Region	Bias for View Consistency	Accuracy	Avg. F_1 score
R1	Probability-Based Bias	95.04	94.61
	Entropy-Based Bias	**96.09**	**95.72**
R2	Probability-Based Bias	93.07	92.82
	Entropy-Based Bias	**93.86**	**93.60**
R5	Probability-Based Bias	94.09	93.61
	Entropy-Based Bias	**94.33**	**93.87**

Table 5: Computing pixel-level view consistency score using probability- and entropy-based biases. We find that entropy-based bias yields better accuracy and average F_1 score.

Region	Bias Agg.	Pixel Agg.	Accuracy	Avg. F_1 score
R1	Avg	Avg	95.75	95.32
	Avg	Max	95.90	95.50
	Max	Avg	**96.09**	**95.72**
	Max	Max	94.68	94.11
R2	Avg	Avg	90.64	90.12
	Avg	Max	93.12	92.84
	Max	Avg	**93.86**	**93.60**
	Max	Max	84.26	93.98
R5	Avg	Avg	94.09	93.61
	Avg	Max	91.71	90.75
	Max	Avg	**94.33**	**93.87**
	Max	Max	88.31	86.88

Table 6: Experiments on different aggregation schemes. We tried all possible combinations and found that the Max-Avg scheme yields best accuracy and average F_1 score.

Appendix

We run extensive experiments to tune the various design dimensions and parameters of ALFA to find the best configuration. Without loss of generality, we run these experiments on two regions from North Carolina, R1 and R2, and one region from Texas, R5.

Pixel-level View Inconsistency Score: Probability-based v.s. Entropy-based. Recall that the view inconsistency scores are computed with ρ_{avg} in Eq (5) or ρ_{max} in Eq (6) defined with the entropy-based bias given by Eq (4). An alternative approach is to replace the bias with the probability-based one given by Eq (3) when computing the view inconsistency scores. We compare the effect of using entropy-based bias v.s. probability-based bias to compute the view inconsistency score with ρ_{max} (the next set of experiments will show that it is better than ρ_{avg}). Without loss of generality, we report the results on three of the test regions in R1, R2, and R5 in Table 5. We can see that the entropy method consistently yields better results both in terms of accuracy and average F_1 score. Therefore, we use entropy method for pixel-level consistency score computation by default in all our experiments.

Effect of Aggregation Schemes. There are two places of score aggregation when we compute the uncertainty score for superpixels.

Region	λ_1	λ_2	λ_3	β_1	β_2	Avg. F_1 score
R1	0.1	0.05	0.1	0.001	0.005	93.37
	0.15	0.1	0.1	0.001	0.005	94.47
	0.1	0.05	0.1	0.05	0.05	**95.36**
	0.1	0.05	0.1	0.05	0.1	95.33
	0.15	0.1	0.1	0.05	0.1	94.41
	0.1	0.05	0.1	0.1	0.05	95.35
	0.15	0.1	0.1	0.1	0.1	94.74
R2	0.1	0.05	0.1	0.001	0.005	90.79
	0.15	0.1	0.1	0.001	0.005	87.97
	0.1	0.05	0.1	0.05	0.05	**91.35**
	0.1	0.05	0.1	0.05	0.1	90.64
	0.15	0.1	0.1	0.05	0.1	80.35
	0.1	0.05	0.1	0.1	0.05	90.85
	0.15	0.1	0.1	0.1	0.1	90.92
R5	0.1	0.05	0.1	0.001	0.005	93.60
	0.15	0.1	0.1	0.001	0.005	93.55
	0.1	0.05	0.1	0.05	0.05	**93.91**
	0.1	0.05	0.1	0.05	0.1	93.60
	0.15	0.1	0.1	0.05	0.1	93.57
	0.1	0.05	0.1	0.1	0.05	93.48
	0.15	0.1	0.1	0.1	0.1	93.56

Table 7: Hyperparameter tuning for λ_1, λ_2, λ_3, β_1, and β_2. We tried many combinations of their values and found a configuration as highlighted in boldface that consistently yields the best average F_1 score.

- The first is for bias aggregation when the computing pixel-level view consistency score. We can use either the average scheme ρ_{avg} in Eq (5) or the maximum scheme ρ_{max} in Eq (6). Since Table 5 has shown that entropy-based bias is better, it is adopted to compute ρ_{avg} and ρ_{max}.
- The second is for superpixel-level aggregation of pixel uncertainty scores. We can use either the average scheme in Eq (9) or the maximum scheme in Eq (10).

This gives us 4 scheme combinations, and Table 6 reports the performance of ALFA with these 4 configurations. We can see that the combination of maximum scheme for bias aggregation and average scheme for pixel-uncertainty-score aggregation consistently gives the best accuracy and average F_1 score. We, therefore, use this combination of aggregation scheme by default in all our experiments.

Hyperparameter Tuning. We tried many combinations of values for λ_1, λ_2, λ_3, β_1, and β_2 to find the set of values which consistently gives the best average F_1 score to be used as the default setting. Table 7 shows some of the most competitive combinations, and we can see that the best setting is $\lambda_1 = 0.1$, $\lambda_2 = 0.05$, $\lambda_3 = 0.1$, $\beta_1 = 0.05$, and $\beta_2 = 0.05$.