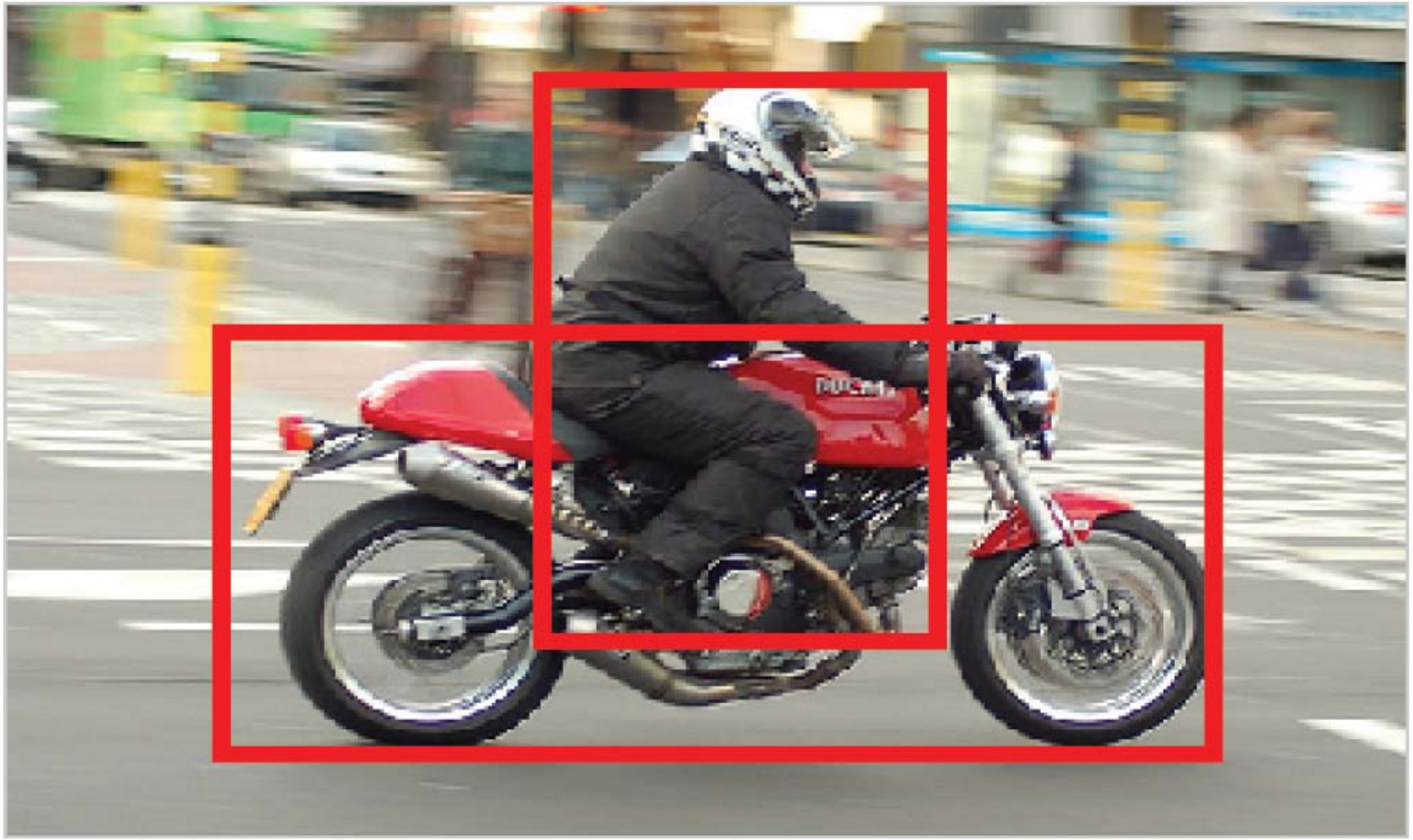


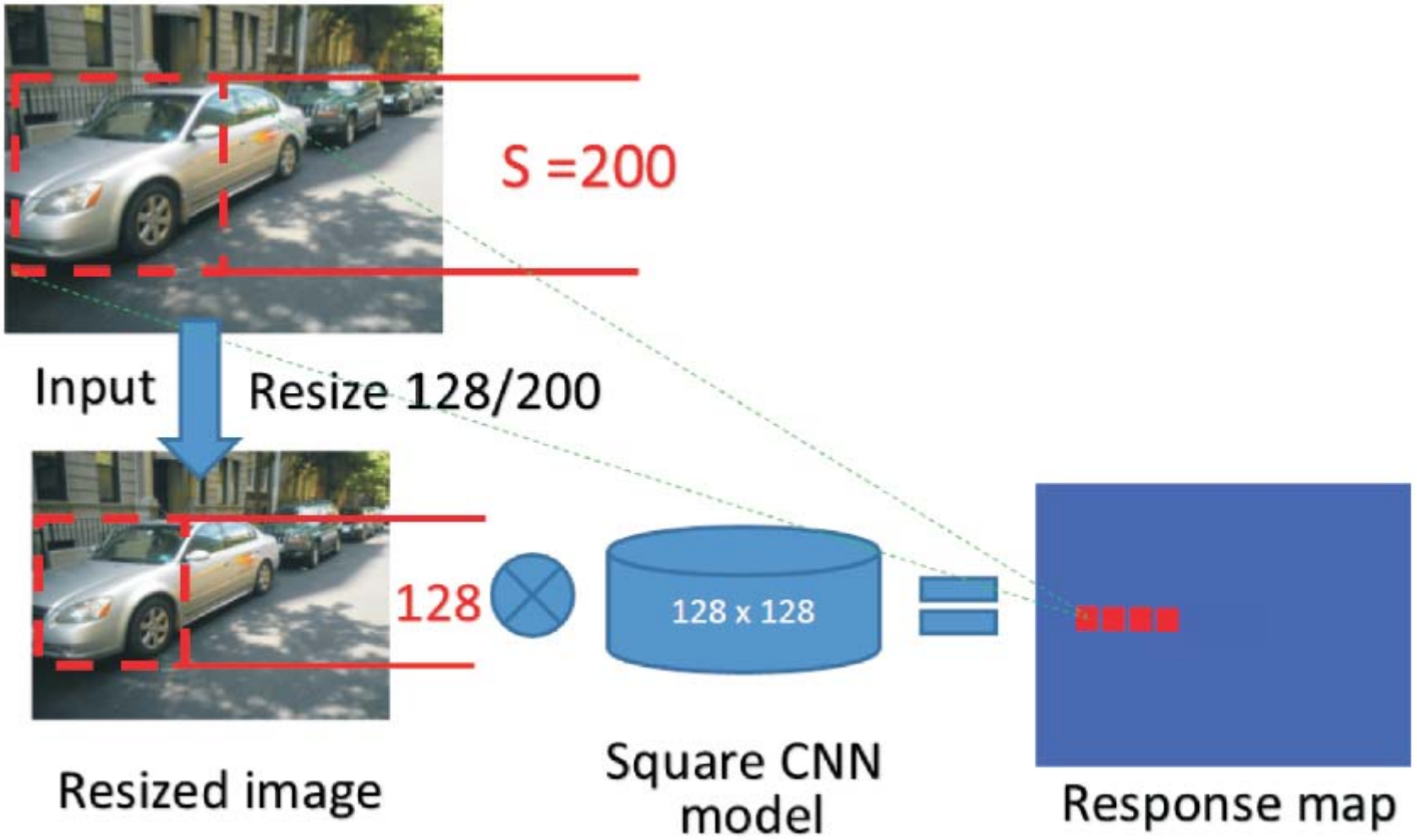

SWIFT

Object Image

Compact Square Object Image

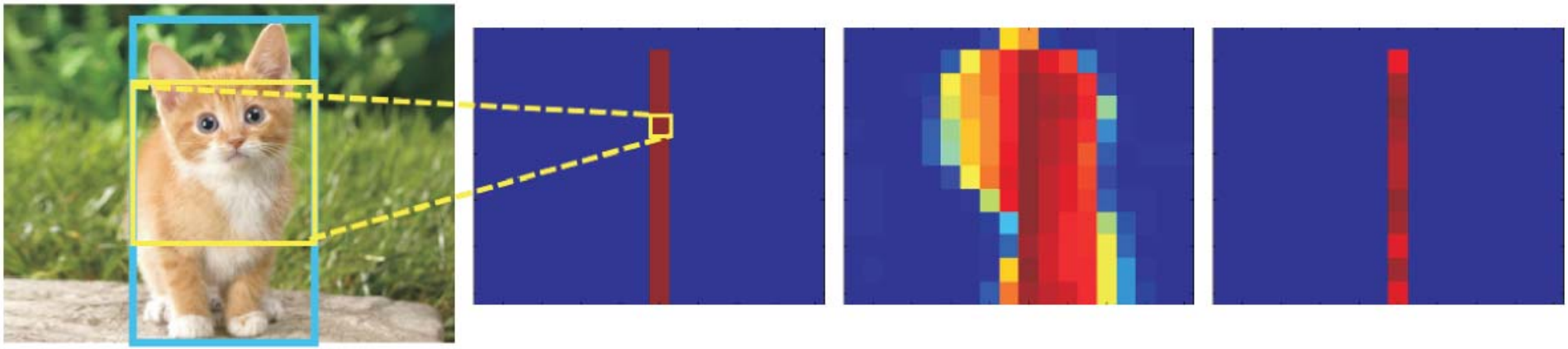
Accuracy	CNN model	Human
whole image	87.3%	97.6%
compact square image	84.6%	95.4%

Table 1. Empirical recognition accuracy of the CNN model [14] and human recognition on the VOC 2007 dataset. We ask people to label the category based on compact square object (CSO) images. With more careful parameter tuning of the CNN we believe the above accuracy can be further improved.


S =200
Input
Resize 128/200
128
128 x 128
Resized image
Square CNN model
Response map

Algorithm 1 Pseudo-code of Scale Selection Function

Function $scaleSelection(S, L)$
if $d(S, L) < \varepsilon$ or $L = S + 1$ **then**
  return
else
  selecting scale $M = \sqrt{LS}$
  call $scaleSelection(S, M)$
  call $scaleSelection(M + 1, L)$
end if



VOC 2007 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
R-CNN (SS)	64.2	69.7	50.0	41.9	32.0	62.6	71.0	60.7	32.7	58.5	46.5	56.1	60.6	66.8	54.2	31.5	52.8	48.9	57.9	64.7	54.2
R-CNN BB (SS)	68.1	72.8	**56.8**	43.0	36.8	66.3	74.2	67.6	34.4	63.5	54.5	61.2	69.1	68.6	58.7	33.4	62.9	51.1	62.5	64.8	58.5
R-CNN (EB)	63.7	64.8	50.3	42.2	31.4	59.4	68.4	61.3	31.2	54.7	42.6	58.6	63.1	64.2	56.3	30.2	53.1	45.2	58.6	66.2	53.2
Ours	**73.1**	**76.1**	55.8	**45.6**	**40.6**	**71.4**	**80.0**	**69.4**	**39.1**	**67.4**	**57.6**	**64.7**	**72.5**	**72.3**	**62.6**	**35.7**	**64.6**	**56.2**	**65.6**	**69.7**	**62.0**

VOC 2012 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
R-CNN (SS)	68.1	63.8	46.1	29.4	27.9	56.6	57.0	65.9	26.5	48.7	39.5	66.2	57.3	65.4	53.2	26.2	54.5	38.1	50.6	51.6	49.6
R-CNN BB (SS)	71.8	65.8	52.0	34.1	32.6	59.6	60.0	69.8	27.6	52.0	**41.7**	69.6	61.3	68.3	57.8	29.6	57.8	40.9	59.3	54.1	53.3
R-CNN (EB)	68.3	63.8	46.5	29.6	27.8	56.6	54.5	63.3	26.2	45.3	37.8	66.4	57.4	64.3	50.5	24.7	52.8	39.3	50.4	52.8	48.9
Ours	**76.8**	**71.2**	**61.2**	**45.1**	**35.9**	**62.5**	**60.9**	**75.5**	**31.3**	**58.3**	39.4	**73.8**	**68.6**	**73.1**	**60.9**	**33.1**	**59.1**	**41.0**	**66.3**	**57.6**	**57.7**

Sample detection results on the VOC 2012 dataset.