1-1-2011

Conditional Entropies as Over-Segmentation and Under-Segmentation Metrics for Multi-Part Image Segmentation

Haifeng Gong
University of Pennsylvania, hfgong@seas.upenn.edu

Jianbo Shi
University of Pennsylvania, jshi@cis.upenn.edu

Recommended Citation

Haifeng Gong and Jianbo Shi, "Conditional Entropies as Over-Segmentation and Under-Segmentation Metrics for Multi-Part Image Segmentation", . January 2011.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-11-17.

Conditional Entropies as Over-Segmentation and Under-Segmentation Metrics for Multi-Part Image Segmentation

Abstract

In this paper, we define two conditional entropy measures for performance evaluation of general image segmentation. Given a segmentation label map and a ground truth label map, our measures describe their compatibility in two ways. The first one is the conditional entropy of the segmentation given the ground truth, which indicates the oversegmentation rate. The second one is that of the ground truth given the segmentation, which indicates the under-segmentation rate. The two conditional entropies indicate the trade-off between smaller and larger granularities like false positive rate and false negative rate in ROC, and precision and recall in PR curve. Our measures are easy to implement, and involve no threshold or other parameter, have very intuitive explanation and many good theoretical properties, e.g., good bounds, monotonicity, continuity. Experiments show that our measures work well on Berkeley Image Segmentation Benchmark using three segmentation algorithms, Efficient Graph- Based segmentation, Mean Shift and Normalized Cut. We also give an asymmetric similarity measure based on the two entropies and compared it with Variation of Information. The comparison revealled that our method has advantages in many situations.We also checked the coarse-to-fine compatibility of segmentation results with changing parameters and ground truths from different annotators.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-11-17.

Noname manuscript No.
(will be inserted by the editor)

Conditional entropies as over-segmentation and under-segmentation metrics for multi-part image segmentation

Haifeng Gong, Jianbo Shi

the date of receipt and acceptance should be inserted later

Abstract In this paper, we define two conditional entropy measures for performance evaluation of general image segmentation. Given a segmentation label map and a ground truth label map, our measures describe their compatibility in two ways. The first one is the conditional entropy of the segmentation given the ground truth, which indicates the over-segmentation rate. The second one is that of the ground truth given the segmentation, which indicates the under-segmentation rate. The two conditional entropies indicate the trade-off between smaller and larger granularities like false positive rate and false negative rate in ROC, and precision and recall in PR curve. Our measures are easy to implement, and involve no threshold or other parameter, have very intuitive explanation and many good theoretical properties, e.g., good bounds, monotonicity, continuity. Experiments show that our measures work well on Berkeley Image Segmentation Benchmark using three segmentation algorithms, Efficient Graph-Based segmentation, Mean Shift and Normalized Cut. We also give an asymmetric similarity measure based on the two entropies and compared it with Variation of Information. The comparison revealled that our method has advantages in many situations. We also checked the coarse-to-fine compatibility of segmentation results with changing parameters and ground truths from different annotators.

1 Introduction

Evaluation of an image segmentation algorithm on a dataset with ground truths is an important topic in computer vision. We consider general multipart image segmentation, in which an image is segmented into multiple parts, the resultant segmentation is validated with one or multiple ground truths, and no semantic label is involved. In this setting, the validation of an segmentation with ground truths is not so straightforward, because the labels in the segmentation and ground truths are not in correspondence, and ground truths come in varying granularities. The performance evaluation of foreground/background segmentation (Ge et al, 2007) and semantic segmentation (Everingham et al, 2010) is simpler and not in the scope of this paper.

Good image segmentation metrics should indicate over-segmentation and under-segmentation explicitly. Qualitatively, there are four possible outputs of comparing a segmentation to a ground truth — almost perfect segmentation, good over-segmentation, good under-segmentation and unacceptable bad segmentation. Like false alarm and missed detection in object detection, over-segmentation and under-segmentation are two types of errors that can happen simultaneously. For example, an algorithm may give under-segmented results at the left side of an image, while giving over-segmented results on the right. All segmentation algorithms have parameters to trade off these two type of errors. For different applications, one might prefer over-segmentation or under-segmentation. For example, if we use segments as preprocessing of an automatic object segmentation or tracking algorithm, we may prefer over-segmentation, because it can be further merged into better and bigger segments using other cues that have not been used in segmentation and it is not easy to split under-segmentation for automatic algorithms. If we use segments for an interactive object cropping, we may prefer under-segmentation, because we can ask users to further split big segments, while too many small segments will need more human interaction to merge. Therefore, it is necessary for segmentation metrics to describe the over-segmentation and under-segmentation rates explicitly and separately.

In addition to the above mentioned requirement, a good measure must fulfill the following conditions:

Department of Computer and Information Science, University of Pennsylvania

1. Well bounded. The measurement should score a perfect segmentation as zero error, and the degenerate segmentations (the segmentation where all pixels share the same label and the segmentation where each pixel has an individual label) as the largest error.
2. Good numeric stability. This requirement comes in two aspects. 1) Changing the label of one pixel should result in a small change in the metric scores. This rules out the simple idea of first establishing label correspondences greedily, then counting the overlapping areas, because small changes in the segmentation may result in very different correspondences. 2) If we gradually change the algorithm parameter to produce segmentations at different granularities, the scores should also change gradually.
3. Easy to implement. It should have no or a very small number of parameters to tune. The algorithm to compute the measure should also be simple and efficient.

1.1 Related Work

The most popular measures are Berkeley image segmentation benchmark (Martin et al, 2001), Normalized Probabilistic Rand (NPR) index (Unnikrishnan et al, 2007) and Variation of Information (VI) (Meilă, 2005)(Meilă, 2007). In Berkeley image segmentation benchmark (Martin et al, 2001), two sets of measures are proposed, a) F-measure, b) Global consistency error (GCE) and Local consistency error (LCE). The F-measure is defined on probability edge maps. It matches segmentation boundaries by bipartite graph matching and produces a precision-recall curve by changing the edge blurring bandwidth. The PR curve only shows the error of boundary localization, and doesn't indicate over-segmentation or under-segmentation. Other boundary based measures (Estrada and Jepson, 2009) have similar problems. As pointed out in (Martin et al, 2001), the GCE and LCE suffer a degeneracy problem, at two trivial segmentations, where each pixel has an individual label or the whole image shares the same label. NPR index (Unnikrishnan et al, 2007) is the normalized version of of Probability Rand (PR) index, which first computes the probability of label correlation of all pairs of points in a set of ground truths, then uses the likelihood of the segmentation with respect to this probability as output score. The normalization is introduced to make the index value comparable across different images. Though it has many good properties, such as comparable scores and accommodation of refinement, it does not indicate over-segmentation and under-segmentation explicitly. Variation of Information (Meilă, 2005)(Meilă, 2007) is closely related to our measure. It uses the sum of the two conditional entropies of the ground truth and the segmentation as a score. We use the two entropies as two individual metrics, to measure under-segmentation and over-segmentation separately.

Although there are many metrics that measures how a segmentation is consistent with a set of ground truths, to the best of our knowledge, no work in the literature has attempted to indicate the over-segmentation and under-segmentation explicitly and separately.

Our work is motivated by the Variance of Information score (Meilă, 2005) and therefore inherits many of its advantages, e.g., 1) many good theoretical properties, 2) no parameter to tune, very easy to implement. In addition, our two metrics can describe over-segmentation and under-segmentation in an information theoretical manner, and therefore, have intuitive explanation in term of information encoding.

1.2 Notations

Let X and Y be two discrete random variables, which follow distributions $P(X)$ and $P(Y)$ respectively. We denote the entropy of X by

$$\mathcal{H}\{X\} = -\sum_{X} P(X)\log P(X) \tag{1}$$

and the conditional entropy of X given Y by

$$\mathcal{H}\{X|Y\} = -\sum_{X,Y} P(X,Y)\log P(X|Y). \tag{2}$$

The enrtopy $\mathcal{H}\{X\}$ measures how much information we need to encode X, while $\mathcal{H}\{X|Y\}$ measure how much information we need to encode X if we have encoded Y and want to further encode X. The entropy of conditional probability of X given $Y = y$ is

$$\mathcal{H}\{X|Y=y\} = -\sum_{X} P(X|Y=y)\log P(X|Y=y). \tag{3}$$

Note that $\mathcal{H}\{X|Y\}$ is the expected value of $\mathcal{H}\{X|Y=y\}$

$$\mathcal{H}\{X|Y\} = \sum_{y} P(Y=y)\mathcal{H}\{P(X|Y=y)\}. \tag{4}$$

In the proof of the theorems in the rest of this paper, we need mutual information. We denote the mutual information between X and Y by

$$\mathcal{M}(X,Y) = \sum_{X,Y} P(X,Y)\log\frac{P(X,Y)}{P(X)P(Y)}, \tag{5}$$

and the conditional mutual information between X and Y given Z by

$$\mathcal{M}(X,Y|Z) = \sum_{X,Y,Z} P(X,Y,Z)\log\frac{P(X,Y|Z)}{P(X|Z)P(Y|Z)}. \tag{6}$$

All the above mentioned entropies and mutual informations are non-negative.

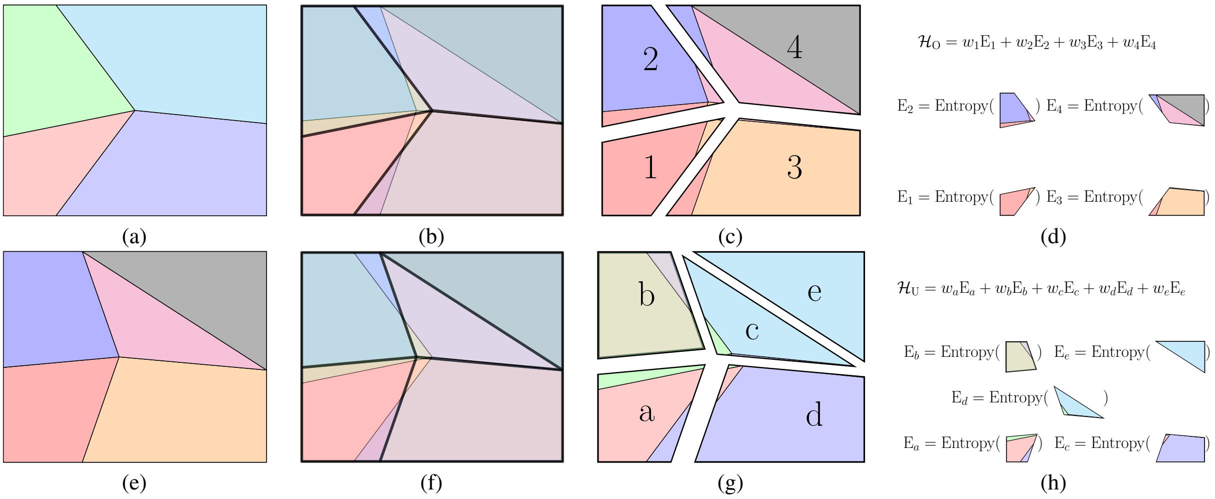

2
4
1
3
$\mathcal{H}_{\mathrm{O}} = w_1\mathrm{E}_1 + w_2\mathrm{E}_2 + w_3\mathrm{E}_3 + w_4\mathrm{E}_4$
$\mathrm{E}_2 = \mathrm{Entropy}($ $)$ $\mathrm{E}_4 = \mathrm{Entropy}($ $)$
$\mathrm{E}_1 = \mathrm{Entropy}($ $)$ $\mathrm{E}_3 = \mathrm{Entropy}($ $)$
(a)
(b)
(c)
(d)
b
e
c
a
d
$\mathcal{H}_{\mathrm{U}} = w_a\mathrm{E}_a + w_b\mathrm{E}_b + w_c\mathrm{E}_c + w_d\mathrm{E}_d + w_e\mathrm{E}_e$
$\mathrm{E}_b = \mathrm{Entropy}($ $)$ $\mathrm{E}_e = \mathrm{Entropy}($ $)$
$\mathrm{E}_d = \mathrm{Entropy}($ $)$
$\mathrm{E}_a = \mathrm{Entropy}($ $)$ $\mathrm{E}_c = \mathrm{Entropy}($ $)$
(e)
(f)
(g)
(h)

Fig. 1 A pair of conditional entropy measures. (a) A Ground truth with 4 segments. (b) A Segmentation overlaid on the ground truth. (c) The segmentation in each segment of the ground truth. (d) Explanation of the over-segmentation entropy. Each segment in the ground truth is further segmented into multiple subsegments by the segmentation. The over-segmentation rate of each segment $1 \sim 4$ can be measured by its entropy $E_i = \mathcal{H}\{S|T = i\}$, because higher entropy in each segment means that each ground truth segment is covered by more segmentation labels, or by the same number of subsegments, but in a more uniform manner. The over-segmentation entropy $\mathcal{H}_O$ is the weighted sum of these entropies. (e) THe segmentation with 5 segments. (f) The ground truth overlaid on the segmentation. (g) The ground truth in each segment of the segmentation. (h) Explanation of the under-segmentation entropy. Each segment in the segmentation is further segmented into multiple parts by the ground truth. The under-segmentation rate of each segment $a \sim e$ can be measured by its entropy. The under-segmentation entropy $\mathcal{H}_U$ is the weighted sum of the entropies of segment $a \sim e$.

2 Conditional Entropies

Let T be the label map of ground-truth, and S be the label map of segmentation. These are integer matrices, $T(x, y) \in \{1, \cdots, N_T\}$ and $S(x, y) \in \{1, \cdots, N_S\}$, in which N_T and N_S are number of segments in the ground truth and the segmentation correspondingly. There is no correspondence between their labels. Let Λ be the image lattice, $\Lambda_{T,i}$ be the lattice occupied by the i-th segment in T. By counting the histograms of labels in T and S, we can define a set of probabilities

$$P(T = i) = \frac{|\Lambda_{T,i}|}{|\Lambda|} \tag{7}$$

$$P(S = j) = \frac{|\Lambda_{S,j}|}{|\Lambda|} \tag{8}$$

$$P(T = i, S = j) = \frac{|\Lambda_{T,i} \cap \Lambda_{S,j}|}{|\Lambda|} \tag{9}$$

from which conditional probability $P(T = i|S = j)$ and $P(S = j|T = i)$ can be computed. Now we are ready to define our two conditional entropies.

If we superimpose S on T, each segment in T is divided into smaller subsegments by S. In i-th segment in T, if the number of the subsegments is smaller, or there is only one dominant subsegment, S and T are more consistent. This property can be measured by entropy $\mathcal{H}\{S|T = i\}$. For the whole image, we combine these entropies using the segment sizes as weights, $\sum_i P(T = i)\mathcal{H}\{S|T = i\} = \mathcal{H}\{S|T\}$, following Equation 4. See Figure 1 for a more detailed explanation.

Definition 1 Over-segmentation entropy (OSE) of segmentation S with respect to ground truth T is the conditional entropy of S given T

$$\mathcal{H}_{\mathrm{O}} = \mathcal{H}\{S|T\} = -\sum_{S,T} P(T, S) \log P(S|T). \tag{10}$$

If we superimpose T on S, similarly, we can define under-segmentation entropy.

Definition 2 Under-segmentation entropy (USE) of segmentation S with respect to ground truth T is the conditional entropy of T given S

$$\mathcal{H}_{\mathrm{U}} = \mathcal{H}\{T|S\} = -\sum_{T,S} P(T, S) \log P(T|S). \tag{11}$$

If multiple ground truths are provided, we use the means of the two entropies as the final metrics. Let $T_1, \cdots, T_L$ be the set of ground truths.

Definition 3 Mean over-segmentation entropy (MOSE) and mean under-segmentation entropy (MUSE) of segmentation S with respect to ground truths $T_1, \cdots, T_L$ are defined by

$$\bar{\mathcal{H}}_{\mathrm{O}} = \frac{1}{L}\sum_i \mathcal{H}\{S|T_i\}, \tag{12}$$

$$\bar{\mathcal{H}}_{\mathrm{U}} = \frac{1}{L}\sum_i \mathcal{H}\{T_i|S\}. \tag{13}$$

Meilă(Meilă, 2005) uses the sum of the over-segmentation and under-segmentation entropies as a distance between the segmentation and ground truth.

Definition 4 Variation of information(Meilă, 2005) between T and S is defined by

$$\mathcal{V}(S,T) = \mathcal{H}_{\mathrm{O}} + \mathcal{H}_{\mathrm{U}}, \tag{14}$$

and for mulitple ground truths, we can define mean variation of information (MVI)

$$\bar{\mathcal{V}}(S,T_1,\cdots,T_L) = \bar{\mathcal{H}}_{\mathrm{O}} + \bar{\mathcal{H}}_{\mathrm{U}}. \tag{15}$$

We will compare MUSE/MOSE with MVI in the experiments. Though this symmetric measure has many good properties, it summarizes the under-segmentation and over-segmentation with equal weights, and cannot tell one from another, or let us prefer one to another.

The intuitive explanation of the two conditional entropies is as follows. The over-segmentation entropy counts how much information we need to further encode $S \cap T$ given T, and the under-segmentation entropy counts how much information we need to further encode $S \cap T$ given S.

If the segmentation splits each segment in the ground truth in half, then the over-segmentation entropy $\mathcal{H}_O$ is $\log 2$. More generally, if the segmentation splits each segment in the ground truth into M equal parts, the over-segmentation entropy is $\log M$. Therefore, $\exp(\mathcal{H}_O)$ is a measure of degree of over segmentation in term of number of uniform splits. Similarly, if each segment in the segmentation contains M segments of ground truths of equal areas, then the under-segmentation entropy is $\log M$. In this sense, both OSE and USE are well normalized and comparable across images. Note that we cannot use counts of segments in each ground truth segment or segmentation segment instead of entropy, because this does not tolerate small boundary mismatches, and gives unnecessarily large scores.

If both conditional entropies are very low, the segmentation is an almost perfect one. If both are high, then it is a bad segmentation. If the over-segmentation entropy is high but the under-segmentation entropy is very low, it is a perfect over-segmentation, and may be useful if we want superpixels. If the under-segmentation entropy is high but the over-segmentation entropy is very low, it is a perfect under-segmentation.

3 Theoretical Analysis

3.1 Bounds

For a perfect segmentation, both over-segmentation and under-segmentation entropies are zero. If S is a perfect under-segmentation of T, that is each segment in T is restricted within a single segment in S, then $\mathcal{H}_O = 0$. Similarly, if S is a perfect over-segmentation of T, that is each segment in S is a part of a segment of T, then $\mathcal{H}_U = 0$. These properties are summarized in the following two theorems.

Theorem 1 *Over-segmentation entropy is bounded, for all possible S,*

$$0 \leqslant \mathcal{H}_{\mathrm{O}} \leqslant \log|\Lambda| - \mathcal{H}\{T\}. \tag{16}$$

Furthermore, $\mathcal{H}_{\mathrm{O}} = 0$ if and only if

$$\forall(x,y),(x',y'),T(x,y) = T(x',y') \Rightarrow S(x,y) = S(x',y'). \tag{17}$$

Proof We first prove Equation (16). Because conditional entropies of discrete random variabls are non-negative, $\mathcal{H}_{\mathrm{O}} \geqslant 0$ is obvious. To prove the upper bound, we write the conditional entropy as the expected value of entropies of conditional probabilities

$$\begin{aligned}\mathcal{H}_{\mathrm{O}} &= \sum_t P(T=t)\mathcal{H}\{S|T=t\} && (18)\\ &\leqslant \sum_t P(T=t)\log|\Lambda_t| && (19)\\ &= \sum_t P(T=t)\left(\log\frac{|\Lambda_t|}{|\Lambda|} + \log|\Lambda|\right) && (20)\\ &= \log|\Lambda| - \mathcal{H}\{T\} && (21)\end{aligned}$$

The equality holds when $P(S|T=t)$ is uniform for all t.

Equation (17) can be proved as follows. $\mathcal{H}_{\mathrm{O}} = 0$ means $\mathcal{H}\{S|T=t\} = 0$ for all t, because $\mathcal{H}_{\mathrm{O}}$ is the weighted sum of $\mathcal{H}\{S|T=t\}$ (see Equation (4)) and $\mathcal{H}\{S|T=t\}$ is non-negative. The entropy of conditional probability is zero means that $P(S=s|T=t)$ is a delta function, that is, for a given t, s has only one possible value. □

Theorem 2 *The under-segmentation entropy is bounded, for all possible S,*

$$0 \leqslant \mathcal{H}_{\mathrm{U}} \leqslant \mathcal{H}\{T\}. \tag{22}$$

Furthermore, $\mathcal{H}_{\mathrm{U}} = 0$ if and only if

$$\forall(x,y),(x',y'),S(x,y) = S(x',y') \Rightarrow T(x,y) = T(x',y'). \tag{23}$$

Proof The lower bound in Equation (22) is obvious. To prove the upper bound, we rewrite the conditional entropy as the following

$$\begin{aligned}\mathcal{H}_{\mathrm{U}} &= \mathcal{H}\{T\} - \mathcal{M}\{S,T\} && (24)\\ &\leqslant \mathcal{H}\{T\}. && (25)\end{aligned}$$

The equality holds when the mutual information $\mathcal{M}\{S,T\}$ is zero, that is, S and T are independent.

Equation (23) can be proved in the same way as Equation (17). □

From the above two theorems, we can see that both conditional entropies are zero (i.e., $\mathcal{H}_\mathrm{O} = 0$ and $\mathcal{H}_\mathrm{U} = 0$), if and only if S and T are equivalent under label number permutations.

3.2 Monotonicity

When we split a segment in the segmentation S into two, the over-segmentation entropy $\mathcal{H}_O$ increases, and the under-segmentation $\mathcal{H}_U$ decreases.

Theorem 3 *If we split a segment in S into two, a new label map S' is obtained. Let $\mathcal{H}'_\mathrm{O}$ be the OSE of S', $\mathcal{H}'_\mathrm{U}$ be its USE. Then we have*

$$\mathcal{H}'_\mathrm{O} \geqslant \mathcal{H}_\mathrm{O} \tag{26}$$

$$\mathcal{H}'_\mathrm{U} \leqslant \mathcal{H}_\mathrm{U}. \tag{27}$$

Proof Because S' is an refinement of S, $P(S|S')$ is a delta function, which implies that $\mathcal{H}\{S|S'\} = \mathcal{H}\{S|S', T\} = 0$. Now we prove the first inequality,

$$\begin{aligned}\mathcal{H}'_\mathrm{O} - \mathcal{H}_\mathrm{O} &= \mathcal{H}\{S'|T\} - \mathcal{H}\{S|T\} &(28)\\ &= \mathcal{H}\{S', T\} - \mathcal{H}\{S, T\} &(29)\\ &= \mathcal{H}\{S', S, T\} - \mathcal{H}\{S, T\} &(30)\\ &= \mathcal{H}\{S'|S, T\} \geqslant 0. &(31)\end{aligned}$$

Similarly, we can prove the second inequality,

$$\begin{aligned}\mathcal{H}'_\mathrm{U} - \mathcal{H}_\mathrm{U} &= \mathcal{H}\{T|S'\} - \mathcal{H}\{T|S\} &(32)\\ &= \mathcal{H}\{S'|S, T\} - \mathcal{H}\{S'|S\} &(33)\\ &= -\mathcal{M}\{S', T|S\} \leqslant 0 &(34)\end{aligned}$$

□

This theorem further confirms that the over-segmentation entropy really represent the degree of over-segmentation, and the under-segmentation entropy really represent the degree of under-segmentation.

Note that Theorem 3 can be described in an alternative way: when merging two segments in the segmentation, OSE decreases and USE increases. This description is helpful in the proof of continuity in the next subsection.

3.3 Continuity

The two entropies are insensitive to boundary perturbation. This point can also be proven theoretically, by considering one pixel perturbation of the segmentation.

Theorem 4 *If we change the label for one pixel in S, a new label map S' is obtained. Let $\mathcal{H}'_\mathrm{O}$ and $\mathcal{H}'_\mathrm{U}$ be the OSE and USE of S' respectively, and $\Delta\mathcal{H}_\mathrm{O} = \mathcal{H}'_\mathrm{O} - \mathcal{H}_\mathrm{O}$ and $\Delta\mathcal{H}_\mathrm{U} = \mathcal{H}'_\mathrm{U} - \mathcal{H}_\mathrm{U}$ be the change. Then we have*

$$\lim_{|\Lambda|\to\infty} \Delta\mathcal{H}_\mathrm{O} = 0. \tag{35}$$

$$\lim_{|\Lambda|\to\infty} \Delta\mathcal{H}_\mathrm{U} = 0. \tag{36}$$

Proof We first prove the situation where the pixel is assigned a new label to become an isolated segment, or an isolated pixel is merged into an another segment. From the proof of Theorem 3, we have $\Delta\mathcal{H}_\mathrm{O} = \mathcal{H}\{S'|S, T\}$ and $\Delta\mathcal{H}_\mathrm{U} = \mathcal{H}\{S'|S, T\} - \mathcal{H}\{S'|S\}$. When $\frac{1}{|\Lambda|} \to 0$, $P(S'|S)$ approaches delta function, and therefore both $\mathcal{H}\{S'|S, T\}$ and $\mathcal{H}\{S'|S\}$ approach zero. So we have $\Delta\mathcal{H}_\mathrm{O} \to 0$ and $\Delta\mathcal{H}_\mathrm{U} \to 0$.

For the situation where the pixel is switched to an existed label, we regard it as two steps: first assigning a new label to create an isolated segment, then merging this isolated pixel to another segment. Then the entropy increments in both steps approach zero when $\frac{1}{|\Lambda|} \to 0$. □

This property ensures that small change in S only results in small change in $\mathcal{H}_O$ and $\mathcal{H}_U$. That is, our measures have good numerical stability.

3.4 GCE and Conditional Entropies

In (Martin et al, 2001), the Global Consistency Error (GCE) is defined in the following way. If $R(S, i)$ is the segment in segmentation S that contains the i-th pixel, the local refinement error from S to T is defined as

$$E(S, T, i) = 1 - \frac{|R(S, i) \cap R(T, i)|}{|R(T, i)|}, \tag{37}$$

and GCE is defined as

$$\mathrm{GCE}(S, T) = \min\left\{\frac{1}{|\Lambda|}\sum_i E(S, T, i),\ \frac{1}{|\Lambda|}\sum_i E(T, S, i)\right\}. \tag{38}$$

In fact, $E(S, T, i)$ is a linear monotonically decreasing function of $P(S|T)$ because in our notation, $\frac{|R(S,i)\cap R(T,i)|}{|R(T,i)|} = P(S = s_i|T = t_i)$, where s_i denotes the label of the i-th pixel in S and t_i the label of the i-th pixel in T. If we choose an alternative monotonically decreasing function, and define E as

$$E(S, T, i) = -\log\frac{|R(S, i) \cap R(T, i)|}{|R(T, i)|}, \tag{39}$$

then our new GCE is

$$\mathrm{GCE}(S,T) = \min\left\{-\frac{1}{|A|}\sum_i \log P(S|T),\ -\frac{1}{|A|}\sum_i \log P(T|S)\right\} \quad (40)$$

$$= \min\left\{\mathcal{H}_O,\ \mathcal{H}_U\right\}. \quad (41)$$

Now, one can see that GCE differs from VI in two ways: it chooses a different penalty function, and a different way to combine the two types of errors.

4 Experiments

We test the conditional entropy metrics on Berkeley Image Segmentation Database(Martin et al, 2001), using three segmentation algorithms, the efficient graph based segmentation (EGB) (Felzenszwalb and Huttenlocher, 2004), Mean Shift(Comanicu and Meer, 2002) (MS) and an accelerated version of Normalized Cut (NCUT) (Shi and Malik, 2000) based on Constrained Delaunay Triangulation(Seidel, 1988; Wu and Yu, 2003). The database contains 300 images, and for each image, about 5 ground truths are provided. For each algorithm and each image, we use $18 \sim 20$ sets of parameters to obtain segmentations at varying granularities. The over-segmentation and under-segmentation entropies are computed for all these segmentations, with respect to all the ground truths. The results are shown in the following subsections.

4.1 Multi-scale segmentation

First, we show an example of conditional entropy curve of multiple segmentations of a image with respect to a single ground truth, in Figure 2. The conditional entropy curves look similar to an upside-down ROC curve. One can see that for the two algorithms illustrated in the figure, the conditional entropy curves reveal the trade-off of over-segmentation and under-segmentation while granularity changes. In this illustration, the NCUT is better than EGB, because at a reasonable range of OSE $[0, 3]$, given the same OSE, NCUT achieve smaller USE. This conclusion can be verified by the segments superimposed on the curves. Note that the over-segmentation entropy has larger domain because it is bounded by the number of pixels.

Figure 3 shows the changes of the conditional entropy curves with respect to ground truths. For similar ground truths, the curves are similar. For ground truths that differ a lot, the curves also differ a lot. For ground truths with more segments, the curves rise higher at left part.

Figure 4 shows the mean conditional entropy curves with respect to multiple ground truths. For each segment, we compute the over-segmentation and under-segmentation entropies with respect to all ground truths, and use the mean of over-segmentation entropy as horizontal axis, and the mean of under-segmentation entropy as vertical axis. One can see that the mean conditional entropy curve successfully reveals the image difficulties and the relative performances of the algorithms on different images.

4.2 Algorithm performance analysis

We compute the overall performance of all the three algorithms on the whole dataset. First, we compute the mean conditional entropies of each image with respect to all ground truths. Then we average the mean conditional entropy curve to obtain an overall conditional entropy curve. Figure 5 shows the results. One can see that the conditional entropy curve reveals that NCUT is better with a smaller number of segments, EGB is better at superpixel level, and Mean Shift is good trade-off of the two. This agrees with our experiences of using these algorithms. The metrics in (Unnikrishnan et al, 2007)(Meilă, 2005) cannot reveal this point.

4.3 Comparison with Variation of Information

To further illustrate the benefits of the two conditional entropies, we define a new similarity measure based on the two, and compare it with Variation of Information(Meilă, 2005)(Meilă, 2007). For general application tasks, we have bias over the two types of errors. That is, over-segmentation is easier to remedy by further processing, and thus should be discounted. The human annotators trend to split an image into a smaller number of segments, and the real ground truth may be a refinement of the annotated one. Therefore, we define the following biased variation of information.

Definition 5 Biased variation of information (BVI) and mean biased variation of information (MBVI) are defined by

$$\mathcal{B}\{S,T\} = w_O \mathrm{Sat}(\mathcal{H}_\mathrm{O} - b_O) + w_U \mathrm{Sat}(\mathcal{H}_\mathrm{U} - b_U),$$
$$\bar{\mathcal{B}}\{S,T_1,\cdots,T_L\} = w_O \mathrm{Sat}(\bar{\mathcal{H}}_\mathrm{O} - b_O) + w_U \mathrm{Sat}(\bar{\mathcal{H}}_\mathrm{U} - b_U),$$

where $\mathrm{Sat}(x)$ is a bottom saturation function

$$\mathrm{Sat}(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leqslant 0 \end{cases}, \quad (42)$$

w_O and w_U are the weights to balance the two term, and b_O and b_U are used to further tolerate small errors. This is an asymmetric similarity.

Note that this is an asymmetric similarity.

We use both MBVI and MVI to select the best segmentation for each image among all possible segmentations at different granularities produced by different algorithms. For multiple ground truths, we compute the mean of MVI. We use $w_O = 0.8$, $w_U = 1$, $b_O = 0.05$ and $b_U = 0$ to reflect our preference to over-segmentation error. We find that

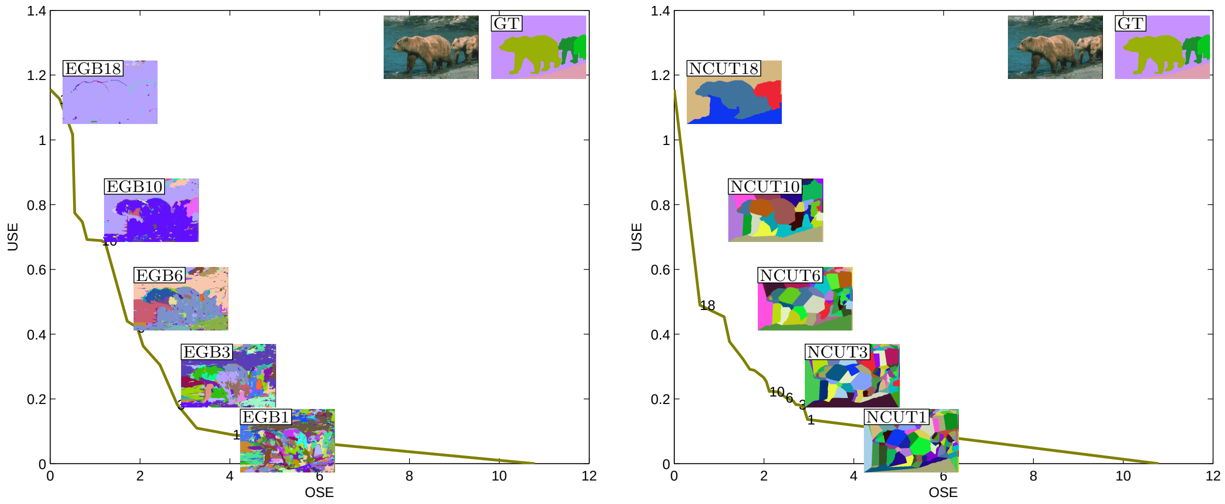


Fig. 2 Examples of conditional entropies curve with respect to one ground truth: left, EGB algorithm; right, NCUT algorithm. Horizontal axis is OSE; vertical axis is USE. The original image and the ground truth are shown on the top right corner. Segmentations at varying scales are shown along the curves. One can see that the curves reflect the trade-off between the two entropies as segmentation granularity changes.

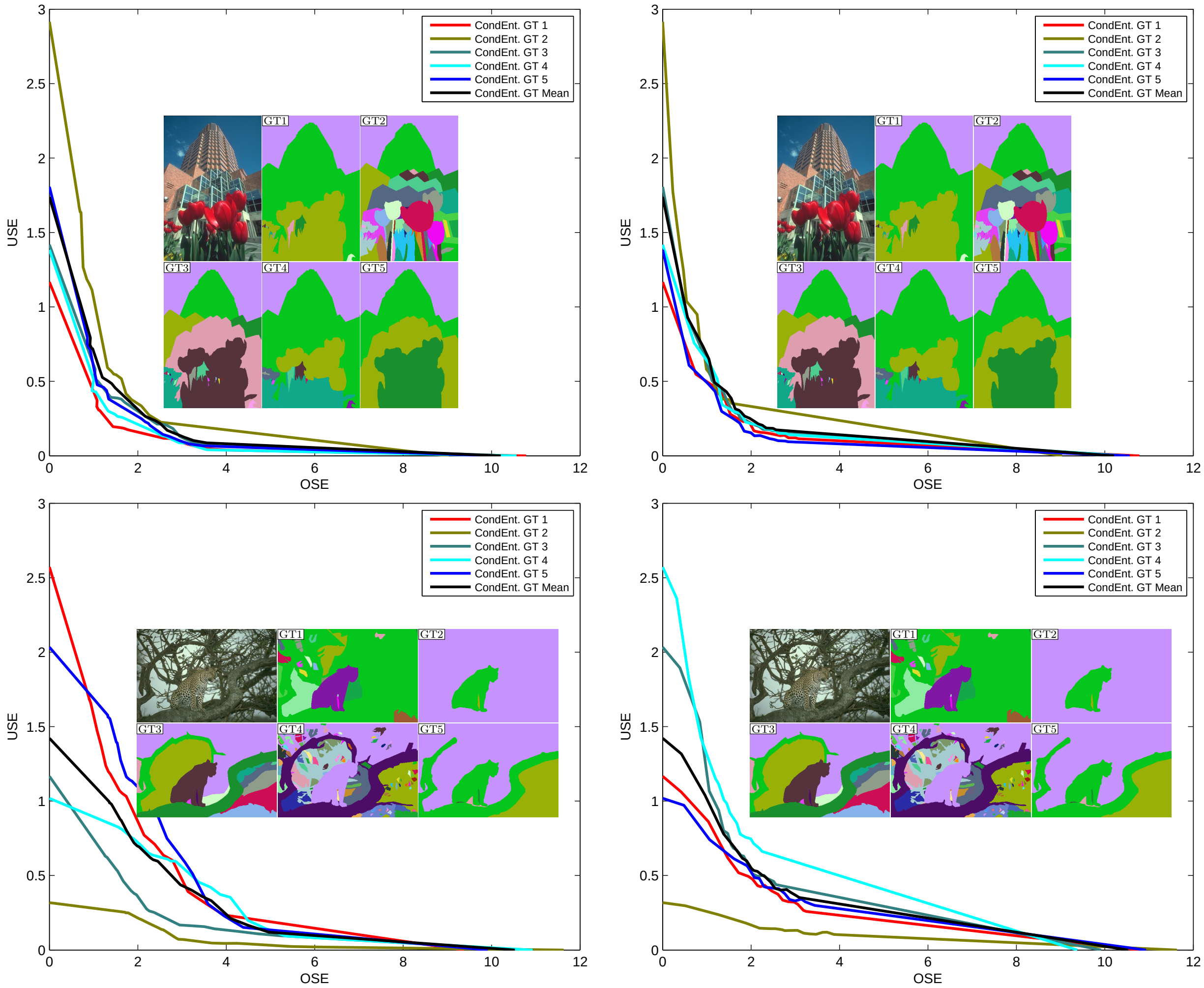


Fig. 3 The conditional entropy curves for an image with respect to varying ground truths. Left column, the results of EGB. Right column, the results of NCUT. The images and the ground truths are shown in the middle of each axis. 1st row, ground truths differ moderately, and the curves also differ moderately. 2nd row, ground truths differ considerably, and the curves also differ considerably.

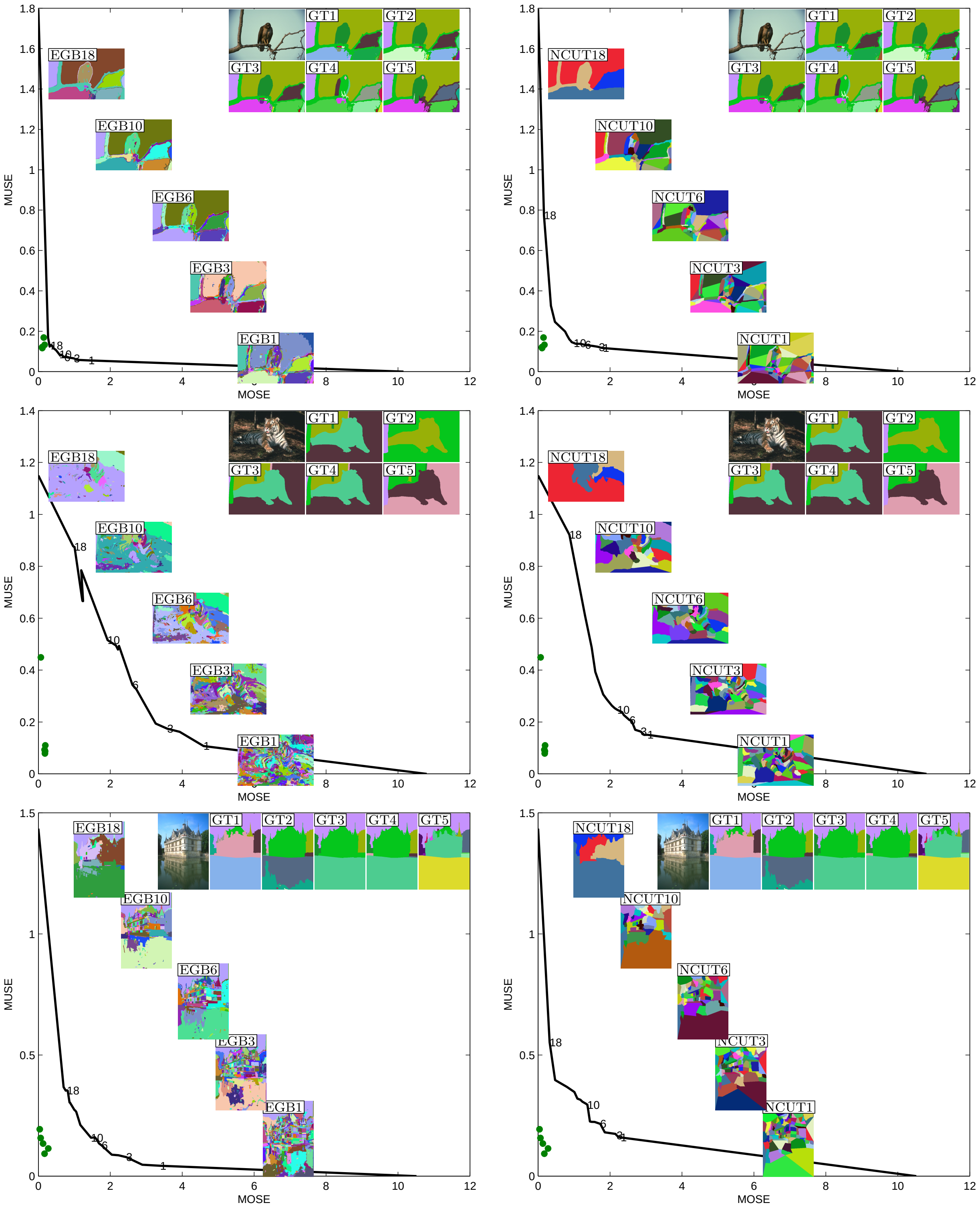


Fig. 4 The mean conditional entropy curves for an image with multiple ground truths. Each row shows the results of an individual image. Left column is the results of EGB; right column is the results of NCUT. Horizontal axis is MOSE; vertical axis is MUSE. The original image and the ground truths are shown on the top right corner. Segmentations at varying scales are shown along the curves. Green dots are the MOSE and MUSE of ground truths. 1st row, an image of low difficulty, EGB achieves the perfect result (EGB18). 2nd row, an image of high difficulty, NCUT works better than EGB, because at a low OSE, it has lower corresponding USE. 3rd row, an image of medium difficulty, EGB is better at over-segmentation and NCUT is better at under-segmentation. One can see that the conditional entropies reflect the difficulties of the three images well, in a sense similar to an ROC curve.

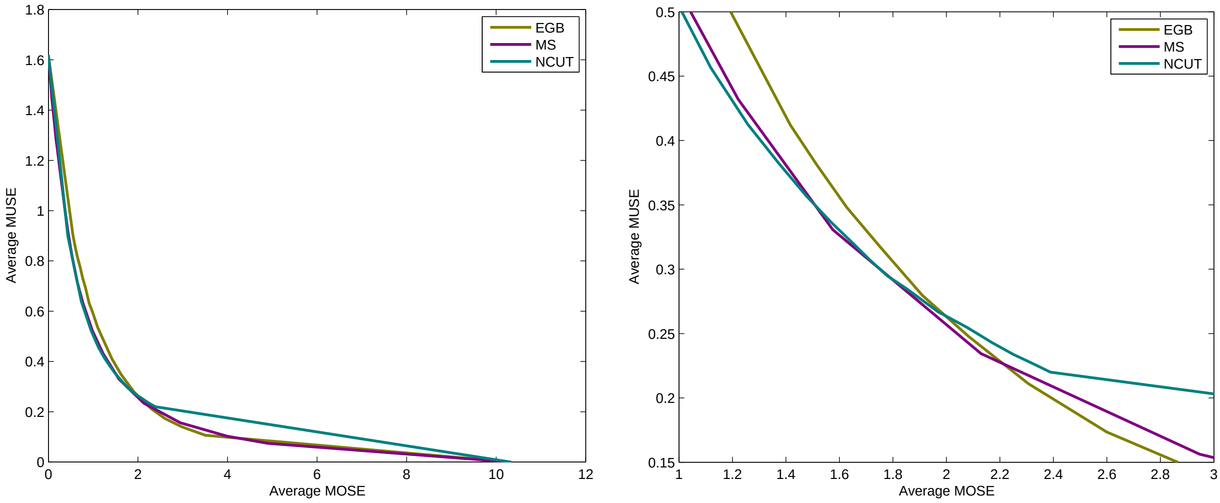


Fig. 5 Comparison of three algorithms — EGB, Mean Shift and NCUT using conditional entropy curves. Left, the mean conditional entropy curves of the three algorithms over the whole dataset. Right, the zoomed version of the left column. EGB is the best at high over-segmentation domain, which indicates that it is good at superpixel. NCUT is the best at high under-segmentation domain, which indicates that it is good at object level segmentation. Mean Shift is a good trade-off of the two types of errors.

MBVI works consistently well and Figure 6 shows some of the results. MVI sometimes chooses degenerate segmentation, because the range of MUSE is much smaller than that of MOSE. From Figure 5, one can see that the maximum of average MUSE is about 1.6, but the maximum of average MOSE is about 10. This also suggests that it may not be a good idea to combine them with equal weights. Figure 6 also shows that BVI can tolerate occasional bad ground truths.

4.4 Coarse-to-fine Compatibility

Let $S_1, \cdots, S_L$ be the segmentations given by an algorithm at changing granularities, from smaller segments to larger ones. Now we consider the conditional entropies between each pair of them. Ideally, if $i \geqslant j$, $\mathcal{H}\{S_i|S_j\} = 0$, otherwise, $\mathcal{H}\{S_i|S_j\}$ increases monotonically as j increases from $j = i$, and decreases monotonically as i increases until $i = j$. That is to say, if we let $M_{i,j} = \mathcal{H}\{S_i|S_j\}$ be a matrix, all its entries above the diagonal are zeros, and below the diagonal, entry values increases from top to bottom and right to left. We call this property coarse-to-fine compatibility. We compute the coarse-to-fine matrices of NCUT on all images in Berkeley Image Segmentation Benchmark, and find that almost all images demonstrate good coarse-to-fine compatibilities. Figure 7 $\sim$ 10 show 4 examples. For EBS and Mean-Shift, similar results are observed.

We also conjectured that ground truths of those images given by different annotators demonstrate the same good properties. Unfortunately, this is only true for a small number of them. Figure 11 $\sim$ 12 show the matrices for the 4 images used in Figure 7 $\sim$ 10. Figure 11 shows two examples that have good coarse-to-fine compatibilities. Figure 12 shows two examples that have bad coarse-to-fine compatibilities. The reason is that some annotators prefer more details on foreground objects while some prefer more details on backgrounds.

Note that none of F-measure, VI and NPR can reveal these properties of segmentations and ground truths.

5 Conclusion

In this paper, we proposed the use of the two conditional entropies between a segmentation and a ground truth as metrics of image segmentation. The two conditional entropies represent the degrees of over-segmentation and under-segmentation separately, and therefore successfully reveal the performance of algorithms at difference granularities. By combining the two in a biased way, we can also show our preference to over-segmentation or under-segmentation.

References

Comanicu D, Meer P (2002) Mean shift: A robust approach toward feature space analysis. PAMI 22(5):603–619

Estrada FJ, Jepson AD (2009) Benchmarking image segmentation algorithms. IJCV 85:167–181

Everingham M, van Gool LJ, Williams CKI, Winn JM, Zisserman A (2010) The pascal visual object classes (voc) challenge. IJCV 88(2):303–338

Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. IJCV 59(2)

Ge F, Wang S, Liu T (2007) New benchmark for image segmentation evaluation. Journal of Electronic Imaging 16(3)

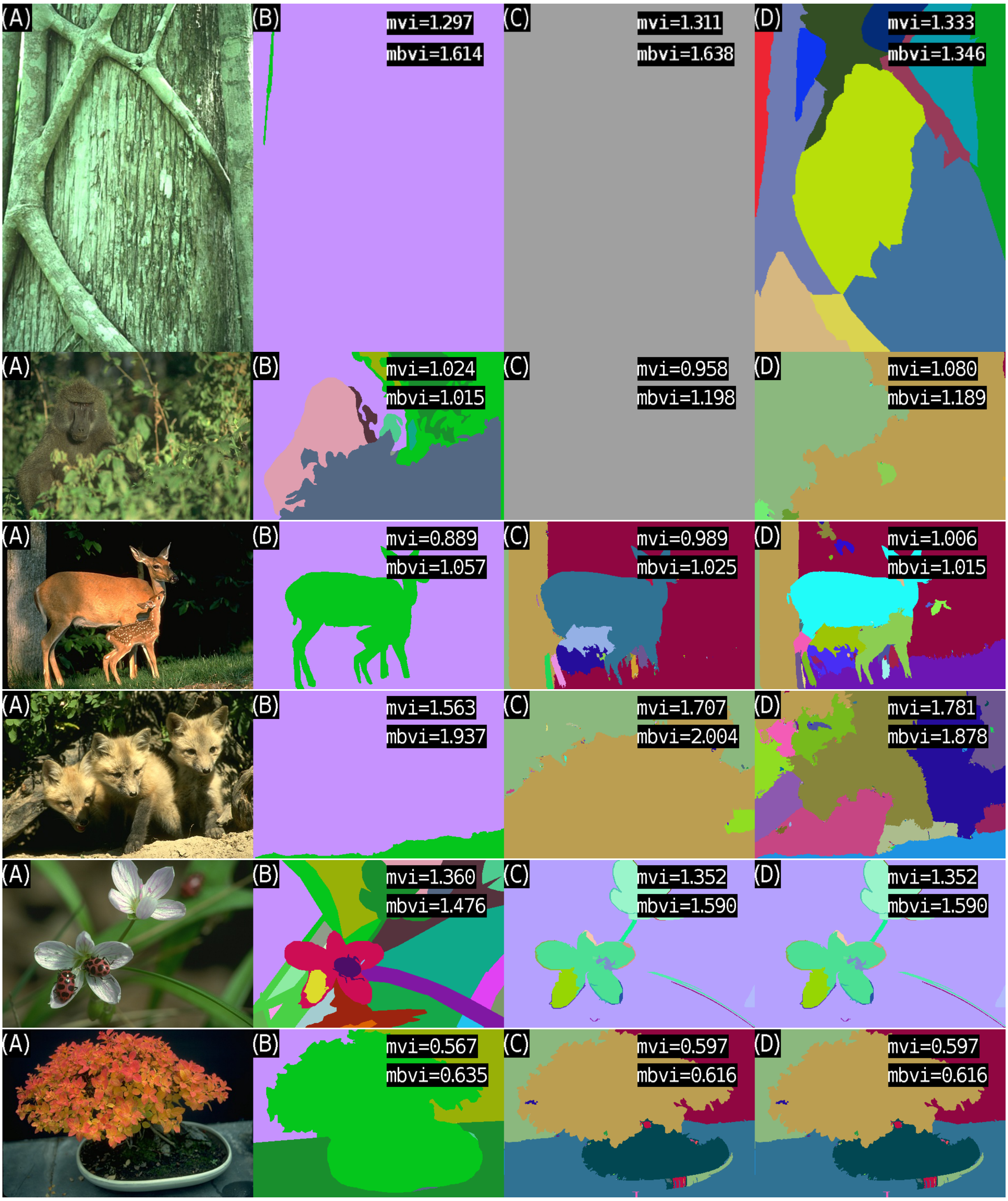
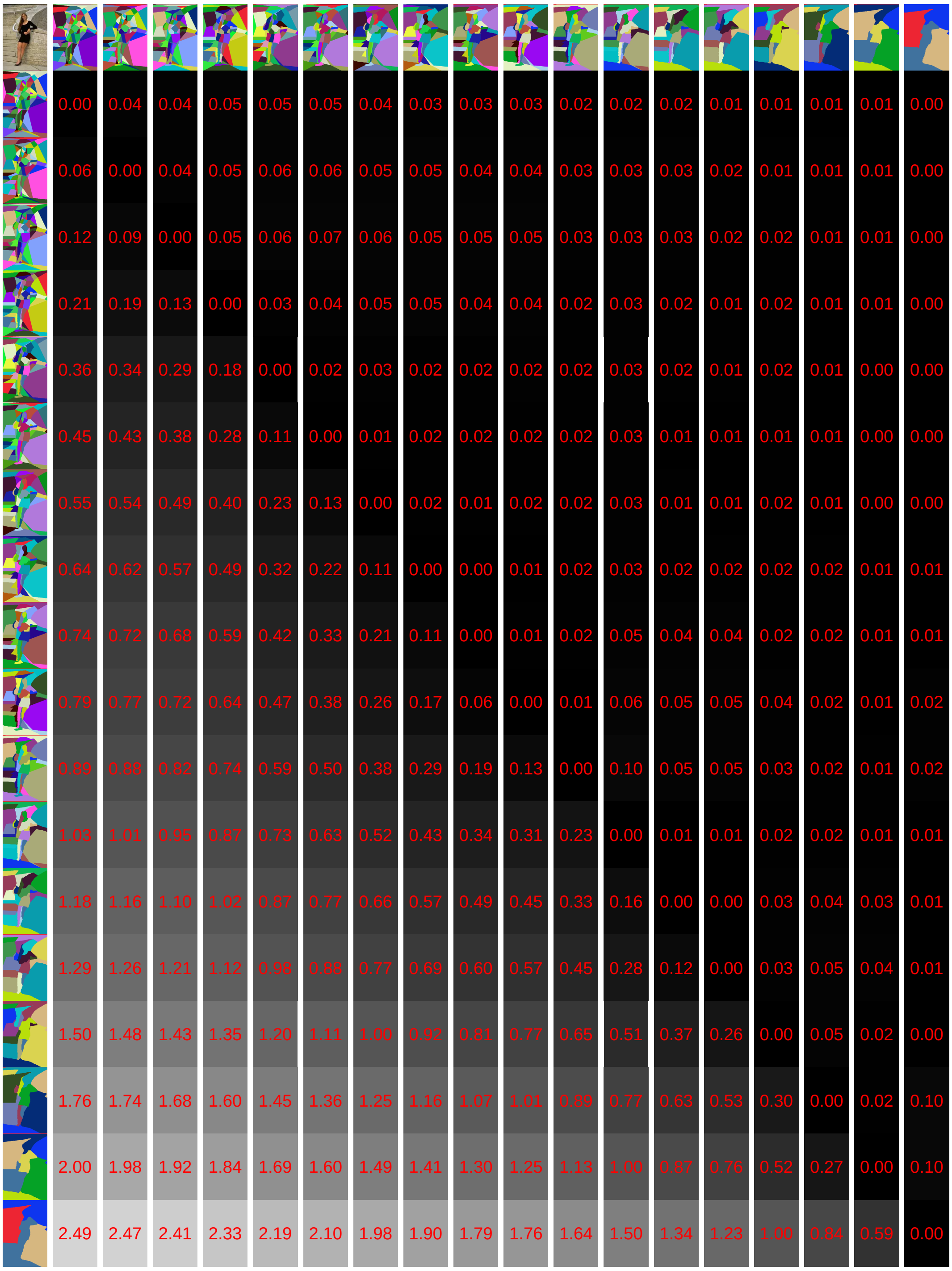


Fig. 6 Comparison of MBVI and MVI. (A) the original image. (B) the worst ground truth chosen by MVI. (C) the best segmentation chosen by MVI. (D) the best segmentation chosen by MBVI. Row 1 $\sim$ 4, MBVI helps to choose better segmentation. Row 5 and 6, MVI and MBVI choose the same segmentations, but MVI gives more reasonable scores. Row 1, MVI chooses the degenerate segmentation as the best one, because of one of the ground truth is not good enough. MBVI choose (D) as the best segmentation, which is more reasonable. Row 2, MVI chooses the degenerate segmentation as the best one, and indicates that the ground truth (B) is worse than segmentation (C) and (D). Row 3, MBVI helps choose a slightly better segmentation. Row 4, MBVI helps choose a more reasonable segmentation. Row 5, MVI and MBVI choose the same best segmentation, but the MVI value of (C) is better than the ground truth (B), which is not reasonable. The MBVI indicates that the ground truth (B) is better than the segmentation (C). Row 6, MVI and MBVI choose the same best segmentation. The MBVI correctly indicates that the almost perfect segmentation (C) is slightly better than the ground truth (B).



0.00	0.04	0.04	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00
0.06	0.00	0.04	0.05	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.00
0.12	0.09	0.00	0.05	0.06	0.07	0.06	0.05	0.05	0.05	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.00
0.21	0.19	0.13	0.00	0.03	0.04	0.05	0.05	0.04	0.04	0.02	0.03	0.02	0.01	0.02	0.01	0.01	0.00
0.36	0.34	0.29	0.18	0.00	0.02	0.03	0.02	0.02	0.02	0.02	0.03	0.02	0.01	0.02	0.01	0.00	0.00
0.45	0.43	0.38	0.28	0.11	0.00	0.01	0.02	0.02	0.02	0.02	0.03	0.01	0.01	0.01	0.01	0.00	0.00
0.55	0.54	0.49	0.40	0.23	0.13	0.00	0.02	0.01	0.02	0.02	0.03	0.01	0.01	0.02	0.01	0.00	0.00
0.64	0.62	0.57	0.49	0.32	0.22	0.11	0.00	0.00	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.01	0.01
0.74	0.72	0.68	0.59	0.42	0.33	0.21	0.11	0.00	0.01	0.02	0.05	0.04	0.04	0.02	0.02	0.01	0.01
0.79	0.77	0.72	0.64	0.47	0.38	0.26	0.17	0.06	0.00	0.01	0.06	0.05	0.05	0.04	0.02	0.01	0.02
0.89	0.88	0.82	0.74	0.59	0.50	0.38	0.29	0.19	0.13	0.00	0.10	0.05	0.05	0.03	0.02	0.01	0.02
1.03	1.01	0.95	0.87	0.73	0.63	0.52	0.43	0.34	0.31	0.23	0.00	0.01	0.01	0.02	0.02	0.01	0.01
1.18	1.16	1.10	1.02	0.87	0.77	0.66	0.57	0.49	0.45	0.33	0.16	0.00	0.00	0.03	0.04	0.03	0.01
1.29	1.26	1.21	1.12	0.98	0.88	0.77	0.69	0.60	0.57	0.45	0.28	0.12	0.00	0.03	0.05	0.04	0.01
1.50	1.48	1.43	1.35	1.20	1.11	1.00	0.92	0.81	0.77	0.65	0.51	0.37	0.26	0.00	0.05	0.02	0.00
1.76	1.74	1.68	1.60	1.45	1.36	1.25	1.16	1.07	1.01	0.89	0.77	0.63	0.53	0.30	0.00	0.02	0.10
2.00	1.98	1.92	1.84	1.69	1.60	1.49	1.41	1.30	1.25	1.13	1.00	0.87	0.76	0.52	0.27	0.00	0.10
2.49	2.47	2.41	2.33	2.19	2.10	1.98	1.90	1.79	1.76	1.64	1.50	1.34	1.23	1.00	0.84	0.59	0.00

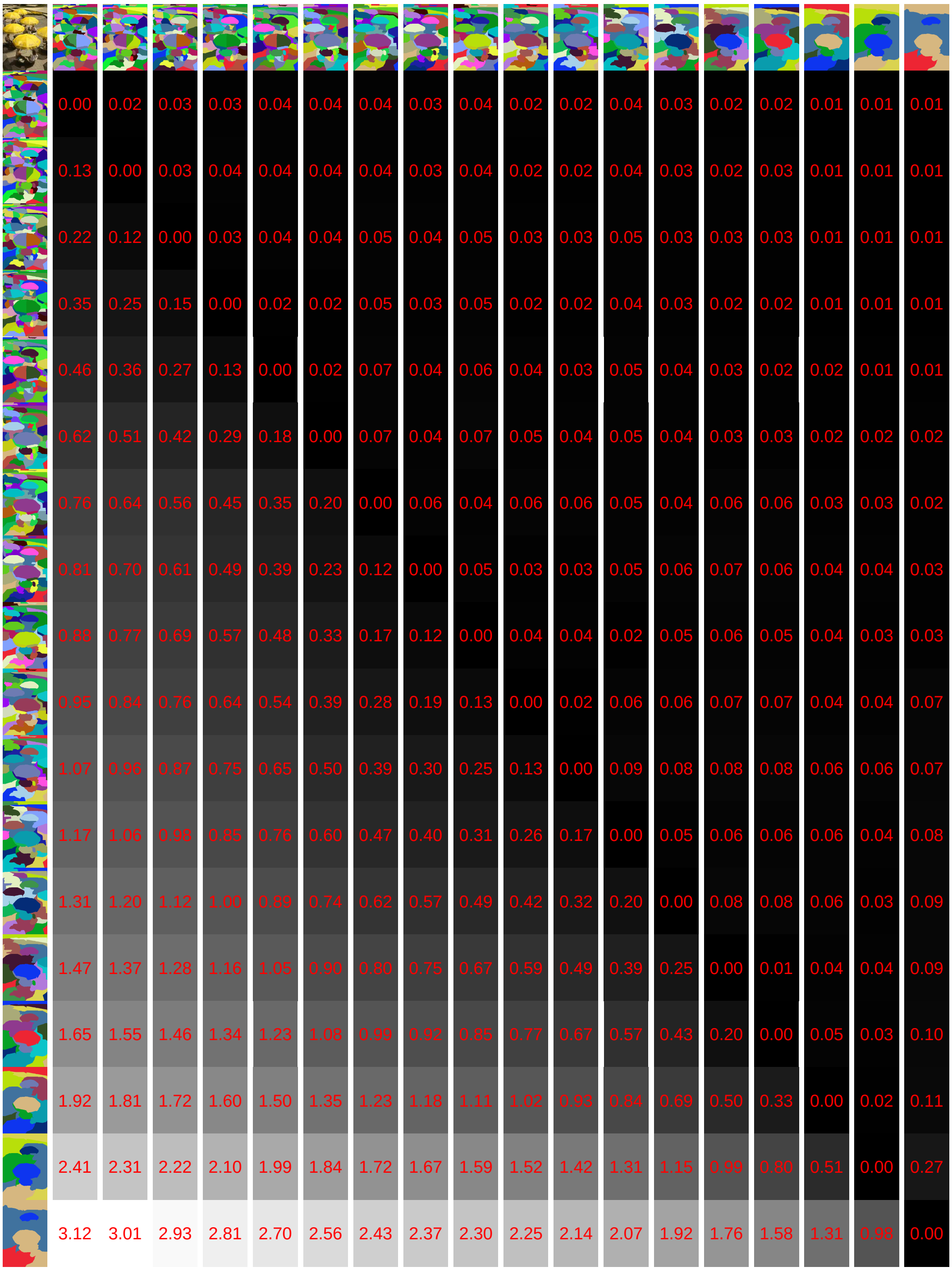
Fig. 7 Conditional entropies between segmentations at different granularities, which demonstrate good coarse-to-fine compatibilities. Top left corner is the image. First row and first column are segmentations at varying granularities.

Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV, vol 2, pp 416–423

Meilă M (2005) Comparing clusterings: An axomitic view. In: ICML

Meilă M (2007) Comparing clusterings — an information based distance. Journal of Multivariate Analysis 98:873–895

Seidel R (1988) Constrained delaunay triangulations and voronoi diagrams with obstacles. Tech. Rep. 260, Inst. for Information Processing, Graz, Austria



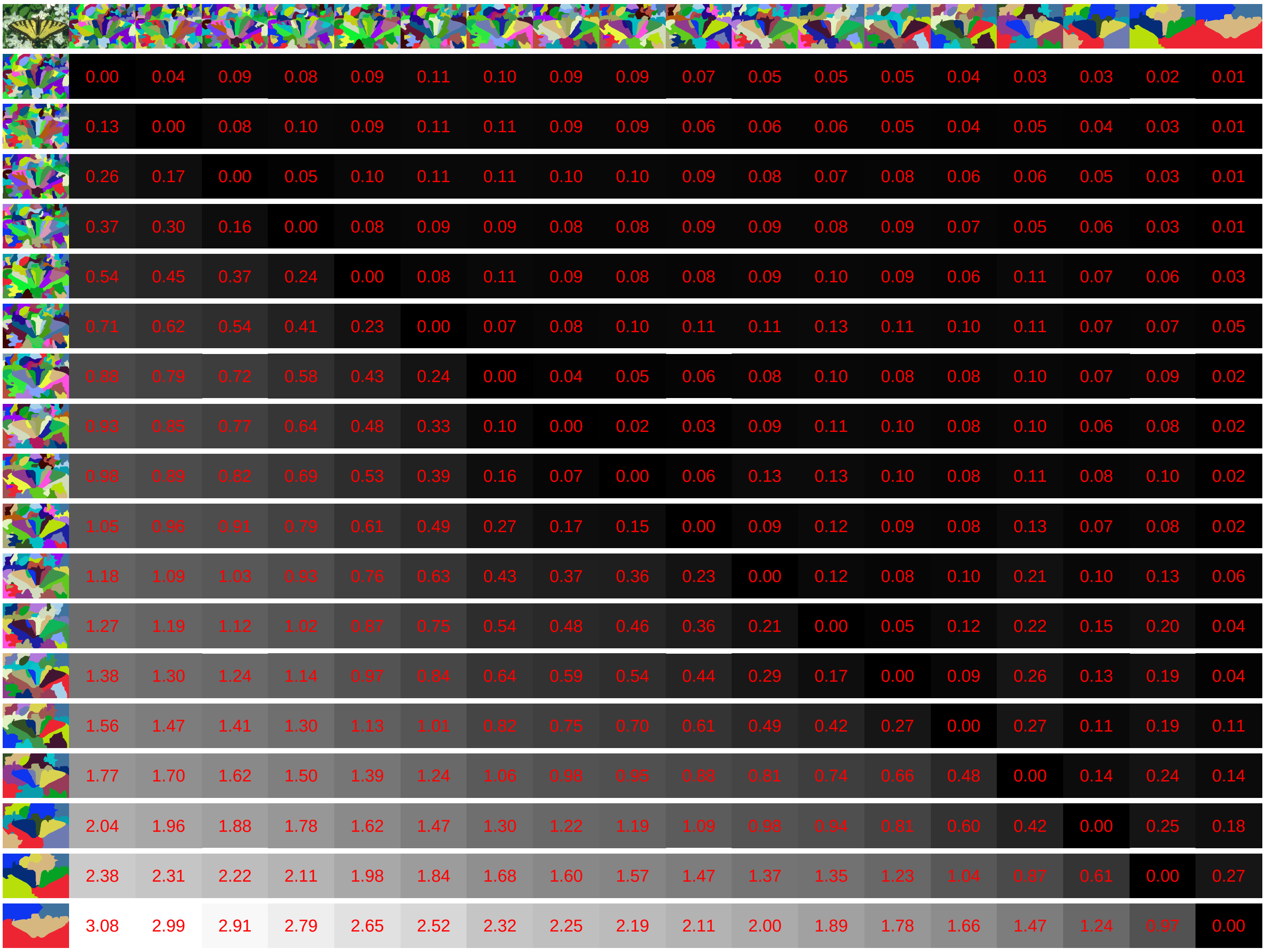
0.00	0.02	0.03	0.03	0.04	0.04	0.04	0.03	0.04	0.02	0.02	0.04	0.03	0.02	0.02	0.01	0.01	0.01
0.13	0.00	0.03	0.04	0.04	0.04	0.04	0.03	0.04	0.02	0.02	0.04	0.03	0.02	0.03	0.01	0.01	0.01
0.22	0.12	0.00	0.03	0.04	0.04	0.05	0.04	0.05	0.03	0.03	0.05	0.03	0.03	0.03	0.01	0.01	0.01
0.35	0.25	0.15	0.00	0.02	0.02	0.05	0.03	0.05	0.02	0.02	0.04	0.03	0.02	0.02	0.01	0.01	0.01
0.46	0.36	0.27	0.13	0.00	0.02	0.07	0.04	0.06	0.04	0.03	0.05	0.04	0.03	0.02	0.02	0.01	0.01
0.62	0.51	0.42	0.29	0.18	0.00	0.07	0.04	0.07	0.05	0.04	0.05	0.04	0.03	0.03	0.02	0.02	0.02
0.76	0.64	0.56	0.45	0.35	0.20	0.00	0.06	0.04	0.06	0.06	0.05	0.04	0.06	0.06	0.03	0.03	0.02
0.81	0.70	0.61	0.49	0.39	0.23	0.12	0.00	0.05	0.03	0.03	0.05	0.06	0.07	0.06	0.04	0.04	0.03
0.88	0.77	0.69	0.57	0.48	0.33	0.17	0.12	0.00	0.04	0.04	0.02	0.05	0.06	0.05	0.04	0.03	0.03
0.95	0.84	0.76	0.64	0.54	0.39	0.28	0.19	0.13	0.00	0.02	0.06	0.06	0.07	0.07	0.04	0.04	0.07
1.07	0.96	0.87	0.75	0.65	0.50	0.39	0.30	0.25	0.13	0.00	0.09	0.08	0.08	0.08	0.06	0.06	0.07
1.17	1.06	0.98	0.85	0.76	0.60	0.47	0.40	0.31	0.26	0.17	0.00	0.05	0.06	0.06	0.06	0.04	0.08
1.31	1.20	1.12	1.00	0.89	0.74	0.62	0.57	0.49	0.42	0.32	0.20	0.00	0.08	0.08	0.06	0.03	0.09
1.47	1.37	1.28	1.16	1.05	0.90	0.80	0.75	0.67	0.59	0.49	0.39	0.25	0.00	0.01	0.04	0.04	0.09
1.65	1.55	1.46	1.34	1.23	1.08	0.99	0.92	0.85	0.77	0.67	0.57	0.43	0.20	0.00	0.05	0.03	0.10
1.92	1.81	1.72	1.60	1.50	1.35	1.23	1.18	1.11	1.02	0.93	0.84	0.69	0.50	0.33	0.00	0.02	0.11
2.41	2.31	2.22	2.10	1.99	1.84	1.72	1.67	1.59	1.52	1.42	1.31	1.15	0.99	0.80	0.51	0.00	0.27
3.12	3.01	2.93	2.81	2.70	2.56	2.43	2.37	2.30	2.25	2.14	2.07	1.92	1.76	1.58	1.31	0.98	0.00

Fig. 8 Conditional entropies between segmentations at different granularities, which demonstrate good coarse-to-fine compatibilities.

Shi J, Malik J (2000) Normalized cuts and image segmentation. PAMI 22(8)

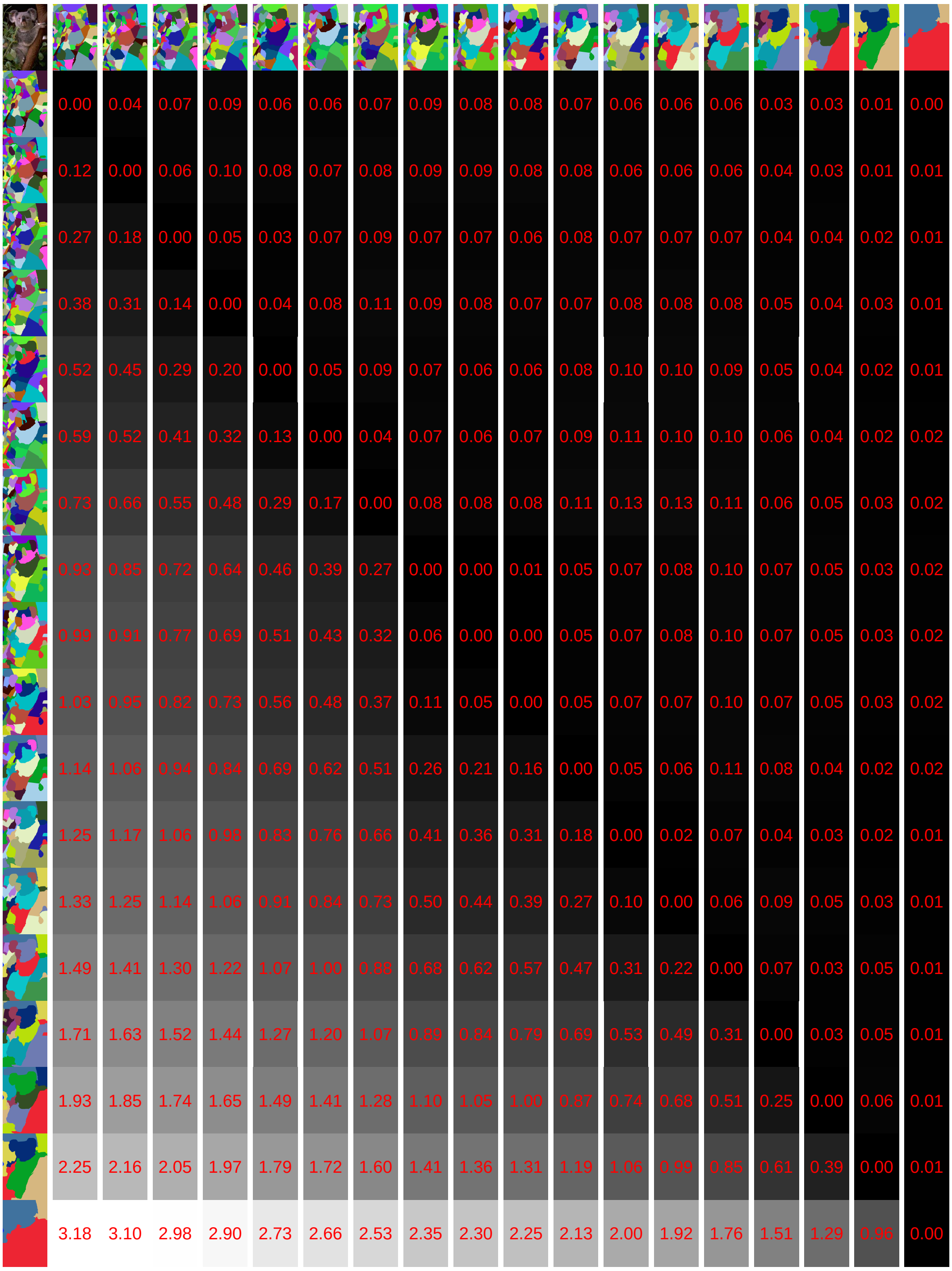
Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. PAMI 29(6):929–944

Wu Q, Yu Y (2003) Two-level image segmentation based on region andedgeintegration. In: Proc. of DICTA



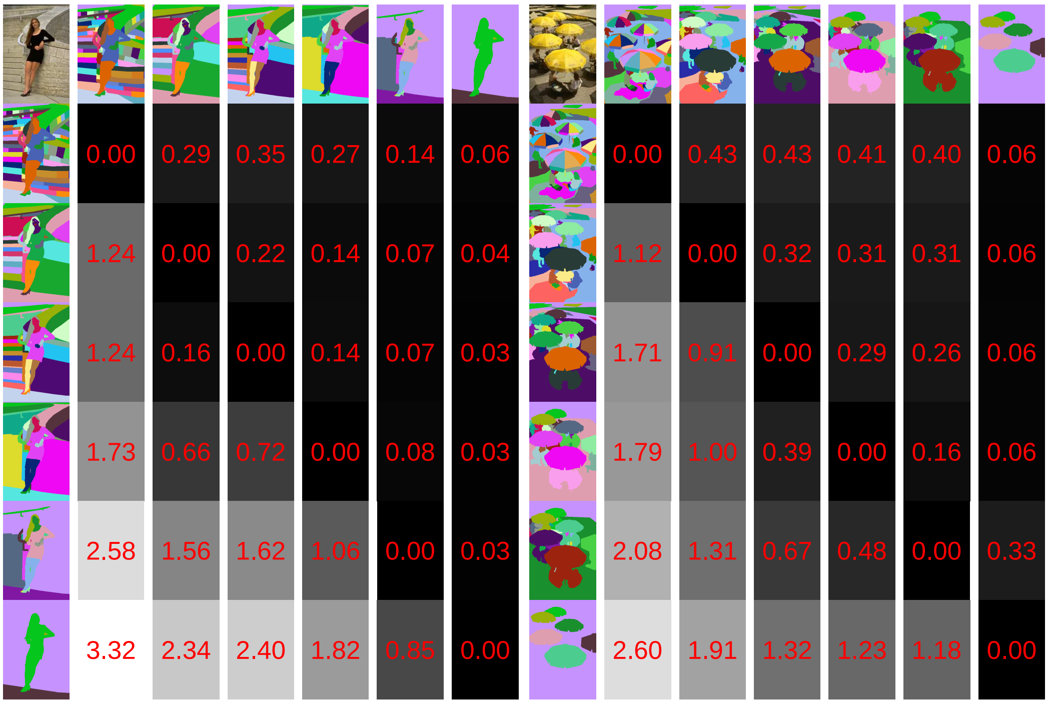
0.00	0.04	0.09	0.08	0.09	0.11	0.10	0.09	0.09	0.07	0.05	0.05	0.05	0.04	0.03	0.03	0.02	0.01
0.13	0.00	0.08	0.10	0.09	0.11	0.11	0.09	0.09	0.06	0.06	0.06	0.05	0.04	0.05	0.04	0.03	0.01
0.26	0.17	0.00	0.05	0.10	0.11	0.11	0.10	0.10	0.09	0.08	0.07	0.08	0.06	0.06	0.05	0.03	0.01
0.37	0.30	0.16	0.00	0.08	0.09	0.09	0.08	0.08	0.09	0.09	0.08	0.09	0.07	0.05	0.06	0.03	0.01
0.54	0.45	0.37	0.24	0.00	0.08	0.11	0.09	0.08	0.08	0.09	0.10	0.09	0.06	0.11	0.07	0.06	0.03
0.71	0.62	0.54	0.41	0.23	0.00	0.07	0.08	0.10	0.11	0.11	0.13	0.11	0.10	0.11	0.07	0.07	0.05
0.88	0.79	0.72	0.58	0.43	0.24	0.00	0.04	0.05	0.06	0.08	0.10	0.08	0.08	0.10	0.07	0.09	0.02
0.93	0.85	0.77	0.64	0.48	0.33	0.10	0.00	0.02	0.03	0.09	0.11	0.10	0.08	0.10	0.06	0.08	0.02
0.98	0.89	0.82	0.69	0.53	0.39	0.16	0.07	0.00	0.06	0.13	0.13	0.10	0.08	0.11	0.08	0.10	0.02
1.05	0.96	0.91	0.79	0.61	0.49	0.27	0.17	0.15	0.00	0.09	0.12	0.09	0.08	0.13	0.07	0.08	0.02
1.18	1.09	1.03	0.93	0.76	0.63	0.43	0.37	0.36	0.23	0.00	0.12	0.08	0.10	0.21	0.10	0.13	0.06
1.27	1.19	1.12	1.02	0.87	0.75	0.54	0.48	0.46	0.36	0.21	0.00	0.05	0.12	0.22	0.15	0.20	0.04
1.38	1.30	1.24	1.14	0.97	0.84	0.64	0.59	0.54	0.44	0.29	0.17	0.00	0.09	0.26	0.13	0.19	0.04
1.56	1.47	1.41	1.30	1.13	1.01	0.82	0.75	0.70	0.61	0.49	0.42	0.27	0.00	0.27	0.11	0.19	0.11
1.77	1.70	1.62	1.50	1.39	1.24	1.06	0.98	0.95	0.88	0.81	0.74	0.66	0.48	0.00	0.14	0.24	0.14
2.04	1.96	1.88	1.78	1.62	1.47	1.30	1.22	1.19	1.09	0.98	0.94	0.81	0.60	0.42	0.00	0.25	0.18
2.38	2.31	2.22	2.11	1.98	1.84	1.68	1.60	1.57	1.47	1.37	1.35	1.23	1.04	0.87	0.61	0.00	0.27
3.08	2.99	2.91	2.79	2.65	2.52	2.32	2.25	2.19	2.11	2.00	1.89	1.78	1.66	1.47	1.24	0.97	0.00

Fig. 9 Conditional entropies between segmentations at different granularities, which demonstrate good coarse-to-fine compatibilities.



0.00	0.04	0.07	0.09	0.06	0.06	0.07	0.09	0.08	0.08	0.07	0.06	0.06	0.06	0.03	0.03	0.01	0.00
0.12	0.00	0.06	0.10	0.08	0.07	0.08	0.09	0.09	0.08	0.08	0.06	0.06	0.06	0.04	0.03	0.01	0.01
0.27	0.18	0.00	0.05	0.03	0.07	0.09	0.07	0.07	0.06	0.08	0.07	0.07	0.07	0.04	0.04	0.02	0.01
0.38	0.31	0.14	0.00	0.04	0.08	0.11	0.09	0.08	0.07	0.07	0.08	0.08	0.08	0.05	0.04	0.03	0.01
0.52	0.45	0.29	0.20	0.00	0.05	0.09	0.07	0.06	0.06	0.08	0.10	0.10	0.09	0.05	0.04	0.02	0.01
0.59	0.52	0.41	0.32	0.13	0.00	0.04	0.07	0.06	0.07	0.09	0.11	0.10	0.10	0.06	0.04	0.02	0.02
0.73	0.66	0.55	0.48	0.29	0.17	0.00	0.08	0.08	0.08	0.11	0.13	0.13	0.11	0.06	0.05	0.03	0.02
0.93	0.85	0.72	0.64	0.46	0.39	0.27	0.00	0.00	0.01	0.05	0.07	0.08	0.10	0.07	0.05	0.03	0.02
0.99	0.91	0.77	0.69	0.51	0.43	0.32	0.06	0.00	0.00	0.05	0.07	0.08	0.10	0.07	0.05	0.03	0.02
1.03	0.95	0.82	0.73	0.56	0.48	0.37	0.11	0.05	0.00	0.05	0.07	0.07	0.10	0.07	0.05	0.03	0.02
1.14	1.06	0.94	0.84	0.69	0.62	0.51	0.26	0.21	0.16	0.00	0.05	0.06	0.11	0.08	0.04	0.02	0.02
1.25	1.17	1.06	0.98	0.83	0.76	0.66	0.41	0.36	0.31	0.18	0.00	0.02	0.07	0.04	0.03	0.02	0.01
1.33	1.25	1.14	1.06	0.91	0.84	0.73	0.50	0.44	0.39	0.27	0.10	0.00	0.06	0.09	0.05	0.03	0.01
1.49	1.41	1.30	1.22	1.07	1.00	0.88	0.68	0.62	0.57	0.47	0.31	0.22	0.00	0.07	0.03	0.05	0.01
1.71	1.63	1.52	1.44	1.27	1.20	1.07	0.89	0.84	0.79	0.69	0.53	0.49	0.31	0.00	0.03	0.05	0.01
1.93	1.85	1.74	1.65	1.49	1.41	1.28	1.10	1.05	1.00	0.87	0.74	0.68	0.51	0.25	0.00	0.06	0.01
2.25	2.16	2.05	1.97	1.79	1.72	1.60	1.41	1.36	1.31	1.19	1.06	0.99	0.85	0.61	0.39	0.00	0.01
3.18	3.10	2.98	2.90	2.73	2.66	2.53	2.35	2.30	2.25	2.13	2.00	1.92	1.76	1.51	1.29	0.96	0.00

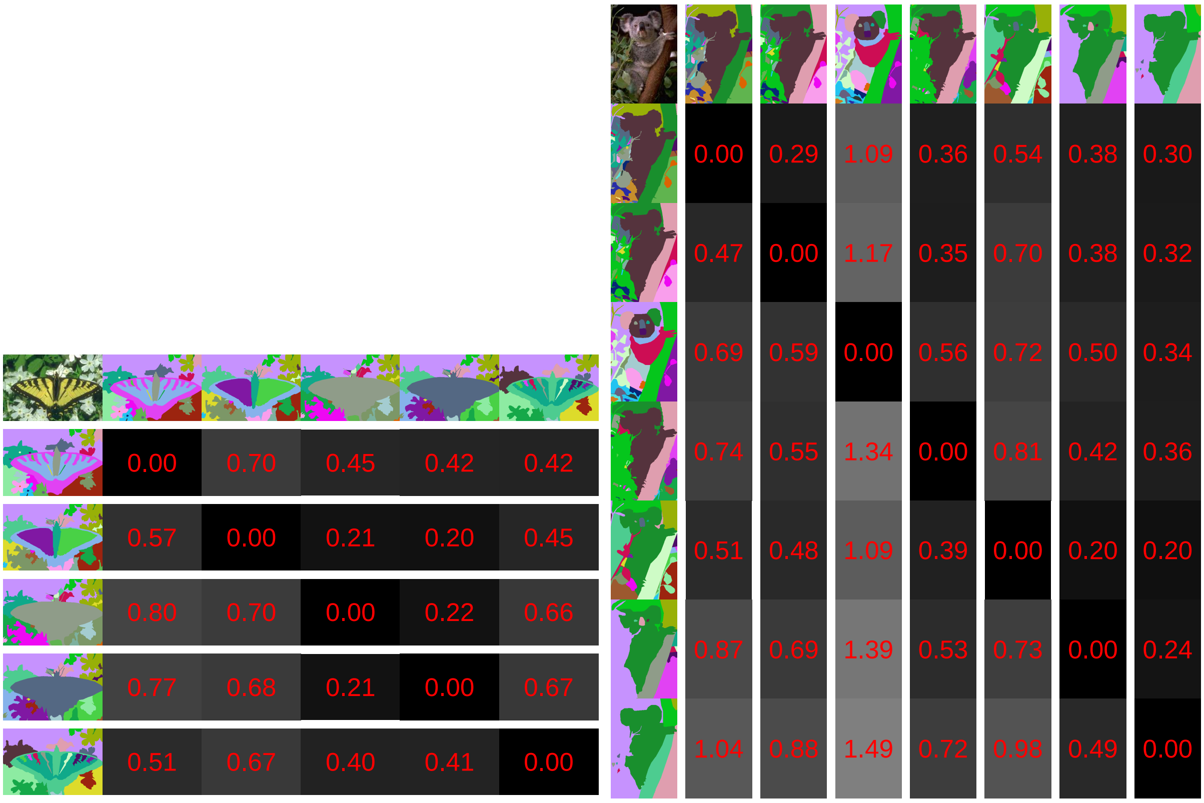
Fig. 10 Conditional entropies between segmentations at different granularities, which demonstrate good coarse-to-fine compatibilities.



0.00	0.29	0.35	0.27	0.14	0.06
1.24	0.00	0.22	0.14	0.07	0.04
1.24	0.16	0.00	0.14	0.07	0.03
1.73	0.66	0.72	0.00	0.08	0.03
2.58	1.56	1.62	1.06	0.00	0.03
3.32	2.34	2.40	1.82	0.85	0.00

0.00	0.43	0.43	0.41	0.40	0.06
1.12	0.00	0.32	0.31	0.31	0.06
1.71	0.91	0.00	0.29	0.26	0.06
1.79	1.00	0.39	0.00	0.16	0.06
2.08	1.31	0.67	0.48	0.00	0.33
2.60	1.91	1.32	1.23	1.18	0.00

Fig. 11 Conditional entropies between ground truths, which demonstrate good coarse-to-fine compatibilities.



0.00	0.70	0.45	0.42	0.42
0.57	0.00	0.21	0.20	0.45
0.80	0.70	0.00	0.22	0.66
0.77	0.68	0.21	0.00	0.67
0.51	0.67	0.40	0.41	0.00

0.00	0.29	1.09	0.36	0.54	0.38	0.30
0.47	0.00	1.17	0.35	0.70	0.38	0.32
0.69	0.59	0.00	0.56	0.72	0.50	0.34
0.74	0.55	1.34	0.00	0.81	0.42	0.36
0.51	0.48	1.09	0.39	0.00	0.20	0.20
0.87	0.69	1.39	0.53	0.73	0.00	0.24
1.04	0.88	1.49	0.72	0.98	0.49	0.00

Fig. 12 Conditional entropies between ground truths, which demonstrate bad coarse-to-fine compatibilities. Some annotators prefer more details on foreground objects, while some annotators prefer more details on backgrounds.