Geometry Constrained Weakly Supervised Object Localization

Weizeng Lu[1], Xi Jia[2], Weicheng Xie[1], Linlin Shen[1*], Yicong Zhou[3], Jinming Duan[2]

[1]Computer Vision Institute, School of Computer Science and Software Engineering,Shenzhen University, Shenzhen, China

[2]University of Birmingham, United Kingdom

[3]University of Macau, Macao, China

Abstract

We propose a geometry constrained network, termed GC-Net, for weakly supervised object localization. GC-Net consists of three modules: a detector, a generator and a classifier. The detector predicts the object location represented by a set of coefficients, which is constrained by the mask produced by the generator. The classifier takes the resulting masked images as input and performs two complementary classification tasks (object and background). To make the mask more accurate, we propose a novel multi-task loss function that takes into account area of the mask, the categorical cross entropy and the negative entropy. Extensive experiments on the CUB-200-2011 and ILSVRC2012 datasets show that GC-Net outperforms state-of-the-art methods.

Motivation

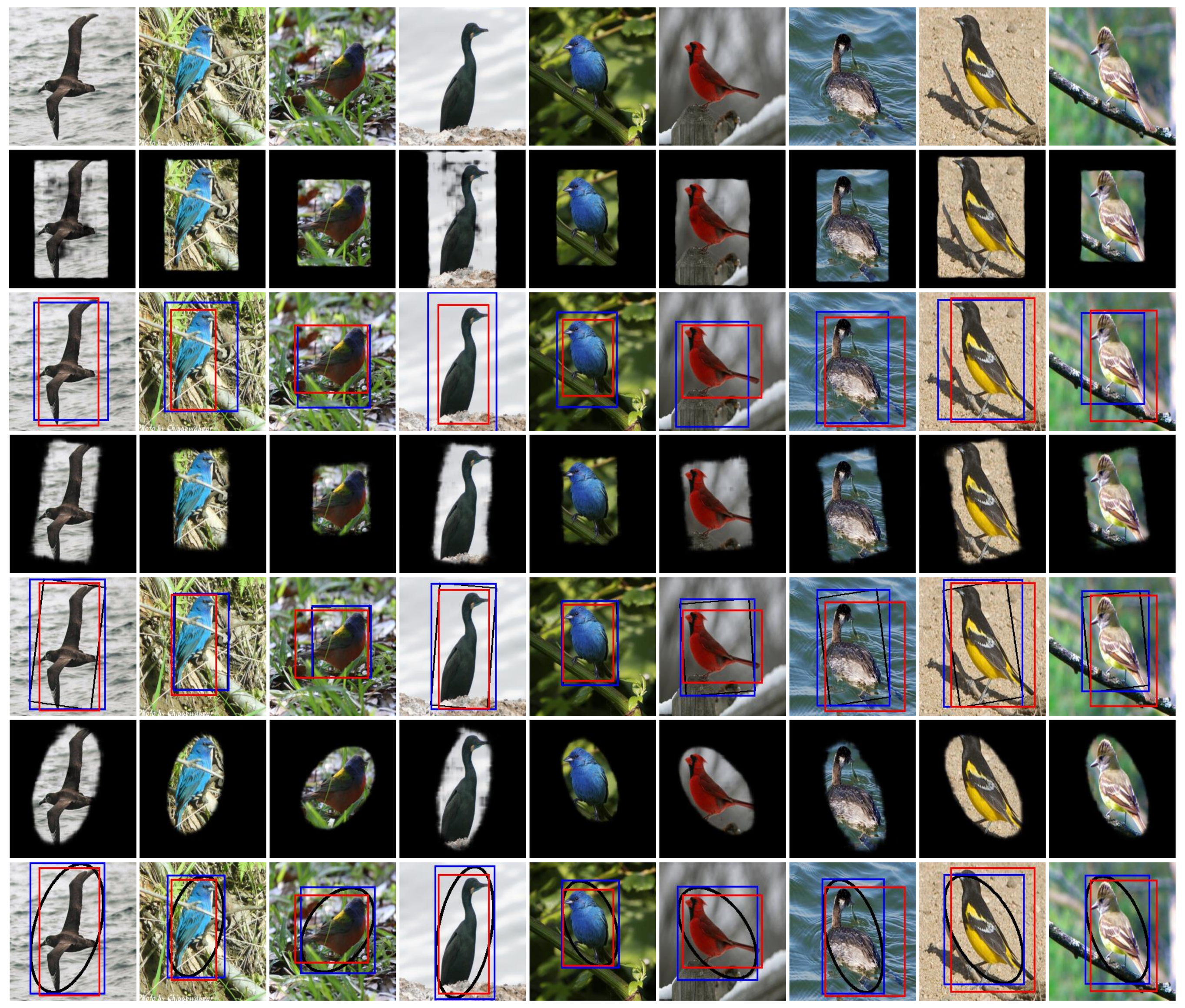

Fig. 1: Some example results from CUB-200-2011 dataset

- Weakly supervised object localization (WSOL)
- We propose a novel method for WSOL, termed GC-Net.
- GC-Net is end-to-end training and without a post-processing step.
- GC-Net able to predict a rough rotation angle of the object

Method

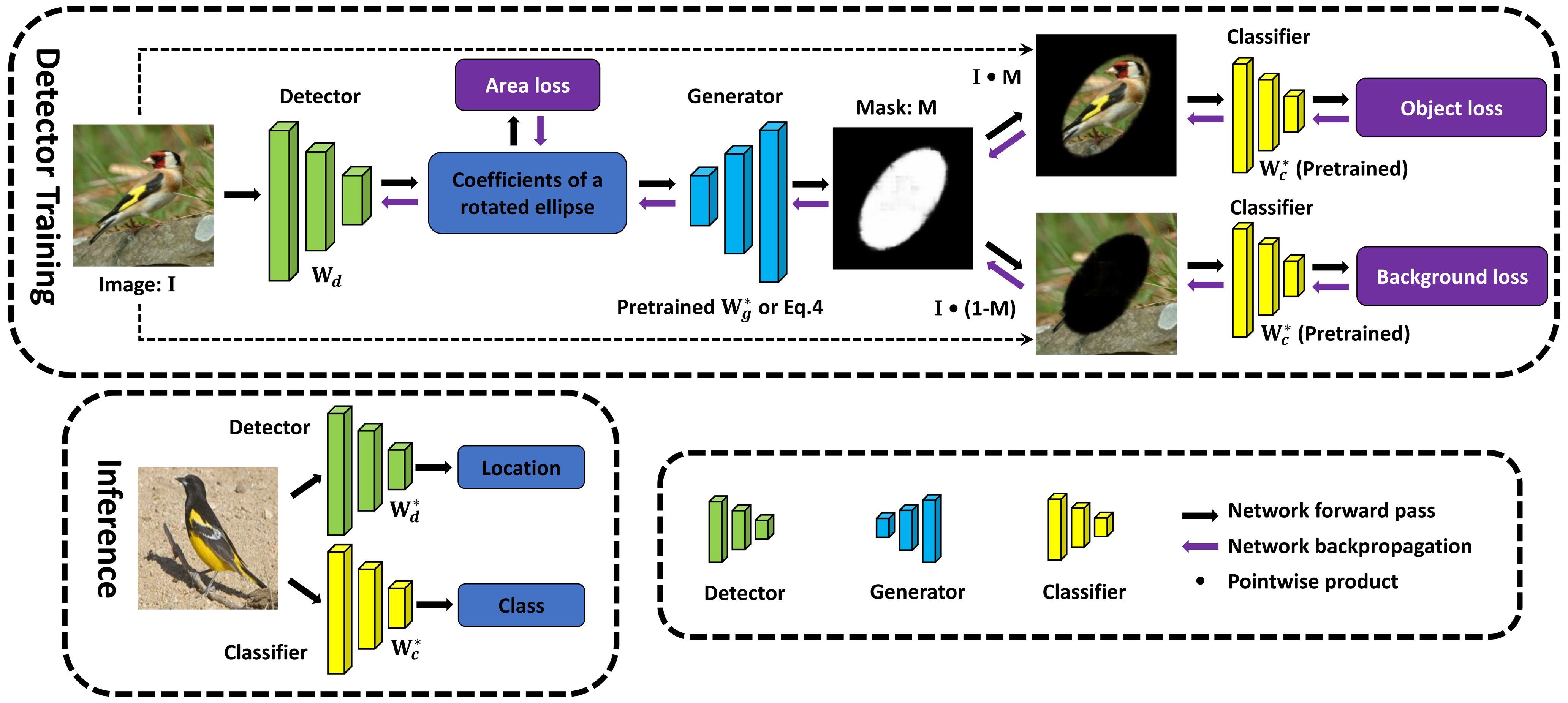

Fig. 2: The architecture of the proposed GC-Net including the detector, generator and classifier.

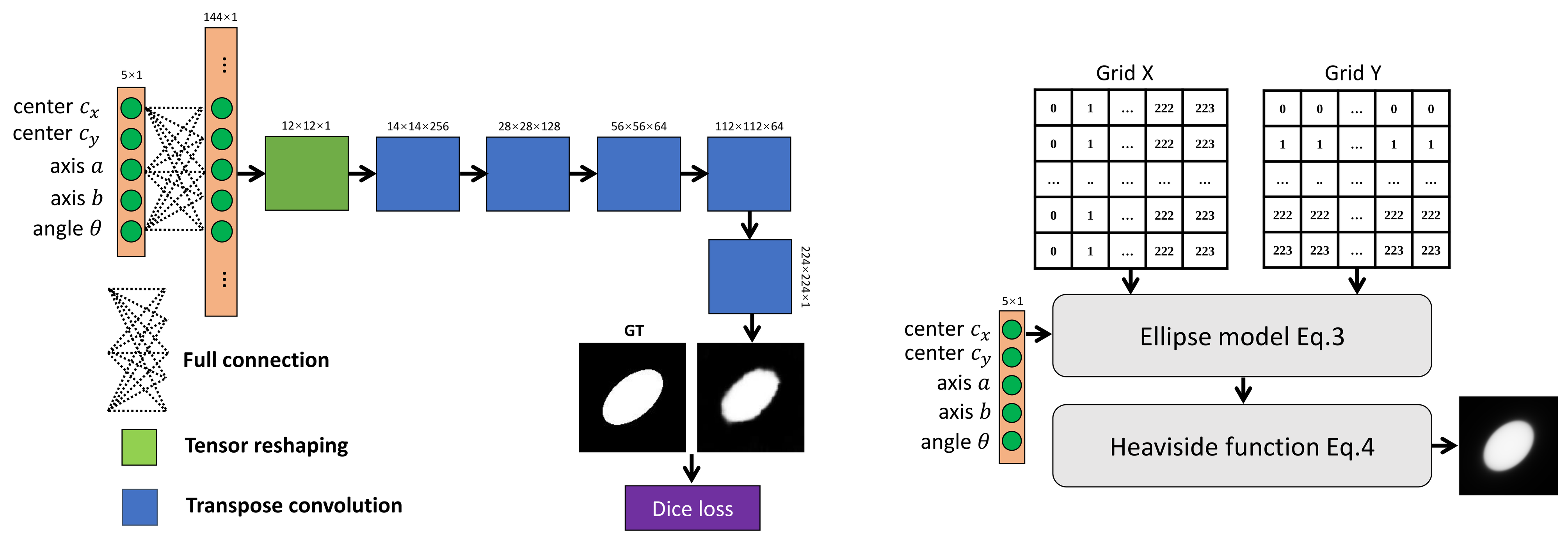

Fig. 3: Object mask generation using learning-driven (left) and model-driven (right) methods.

Results

Table 1: Comparison of the performance between GC-Net and the state-of-the-art on the CUB-200-2011 test set

	ClsErr		LocErr		
Methods compared	Top1	Top5	Top1	Top5	CorLoc
CAM-VGG [22]	23.4	**7.5**	55.85	47.84	56.0
ACoL-VGG [20]	28.1	-	54.08	43.49	54.1
SPG-VGG [21]	24.5	7.9	51.07	42.15	58.9
TSC-VGG [5]	-	-	-	-	65.5
DA-Net-VGG [19]	24.6	7.7	47.48	38.04	67.7
GC-Net-Elli-VGG (ours)	**23.2**	7.7	**41.15**	**30.10**	**74.9**
GC-Net-Rect-VGG (ours)	**23.2**	7.7	**36.76**	**24.46**	**81.1**
CAM-GoogLeNet [22]	26.2	8.5	58.94	49.34	55.1
Friend or Foe-GoogLeNet [18]	-	-	-	-	56.5
SPG-GoogLeNet [21]	-	-	53.36	42.28	-
DA-Net-Inception-V3 [19]	28.8	9.4	50.55	39.54	67.0
GC-Net-Elli-GoogLeNet (ours)	**23.2**	**6.6**	**43.46**	**31.58**	**72.6**
GC-Net-Rect-GoogLeNet (ours)	**23.2**	**6.6**	**41.42**	**29.00**	**75.3**

Table 2: Comparison of the performance between GC-Net and the state-of-the-art on the ILSVRC2012 validation set

	ClsErr		LocErr	
Methods compared	Top1	Top5	Top1	Top5
Backprop-VGG [11]	-	-	61.12	51.46
CAM-VGG [22]	33.4	12.2	57.20	45.14
ACol-VGG [20]	32.5	12.0	54.17	40.57
Backprop-GoogLeNet [11]	-	-	61.31	50.55
GMP-GoogLeNet [22]	35.6	13.9	57.78	45.26
CAM-GoogLeNet [22]	35.0	13.2	56.40	43.00
HaS-32-GoogLeNet [13]	-	-	54.53	-
ACol-GoogLeNet [20]	29.0	11.8	53.28	42.58
SPG-GoogLeNet [21]	-	-	51.40	40.00
DA-Net-InceptionV3 [19]	27.5	8.6	52.47	41.72
GC-Net-Elli-Inception-V3 (ours)	**22.6**	**6.4**	51.47	42.58
GC-Net-Rect-Inception-V3 (ours)	**22.6**	**6.4**	**50.94**	41.91

Ablation Study

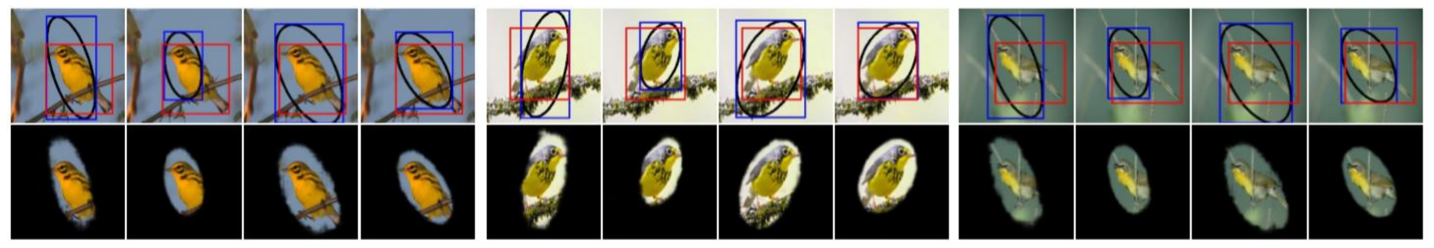

Fig. 4: For each example, from left to right the losses are Lo, Lo+La, Lo+Lb and La+Lo+Lb, respectively.

Table 3: Comparison of the object localization performance on CUB200-2011 using different losses.

	LocErr		
Loss functions	Top1	Top5	CorLoc
$\mathcal{L}_o$	59.22	51.75	51.69
$\mathcal{L}_o+\mathcal{L}_a$	69.89	63.12	39.89
$\mathcal{L}_o+\mathcal{L}_b$	47.03	37.69	66.52
$\mathcal{L}_a+\mathcal{L}_o+\mathcal{L}_b$	**41.15**	**30.10**	**74.89**

More Examples

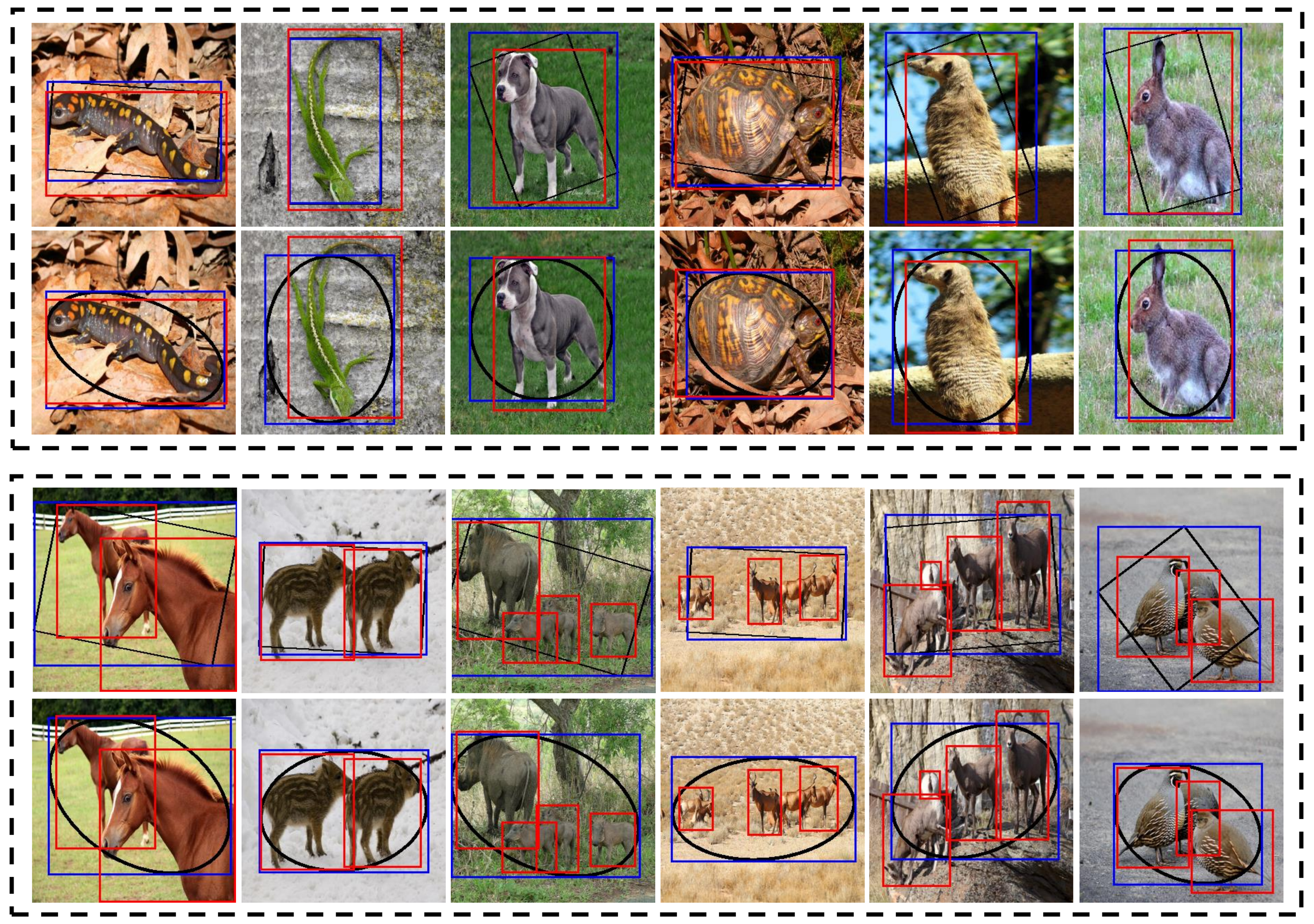

Fig. 5: Localization results on some images from the ILSRC2012 dataset using GC-Net. Top: single object localization. Bottom: multiple object localization. GC-Net tends to predict a bbox that contains all target objects.