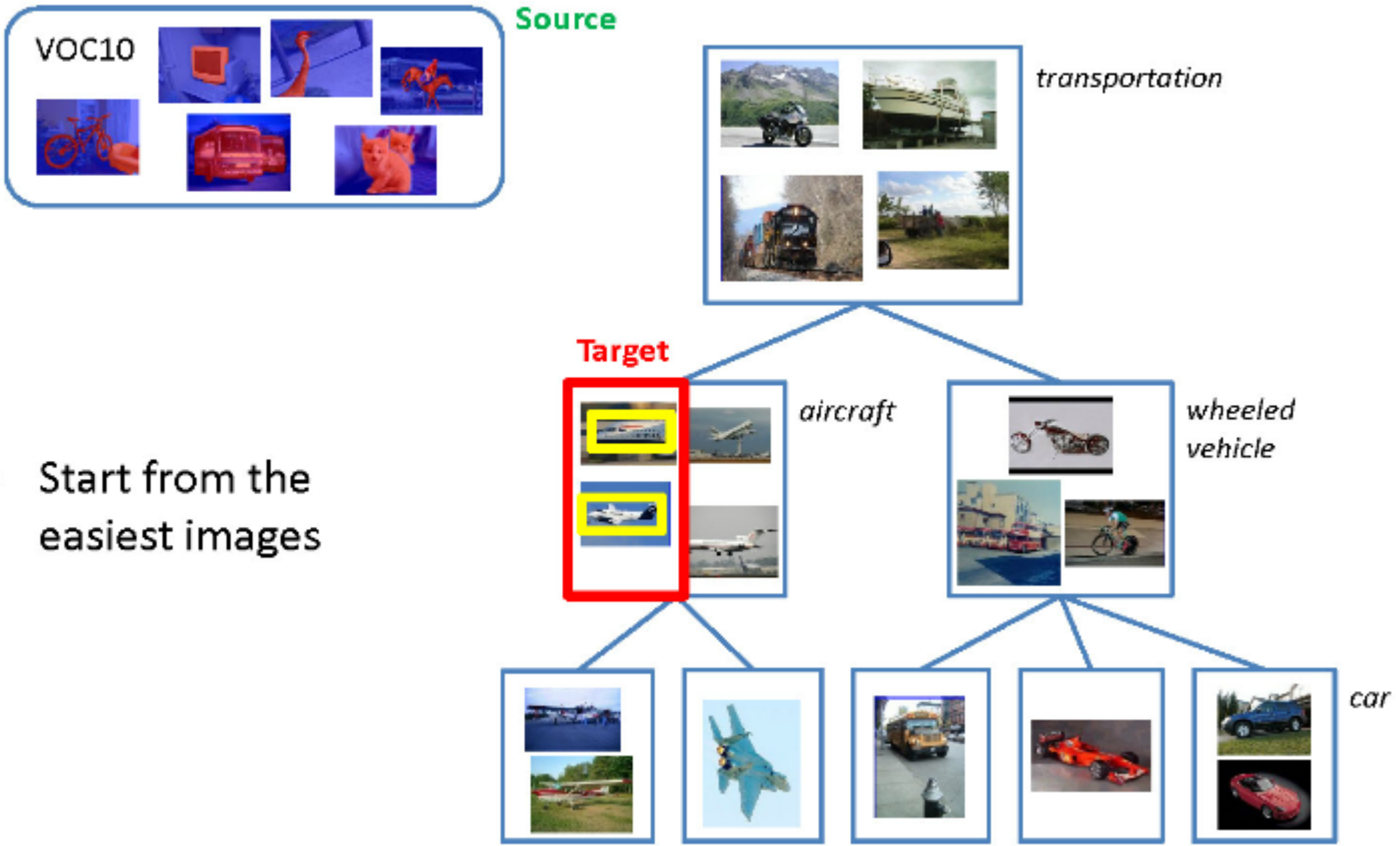

segmented images (*source*) help segmenting new ones (*target*): *segmentation transfer*

Proceed recursively: *propagation*




slide credit: V. Ferrari

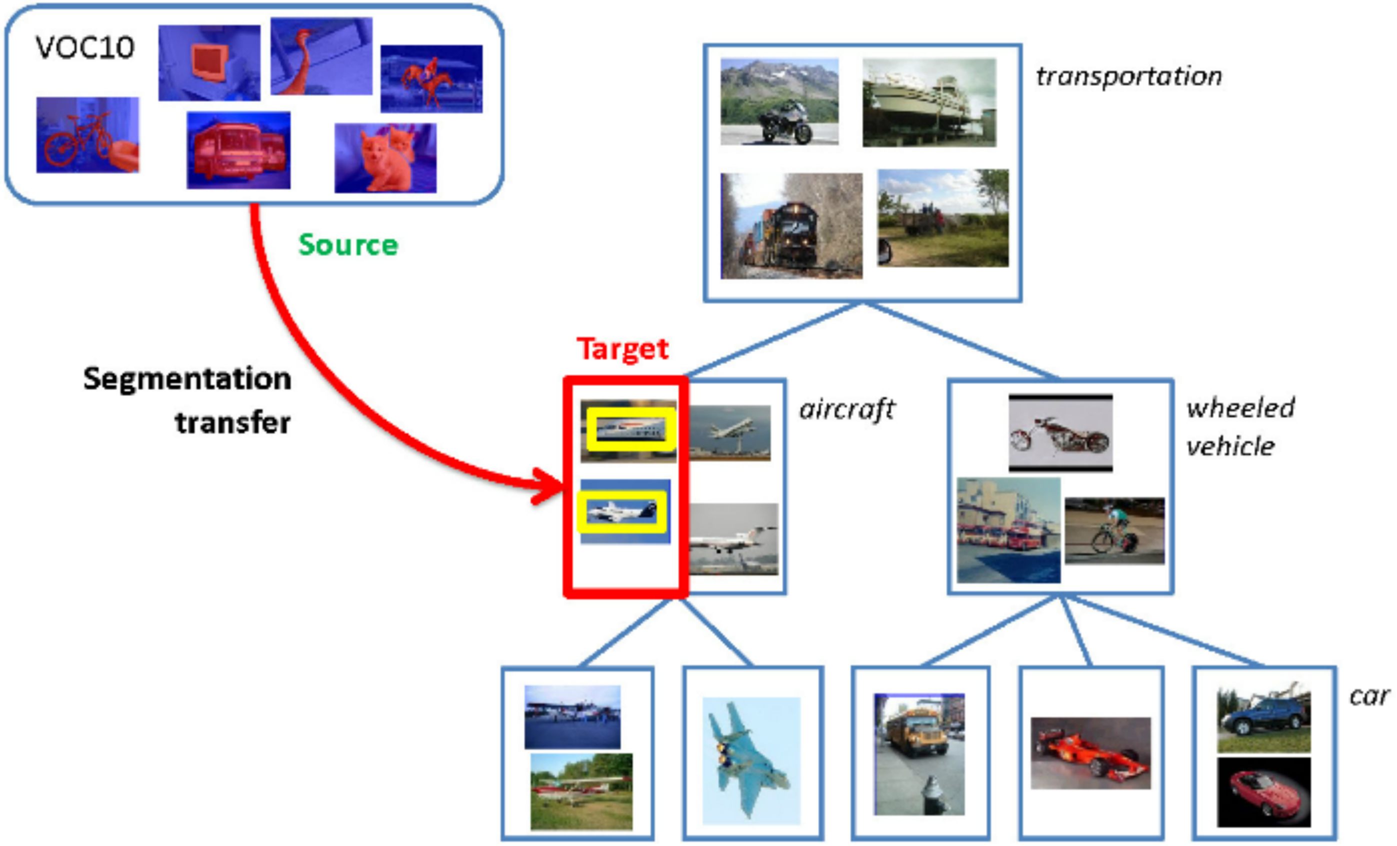
# Segmentation propagation in a hierarchy




Start from the
easiest images

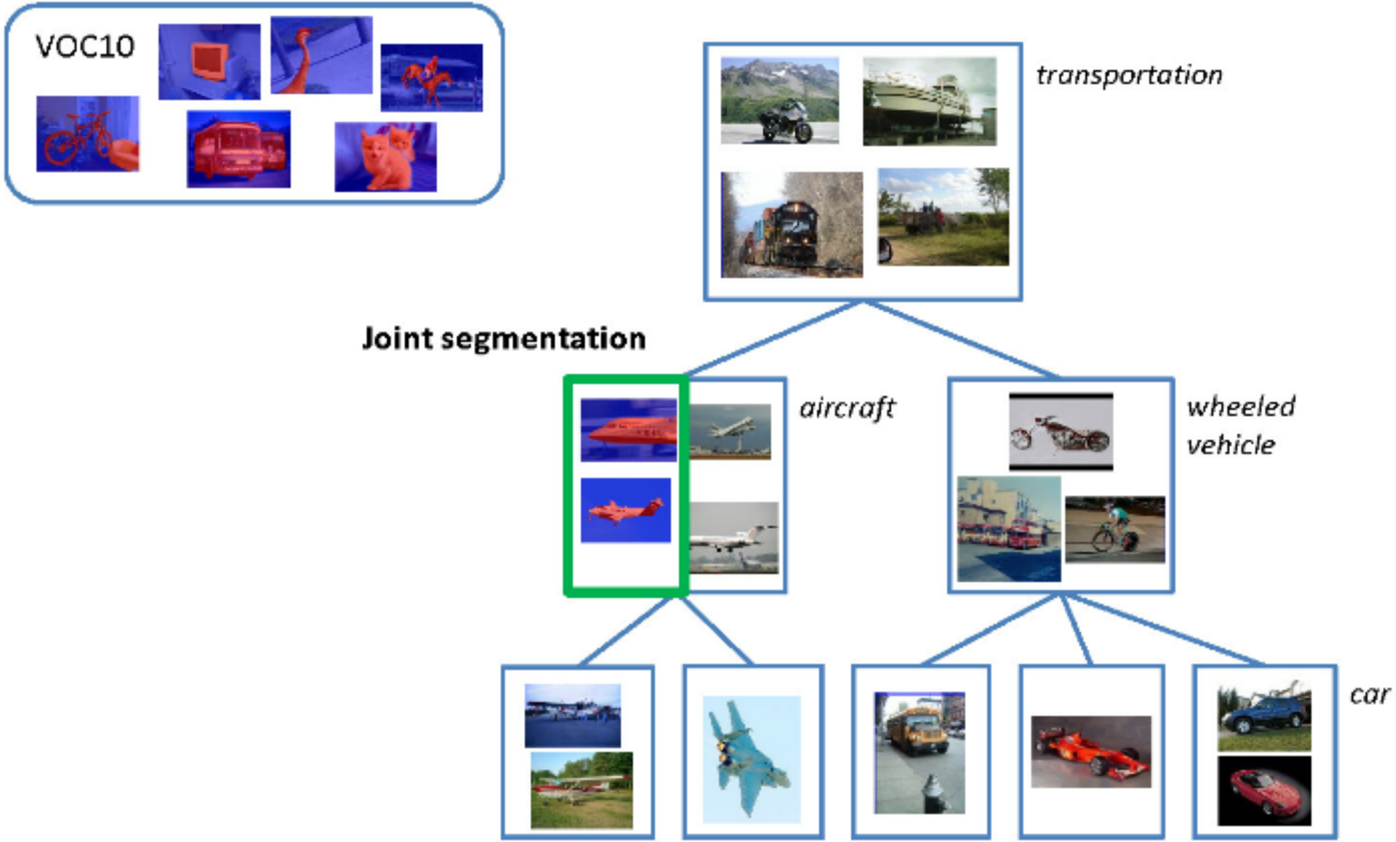
slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



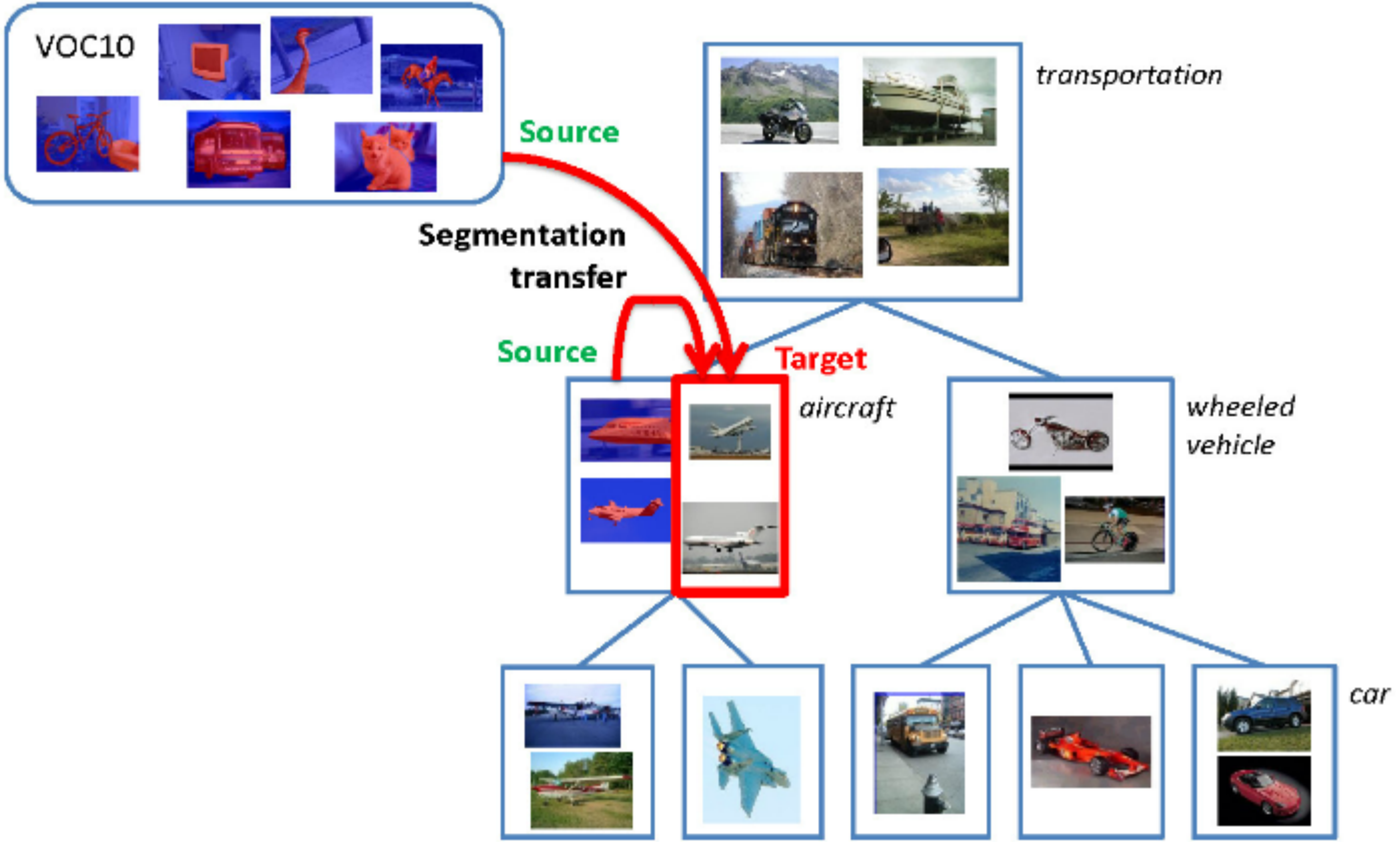

slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



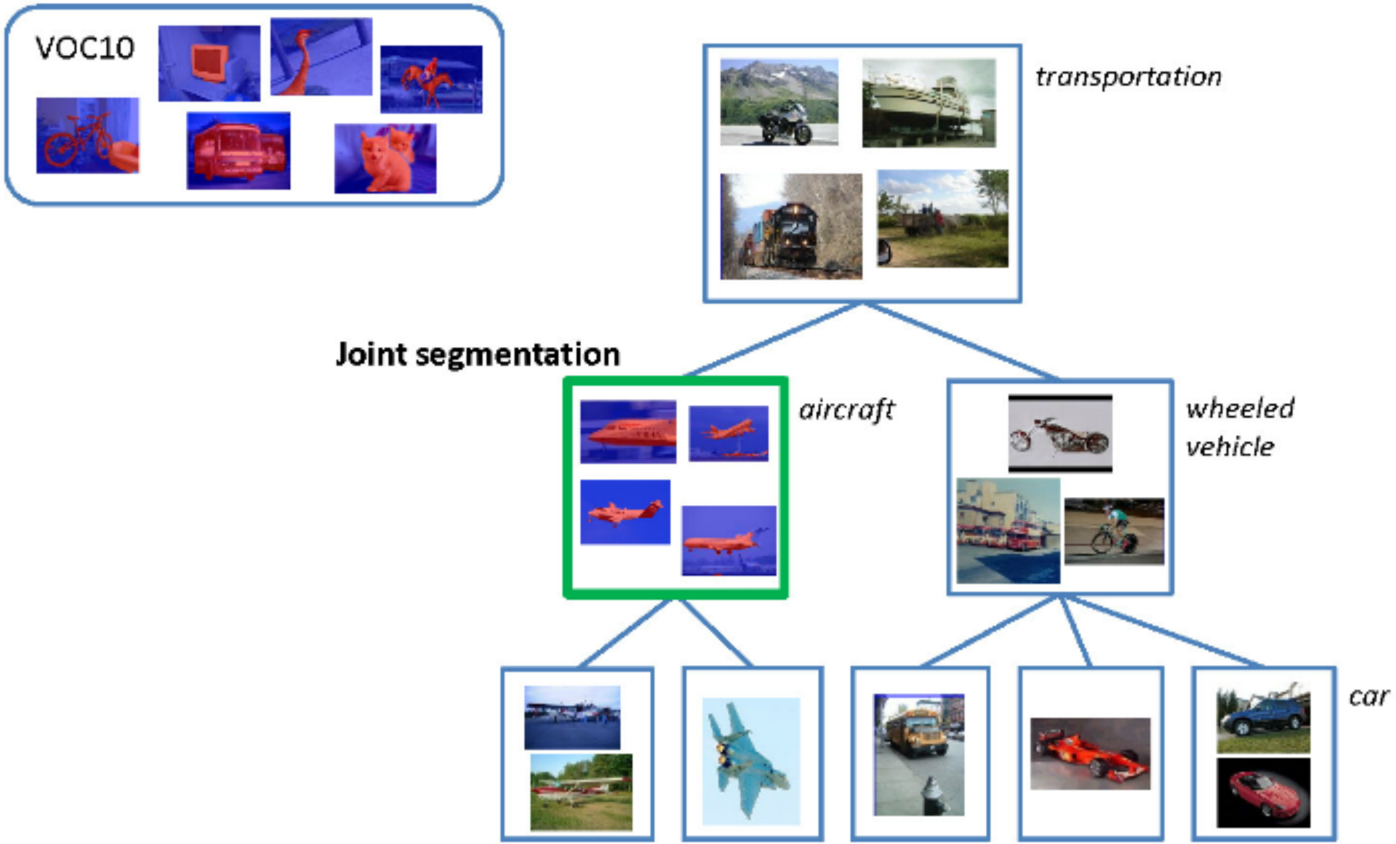

slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



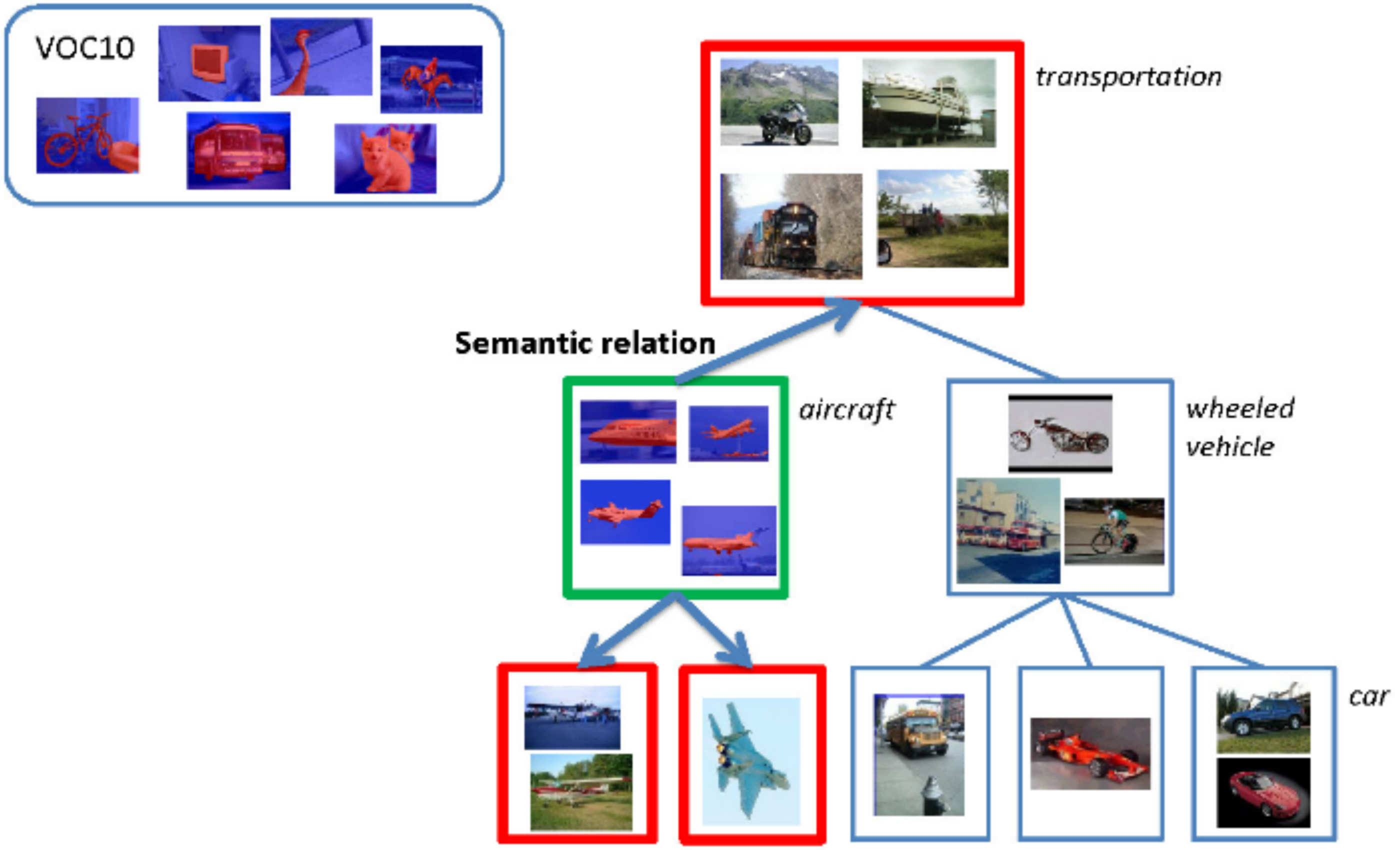

slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



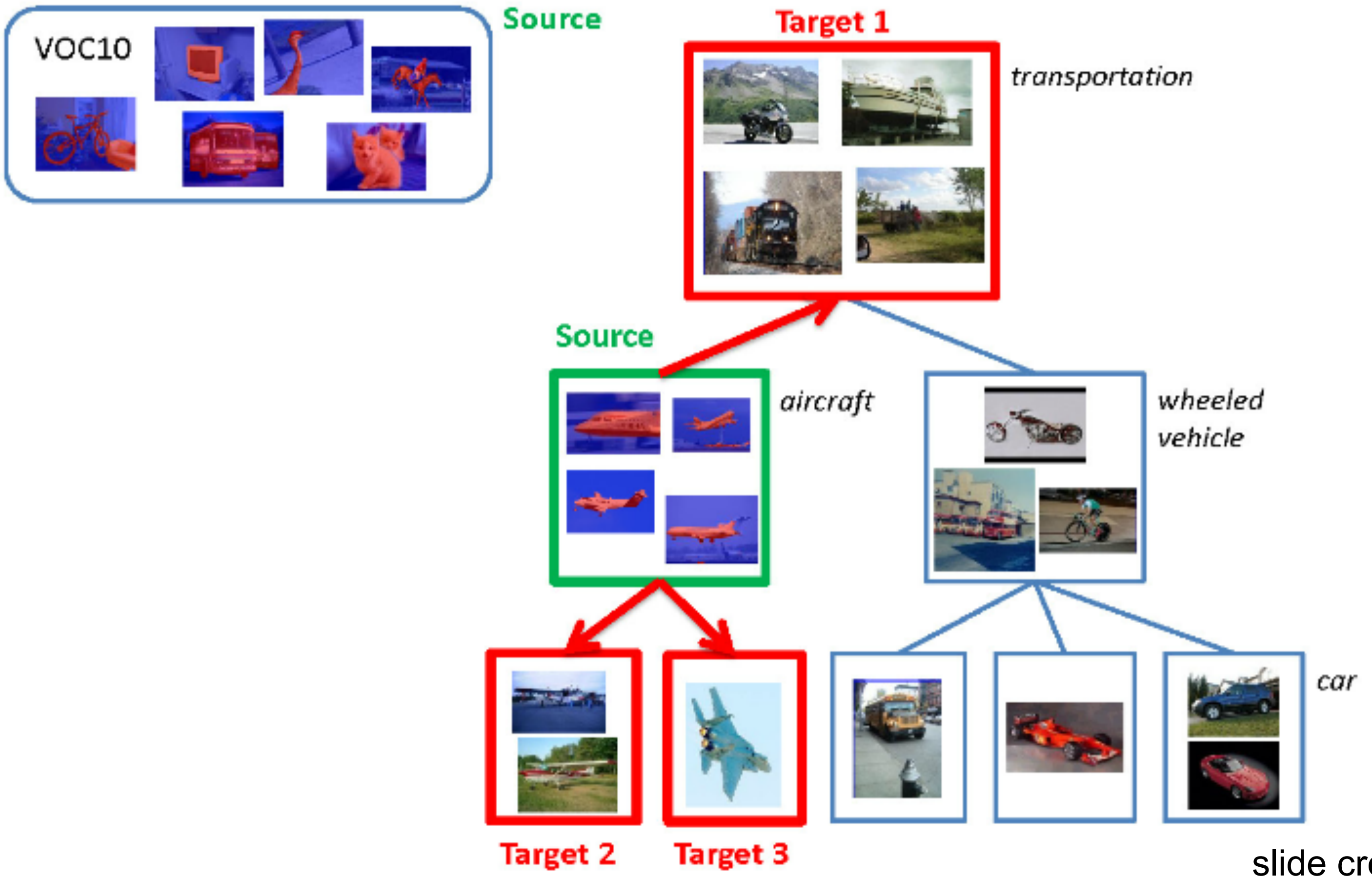

slide credit: V. Ferrari

# Segmentation propagation in a hierarchy



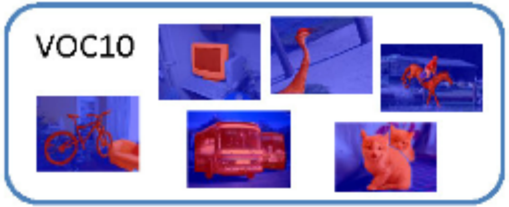

slide credit: V. Ferrari

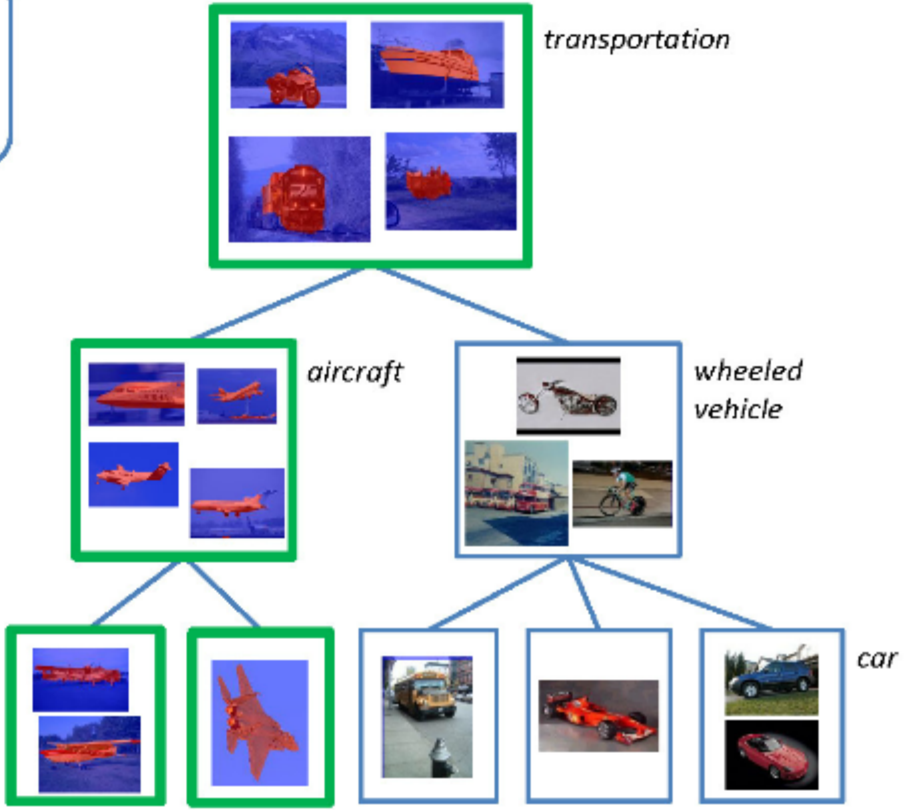
# Segmentation propagation in a hierarchy




slide credit: V. Ferrari

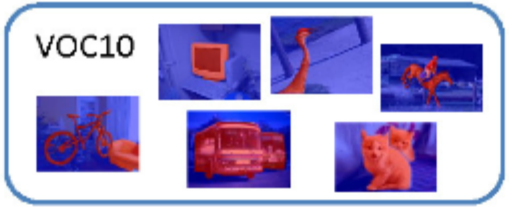
# Segmentation propagation in a hierarchy

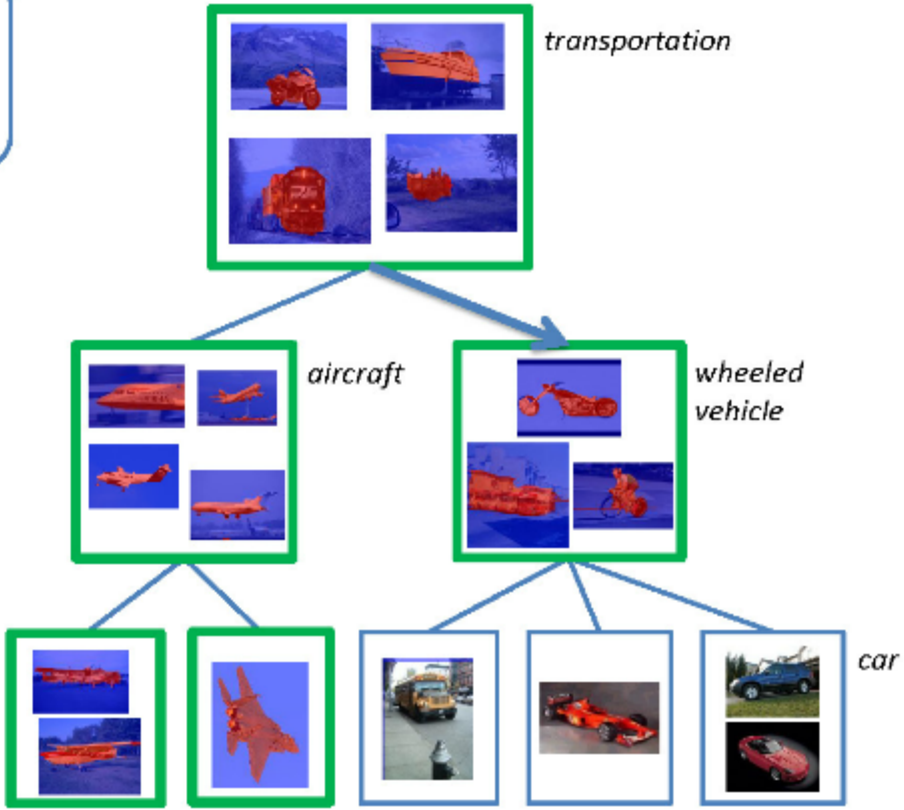



slide credit: V. Ferrari

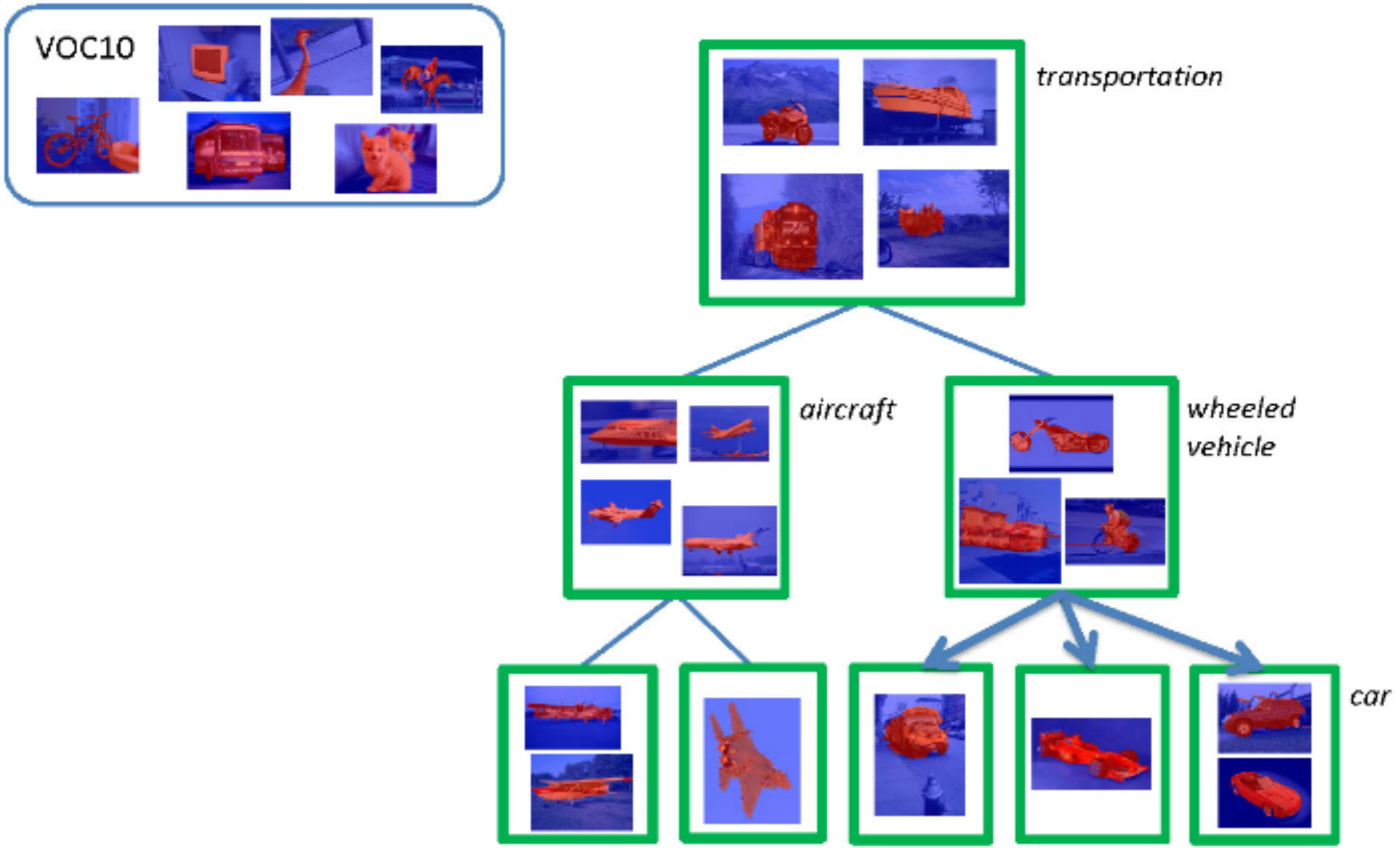
# Segmentation propagation in a hierarchy




slide credit: V. Ferrari

# Segmentation propagation in a hierarchy




slide credit: V. Ferrari

# Outline

Goal & Motivation

System Overview

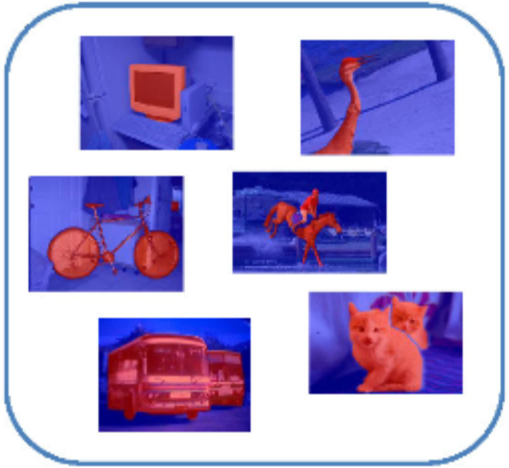
**Segmentation Transfer**

Joint Segmentation

Results

# Segmentation transfer

Source

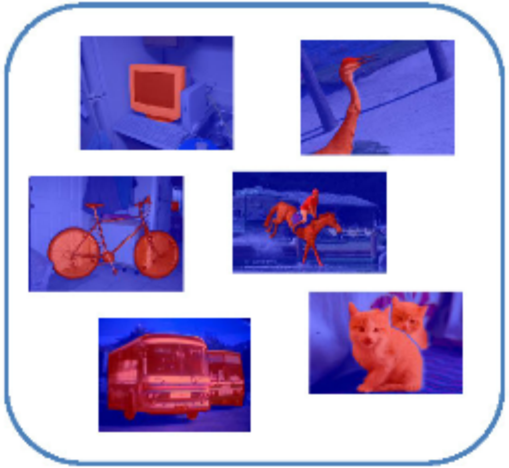


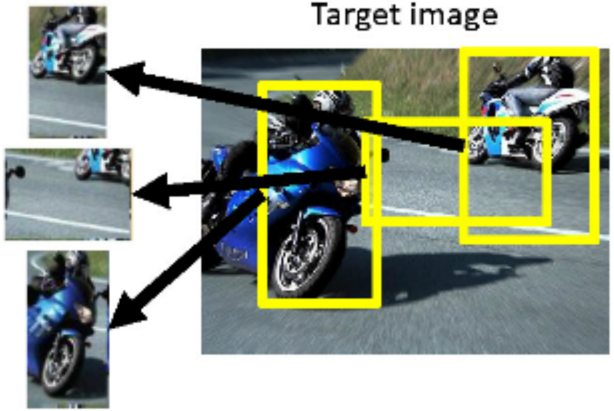
Target image

- Related to earlier annotation transfer works
  [Russel NIPS07, Liu CVPR09, Guillaumin ICCV09,
  Rosenfeld ICCV11, Kuettel CVPR12, Rubinstein ECCV12]

slide credit: V. Ferrari

# Segmentation transfer



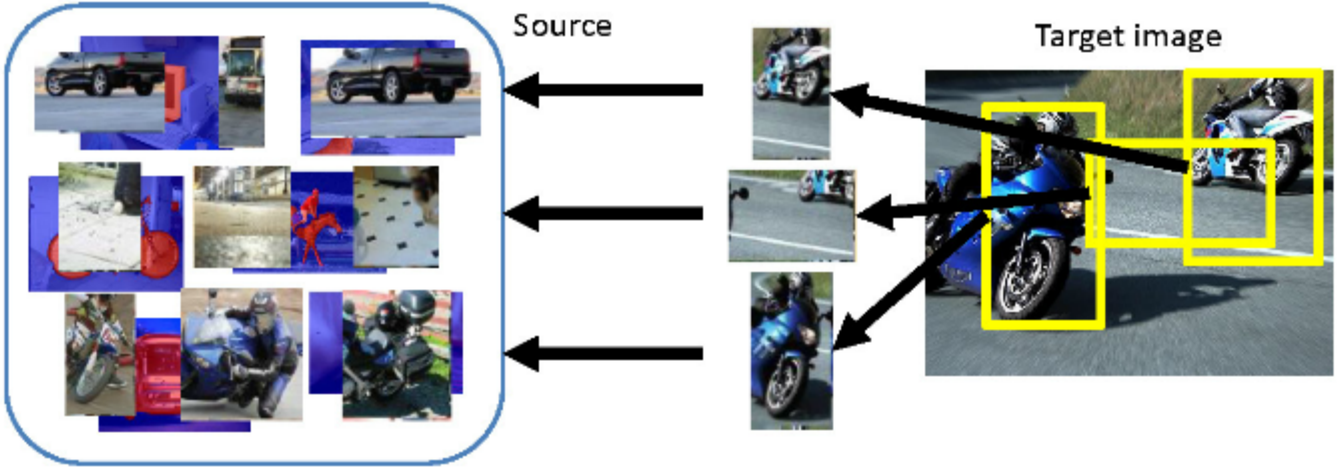

1. Sample windows on objects

Objectness sampling

[Alexe CVPR10]

slide credit: V. Ferrari

# Segmentation transfer



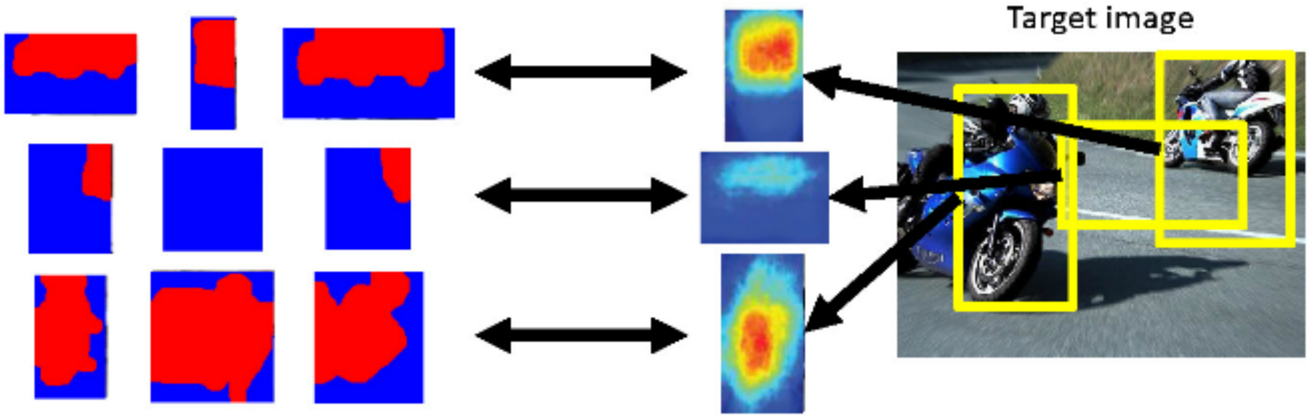

1. Sample windows on objects
2. Find visually similar windows

HOG + compact binary code for efficient retrieval

[Dalal CVPR05, Torralba CVPR08, Gong CVPR11]

slide credit: V. Ferrari

# Segmentation transfer

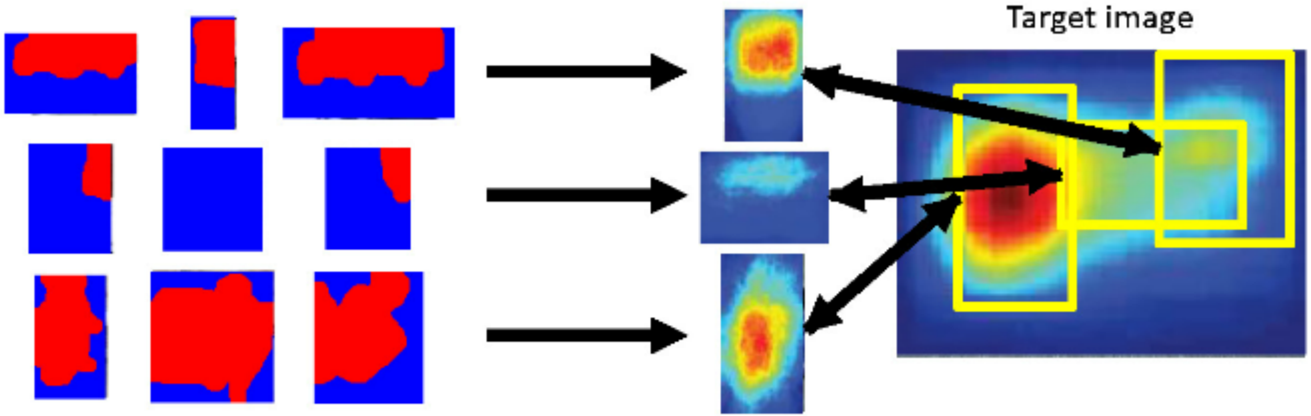



1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations

slide credit: V. Ferrari

[Kuettel CVPR12]

# Segmentation transfer

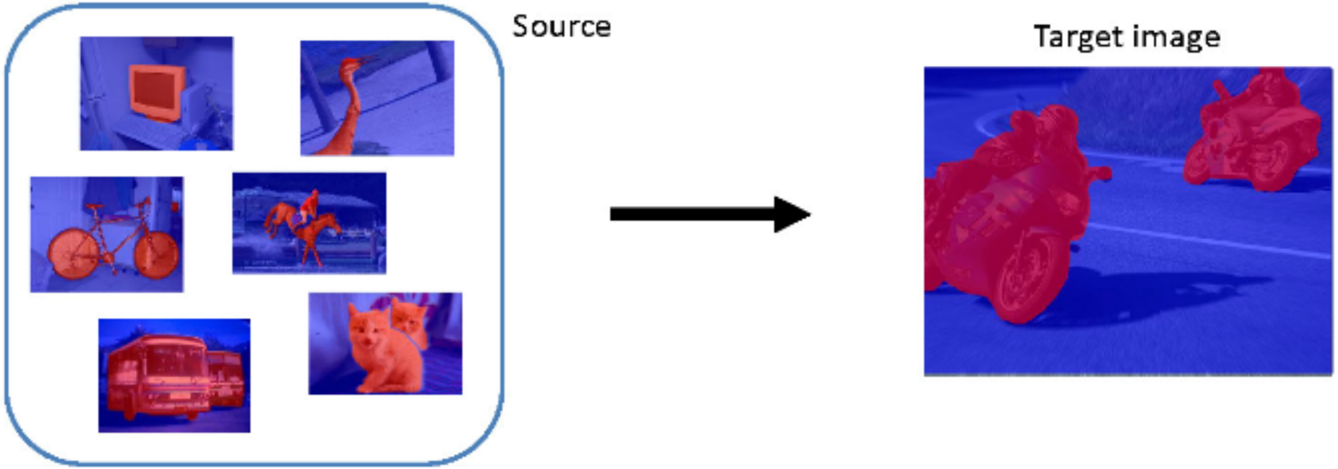



1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations

slide credit: V. Ferrari

[Kuettel CVPR12]

# Segmentation transfer




1. Sample windows on objects
2. Find visually similar windows
3. Aggregate their segmentations
4. Initialize and run GrabCut

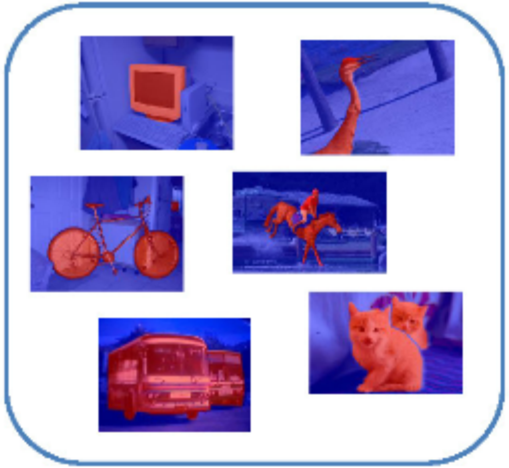
slide credit: V. Ferrari

[Rother SIGGRAPH04]

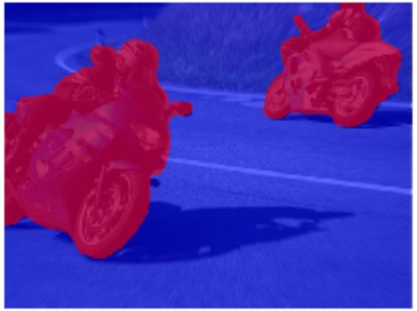
# Digression: Grab Cut

# Segmentation transfer

Source



Target image



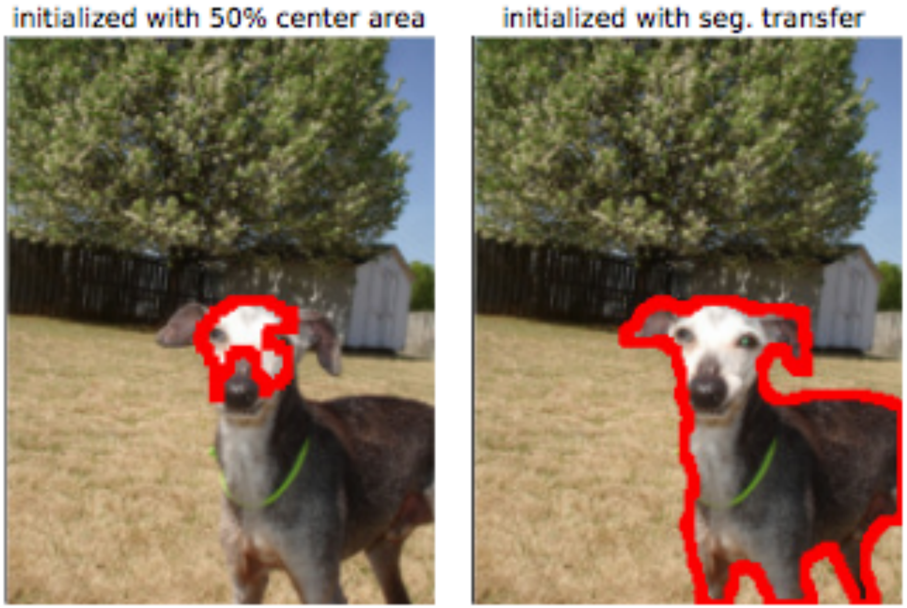
- Window-level: compositionality
- But: increases the number of descriptors in the source
  -> 1M images X 100 windows = 100M descriptors
  -> Compact binary codes [Gong CVPR11]
  200x faster, 500x smaller

Aside: “Deep Hashing for Compact Binary Codes Learning” CVPR 2015 :)

slide credit: V. Ferrari

[Gong CVPR11, Kuettel CVPR12, Liu CVPR09]

# Segmentation Transfer




[3]

# Outline

Goal & Motivation

System Overview

Segmentation Transfer

Joint Segmentation

Results

# Joint Segmentation

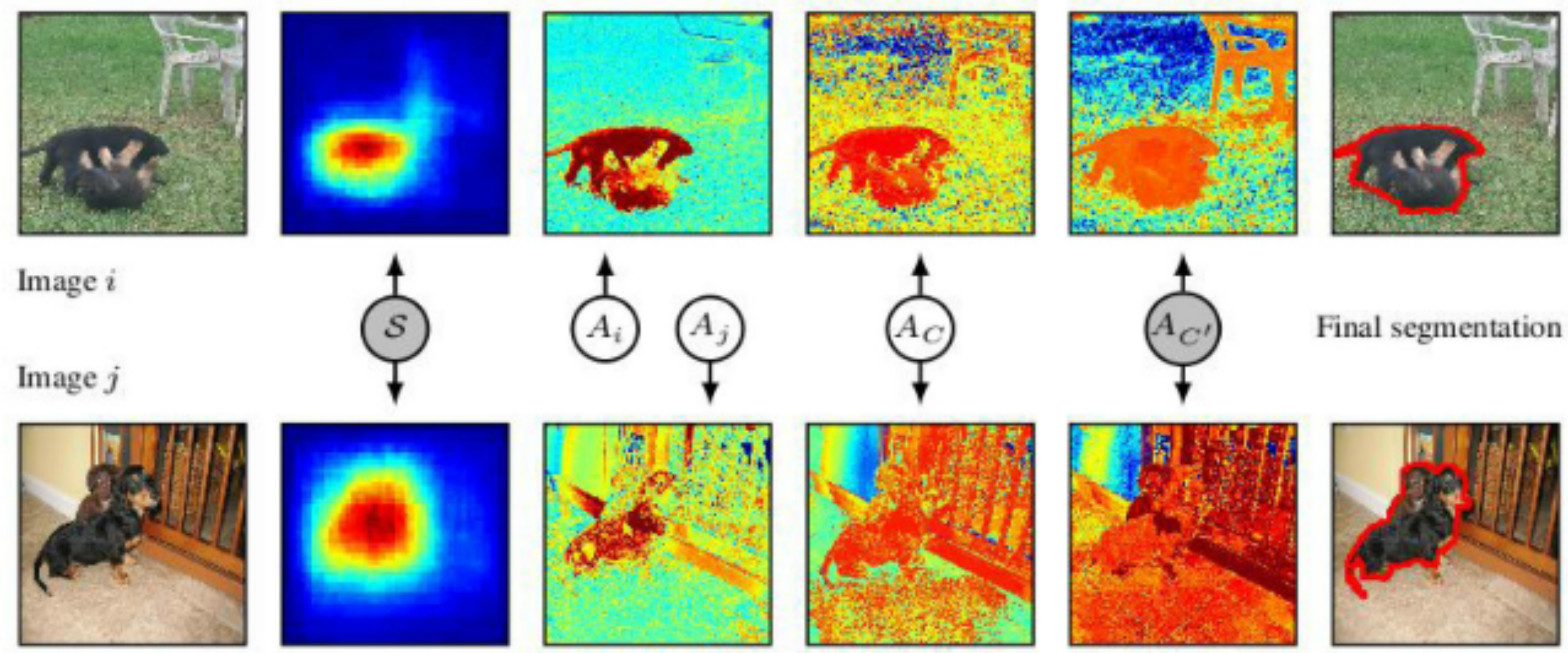
[4] Batra

(a) Group of related images + multiple scribbles (b) Current segmentations (c) Next scribble region recommendation

[5] Rother

Input Image pair GrabCut on individual images Cosegmentation, no spatial consistency Cosegmentation

# Joint Segmentation with Shared Appearance




slide credit: V. Ferrari

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p E_{ip}\left(x_{ip}; A, S\right) + \sum_{p,q} E_{ipq}\left(x_{ip}, x_{iq}\right) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

# Joint Segmentation with Shared Appearance

$$E(\mathbf{x}; A, S) = \sum_i \left( \sum_p E_{ip}\left(x_{ip}; A, S\right) + \sum_{p,q} \boxed{E_{ipq}\left(x_{ip}, x_{iq}\right)} \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ipq}\left(x_{ip}, x_{iq}\right) = \delta\left(x_{ip} \neq x_{iq}\right) \cdot \frac{\exp\left(-\gamma ||c_{ip} - c_{iq}||^2\right)}{d\left(i, p, q\right)}$$

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p \boxed{E_{ip}\left(x_{ip}; A, S\right)} + \sum_{p,q} E_{ipq}\left(x_{ip}, x_{iq}\right) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}\left(x_{ip}; A, S\right) = -\alpha_I \log p\left(x_{ip}; c_{ip}, A_i\right) - \alpha_C \log p\left(x_{ip}; c_{ip}, A_C\right) - \alpha_M \log M_{ip}\left(x_{ip}; S\right)$$

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p \boxed{E_{ip}(x_{ip}; A, S)} + \sum_{p,q} E_{ipq}(x_{ip}, x_{iq}) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}(x_{ip}; A, S) = \boxed{-\alpha_I \log p(x_{ip}; c_{ip}, A_i)} - \alpha_C \log p(x_{ip}; c_{ip}, A_C) - \alpha_M \log M_{ip}(x_{ip}; S)$$

1. Appearance model for image i.

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p \boxed{E_{ip}(x_{ip}; A, S)} + \sum_{p,q} E_{ipq}(x_{ip}, x_{iq}) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}(x_{ip}; A, S) = -\alpha_I \log p(x_{ip}; c_{ip}, A_i) - \boxed{\alpha_C \log p(x_{ip}; c_{ip}, A_C)} - \alpha_M \log M_{ip}(x_{ip}; S)$$

1. Appearance model for image i.
2. Appearance model for class C

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p \boxed{E_{ip}\left(x_{ip}; A, S\right)} + \sum_{p,q} E_{ipq}\left(x_{ip}, x_{iq}\right) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}\left(x_{ip}; A, S\right) = -\alpha_I \log p\left(x_{ip}; c_{ip}, A_i\right) - \alpha_C \log p\left(x_{ip}; c_{ip}, A_C\right) - \boxed{\alpha_M \log M_{ip}\left(x_{ip}; S\right)}$$

1. Appearance model for image i.
2. Appearance model for class C
3. Transferred mask from source S to image i

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p \boxed{E_{ip}\left(x_{ip}; A, S\right)} + \sum_{p,q} E_{ipq}\left(x_{ip}, x_{iq}\right) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}\left(x_{ip}; A, S\right) = -\alpha_I \log p\left(x_{ip}; c_{ip}, A_i\right) - \alpha_C \log p\left(x_{ip}; c_{ip}, A_C\right) - \boxed{\alpha_M \log M_{ip}\left(x_{ip}; S\right)}$$

$$\boxed{M_{ip}(x_{ip}; S) = M_{ip}^{x_{ip}}\left(1 - M_{ip}\right)^{1 - x_{ip}}}$$

3. Transferred mask from source S to image i

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p E_{ip}(x_{ip}; A, S) + \sum_{p,q} E_{ipq}(x_{ip}, x_{iq}) \right)$$

set of image $i \in I$ of a class C in ImageNet
$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}(x_{ip}; A, S) = -\alpha_I \log p(x_{ip}; c_{ip}, A_i) - \alpha_C \log p(x_{ip}; c_{ip}, A_C) - \alpha_M \log M_{ip}(x_{ip}; S)$$

1. Appearance model for image i.
2. Appearance model for class C
3. Transferred mask from source S to image i

# Joint Segmentation with Shared Appearance

$$E(\mathrm{x}; A, S) = \sum_i \left( \sum_p E_{ip}(x_{ip}; A, R(C)) + \sum_{p,q} E_{ipq}(x_{ip}, x_{iq}) \right)$$

set of image $i \in I$ of a class C in ImageNet

$x_{ip}$ label for pixel $p$ in image $i$

$$E_{ip}(x_{ip}; A, R(C)) = -\alpha_I \log p(x_{ip}; c_{ip}, A_i) - \alpha_C \log p(x_{ip}; c_{ip}, A_C)$$

$$-\alpha_M \log M_{ip}(x_{ip}; R(C)) - \frac{\alpha_R}{|R(C)|} \sum_{C' \in R(C)} \log p(x_{ip}; c_{ip}, A_{C'})$$

4. Appearance model for related classes related classes $C' \in R(C)$

# Outline

Goal & Motivation

System Overview

Segmentation Transfer

Joint Segmentation

**Results**

# Experiments on iCoseg

- 38 classes, 643 images. No propagation, no related classes

| | Joulin 2010 | Vicente 2011 | Image only (~Grabcut) | Class only (~Batra) | Image +class | Image +transfer (~Kuettel) | Image +transfer +class |
|---|---|---|---|---|---|---|---|
| Accuracy | 78.9% | 85.4% | 82.4% | 83.6% | 88.2% | 87.6% | **91.4%** |

- Within-class appearance sharing helps
- Segmentation transfer helps
- Outperforms Joulin10, Vicente11 and much faster
  (using unrelated segmented images from PASCAL VOC10)

slide credit: V. Ferrari

[Rother SIGGRAPH04, Joulin CVPR10, Vicente CVPR11, Batra IJCV11, Kuettel CVPR12]

# Experiments on ImageNet

animal, instruments subtrees

60k bounding boxes

440k only class labels

4k manually annotated over 450 classes

# Experiments on ImageNet


slide credit: V. Ferrari

# Experiments on ImageNet

Sail boat

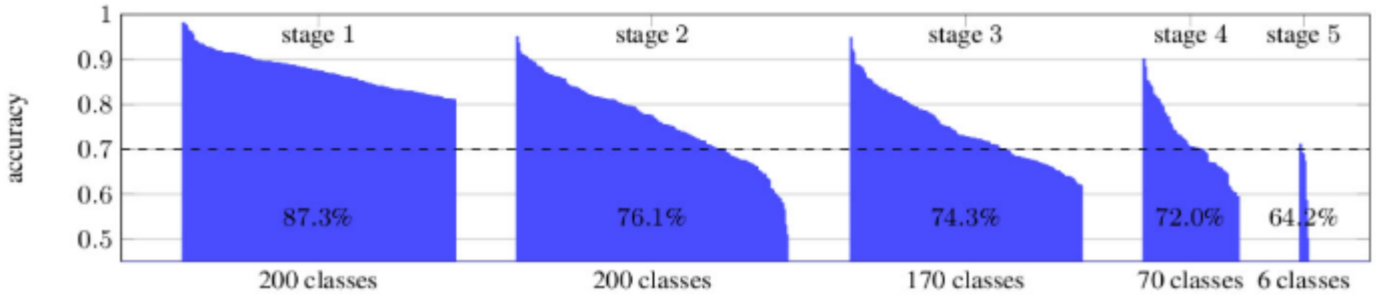
Megaphone

Monitor

slide credit: V. Ferrari

# Experiments on ImageNet

- Overall experiment: 0.5M images, 577 classes
- Full propagation scheme
- 10 images X 446 classes annotated with AMT
- 77.1% accuracy vs 71.0% for GrabCut




- Performance degrades gracefully over stages

slide credit: V. Ferrari

# Conclusion

automatic

large-scale

exploits class structure

extends segmentation datasets

# References

[1] A. Rosenfeld and D. Weinshall. Extracting Foreground Masks towards Object Recognition. In *Proceedings IEEE International Conference on Computer Vision*, 2011.
[2] D. Kuettel and V. Ferrari. Figure-ground segmentation by transferring window masks. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. 2012. p. 558-565.
[3] M. Guillamin, D. Kuettel, V. Ferrari. ImageNet Auto-Annotation with Segmentation Propagation. International Journal of Computer Vision. 2014.
[4] Batra, D.; Kowdle, A.; Parikh, D.; Jiebo Luo; Tsuhan Chen, "iCoseg: Interactive co-segmentation with intelligent scribble guidance," *Computer Vision and Pattern Recognition (CVPR), 2010*
[5] Rother, C.; Minka, T.; Blake, A.; Kolmogorov, V., "Cosegmentation of Image Pairs by Histogram Matching - Incorporating a Global Constraint into MRFs," *Computer Vision and Pattern Recognition, 2006*