Appendix to Exploring Uncertainty in Image Segmentation Ensembles

Submission-ID: 1025

1 Dataset description

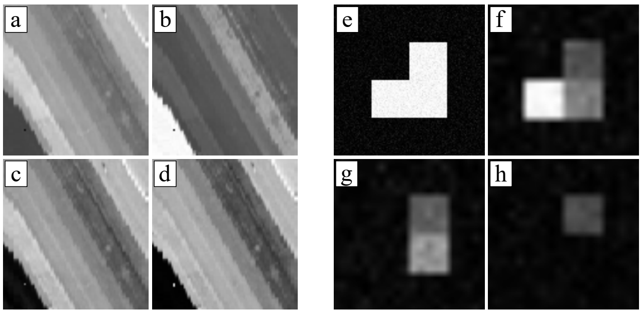

Figure 1: Analyzed datasets: Channels (a) 26, (b) 46, (c) 125 and (d) 176 of SalinasA hyperspectral dataset. Middle slices of (e) a synthetic computed tomography dataset and (f-h) three synthetic absorptiometry dataset of the same object.

The dataset we analyzed in the evaluation section of the main paper is a synthetic, combined computed tomography and K-edge absorptiometry dataset, it can be seen in Figure 1(a-d). It consists of a conventional attenuation image shown in (e) as well as three channels showing the concentration of a specific element, shown in (f-h).

2 Evaluation of Hyperspectral dataset

We use the SalinasA scene [HYP], a hyperspectral image of fields with different crops, Figure 1(e-h) shows four of its 204 channels. For a sub-ensemble of size 49, created by filtering on the mean member uncertainty, we plot neighborhood versus algorithm uncertainty in Figure 2(d). We want to analyze situations where the neighborhood uncertainty is low, but where the algorithm uncertainty is high. The goal is to know where the algorithm was uncertain despite a homogeneous segmentation result. Therefore, we select all pixels with an algorithm uncertainty between approximately 0.5 and 0.85, the upper half of the algorithm uncertainty range, and a neighborhood uncertainty between approximately 0 and 0.03 (roughly the ten lowest percent of the neighborhood uncertainty range). The selected pixels are also highlighted in the ensemble uncertainty image (e). There, a single strip starting in the lower right corner stands out, in which most of the selected pixels are located. For this strip the algorithm was uncertain, even though it is nearly homogeneously segmented. The ensemble uncertainty in (e) shows this strip in dark color, signifying low uncertainty. So, while the individual algorithms often were uncertain, they still all agreed on a specific label. We further investigate this through probability probing. The charts (i-l) show details of the pixel marked with a red cross in (e). In (i) we see that the algorithm uncertainty is rather high in all members. The large majority still assigned label 5 to it, as shown in (j). Consulting the ground truth (f) reveals that this is correct. We see in (k) and (l) that while there are relatively high probabilities for label 5, label 0 also got probabilities of up to 0.5 from the probabilistic segmentation algorithm. From the algorithm uncertainty alone we might argue that the algorithm needs to be tweaked to better recognize label 5, yet the ensemble and neighborhood uncertainty show that there is no need for refinement.

3 Segmentation algorithm, parameters and ranges

For segmentation, we have used a framework using the Support Vector Machine (SVM) classification algorithm and the Extended Random Walker (ERW) segmentation algorithm. This pipeline, proposed by Kang et al. [KLF*15] was run with 100 different input parameter combinations to generate our ensembles. The pipeline has eight parameters, which are summarized in Table 1. The parameter ranges over which we sampled for each dataset we evaluated are shown in Table 2. In addition to input parameters, the SVM requires data points for learning each label as input. We have used the same set of data points, predetermined by the user, for each segmentation run in a sampling.

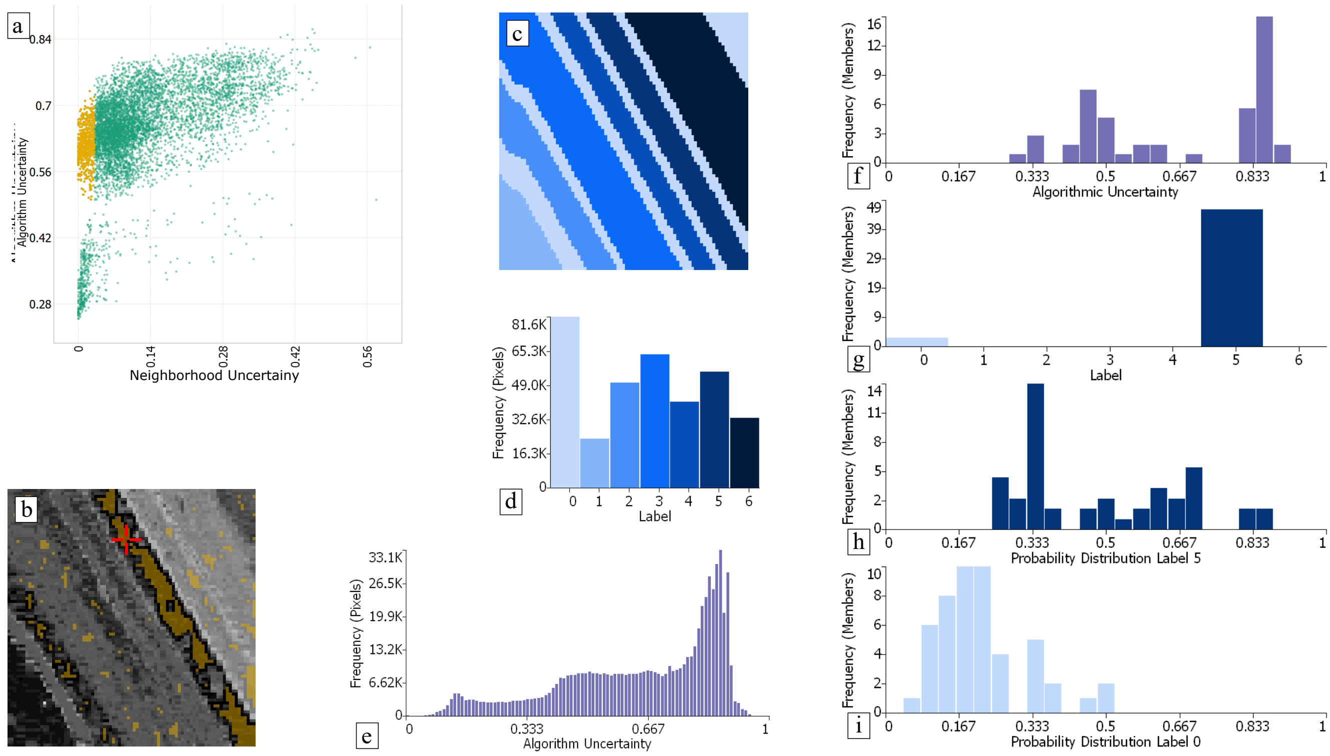

Figure 2: Evaluation a hyperspectral segmentation ensemble. (a) Algorithm vs. Neighborhood uncertainty. (b) Neighborhood uncertainty, highlighted are pixels selected in (a). (c) Ground truth segmentation (d) Label distribution over all pixels and members. (e) Uncertainty distribution over all pixels and members. (f-i) Histogram of distributions over members for pixel marked in (b), of: (f) Algorithm Uncertainty, (h) Labels, (h) Probabilities for label 5, (i) Probabilities for label 0.

Algorithm	Parameter Name	Description
SVM	C_{svm}	Soft classification penalty
	γ_{svm}	Gaussian RBF kernel width parameter
	n_{svm}	Number of channels to consider
ERW	β_{erw}	Normalization neighborhood weight
	γ_{erw}	Weight of prior model vs. neighborhood
	m_{erw}	Maximum number of iterations in linear solver
	$dist_{erw}$	Metric used for distances between neighboring pixel values
PCA	c_{pca}	Number of PCA components considered

Table 1: Parameters to our segmentation framework.

References

[HYP] Hyperspectral remote sensing scenes. `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`. Online. Retrieved Feb. 16, 2018.

[KLF*15] KANG X., LI S., FANG L., LI M., BENEDIKTSSON J. A.: Extended random walker-based classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing 53*, 1 (2015), 144–153. doi:10.1109/TGRS.2014.2319373.

	SalinasA	Synthetic CT
Size	83x86	120x120x8
Channels	204	4
Labels	7	4
# of Samples	100	100
C_{svm}	$0.01..10{,}000^l$	$0.1..10{,}000^l$
γ_{svm}	$10^{-11}..10^l$	$10^{-4}..10^l$
n_{svm}	1..204	3..4
β_{erw}	$10^{-5}..10{,}000^l$	$10^{-5}..10{,}000^l$
γ_{erw}	$0.5..100^l$	$0.5..100^l$
m_{erw}	100..10,000	1,000..10,000
$dist_{erw}$	l_1-norm,l_2-norm,l_∞-norm, squared sum, cosine distance, kullback-leibler divergence, jensen-shannon divergence, earth movers distance, chi-square distance	squared sum
c_{pca}	1..204	4

Table 2: Properties and parameter ranges over which we sampled for the datasets used in our evaluation. ^{l} indicates a logarithmic sampling scale. The sampling for the synthetic dataset was done using a fixed prior model derived from thresholding instead of SVM, thus here no SVM parameters are listed.