Supervised Learning
Weak-supervised Learning
Unsupervised Learning

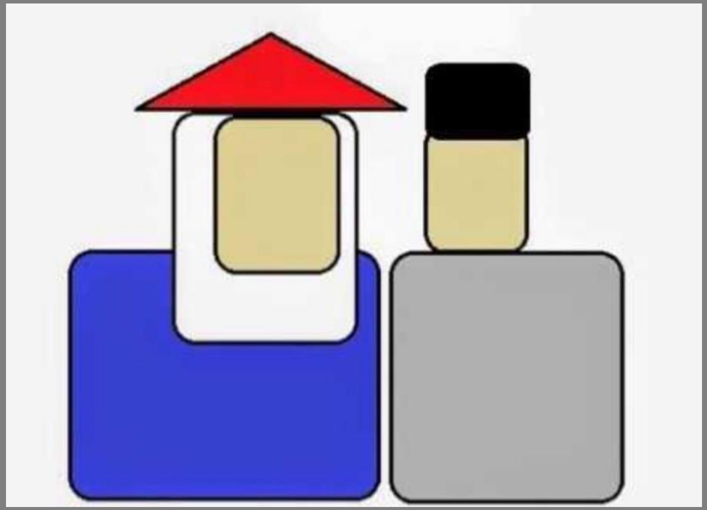
Via two papers

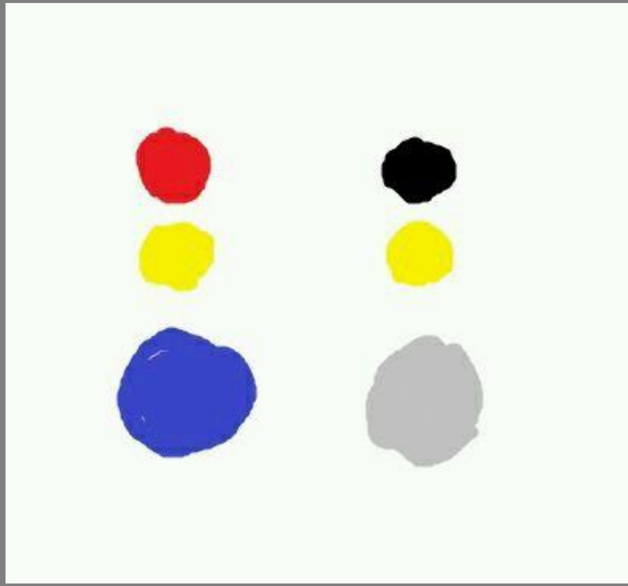
Zhedong Zheng

What can we learn from

(from Stephen Chow’s film)


我全都要





Related Works

1. Simple Does It: Weakly Supervised Instance and Semantic Segmentation **(CVPR 2017)** **Weak**

2. Colorful Image Colorization **(ECCV 2016 oral)** **Self**

Related Works

1. Simple Does It: Weakly Supervised Instance and Semantic Segmentation **(CVPR 2017)** **Weak**

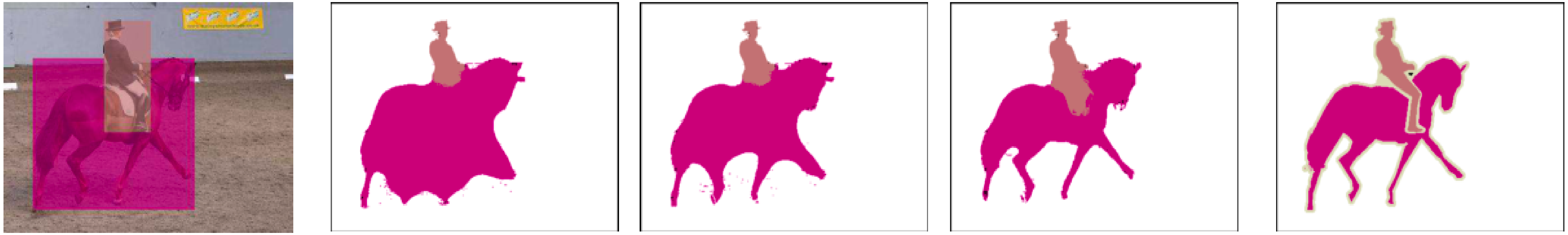
2. Colorful Image Colorization **(ECCV 2016 oral)** **Self**

What

Training sample, with box annotations | Test image, fully supervised result | Test image, weakly supervised result

Figure 1: We propose a technique to train semantic labelling from bounding boxes, and reach 95% of the quality obtained when training from pixel-wise annotations.

How



Example input rectangles | Output after 1 training round | After 5 rounds | After 10 rounds | Ground truth

Start from object bounding box annotations

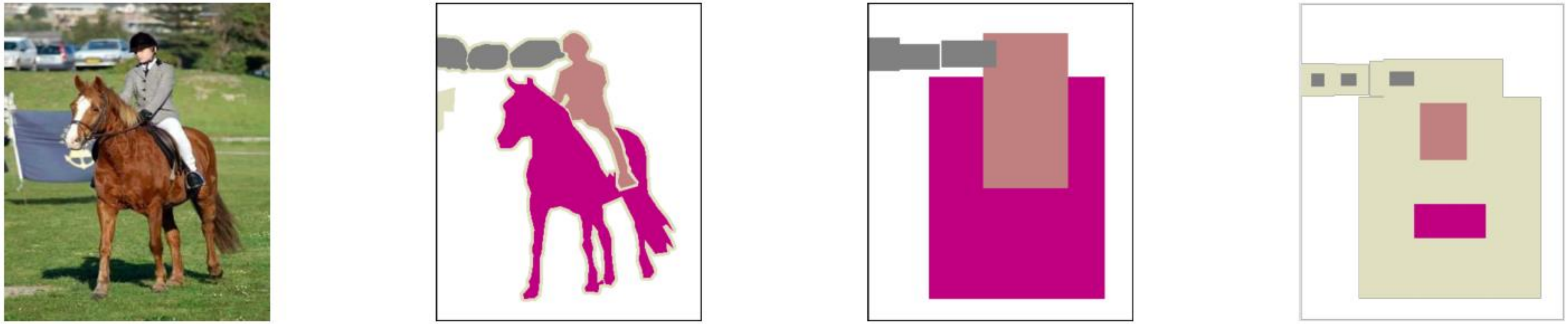
Recall Several Rules

1. **Background :** No bounding box -> background

2. **Object Extent :** Bboxes are instance-level, provide information

3. **Objectness :** Spatial Continuity / Contrasting boundary

How to begin?



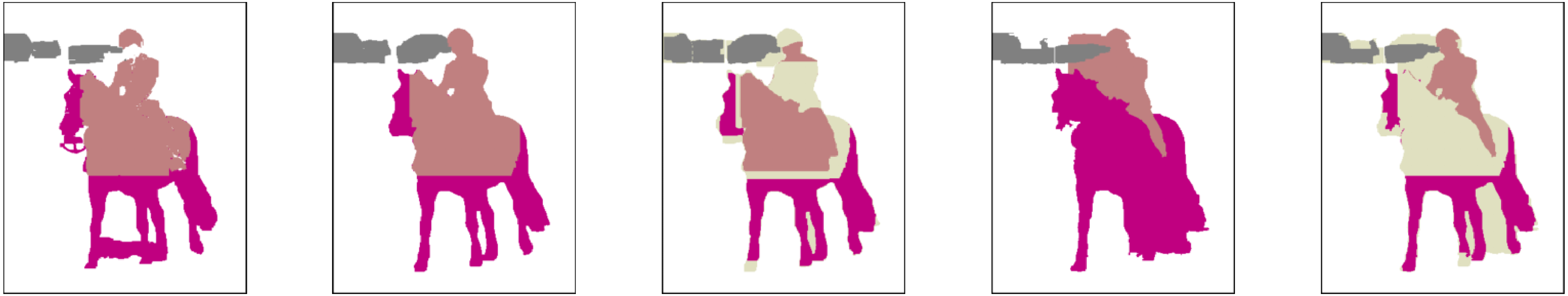
(a) Input image (b) Ground truth (c) Box (d) $\mathtt{Box}^{\mathtt{i}}$

If two boxes overlap, we assume **the smaller one is in front**.

How to begin?

Building upon the insights from the baselines in Section 3.1 and 3.2, we use the MCG segment proposals to supplement GrabCut+. Inside the annotated boxes, we mark as foreground pixels where both MCG and GrabCut+ agree; the remaining ones are marked as ignore. We denote this approach as MCG ∩ GrabCut+ or M ∩ G+ for short.

Because MCG and GrabCut+provide complementary information, we can think of M ∩ G+ as an improved version of GrabCut+[i] providing a different trade-off between precision and recall on the generated labels (see Figure 3i).



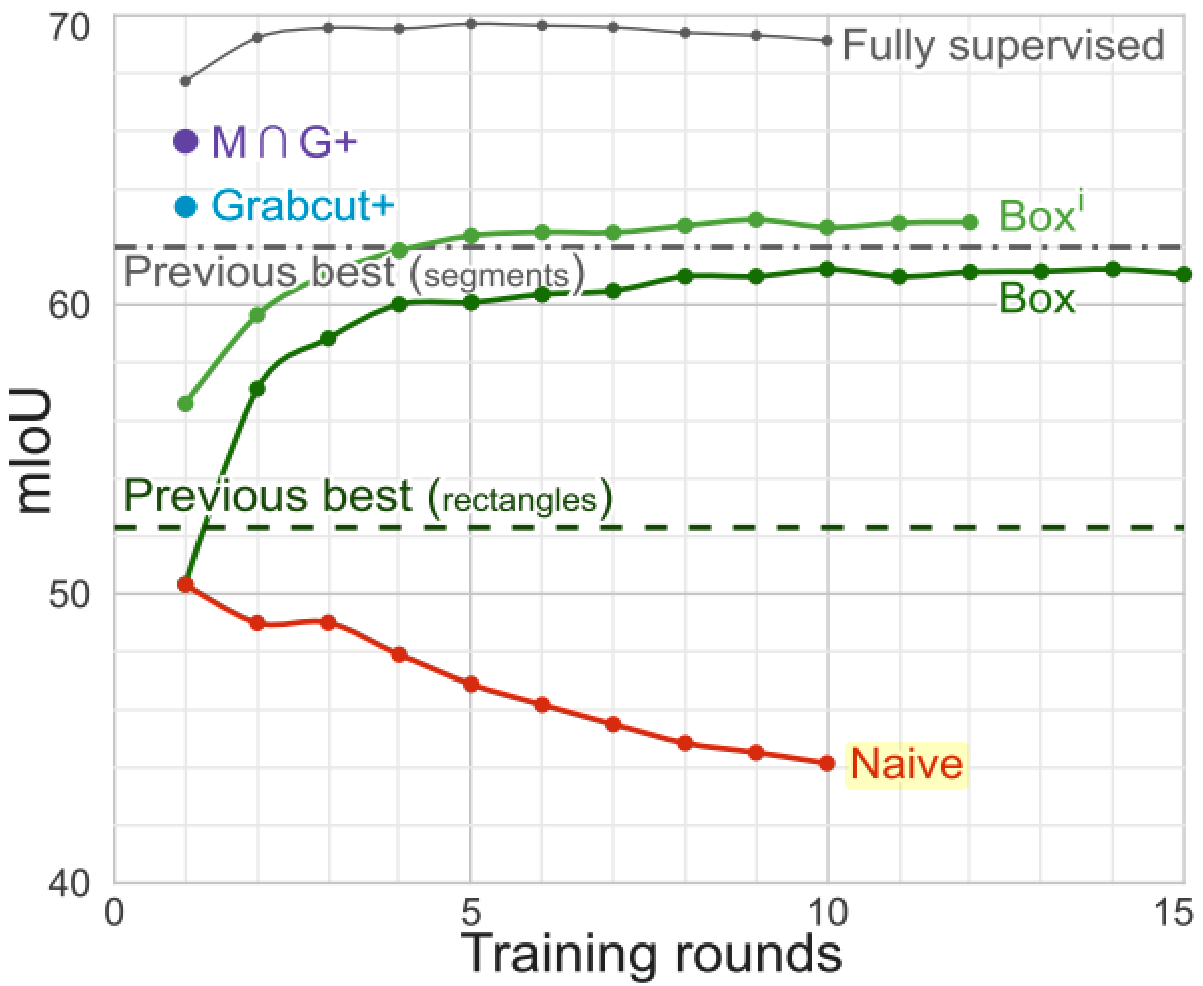
(e) GrabCut (f) GrabCut+ (g) GrabCut+[i] (h) MCG (i) M ∩ G+

Post-Process

- Any pixel outside bbox is discard.
- If IoU<50%, re-inital
- DenseCRF

2. If the area of a segment is too small compared to its corresponding bounding box (e.g. IoU$<$ 50%), the box area is reset to its initial label (fed in the first round). This enforces a minimal area (cue C2).

Result



Naïve is without post-processing.

Result

Method		val. mIoU
-	Fast-RCNN	44.3
	GT Boxes	62.2
Weakly supervised	Box	61.2
	$\texttt{Box}^{\texttt{i}}$	62.7
	MCG	62.6
	GrabCut+	63.4
	$\texttt{GrabCut+}^{\texttt{i}}$	64.3
	$\texttt{M} \cap \texttt{G+}$	**65.7**
Fully supervised	DeepLab_{ours} [5]	69.1

Table 1: Weakly supervised semantic labelling results for our baselines. Trained using Pascal VOC12 bounding boxes alone, validation set results. DeepLab_{ours} indicates our fully supervised result.

Result

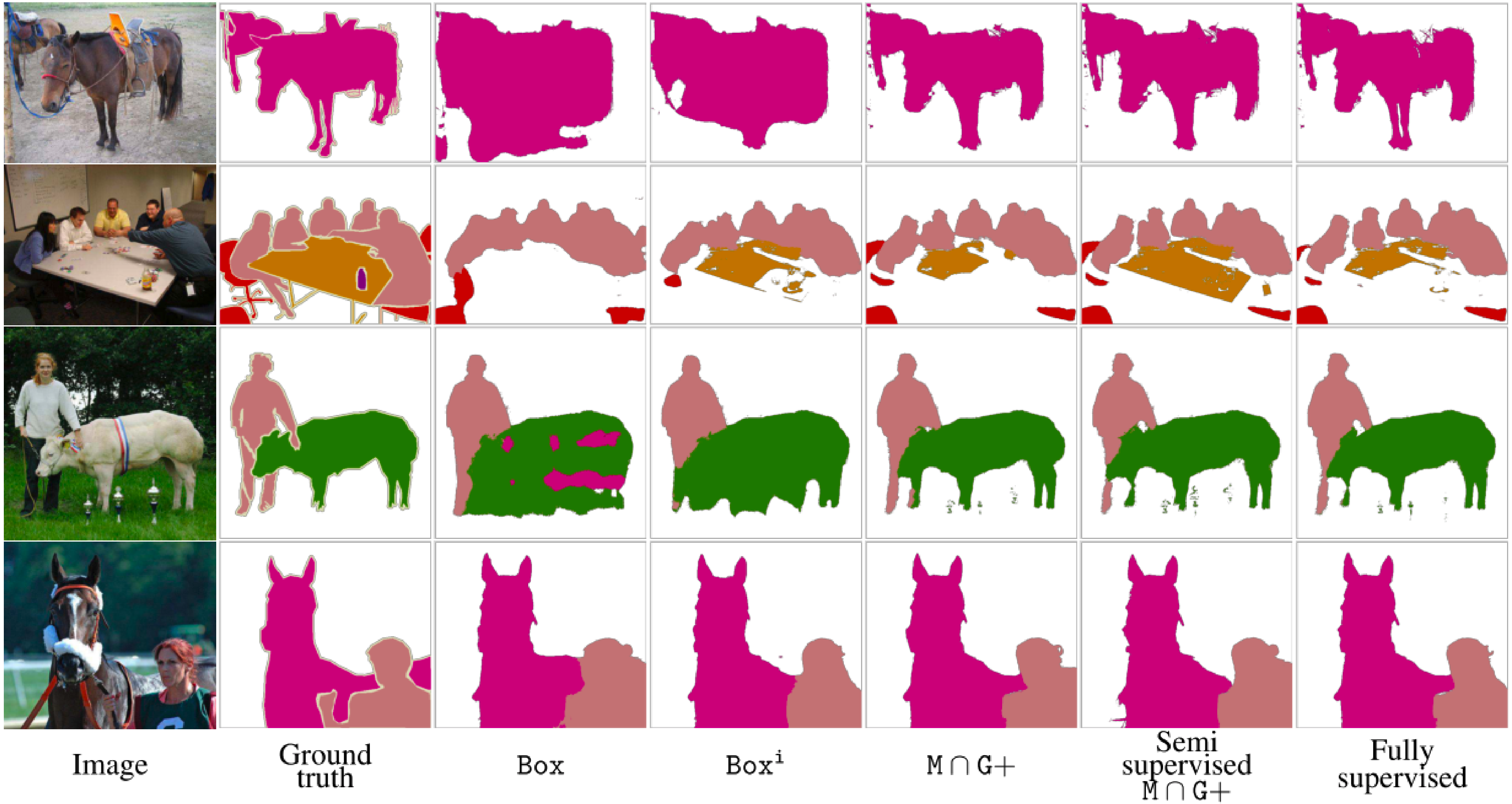
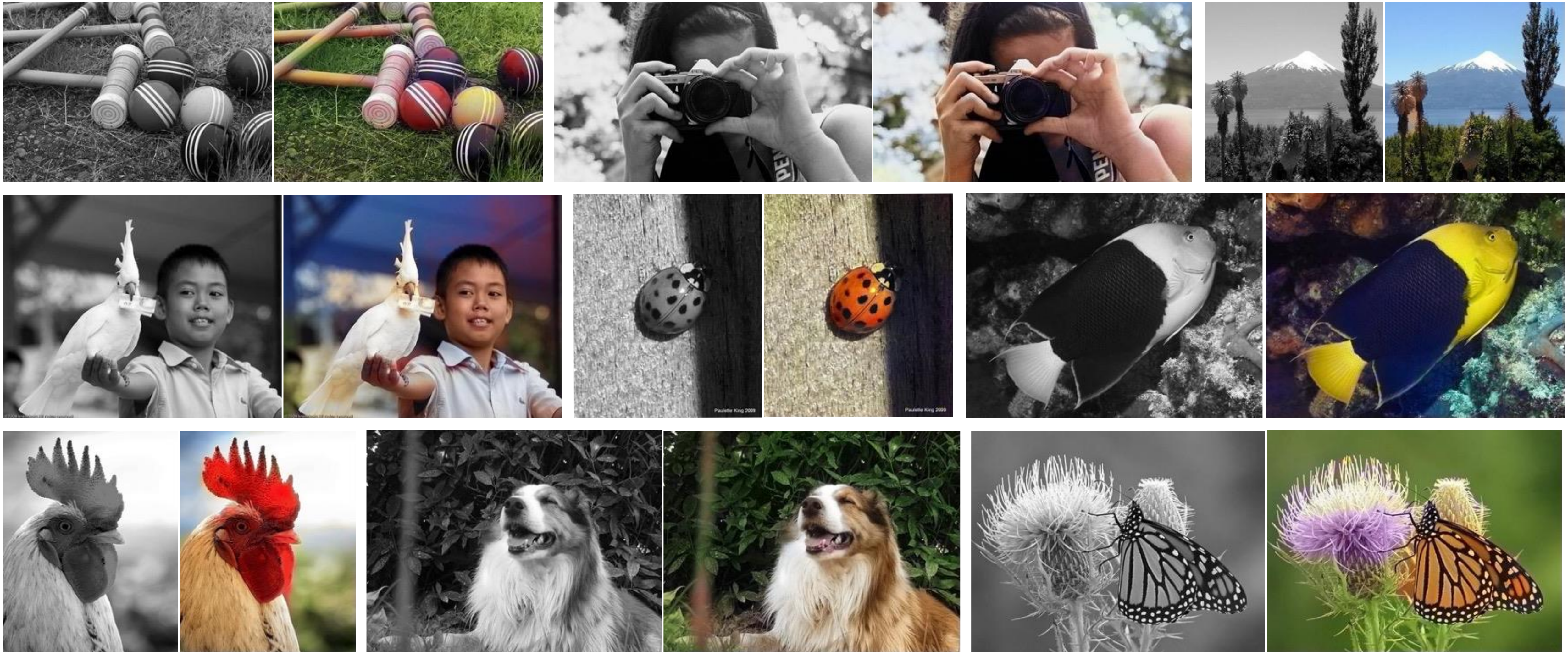


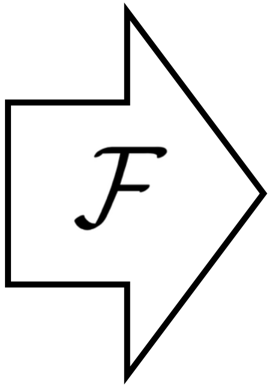
Image
Ground truth
Box
Box^i
$M \cap G+$
Semi supervised $M \cap G+$
Fully supervised

Related Works

1. Simple Does It: Weakly Supervised Instance and Semantic Segmentation **(CVPR 2017)** **Weak**

2. Colorful Image Colorization **(ECCV 2016 oral)** **Self**



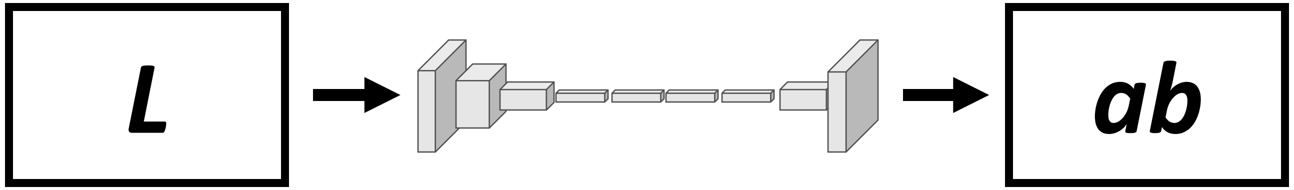


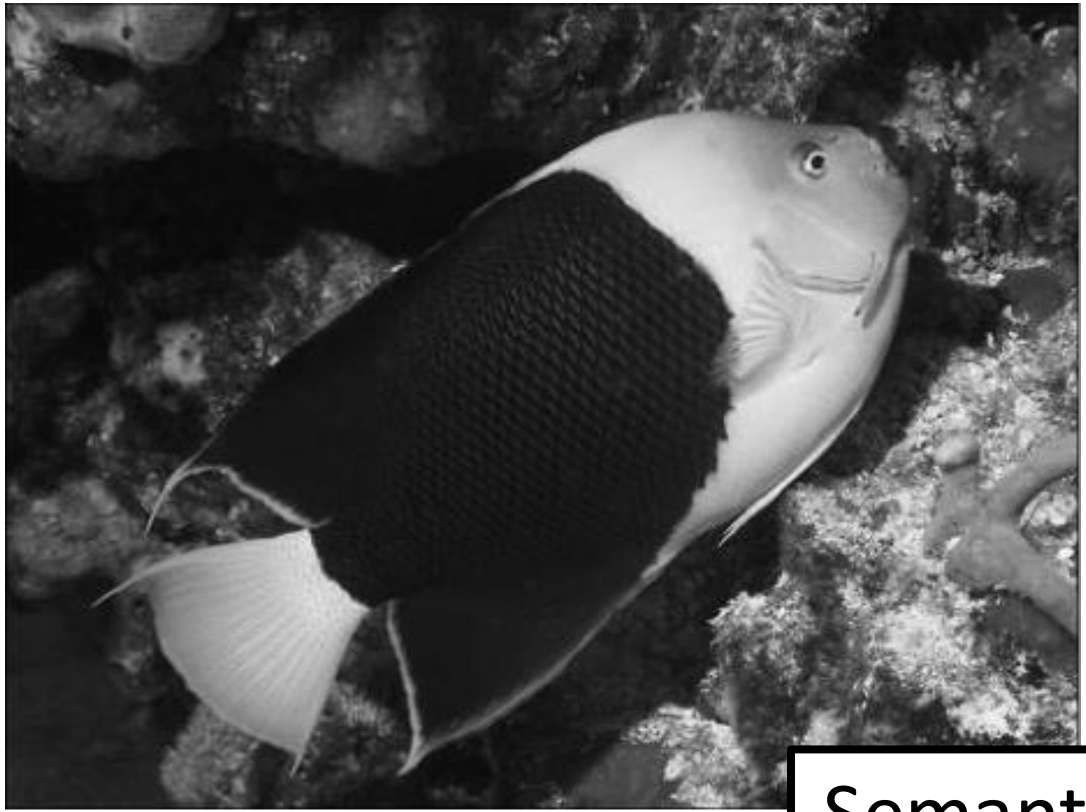
Grayscale image: *L* channel

$$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$$

Color information: *ab* channels

$$\widehat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times 2}$$





$\mathcal{F}$

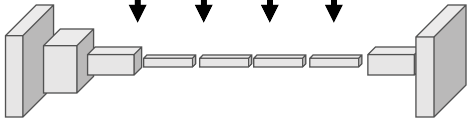
Semantics? Higher-level abstraction?

Grayscale image: *L* ch

$\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$

ncatenate (L,ab)

$(\mathbf{X}, \widehat{\mathbf{Y}})$

L →  → ***ab*** ← "Free" supervisory signal

Inherent Ambiguity

Grayscale

Inherent Ambiguity

Our Output

Ground Truth

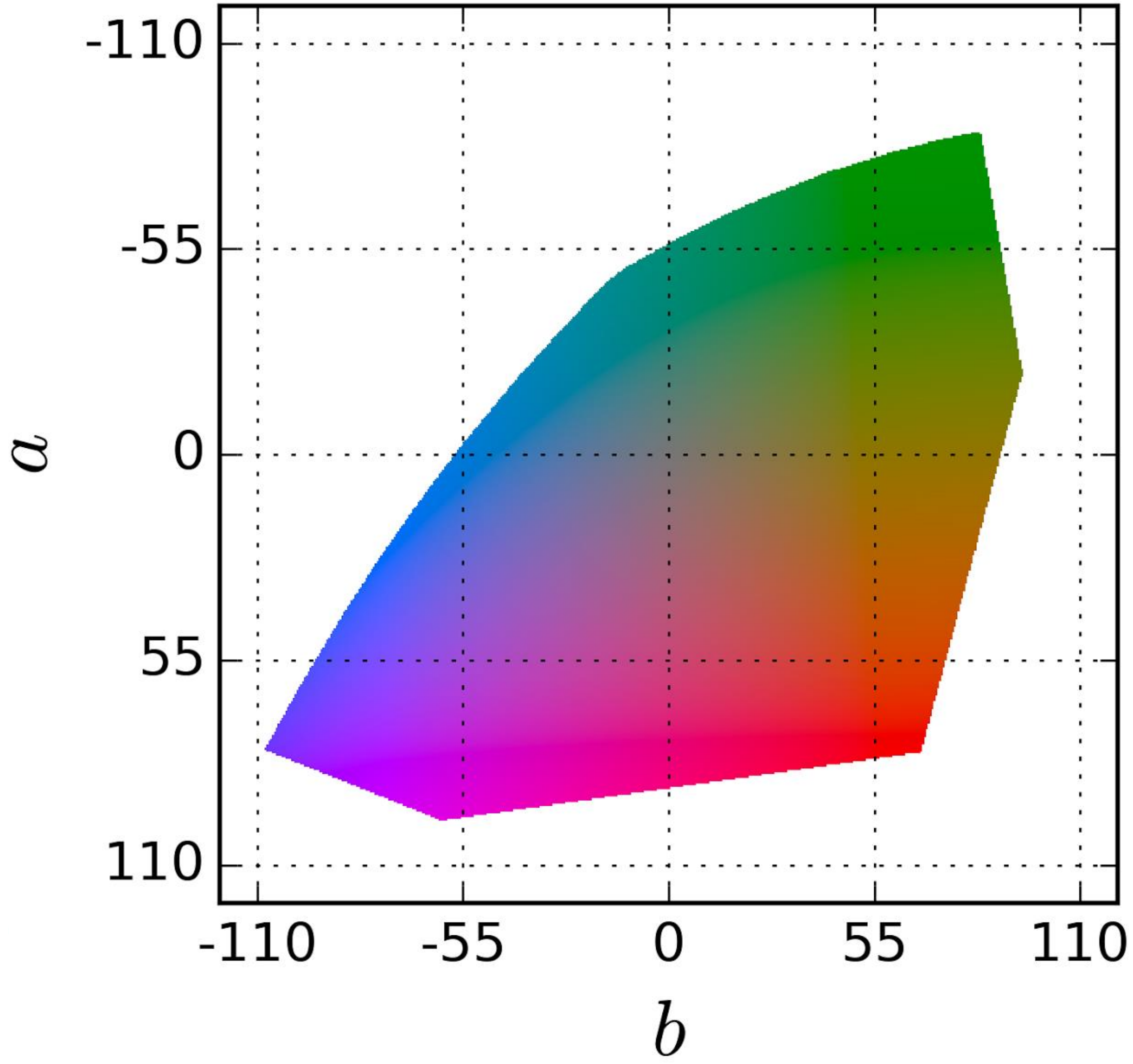
Better Loss Function

- Regression with L2 loss inadequate

$$L_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

Colors in *ab* space

(continuous)



Better Loss Function

- Regression with L2 loss inadequate

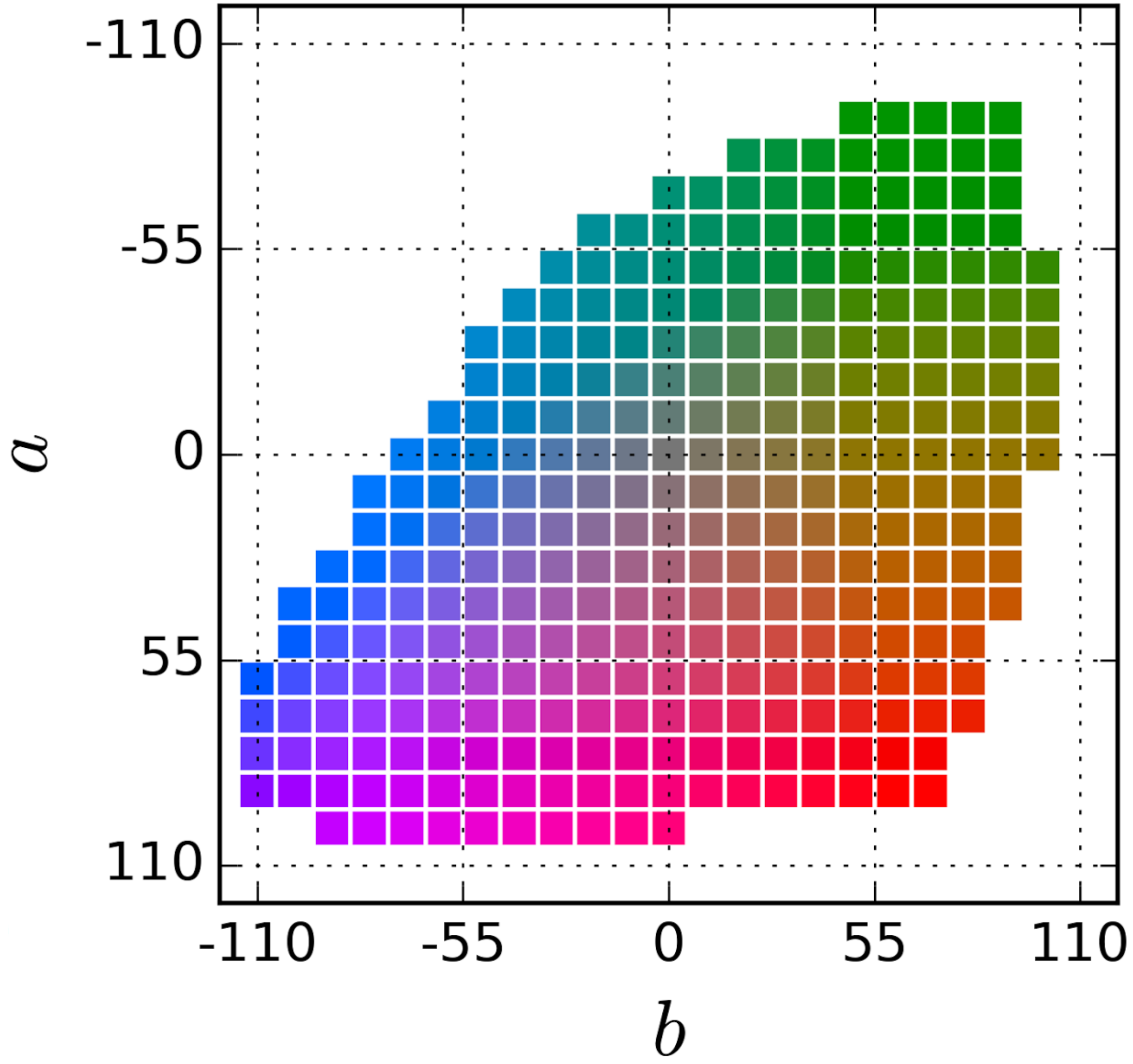
$$\mathrm{L}_2(\widehat{\mathbf{Y}}, \mathbf{Y}) = \frac{1}{2} \sum_{h,w} \|\mathbf{Y}_{h,w} - \widehat{\mathbf{Y}}_{h,w}\|_2^2$$

- Use **multinomial classification**

$$\mathrm{L}(\widehat{\mathbf{Z}}, \mathbf{Z}) = -\frac{1}{HW} \sum_{h,w} \sum_{q} \mathbf{Z}_{h,w,q} \log(\widehat{\mathbf{Z}}_{h,w,q})$$

Colors in *ab* space

(discrete)



Failure Cases

Biases

Evaluation

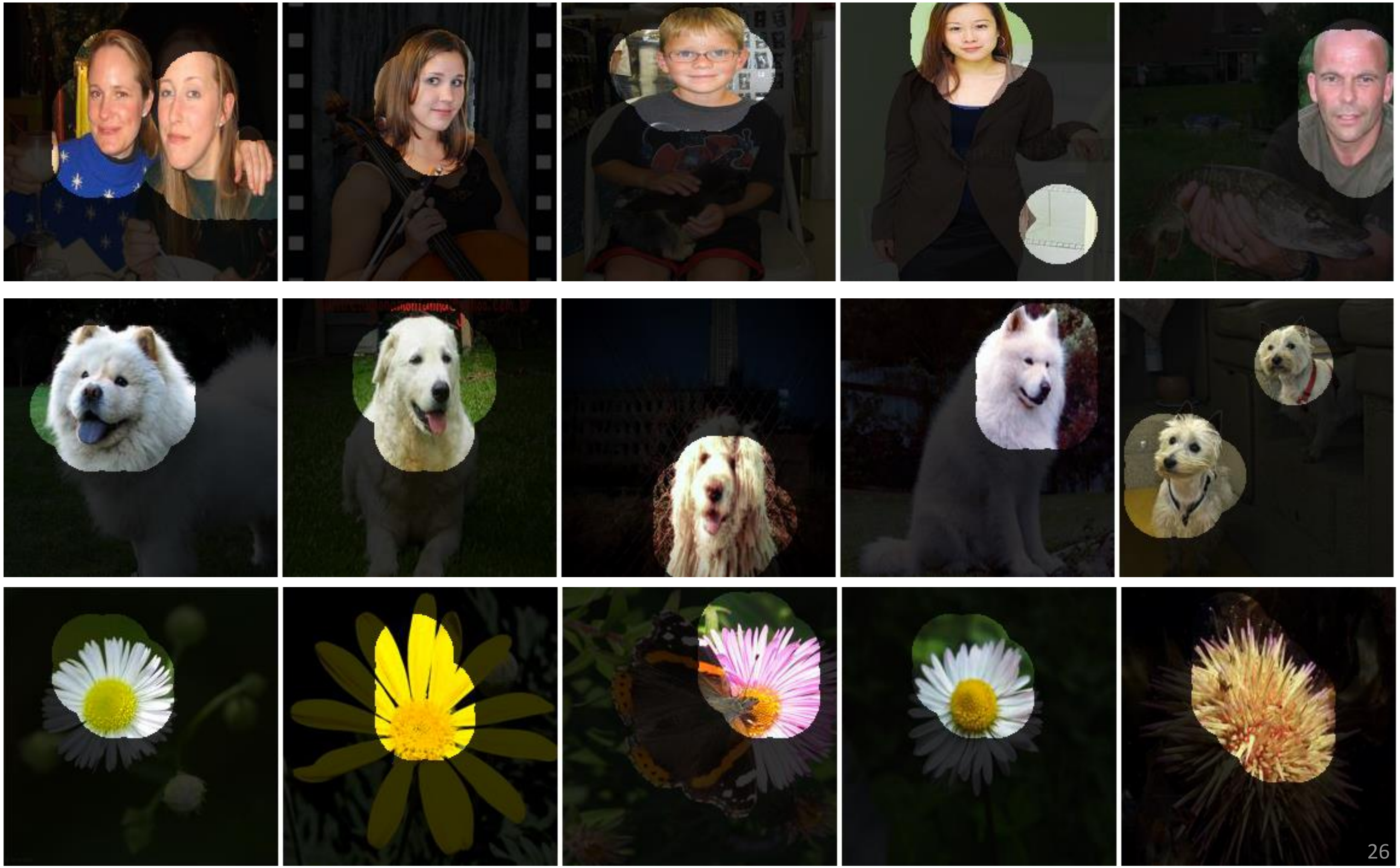
	Visual Quality	Representation Learning
Quantitative	Per-pixel accuracy Perceptual realism Semantic interpretability	Task generalization ImageNet classification Task & dataset generalization PASCAL classification, detection, segmentation
Qualitative	Low-level stimuli Legacy grayscale photos	Hidden unit activations

Hidden Unit (conv5) Activations

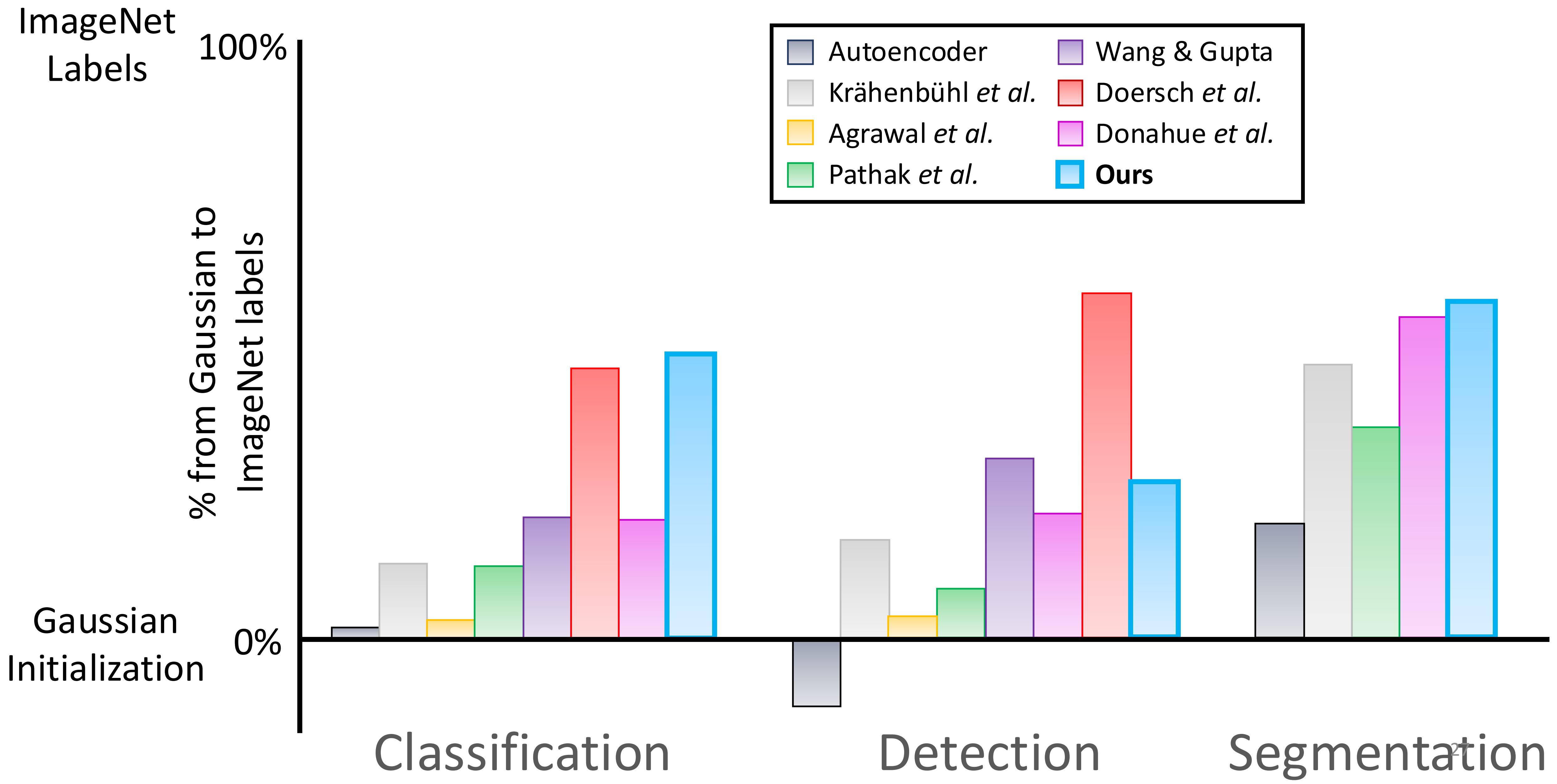
faces

dog faces

flowers

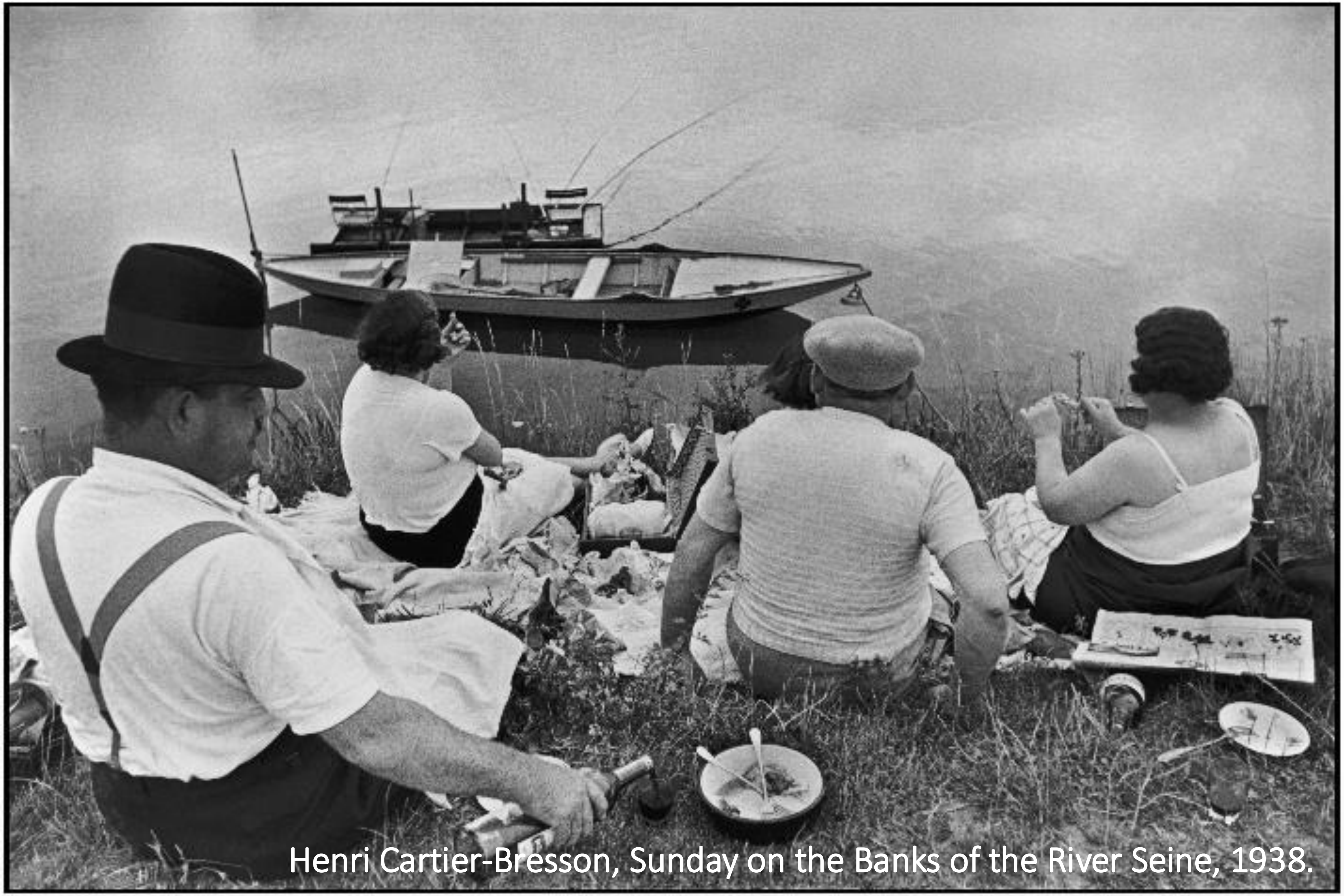


Dataset & Task Generalization on PASCAL VOC



Amateur Family Photo, 1956.

Amateur Family Photo, 1956.



Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.

Henri Cartier-Bresson, Sunday on the Banks of the River Seine, 1938.