Informative Object Annotations - Supplementary Material

Lior Bracha
Bar-Ilan University
lior.bracha@live.biu.ac.il

Gal Chechik
Bar-Ilan University, NVIDIA Research
gal.chechik@biu.ac.il

Abstract

Supplementary materials to the main paper.

1. Robustness to hyper parameters

We tested the robustness of our approach with respect to two hyper parameters of the model: (1) number of trees on which we averaged to create Tree-Mixture-Model. (2) vocabulary size.

For the first, we computed all scoring functions for tree mixtures with 1,3,5 and 10 trees, and found only a 3% difference in the p@1 of $cw\Delta H$.

Second, we tested robustness to the number of words in our vocabulary. The vocabulary size is important because our analysis was performed over the most frequent labels in the corpus. As a result, the size of the vocabulary could have affected precision, because entry-level terms (*dog, car*) tend to be more frequent than more fine-grained terms (*e.g.* Labrador, Toyota). We repeated our analysis with different thresholds on the minimum label frequency included in the vocabulary (threshold for values of 50, 100:1000). Figure S 1 plots the precision@1 of the various scoring functions, showing that the analysis is robust to the size of the vocabulary.

2. Expected entropy

The ***expected entropy-reduction*** was defined as

$$E(\Delta H) = q(l_i|I)\Delta H + (1 - q(l_i|I)) \cdot 0$$

When an incorrect label is transmitted, we assume here that no information is passed to the listener. However, this does have an impact on the listener entropy and knowledge scheme. Transmission of a false message could be thought of as negative information as it is misleading for the listener.

3. Implementation Details

Algorithm 1 describes in detail the steps to compute the $cw\Delta H$ scores for a set of labels. Algorithm 2 describes the inference phase, where the computed scores provide an information-based ranking of the image annotations. Here, we do not specify whether we take a single label as ground truth (by majority) or multiple labels (see Sec 4.2) but give a general framework.

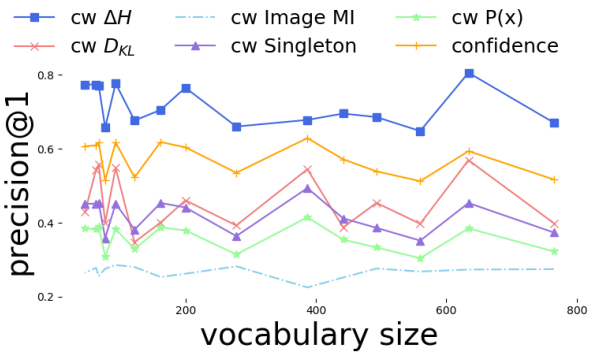


Figure S 1. **Robustness to vocabulary size**. Different thresholds for the minimum number of label occurrence were tested. The precision of $cw\Delta H$ remains very high for a large range of vocabulary sizes. The relationship between the different scoring functions is consistent as well.

4. Qualitative Results

Figure S2 illustrates the annotations ranking for some images from OID test-set. In these examples we give the full, raw output of our experiments, showing results from all scoring-functions, with or without the confidence weights. "verification" column specifies whether the label was verified by OID raters as correct. "R*" columns present our raters response (see Sec. 4.1) and "y-true" column is the ground truth determined by majority. R* columns in which no entry is marked "1", means that the rater's label was not in the vocabulary.

Algorithm 1 $cw\Delta H$ scoring-function

Input Annotated images

Output $cw\Delta H$ label-scores (averaged on k trees)

for tree **do**
  Sample uniformly from the label-distribution D (bootstrap)
  if $labelfrequency > threshold$ **then**
    $vocab \leftarrow label$
  end if

  for label pair $(i, j) \in vocab$ **do**
    Compute pair (2x2) joint distribution $p(i, j)$
    Compute the mutual information of i,j (Eq. 7)
  end for

  Create Graph $G(V, E)$
  Assign MI_{ij} as the weight of the edge connecting label i and label j
  Find a maximum weight spanning tree (MST)
  Sort the graph such that each node has a single parent
  Compute tree entropy (Eq.8)

  for each label l_i in $l_1, ..., l_d$ **do**
    Set l_i as root
    Set root marginal distribution as $p(l_i) = [0, 1]$
    Reverse edges direction such that all $e \in E$ will be descendents of l_i
    Propagate the message $p(l_i)$ throughout the tree and update CPTs.
    Compute new tree entropy (Eq. 8)
    Compute $\Delta H(l_i)$ (Eq. 3)
  end for
end for

Average $\Delta H(l_i)$ over trees for all labels $(l_1...l_d)$

Algorithm 2 Ranking and evaluation

Input $\Delta H(l_i)$ label scores; ground-truth data

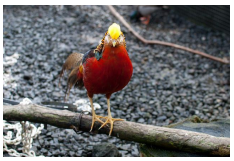
Output Ranking of image annotations, precision and recall

for image **do**
  $cw\Delta H \leftarrow confidence(l_i) \cdot \Delta H(l_i)$
  Rank image annotations by $cw\Delta H$.
  Evaluate against ground-truth label.
  Compute precision and recall.
end for
Average precision and recall across images.

Label Name	Display Name	Confidence	px	singleton	H	dkl	mi	random	Verification	cH	cDKL	cMI	cPX	cSingleton	y_true	R1	R2	R3
/m/0bt9lr	dog	1.00	0.05	0.27	71.65	14.24	4.51	0.51	1	71.65	14.24	4.51	0.05	0.27	1	1	1	1
/m/068hy	pet	1.00	0.06	0.33	71.42	11.84	4.05	0.52	1	71.42	11.84	4.05	0.06	0.33	0	0	0	0
/m/0kpmf	dog breed	0.90	0.03	0.21	71.41	14.15	4.92	0.49	1	64.27	12.73	4.43	0.03	0.19	0	0	0	0
/m/04rky	mammal	0.90	0.10	0.45	70.92	9.61	3.44	0.50	1	63.83	8.65	3.09	0.09	0.40	0	0	0	0
/m/0jbk	animal	0.90	0.16	0.63	69.94	7.36	2.68	0.49	1	62.95	6.62	2.41	0.15	0.56	0	0	0	0
/m/05zmzn1	street dog	0.80	0.00	0.04	71.57	17.68	7.80	0.51	0	57.26	14.14	6.24	0.00	0.03	0	0	0	0
/m/0wfrc8w	animal sports	0.70	0.01	0.09	71.97	13.21	6.54	0.48	0	50.38	9.24	4.58	0.01	0.06	0	0	0	0
/m/0gzt5	hound	0.60	0.01	0.05	71.91	18.01	7.53	0.50	0	43.14	10.80	4.52	0.00	0.03	0	0	0	0
/m/0wgq8gt	dog sports	0.60	0.00	0.04	71.61	16.34	8.12	0.48	0	42.96	9.80	4.87	0.00	0.02	0	0	0	0
/m/06ntj	sports	0.60	0.10	0.45	70.23	7.31	3.45	0.49	0	42.14	4.39	2.07	0.06	0.27	0	0	0	0
/m/017y8_	terrier	0.50	0.01	0.07	71.42	16.62	6.90	0.53	0	35.71	8.31	3.45	0.00	0.04	0	0	0	0
/m/011rl	carnivor	0.50	0.03	0.17	70.85	13.22	5.35	0.48	1	35.43	6.61	2.68	0.01	0.08	0	0	0	0

LabelName	Display Name	Confidence	px	singleton	H	dkl	mi	random	Verification	cH	cDKL	cMI	cPX	cSingleton	y_true	R1	R2	R3
/m/07j7r	tree	1.00	0.03	0.19	71.69	9.66	5.13	0.50	1.00	**71.69**	**9.66**	**5.13**	0.03	0.19	1	1	1	1
/m/05s2s	plant	0.90	0.13	0.54	72.16	8.62	3.03	0.49	0.00	64.94	7.76	2.73	**0.12**	**0.49**	0	0	0	0
/m/09t49	leaf	0.80	0.04	0.23	72.39	10.10	4.78	0.49	1.00	57.91	8.08	3.82	0.03	0.18	0	0	0	0
/m/0b5gs	branch	0.70	0.02	0.16	70.26	8.39	5.41	0.45	1.00	49.18	5.88	3.79	0.02	0.12	0	0	0	0
/m/06z0n	sunlight	0.60	0.01	0.06	70.15	10.36	7.26	0.52	1.00	42.09	6.22	4.36	0.00	0.04	0	0	0	0
/m/0c9ph5	flower	0.50	0.12	0.51	**72.99**	9.14	3.18	0.52	0.00	36.49	4.57	1.59	0.06	0.25	0	0	0	0
/m/0gqbt	shrub	0.50	0.04	0.22	72.67	10.97	4.87	0.51	0.00	36.34	5.48	2.44	0.02	0.11	0	0	0	0
/m/036qh8	produce	0.50	0.14	**0.57**	71.10	7.58	2.89	0.49	1.00	35.55	3.79	1.44	0.07	0.29	0	0	0	0
/m/01341	autumn	0.50	0.01	0.06	71.03	**11.33**	**7.28**	0.48	1.00	35.51	5.66	3.64	0.00	0.03	0	0	0	0

LabelName	Display Name	Confidence	px	singleton	H	dkl	mi	random	Verification	cH	cDKL	cMI	cPX	cSingleton	y_true	R1	R2	R3
/m/05s2s	plant	0.90	0.13	0.54	72.16	8.62	3.03	0.49	1.00	64.94	7.76	2.73	0.12	0.49	1	0	1	1
/m/07j7r	tree	0.90	0.03	0.19	71.69	9.66	5.13	0.50	0.00	64.52	8.70	4.62	0.03	0.17	0	0	0	0
/m/038hg	green	0.90	0.03	0.22	70.44	7.86	4.91	0.47	1.00	63.40	7.07	4.42	0.03	0.19	0	0	0	0
/m/01fnns	vegetation	0.90	0.00	0.03	69.81	12.48	8.35	0.49	1.00	62.82	11.23	7.52	0.00	0.03	0	0	0	0
/m/03bmqb	flora	0.80	0.07	0.37	73.51	11.20	3.86	0.49	1.00	58.81	8.96	3.09	0.06	0.29	0	0	0	0
/m/09t49	leaf	0.80	0.04	0.23	72.39	10.10	4.78	0.49	1.00	57.91	8.08	3.82	0.03	0.18	0	0	0	0
/m/08t9c_	grass	0.80	0.02	0.15	69.93	8.66	5.59	0.49	0.00	55.94	6.93	4.48	0.02	0.12	0	0	0	0
/m/09dv7	soil	0.80	0.02	0.14	69.66	8.24	5.77	0.52	0.00	55.73	6.60	4.61	0.02	0.11	0	0	0	0
/m/0c9ph5	flower	0.70	0.12	0.51	72.99	9.14	3.18	0.52	0.00	51.09	6.40	2.23	0.08	0.35	0	0	0	0
/m/0hnc1	woodland	0.70	0.01	0.06	70.68	12.27	7.33	0.51	0.00	49.48	8.59	5.13	0.00	0.04	0	0	0	0
/m/0gqbt	shrub	0.60	0.04	0.22	72.67	10.97	4.87	0.51	0.00	43.60	6.58	2.92	0.02	0.13	0	0	0	0
/m/01v327	lawn	0.60	0.01	0.10	69.30	9.47	6.41	0.49	0.00	41.58	5.68	3.84	0.01	0.06	0	0	0	0
/m/0b101	garden	0.50	0.01	0.11	72.36	11.40	6.16	0.50	0.00	36.18	5.70	3.08	0.01	0.05	0	0	0	0



LabelName	Display Name	Confidence	px	singleton	H	dkl	mi	random	Verification	cH	cDKL	cMI	cPX	cSingleton	y_true	R1	R2	R3
/m/015p6	bird	1.00	0.03	0.21	72.36	12.52	4.96	0.54	1.00	72.36	12.52	4.96	0.03	0.21	1	1	1	1
/m/01c4rd	beak	0.90	0.03	0.17	70.96	12.71	5.35	0.50	1.00	63.86	11.44	4.82	0.02	0.15	0	0	0	0
/m/05h0n	nature	0.90	0.02	0.14	70.82	8.99	5.72	0.48	1.00	63.74	8.09	5.15	0.02	0.12	0	0	0	0
/m/01280g	wildlife	0.90	0.06	0.34	70.29	9.63	4.02	0.53	1.00	63.26	8.66	3.62	0.06	0.31	0	0	0	0
/m/06fvc	red	0.90	0.02	0.15	69.78	7.98	5.64	0.51	1.00	62.80	7.18	5.08	0.02	0.13	0	0	0	0
/m/03bmqb	flora	0.80	0.07	0.37	73.51	11.20	3.86	0.49	1.00	58.81	8.96	3.09	0.06	0.29	0	0	0	0
/m/0336h	finch	0.80	0.00	0.04	71.22	17.54	7.79	0.50	0.00	56.98	14.03	6.23	0.00	0.03	0	0	0	0
/m/035qhg	fauna	0.80	0.10	0.46	70.70	8.83	3.39	0.49	1.00	56.56	7.07	2.71	0.08	0.37	0	0	0	0
/m/0jbk	animal	0.80	0.16	0.63	69.94	7.36	2.68	0.49	1.00	55.96	5.88	2.15	0.13	0.50	0	0	0	0
/m/0c9ph5	flower	0.70	0.12	0.51	72.99	9.14	3.18	0.52	0.00	51.09	6.40	2.23	0.08	0.35	0	0	0	0
/m/09t49	leaf	0.70	0.04	0.23	72.39	10.10	4.78	0.49	0.00	50.67	7.07	3.34	0.03	0.16	0	0	0	0
/m/02cqfm	close-up	0.60	0.07	0.36	70.83	7.13	3.93	0.51	1.00	42.50	4.28	2.36	0.04	0.21	0	0	0	0
/m/0b5gs	branch	0.60	0.02	0.16	70.26	8.39	5.41	0.45	1.00	42.15	5.04	3.25	0.01	0.10	0	0	0	0

LabelName	Display Name	Confidence	px	singleton	H	dkl	mi	random	Verification	cH	cDKL	cMI	cPX	cSingleton	y_true	R1	R2	R3
/m/0cmf2	airplane	1.00	0.02	0.16	73.62	19.97	5.41	0.50	1.00	73.62	19.97	5.41	0.02	0.16	1.00	1.00	1.00	0.00
/m/0dhz0	airliner	1.00	0.01	0.08	73.25	23.16	6.68	0.49	1.00	73.25	23.16	6.68	0.01	0.08	0.00	0.00	0.00	0.00
/m/018rqw	Wide-body aircraft	1.00	0.00	0.04	73.11	26.94	7.82	0.48	1.00	73.11	26.94	7.82	0.00	0.04	0.00	0.00	0.00	0.00
/m/0vg8	airline	0.90	0.01	0.10	73.64	22.46	6.27	0.50	1.00	66.27	20.22	5.65	0.01	0.09	0.00	0.00	0.00	0.00
/m/0k5j	aircraft	0.9	0.03	0.19	72.88	18.65	5.11	0.49	1.00	65.59	16.79	4.60	0.03	0.17	0.00	0.00	0.00	0.00
/m/025t3bg	air travel	0.90	0.00	0.04	72.40	23.84	7.94	0.51	1.00	65.16	21.46	7.15	0.00	0.04	0.00	0.00	0.00	0.00
/m/0fzyg	aviation	0.80	0.03	0.19	72.73	18.07	5.11	0.47	1.00	58.18	14.45	4.08	0.02	0.16	0.00	0.00	0.00	0.00
/m/015y8h	jet aircraft	0.80	0.01	0.10	72.29	20.25	6.31	0.50	1.00	57.83	16.20	5.05	0.01	0.08	0.00	0.00	0.00	0.00
/m/0m8_b	boeing 777	0.80	0.00	0.03	69.90	25.60	8.15	0.52	0.00	55.92	20.48	6.52	0.00	0.03	0.00	0.00	0.00	0.00
/m/01bq4	boeing 767	0.70	0.00	0.03	69.60	24.54	8.21	0.50	0.00	48.72	17.18	5.75	0.00	0.02	0.00	0.00	0.00	0.00
/m/015z_b	takeoff	0.60	0.00	0.04	72.22	22.30	7.94	0.50	0.00	43.33	13.38	4.76	0.00	0.02	0.00	0.00	0.00	0.00
/m/0w7s	Aerospace engineering	0.60	0.00	0.04	72.09	20.67	7.94	0.50	1.00	43.25	12.40	4.77	0.00	0.02	0.00	0.00	0.00	0.00
/m/0dp7vs	Narrow body aircraft	0.50	0.01	0.05	72.18	23.50	7.51	0.50	0.00	36.09	11.75	3.76	0.00	0.03	0.00	0.00	0.00	0.00
/m/016g8_	Tarmac	0.50	0.00	0.04	67.92	11.26	7.96	0.52	1.00	33.96	5.63	3.98	0.00	0.02	0.00	0.00	0.00	0.00
/m/083s2	wing	0.50	0.03	0.20	67.90	8.15	5.02	0.47	1.00	33.95	4.08	2.51	0.02	0.10	0.00	0.00	0.00	0.00
/m/07yv9	vehicle	0.90	0.24	0.77	26.56	12.44	2.12	0.49	0.00	23.90	11.20	1.91	0.22	0.69	0.00	0.00	0.00	0.00

Figure S 2. **Ranked annotations of images from the test set.** The different scoring-functions were analyzed with (right) or without (left) confidence weights. R* mark the raters response (OID-IOTA-10K, see Sec. 4.1).