Appendix

Section A elaborates on the validation of the normality prior. Section B presents detailed derivations of normality propagation. Section C reports additional experiments and analysis.

A Validation of Normality Prior

To further confirm the applicability of the normality prior to videos beyond standard benchmarks, we collected 50 abnormal videos from YouTube using keywords like "fighting" and "violent acts", and observed that 98% of the start and end frames in these videos are normal. We included the links to these videos at the end of the supplementary material.

B Normality Propagation

Detailed derivations of the proposed normality propagation are presented here, following [4]. Given the construct label vector $\mathbf{y}$ and similarity matrix $\mathbf{S}$, normality propagation aims to estimate normal magnitudes $\mathbf{z}$ of snippets in a video. It iterates $\mathbf{z}(n+1) = \alpha\mathbf{S}\mathbf{z}(n) + (1-\alpha)\mathbf{y}$, where α is a hyper-parameter in the range $(0,1)$, n is the index of an iteration, and $\mathbf{z}(0) = \mathbf{y}$.

By the iteration equation above, we have

$$\begin{aligned}
\mathbf{z}(1) &= \alpha\mathbf{S}\mathbf{y} + (1-\alpha)\mathbf{y}, \\
\mathbf{z}(2) &= (\alpha\mathbf{S})^2\mathbf{y} + (1-\alpha)(1+\alpha\mathbf{S})\mathbf{y}, \\
\mathbf{z}(3) &= (\alpha\mathbf{S})^3\mathbf{y} + (1-\alpha)(1+\alpha\mathbf{S}+(\alpha\mathbf{S})^2)\mathbf{y}, \\
\mathbf{z}(n) &= (\alpha\mathbf{S})^n\mathbf{y} + (1-\alpha)\sum_{i=1}^{n}(\alpha\mathbf{S})^{(i-1)}\mathbf{y},
\end{aligned} \tag{1}$$

Since $0 < \alpha < 1$ and the eigenvalues $|\lambda(\mathbf{S})|$ of $\mathbf{S}$ satisfy $|\lambda(\mathbf{S})| \leq 1$, $\lim\limits_{n\to\infty}(\alpha\mathbf{S})^n = 0$, and $\lim\limits_{n\to\infty}\sum_{i=1}^{n}(\alpha\mathbf{S})^{(i-1)} = (\mathbf{I}-\alpha\mathbf{S})^{-1}$. Thus,

$$\mathbf{z}^* = (1-\alpha)(\mathbf{I}-\alpha\mathbf{S})^{-1}\mathbf{y}. \tag{2}$$

It is clearly equivalent to

$$\mathbf{z}^* = (\mathbf{I}-\alpha\mathbf{S})^{-1}\mathbf{y}. \tag{3}$$

Therefore, this process converges to $\mathbf{z}^* = (\mathbf{I}-\alpha\mathbf{S})^{-1}\mathbf{y}$.

In addition, we also show the solver of the following objective function is equivalent to the vector $\mathbf{z}^*$ as defined by 3.

$$\mathcal{Q}(\mathbf{z}) = \alpha\mathbf{z}^T\mathbf{S}\mathbf{z} + (1-\alpha)(\mathbf{z}-\mathbf{y})^T(\mathbf{z}-\mathbf{y}), \tag{4}$$

Table 1: Ablation study on different r on UCF-Crime.

r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
AUC	73.27	75.45	75.23	76.64	75.45	75.45	75.12	74.14

Table 2: Performance on XD-Violence with I3D features.

Method	RareAnom [2]	C2FPL [1]	Ours
AUC (%)	68.33	80.09	79.82

Differentiating $\mathcal{Q}(\mathbf{z})$ with respect to $\mathbf{z}$, we have

$$\left.\frac{\partial \mathcal{Q}}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}^*} = \alpha(\mathbf{z}^* - \mathbf{S}\mathbf{z}^*) + (1+\alpha)(\mathbf{z}^* - \mathbf{y}) = 0, \tag{5}$$

which can be transformed into

$$\mathbf{z}^* - \alpha\mathbf{S}\mathbf{z}^* - (1-\alpha)\mathbf{y} = 0. \tag{6}$$

Thus, we have

$$\mathbf{z}^* = (1-\alpha)(\mathbf{I} - \alpha\mathbf{S})^{-1}\mathbf{y}. \tag{7}$$

It is equivalent to 3.

C Additional Experiments and Analysis

Ablation Study of Abnormal Ratio r. The ablation study on different values of r is shown in Table 1. The results indicate that our method exhibits robustness to a wide range of r values (i.e., from 0.2 to 0.7), whereas extreme values of r lead to inferior results.

Results on XD-Violence. XD-Violence [3] is a large-scale and multi-scene dataset. It contains 4754 untrimmed videos (2349 normal videos, 2405 abnormal videos) with six physically violent classes. Recent UVAD methods [1] follow the data organization of [3] but do not use annotations in the training set. Specifically, the training split contains 3954 videos and the testing split contains 800 videos. We present the performance on XD-Violence with I3D features in Table 2. As we can see, we have comparable performance with C2FPL [1].

https://www.youtube.com/watch?v=ou83m0KY3MQ
https://www.youtube.com/watch?v=ZfRSZJkLEh8
https://www.youtube.com/watch?v=vHUZWv9Fikk
https://www.youtube.com/watch?v=5VqpWwGGx-8
https://www.youtube.com/watch?v=vxZSAbCafBI
https://www.youtube.com/watch?v=bxKYHntHjAg
https://www.youtube.com/watch?v=MXCxRhMCxfQ
https://www.youtube.com/watch?v=wTu-do3SkGo
https://www.youtube.com/watch?v=hJyQ8TvwvEI
https://www.youtube.com/watch?v=JEo699fsRGQ
https://www.youtube.com/watch?v=LUPxMIPha6Y
https://www.youtube.com/watch?v=_4Bxk73jRQs
https://www.youtube.com/watch?v=4i86SUNFFzs
https://www.youtube.com/watch?v=eMaJOtOu5M8
https://www.youtube.com/watch?v=nyKeJfMdLes
https://www.youtube.com/watch?v=8JdLnCp17Jc
https://www.youtube.com/watch?v=NjvEIHFPcA0
https://www.youtube.com/watch?v=8sUI_9NRXpg
https://www.youtube.com/watch?v=4TpLxUGcnbY
https://www.youtube.com/watch?v=0f7QPAaEvNU
https://www.youtube.com/watch?v=uf0bD6h_UkQ
https://www.youtube.com/watch?v=60urCnfUcME
https://www.youtube.com/watch?v=NgyG0_CY1tE
https://www.youtube.com/watch?v=jBc_WALulgQ
https://www.youtube.com/watch?v=H0WqeDa5ihE
https://www.youtube.com/watch?v=ZvTWx1HPMLE
https://www.youtube.com/watch?v=zBTNTT6chFg
https://www.youtube.com/watch?v=c0ztLtpuwkM
https://www.youtube.com/watch?v=DTq6Gu30-uA
https://www.youtube.com/watch?v=YhQ9vLgplDw
https://www.youtube.com/watch?v=2tGTCDfUAgY
https://www.youtube.com/watch?v=wU84z2JWq08
https://www.youtube.com/watch?v=ZKnyhKyr2Ik
https://www.youtube.com/watch?v=IF59QNrF1VQ
https://www.youtube.com/watch?v=5CN8rM3Yjuo
https://www.youtube.com/watch?v=S8RpgIR2KfA
https://www.youtube.com/watch?v=KWgL9-YBZxg
https://www.youtube.com/watch?v=BiqpH2PPv6w
https://www.youtube.com/watch?v=WDCKCKO3s-8
https://www.youtube.com/watch?v=pQb54DytuP0
https://www.youtube.com/watch?v=rMgfYQMgdHU
https://www.youtube.com/watch?v=Z-NAigzTh0U
https://www.youtube.com/watch?v=yQ4yv7WGVp4
https://www.youtube.com/watch?v=FFgphwEAUhQ
https://www.youtube.com/watch?v=PEyj5mGDqls
https://www.youtube.com/watch?v=FfOGwGjcGVU
https://www.youtube.com/watch?v=5IfCSEX0V44
https://www.youtube.com/watch?v=MTJ1tlfPvjM
https://www.youtube.com/watch?v=1W88aaS3QbE
https://www.youtube.com/watch?v=79-3e9b5-38

References

1. Al-Lahham, A., Tastan, N., Zaheer, M.Z., Nandakumar, K.: A coarse-to-fine pseudo-labeling (c2fpl) framework for unsupervised video anomaly detection. In: WACV. pp. 6793–6802 (2024)
2. Thakare, K.V., Dogra, D.P., Choi, H., Kim, H., Kim, I.J.: Rareanom: A benchmark video dataset for rare type anomalies. PR **140**, 109567 (2023)
3. Wu, P., Liu, J., Shi, Y., Sun, Y., Shao, F., Wu, Z., Yang, Z.: Not only look, but also listen: Learning multimodal violence detection under weak supervision. In: ECCV. pp. 322–339 (2020)
4. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. NeurIPS **16**, 321–328 (2003)