CoSW: Conditional Sample Weighting for Smoke Segmentation with Label Noise

Lujian Yao, Haitao Zhao*, Zhongze Wang, Kaijie Zhao, Jingchao Peng

East China University of Science and Technology

Smoke Segmentation

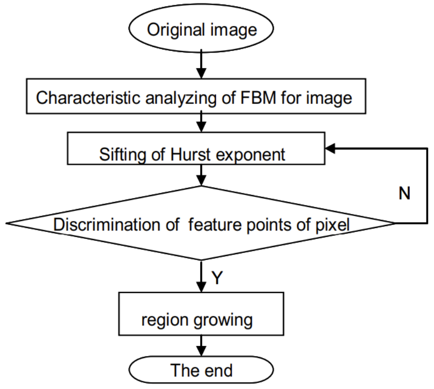
- In wildlife, smoke is an important indicator of **fire**. ESS enables **rapid** identification of the location of the **smoke source**, facilitating the timely extinguishing of the flames by rescue personnel and **preventing the occurrence of large fires.**

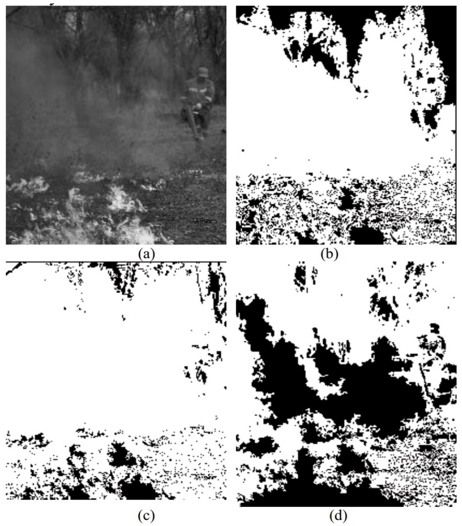
- In industrial production, ESS can also aid in promptly detecting the location of **gas leaks** and prevent **the spread of toxic and harmful gases**.

Traditional Smoke Segmentaion

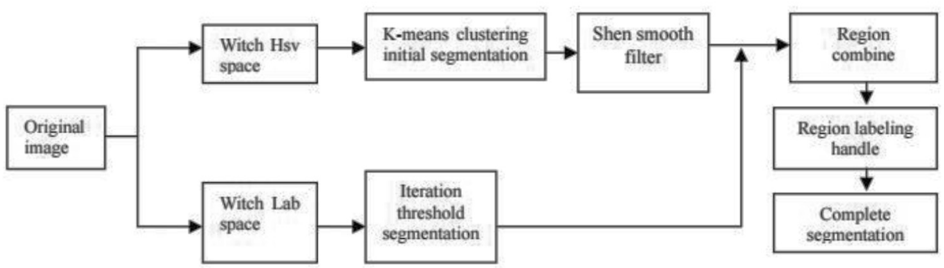
In smoke segmentation, **the color and texture features** of smoke play a significant role.
color enhancement techniques and **color channel analysis** to segment smoke regions

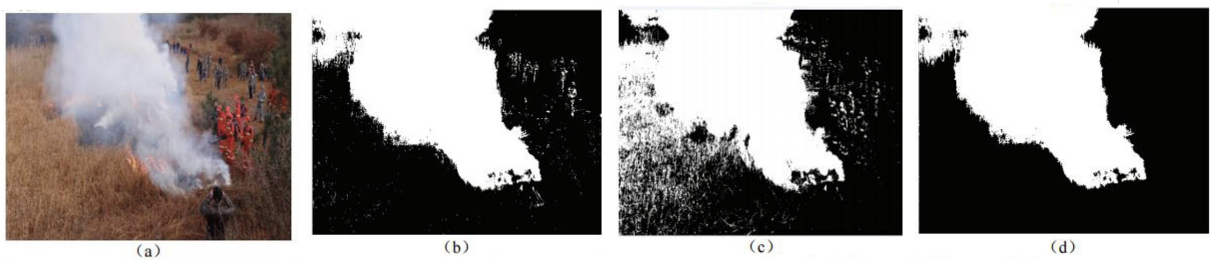
Region Growing [1]





Color Space [2]





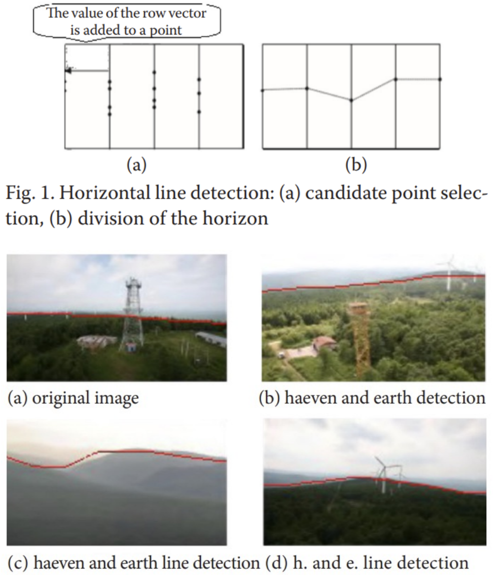
[1] Wang X, Jiang A, Wang Y. A segmentation method of smoke in forest-fire image based on fbm and region growing. 2011 Fourth International Workshop on Chaos-Fractals Theories and Applications. IEEE, 2011: 390-393.
[2] Xing D, Zhongming Y, Lin W, et al. Smoke image segmentation based on color model. Journal on Innovation and Sustainability RISUS, 2015, 6(2): 130-138.

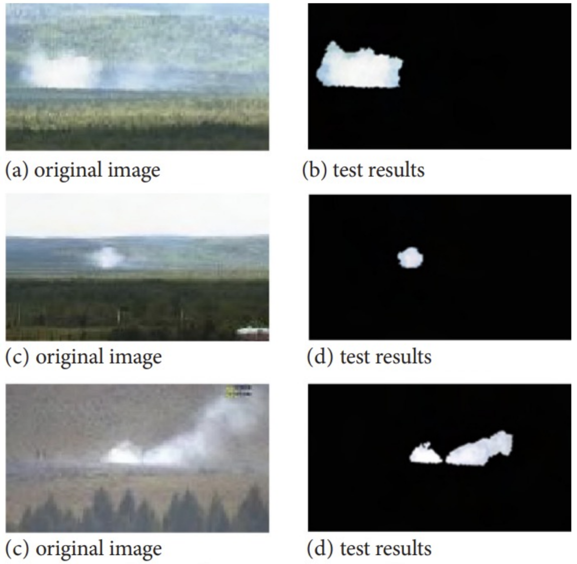
Traditional Smoke Segmentaion

華東理工大學
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

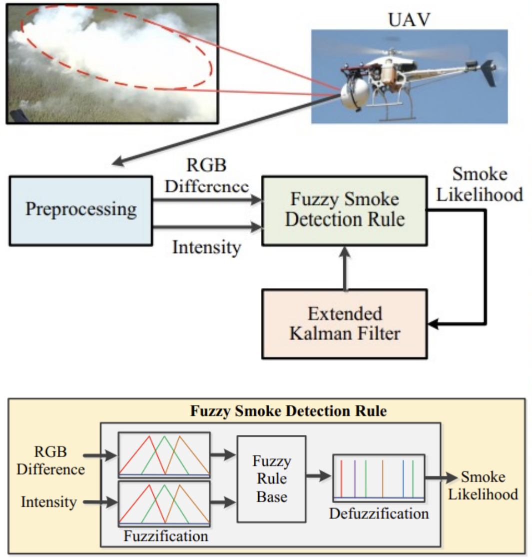
Others focus on **the clustering technique** and **morphology characteristics** of smoke.

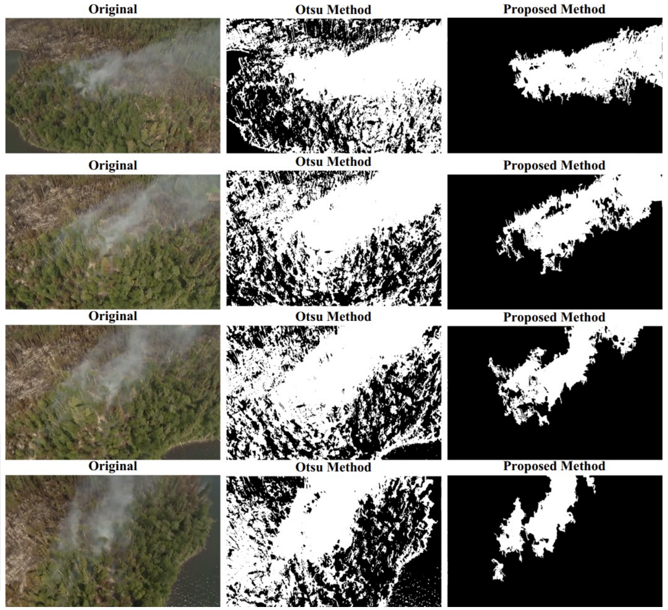
Iterative clustering, superpixel segmentation [3]





Morphology [4]





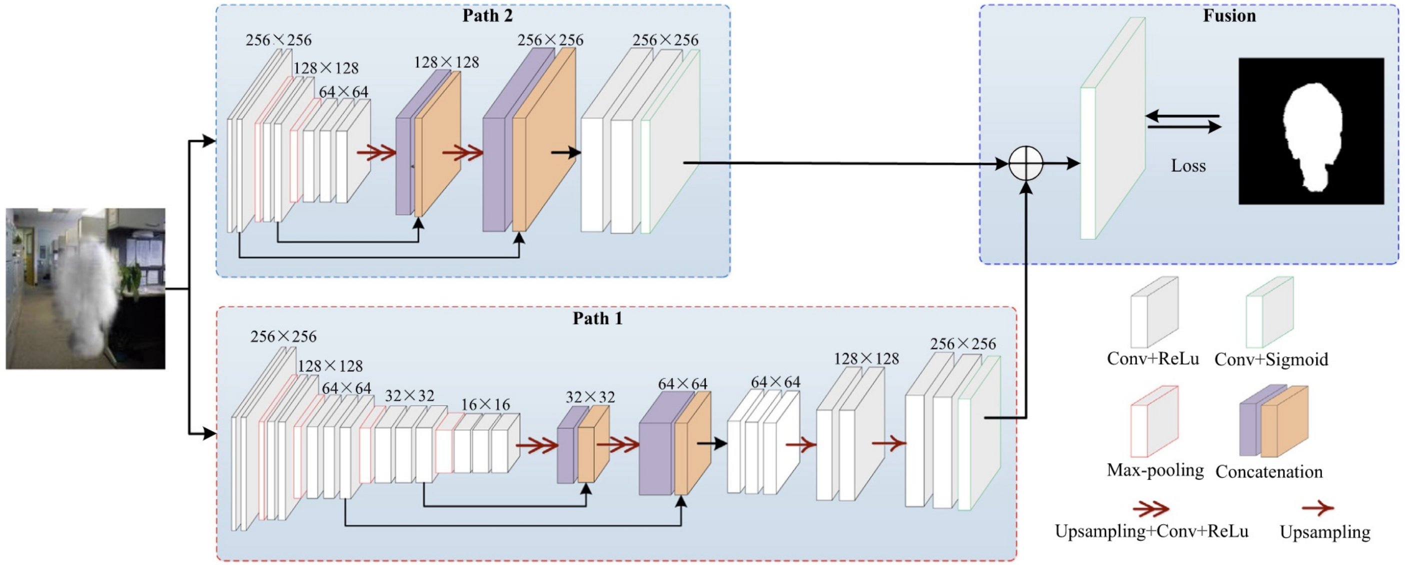
[3] Xiong D, Yan L. Early smoke detection of forest fires based on SVM image segmentation. Journal of Forest Science, 2019, 65(4): 150-159.
[4] Yuan C, Liu Z, Zhang Y. Learning-based smoke detection for unmanned aerial vehicles applied to forest fire surveillance. Journal of Intelligent & Robotic Systems, 2019, 93: 337-349.

Smoke segmentation based on deep learning

- **Larger receptive fields** to cope with the **variability and blurred edges** of smoke.

DSS [5]

- **Dual-channel encoder-decoder**
- One channel is used to coarsely extract **global** information
- The other channel is utilized to finely extract **local** information.

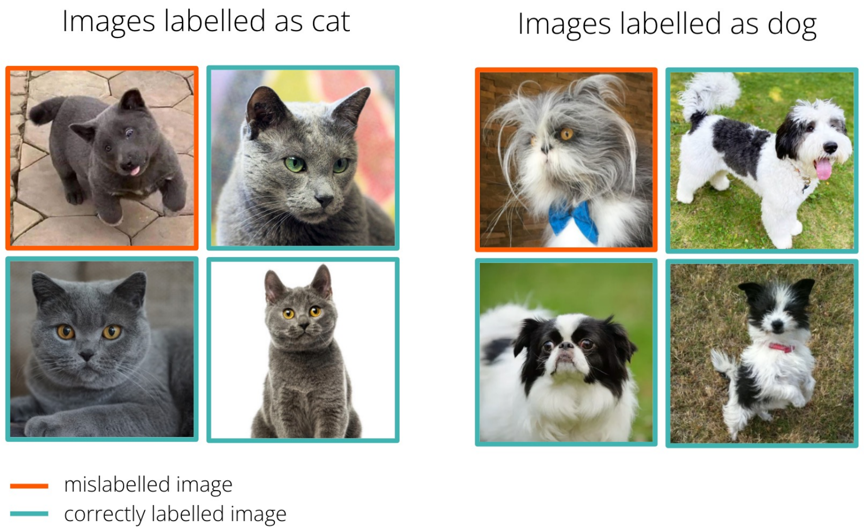


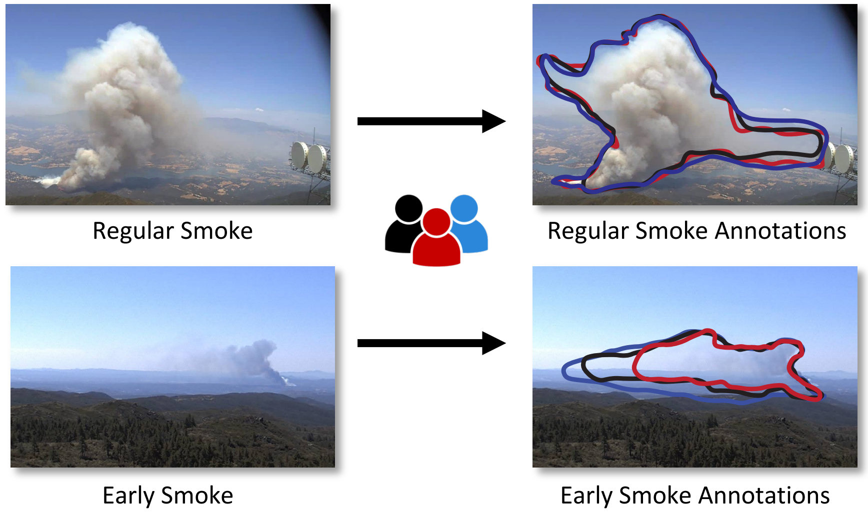
dual-channel encoder-decoder architecture

[5] Yuan F, Zhang L, Xia X, et al. Deep smoke segmentation. Neurocomputing, 2019, 357: 248-260.

Label Noise

華東理工大學
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY





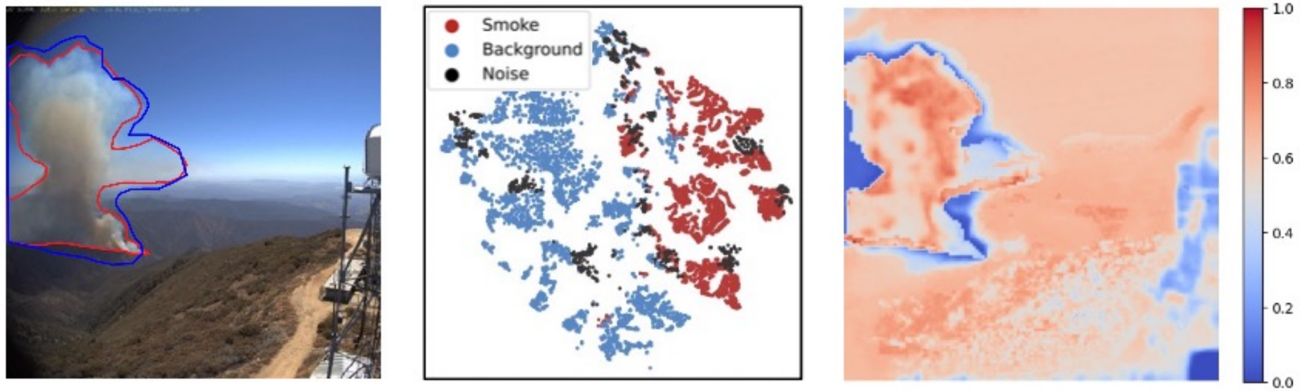
- Noisy labels are almost inevitable in smoke segmentation.
- Smoke edges are complex and blurry, making it hard to distinguish smoke and background.
- Smoke is non-rigid and lacks a fixed shape, making it difficult for annotators to become proficient through practice with the same shape.

Motivation

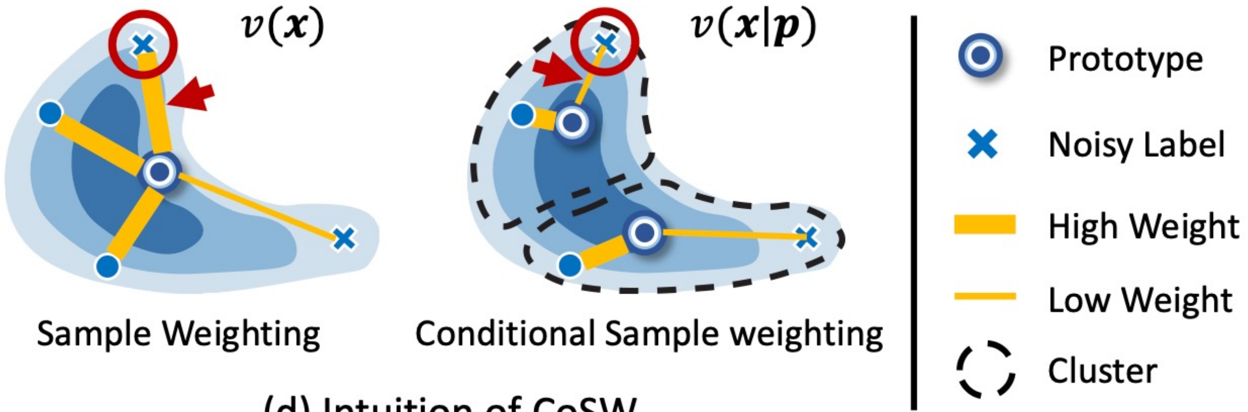
variable transparency ***inconsistent features***

1) How to determine the sample weight through prototypes.

2) How to update prototypes under noisy labels.



(a) Smoke Image (b) Embeddings (c) CoSW



(d) Intuition of CoSW

Total Entropy (based on Shannon entropy)

$$T = -\sum_{k=1}^{\Omega K}\sum_{n=1}^{N^k}\frac{v_n^k}{N}\ln\frac{v_n^k}{N},$$

Within-prototype Entropy (WE)

$$T_w = -\sum_{k=1}^{\Omega K}\frac{N^k}{N}\sum_{n=1}^{N^k}\frac{v_n^k}{N^k}\ln(\frac{v_n^k}{N^k}),$$

Between-prototype Entropy

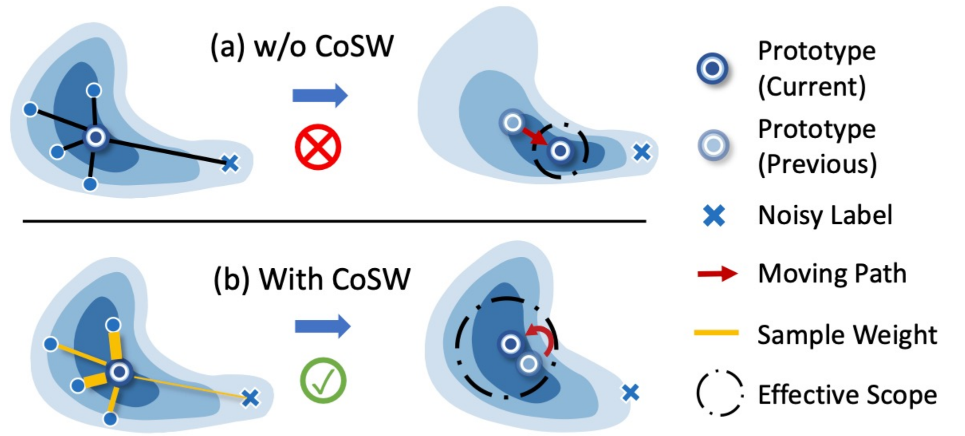
$$T_b = -\sum_{k=1}^{\Omega K}\frac{N^k}{N}\ln\frac{N^k}{N},$$

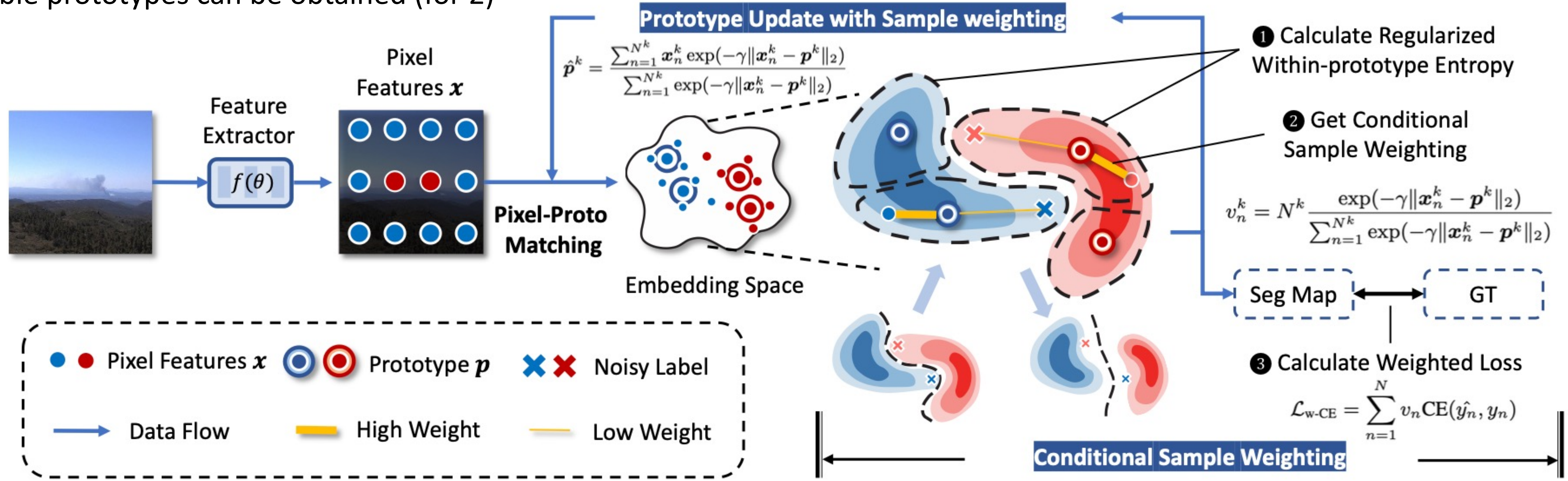
CoSW Overview

Regularized Within-prototype Entropy (RWE)

$$J(P,V) = -\sum_{k=1}^{\Omega K}\frac{N^k}{N}\sum_{n=1}^{N^k}\frac{v_n^k}{N^k}\ln\left(\frac{v_n^k}{N^k}\right) - \gamma\sum_{k=1}^{\Omega K}\frac{N^k}{N}\sum_{n=1}^{N^k}\frac{v_n^k}{N^k}\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2,$$

- Consider the information between all pixels that matched the given prototype and obtain adaptive sample weights (for 1).
- Calculate the expectation of weighting samples, stable prototypes can be obtained (for 2)





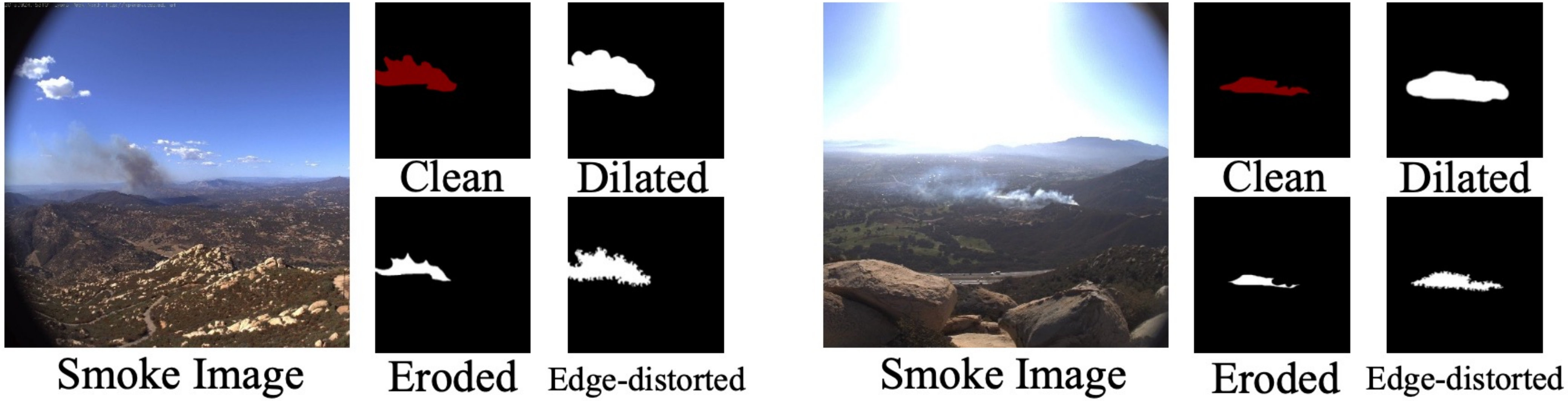
Datasets

SmokeSeg (Real-world Noise)
- **Real-world** smoke segmentation dataset
- Most images are **early smoke.**

SMOKE5K (Real-world Noise)
- 4K synthetic images
- \+ 1K real-world images)

Datasets

NS-1K (Synthetic Noise)



Comparison with State-of-the-Art Methods

	Methods	Backbone	Total		Small		Medium		Large	
			F_1	$mIoU$	F_1	$mIoU$	F_1	$mIoU$	F_1	$mIoU$
Real-time	AFFormer° [8]	AFFormer-B	58.41	47.89	45.98	34.10	64.52	53.20	66.96	54.15
	SeaFormer° [40]	SeaFormer-B	57.58	44.70	41.33	30.35	66.02	53.10	67.16	53.21
	SegFormer° [47]	MiT-B0	60.90	<u>48.53</u>	<u>48.31</u>	<u>35.69</u>	**68.86**	53.34	68.53	55.80
	SC§ [52]	MiT-B0	59.78	47.70	45.76	33.15	68.37	**55.68**	66.13	53.96
	CleanNet* [27]	MiT-B0	<u>61.63</u>	47.00	47.16	33.57	66.23	50.91	**70.96**	**56.22**
	Ours	MiT-B0	**62.98**	**48.62**	**49.85**	**36.64**	<u>68.81</u>	<u>53.87</u>	<u>70.79</u>	<u>55.85</u>
Normal	DeeplabV3+° [5]	ResNet-50	65.92	53.50	54.03	41.07	71.82	58.73	71.87	58.95
	OCRNet° [57]	HRNet-48	64.93	52.45	52.04	39.47	71.04	57.47	70.51	58.19
	SegNeXt° [13]	MSCAN-L	66.71	52.37	58.05	44.16	70.41	55.77	72.97	58.42
	Trans-BVM† [51]	ResNet-50	67.15	53.11	59.02	44.62	71.50	56.99	73.36	58.97
	Ours	ResNet-50	68.49	54.09	61.27	46.28	72.31	58.08	<u>74.78</u>	60.81
	SegFormer° [47]	MiT-B3	67.70	53.37	57.67	45.21	<u>73.87</u>	<u>60.49</u>	72.06	58.52
	Trans-BVM† [51]	MiT-B3	67.68	53.09	60.85	45.87	71.73	57.35	73.51	59.10
	SC§ [52]	MiT-B3	69.55	55.04	<u>62.26</u>	<u>48.47</u>	71.41	57.06	72.91	58.23
	CleanNet* [27]	MiT-B3	<u>70.17</u>	<u>56.94</u>	61.98	48.05	73.06	59.07	74.57	<u>60.93</u>
	Ours	MiT-B3	**72.32**	**59.25**	**64.62**	**50.86**	**74.37**	**61.14**	**75.52**	**62.30**

The best results: **bold**. The second-best results: <u>underline</u>.

Comparison with State-of-the-Art Methods

(a) Comparison on SMOKE5K.

Methods	F_β	$mIoU$
OCRNet° [57]	72.51	63.00
DeeplabV3+° [5]	73.83	64.08
SegNeXt° [13]	76.44	67.08
Trans-BVM† [51]	76.23	67.55
Ours	77.02	67.58
SegFormer° [47]	78.68	68.29
Trans-BVM† [51]	78.91	68.97
SC§ [52]	79.33	69.40
CleanNet* [27]	<u>80.37</u>	<u>70.23</u>
Ours	**81.71**	**71.24**

(b) Comparison on the synthetic noisy dataset NS-1K.

		Trans-BVM† [51]		SC§ [52]		CleanNet* [27]		Ours	
Noise Ratio	Noise Intensity	F_1	$mIoU$	F_1	$mIoU$	F_1	$mIoU$	F_1	$mIoU$
0%	-	**52.79**	**39.12**	51.94	37.84	52.02	38.24	<u>52.59</u>	<u>38.32</u>
20%	*Low*	**51.91**	**38.16**	<u>51.81</u>	<u>37.94</u>	51.21	37.24	51.77	37.58
	High	48.02	34.52	**51.14**	**37.15**	50.34	36.31	<u>50.99</u>	<u>36.80</u>
40%	*Low*	45.40	32.30	<u>49.69</u>	<u>35.89</u>	49.09	35.53	**50.18**	**36.69**
	High	41.09	28.15	45.31	32.16	<u>46.08</u>	<u>33.15</u>	**48.34**	**34.73**
60%	*Low*	42.73	29.01	44.12	30.87	<u>46.33</u>	<u>33.96</u>	**48.38**	**35.34**
	High	40.57	27.50	41.77	28.44	<u>43.69</u>	<u>29.87</u>	**45.86**	**32.08**
80%	*Low*	39.28	26.42	40.52	27.42	<u>42.96</u>	<u>29.37</u>	**44.40**	**31.24**
	High	37.38	25.58	39.09	25.88	<u>40.27</u>	<u>27.21</u>	**42.37**	**29.89**

The best results: **bold**. The second-best results: <u>underline</u>.

Ablation Study

華東理工大學
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

(a) Ablation study of CoSW (dataset: SmokeSeg).

	Proto	Sample Weight	Proto Weight	F_1	$mIoU$
bl				67.70	53.37
(1)	✓			68.17 ↑ 0.47	55.23 ↑ 1.86
(2)	✓	✓		70.04 ↑ 2.34	56.40 ↑ 3.03
(3)	✓		✓	69.38 ↑ 1.68	55.72 ↑ 2.35
(4)	✓	✓	✓	**71.39** ↑ 3.69	**57.68** ↑ 4.31

Table 4: Different entropies.

	F_1	$mIoU$
Kapur's Entropy	71.34	58.38
Burg's Entropy	69.20	55.72
Shannon's Entropy	**72.32**	**59.25**

- Burg's Entropy

$$T^B(\Pi) = -\sum_{i=1}^{N} \ln \pi_i .$$

$$\max_{P,V} J^B(P,V) = -\sum_{k=1}^{\Omega K}\sum_{n=1}^{N^k} \ln\left(\frac{v_n^k}{N^k}\right) - \gamma \sum_{k=1}^{\Omega K} \frac{N^k}{N} \sum_{n=1}^{N^k} \frac{v_n^k}{N^k} \|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2 .$$

$$v_n^k = N^k \frac{(-\gamma\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2)}{\sum_{n=1}^{N^k}(-\gamma\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2)},$$

$$\hat{\boldsymbol{p}}^k = \frac{\sum_{n=1}^{N^k} \boldsymbol{x}_n^k(-\gamma\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2)}{\sum_{n=1}^{N^k}(-\gamma\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2)},$$

- Kapur's Entropy

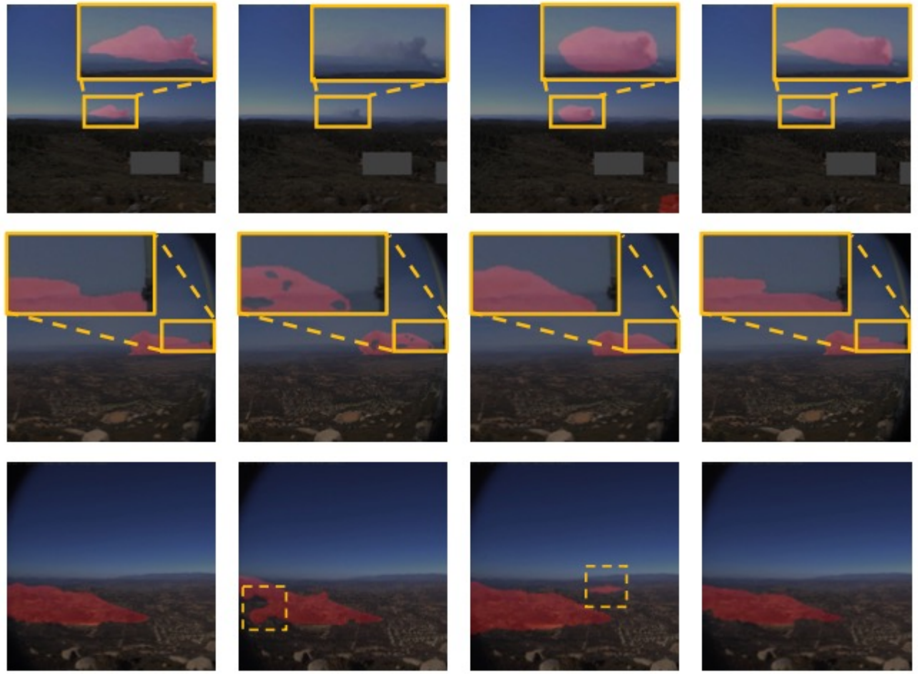
$$T^K(\Pi) = -\sum_{i=1}^{N} \pi_i \ln \pi_i - \sum_{i=1}^{N} (1-\pi_i)\ln(1-\pi_i).$$

$$\max_{P,V} J^K(P,V) = -\sum_{k=1}^{\Omega K} \frac{N^k}{N} \sum_{n=1}^{N^k} \frac{v_n^k}{N^k} \ln\left(\frac{v_n^k}{N^k}\right) - \sum_{k=1}^{\Omega K} \frac{N^k}{N} \sum_{n=1}^{N^k} \left(1 - \frac{v_n^k}{N^k}\right) \ln\left(1 - \frac{v_n^k}{N^k}\right) - \gamma \sum_{k=1}^{\Omega K} \frac{N^k}{N} \sum_{n=1}^{N^k} \frac{v_n^k}{N^k} \|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2 .$$

$$v_n^k = N^k \frac{1}{1+\exp(-\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2 - \lambda_k)^\gamma},$$

$$\hat{\boldsymbol{p}}^k = \sum_{n=1}^{N^k} \frac{\boldsymbol{x}_n^k}{1+\exp(-\|\boldsymbol{x}_n^k - \boldsymbol{p}^k\|_2 - \lambda_k)^\gamma},$$

Visualization



GT Trans-BVM CleanNet Ours

(a) Comparison of segmentation results in different scales.

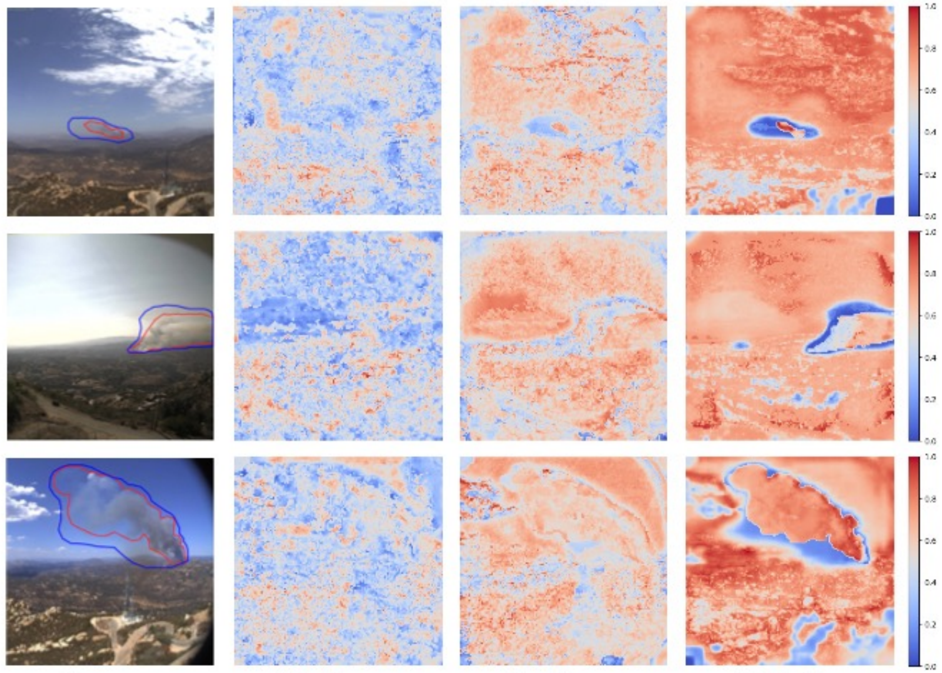


Image 1k Iters 5k Iters 20k Iters

(b) The formation process of CoSW (blue curve for noisy labels, and red for clean in the first image column).

To determine whether the prototype is clean, it is necessary to introduce clean validation, which is not implemented in our work. This is also a direction for further research.

THANK YOU

Lujian Yao