Supplementary Material

A. Experimental details

We summarize the OOD detection evaluation task in Table 6. The OOD test dataset is selected from MS-COCO and nuImages dataset, which contains disjoint labels from the respective ID dataset. For the Youtube-VIS dataset, we use the dataset released in year 2021. Since there are no ground truth labels available for the validation images, we select the last 597 videos in the training set as the in-distribution evaluation dataset. The remaining 2,388 videos are used for training. The BDD100K and Youtube-VIS model are both trained for a total of 52,500 iterations. See detailed ablations on the hyperparameters in Section 4.3 of the main paper.

	Task 1	**Task 2**
ID train dataset	BDD100K train	Youtube-VIS train
ID val dataset	BDD100K val	Youtube-VIS val
OOD dataset	COCO / nuImages	COCO / nuImages
#ID train images	273,406	67,861
#ID val images	39,973	21,889
#OOD images from COCO	1,914	28,922
#OOD images from nuImages	2,100	2,100

Table 6. OOD detection evaluation tasks.

B. In-distribution classes

We provide a detailed description of the in-distribution classes for the two video datasets as follows.

BDD100K dataset contains 8 classes, which are *pedestrian, rider, car, truck, bus, train, motorcycle, bicycle*.

The Youtube-VIS dataset contains 40 classes, which are *airplane, bear, bird, boat, car, cat, cow, deer, dog, duck, earless_seal, elephant, fish, flying_disc, fox, frog, giant_panda, giraffe, horse, leopard, lizard, monkey, motorbike, mouse, parrot, person, rabbit, shark, skateboard, snake, snowboard, squirrel, surfboard, tennis_racket, tiger, train, truck, turtle, whale, zebra*.

C. Software and hardware

We run all experiments with Python 3.8.5 and PyTorch 1.7.0, using NVIDIA GeForce RTX 2080Ti GPUs.

D. Baselines

To evaluate the baselines, we follow the original methods in MSP [17], ODIN [33], Generalized ODIN [20], Mahalanobis distance [31], CSI [59], energy score [36] and gram matrices [54] and apply them accordingly on the classification branch of the object detectors. For ODIN [33], the temperature is set to be $T = 1000$ following the original work. For both ODIN and Mahalanobis distance [31], the noise magnitude is set to 0 because the region-based object detector is not end-to-end differentiable given the existence of region cropping and ROIAlign. For GAN [30], we follow the original paper and use a GAN to generate OOD images. The prediction of the OOD images/objects is regularized to be close to a uniform distribution, through a KL divergence loss with a weight of 0.05. We set the shape of the generated images to be 100×100 and resize them to have the same shape as the real images. We optimize the generator and discriminator using the Adam optimizer [26], with a learning rate of 0.001. For CSI [59], we use the rotations (0°, 90°, 180°, 270°) as the self-supervision task. We set the temperature in the contrastive loss to 0.5. We use the features right before the classification branch (with the dimension to be 1024) to perform contrastive learning. The weights of the losses that are used for classifying shifted instances and instance discrimination are both set to 0.1 to prevent training collapse. For Generalized ODIN [20], we replace and train the classification head of the object detector by the most effective Deconf-C head shown in the original paper.

E. Ablation study on a different backbone architecture

In this section, we evaluate the proposed `STUD` using a different backbone architecture of the Faster-RCNN, which is RegNetX-4.0GF [49]. Similarly, we compare with the same set of OOD detection baselines as stated in the main paper. The

In-distribution $\mathcal{D}$	**Method**	**FPR95** ↓	**AUROC** ↑	**mAP (ID)**↑
		OOD: MS-COCO / nuImages		
BDD100K	MSP [17]	80.09 / 93.05	74.19 / 63.14	32.0
	ODIN [33]	64.74 / 82.08	77.65 / 67.09	32.0
	Mahalanobis [31]	54.02 / 79.85	82.38 / 75.48	32.0
	Gram matrices [54]	63.96 / 63.61	67.56 / 67.47	32.0
	Energy score [36]	64.79 / 81.62	78.78 / 69.43	32.0
	Generalized ODIN [20]	60.76 / 82.00	80.14 / 70.74	32.5
	CSI [59]	52.98 / 80.00	83.57 / 74.91	31.8
	GAN-synthesis [30]	58.35 / 83.65	81.43 / 70.39	31.5
	STUD (ours)	**52.51 / 79.75**	**84.03 / 76.55**	32.3
Youtube-VIS	MSP [17]	89.86 / 97.42	67.04 / 54.02	26.7
	ODIN [33]	89.28 / 96.30	67.54 / 60.82	26.7
	Mahalanobis [31]	90.00 / 94.44	70.47 / 54.83	26.7
	Gram matrices [54]	87.64 / 91.25	69.76 / 61.43	26.7
	Energy score [36]	88.54 / 90.21	67.83 / 58.02	26.7
	Generalized ODIN [20]	85.15 / 98.00	71.57 / 64.23	27.3
	CSI [59]	82.43 / 88.61	71.81 / 54.00	24.2
	GAN-synthesis [30]	85.75 / 93.75	72.95 / 56.94	25.5
	STUD (ours)	**81.14 / 80.77**	**74.82 / 69.52**	27.2

Table 7. Comparison with competitive out-of-distribution detection methods. All baseline methods are based on a model trained on ID data only using RegNetX-4.0GF as the backbone. ↑ indicates larger values are better, and ↓ indicates smaller values are better. All values are percentages. **Bold** numbers are superior results.

results are shown in Table 7.

From Table 7, we demonstrate that STUD is effective on alternative neural network architectures. In particular, using RegNet [49] as backbone yields better OOD detection performance compared with the baselines. Moreover, we show that STUD achieves stronger OOD detection performance while preserving or even slightly increasing the object detection accuracy on ID data (measured by mAP). This is in contrast with CSI, which displays significant degradation, with mAP decreasing by 3% on Youtube-VIS.

F. Additional related work

Video anomaly detection (VAD) aims to identify anomalous events on both the object level [7, 22, 68] and frame level [35, 39, 51] by techniques such as skeleton trajectory modeling [43], weakly supervised learning [69], attention [47], temporal pose graph [38], self-supervised learning [10] and autoencoders [3]. Compared with STUD, the anomalies in VAD do not necessarily have different semantics from the ID training data. Moreover, none of the approaches considered synthesizing unknowns with the help of videos or energy-based model regularization.

G. Additional visualization examples

We provide additional visualization of the detected objects on different OOD datasets with models trained on different in-distribution datasets. The results are shown in Figures 7-10.

Figure 7. Additional visualization of detected objects on the OOD images (from MS-COCO) by a vanilla Faster-RCNN (*top*) and `STUD` (*bottom*). The in-distribution is BDD100K dataset. **Blue**: Objects detected and classified as one of the ID classes. **Green**: OOD objects detected by `STUD`, which reduce false positives among detected objects.

Figure 8. Additional visualization of detected objects on the OOD images (from nuImages) by a vanilla Faster-RCNN (*top*) and STUD (*bottom*). The in-distribution is BDD100K dataset. **Blue**: Objects detected and classified as one of the ID classes. **Green**: OOD objects detected by STUD, which reduce false positives among detected objects.

Figure 9. Additional visualization of detected objects on the OOD images (from MS-COCO) by a vanilla Faster-RCNN (*top*) and STUD (*bottom*). The in-distribution is Youtube-VIS dataset. **Blue**: Objects detected and classified as one of the ID classes. **Green**: OOD objects detected by STUD, which reduce false positives among detected objects.

Figure 10. Additional visualization of detected objects on the OOD images (from nuImages) by a vanilla Faster-RCNN (*top*) and `STUD` (*bottom*). The in-distribution is Youtube-VIS dataset. **Blue**: Objects detected and classified as one of the ID classes. **Green**: OOD objects detected by `STUD`, which reduce false positives among detected objects.

References

[1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 11618–11628, 2020. 5

[2] Tianshi Cao, Chinwei Huang, David Yu-Tung Hui, and Joseph Paul Cohen. A benchmark of medical out of distribution detection. *CoRR*, abs/2007.04250, 2020. 1

[3] Yunpeng Chang, Zhigang Tu, Wei Xie, and Junsong Yuan. Clustering driven deep autoencoder for video anomaly detection. In *European Conference on Computer Vision, ECCV 2020*, pages 329–345, 2020. 13

[4] Kumari Deepshikha, Sai Harsha Yelleni, P. K. Srijith, and C. Krishna Mohan. Monte carlo dropblock for modelling uncertainty in object detection. *CoRR*, abs/2108.03614, 2021. 8

[5] Akshay Raj Dhamija, Manuel Günther, and Terrance E. Boult. Reducing network agnostophobia. In *Advances in Neural Information Processing Systems 31, NeurIPS 2018*, pages 9175–9186, 2018. 8

[6] Akshay Raj Dhamija, Manuel Günther, Jonathan Ventura, and Terrance E. Boult. The overlooked elephant of object detection: Open set. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, pages 1010–1019, 2020. 8

[7] Keval Doshi and Yasin Yilmaz. Any-shot sequential anomaly detection in surveillance videos. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020*, pages 4037–4042, 2020. 13

[8] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. *Proceedings of the International Conference on Learning Representations*, 2022. 1, 4, 8

[9] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059, 2016. 8

[10] Mariana-Iuliana Georgescu, Antonio Barbalau, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. Anomaly detection in video via self-supervised and multi-task learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021*, pages 12742–12752, 2021. 13

[11] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. https://github.com/facebookresearch/detectron, 2018. 5

[12] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Zero-shot detection via vision and language knowledge distillation. *arXiv preprint arXiv:2104.13921*, 2021. 8

[13] David Hall, Feras Dayoub, John Skinner, Haoyang Zhang, Dimity Miller, Peter Corke, Gustavo Carneiro, Anelia Angelova, and Niko Sünderhauf. Probabilistic object detection: Definition and evaluation. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, pages 1020–1029, 2020. 8

[14] Ali Harakeh and Steven L. Waslander. Estimating and evaluating regression predictive uncertainty in deep object detectors. In *International Conference on Learning Representations*, 2021. 8

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pages 770–778, 2016. 5

[16] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 41–50, 2019. 8

[17] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations, ICLR 2017*, 2017. 1, 5, 6, 8, 12, 13

[18] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations*, 2019. 3, 8

[19] Tobias Hinz, Stefan Heinrich, and Stefan Wermter. Generating multiple objects at spatially distinct locations. In *International Conference on Learning Representations, ICLR 2019*, 2019. 8

[20] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10951–10960, 2020. 1, 5, 6, 8, 12, 13

[21] Rui Huang, Andrew Geng, and Yixuan Li. On the importance of gradients for detecting distributional shifts in the wild. In *Advances in Neural Information Processing Systems*, 2021. 8

[22] Radu Tudor Ionescu, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu, and Ling Shao. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pages 7842–7851, 2019. 13

[23] K. J. Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N. Balasubramanian. Towards open world object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*, 2021. 2, 6, 8

[24] K. J. Joseph, Jathushan Rajasegaran, Salman H. Khan, Fahad Shahbaz Khan, Vineeth Balasubramanian, and Ling Shao. Incremental object detection via meta-learning. *arXiv preprint arXiv:2003.08798*, 2020. 8

[25] Dahun Kim, Tsung-Yi Lin, Anelia Angelova, In So Kweon, and Weicheng Kuo. Learning open-world object proposals without learning to classify. *CoRR*, abs/2108.06753, 2021. 8

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015*, 2015. 12

[27] Diederik P. Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems 2018, NeurIPS 2018*, pages 10236–10245, 2018. 8

[28] Polina Kirichenko, Pavel Izmailov, and Andrew G Wilson. Why normalizing flows fail to detect out-of-distribution data. *Advances in Neural Information Processing Systems*, 33, 2020. 8

[29] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, pages 6402–6413, 2017. 1

[30] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *International Conference on Learning Representations*, 2018. 2, 5, 6, 8, 12, 13

[31] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems*, pages 7167–7177, 2018. 1, 5, 6, 8, 12, 13

[32] Yi Li and Nuno Vasconcelos. Background data resampling for outlier-aware classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 13215–13224, 2020. 8

[33] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations, ICLR 2018*, 2018. 1, 5, 6, 8, 12, 13

[34] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755, 2014. 5

[35] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection - A new baseline. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pages 6536–6545, 2018. 13

[36] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 2020. 1, 4, 5, 6, 7, 8, 12, 13

[37] Xialei Liu, Hao Yang, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Multi-task incremental learning for object detection. *CoRR*, abs/2002.05347, 2020. 8

[38] Amir Markovitz, Gilad Sharir, Itamar Friedman, Lihi Zelnik-Manor, and Shai Avidan. Graph embedded pose clustering for anomaly detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 10536–10544, 2020. 13

[39] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pages 935–942, 2009. 13

[40] Dimity Miller, Feras Dayoub, Michael Milford, and Niko Sünderhauf. Evaluating merging strategies for sampling-based uncertainty techniques in object detection. In *International Conference on Robotics and Automation, ICRA 2019*, pages 2348–2354, 2019. 8

[41] Dimity Miller, Lachlan Nicholson, Feras Dayoub, and Niko Sünderhauf. Dropout sampling for robust object detection in open-set conditions. In *2018 IEEE International Conference on Robotics and Automation, ICRA 2018*, pages 1–7, 2018. 8

[42] Sina Mohseni, Mandar Pitale, JBS Yadawa, and Zhangyang Wang. Self-supervised learning for generalizable out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5216–5223, 2020. 8

[43] Romero Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Reda Mansour, and Svetha Venkatesh. Learning regularity in skeleton trajectories for anomaly detection in videos. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pages 11996–12004, 2019. 13

[44] Peyman Morteza and Yixuan Li. Provable guarantees for understanding out-of-distribution detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022. 8

[45] Eric T. Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Görür, and Balaji Lakshminarayanan. Do deep generative models know what they don't know? In *International Conference on Learning Representations, ICLR 2019*, 2019. 8

[46] Donald F. Othmer. Symposium on distillation separation of water from acetic acid by azeotropic distillation. *Industrial & Engineering Chemistry*, 27(3):250–255, 1935. 2

[47] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 14360–14369, 2020. 13

[48] Juan-Manuel Pérez-Rúa, Xiatian Zhu, Timothy M. Hospedales, and Tao Xiang. Incremental few-shot object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 13843–13852, 2020. 8

[49] Ilija Radosavovic, Raj Prateek Kosaraju, Ross B. Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 10425–10433, 2020. 12, 13

[50] Shafin Rahman, Salman H. Khan, and Fatih Porikli. Zero-shot object detection: Joint recognition and localization of novel concepts. *International Journal of Computer Vision*, 128(12):2979–2999, 2020. 8

[51] Mahdyar Ravanbakhsh, Moin Nabi, Hossein Mousavi, Enver Sangineto, and Nicu Sebe. Plug-and-play CNN for crowd motion analysis: An application in abnormal event detection. In *2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, pages 1689–1698, 2018. 13

[52] Jie Ren, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark A. DePristo, Joshua V. Dillon, and Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. In *Advances in Neural Information Processing Systems 2019, NeurIPS 2019*, pages 14680–14691, 2019. 8

[53] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015. 5

[54] Chandramouli Shama Sastry and Sageev Oore. Detecting out-of-distribution examples with gram matrices. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020*, pages 8491–8501, 2020. 5, 6, 8, 12, 13

[55] Robin Schirrmeister, Yuxuan Zhou, Tonio Ball, and Dan Zhang. Understanding anomaly detection with deep invertible networks through hierarchies of distributions and features. In *Advances in Neural Information Processing Systems 33, NeurIPS 2020*, 2020. 8

[56] Vikash Sehwag, Mung Chiang, and Prateek Mittal. SSD: A unified framework for self-supervised outlier detection. In *International Conference on Learning Representations*, 2021. 8

[57] Joan Serrà, David Álvarez, Vicenç Gómez, Olga Slizovskaia, José F. Núñez, and Jordi Luque. Input complexity and out-of-distribution detection with likelihood-based generative models. In *International Conference on Learning Representations, ICLR 2020*, 2020. 8

[58] Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. In *Advances in Neural Information Processing Systems*, 2021. 8

[59] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *Advances in Neural Information Processing Systems*, 2020. 1, 5, 6, 8, 12, 13

[60] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. Conditional image generation with pixelcnn decoders. In *Advances in neural information processing systems*, pages 4790–4798, 2016. 8

[61] Haoran Wang, Weitang Liu, Alex Bocchieri, and Yixuan Li. Can multi-label classification networks know what they don't know? *Advances in Neural Information Processing Systems*, 2021. 8

[62] Xin Wang, Thomas E. Huang, Benlin Liu, Fisher Yu, Xiaolong Wang, Joseph E. Gonzalez, and Trevor Darrell. Robust object detection via instance-level temporal cycle confusion. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV*, 2021. 8

[63] Zhisheng Xiao, Qing Yan, and Yali Amit. Likelihood regret: An out-of-distribution detection score for variational auto-encoder. *Advances in Neural Information Processing Systems*, 33, 2020. 8

[64] Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. *CoRR*, abs/2108.11941, 2021. 8

[65] Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2021. 8

[66] Linjie Yang, Yuchen Fan, and Ning Xu. Video instance segmentation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*, pages 5187–5196, 2019. 2, 4, 5

[67] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 2633–2642, 2020. 1, 2, 5

[68] Guang Yu, Siqi Wang, Zhiping Cai, En Zhu, Chuanfu Xu, Jianping Yin, and Marius Kloft. Cloze test helps: Effective video anomaly detection via learning to complete video events. In *28th ACM International Conference on Multimedia*, pages 583–591, 2020. 13

[69] Muhammad Zaigham Zaheer, Arif Mahmood, Marcella Astrid, and Seung-Ik Lee. CLAWS: clustering assisted weakly supervised learning with normalcy suppression for anomalous event detection. In *16th European Conference on Computer Vision, ECCV 2020*, pages 358–376, 2020. 13

[70] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *6th International Conference on Learning Representations, ICLR 2018*, 2018. 6

[71] Jingyang Zhang, Nathan Inkawhich, Yiran Chen, and Hai Li. Fine-grained out-of-distribution detection with mixup outlier exposure. *CoRR*, abs/2106.03917, 2021. 8