Contrastive Feature Decoupling for Weakly-Supervised Disease Detection

Jhih-Ciang Wu[1,2], Ding-Jie Chen[1(✉)], and Chiou-Shann Fuh[2]

[1] Institute of Information Science, Academia Sinica, Taipei, Taiwan
djchen.tw@gmail.com
[2] National Taiwan University, Taipei, Taiwan

Abstract. Machine learning based Computer-Aided Diagnosis (CAD) aims to assist clinicians in the pathological diagnosis process. While dealing with video pathological diagnosis such as colonoscopy polyp detection, the recent SOTA method employs Weakly-supervised Video Anomaly Detection (WVAD) in the Multiple Instance Learning (MIL) scenarios to concern the temporal correlation within data and to formulate the concept of the interest disease simultaneously. Such a MIL-based WVAD method leverages video-level annotations to detect frame-level diseases and shows promising results. This paper casts the video pathological diagnosis as a MIL-based WVAD task and introduces Contrastive Feature Decoupling (CFD) network to decouple normal and abnormal feature ingredients per snippet. With such decoupled features, we are able to highlight the abnormal feature ingredients for accurately reasoning the disease score per snippet. The core components within our CFD model are the memory bank and contrastive loss. The former is used to learn atoms for representing normal features, and the latter is used to encourage our model to gain robust disease detection. We demonstrate that our CFD network is achieving new SOTA performance on the existing Polyp dataset and the introduced PANDA-MIL dataset. Our dataset are available at https://github.com/Jhih-Ciang/PANDA-MIL.

Keywords: Disease detection · Multiple instance learning · Weakly-supervised learning

1 Introduction

Computer-aided diagnosis utilizes machine learning techniques to conduct a pathological diagnosis concerning biomedical imaging data collected from various pathological modalities, such as computed tomography [19], magnetic resonance imaging [11], ultrasound [23], and angiography [9]. With the assistance of CAD techniques, the clinicians merely need to check the possible pathological regions narrowed down by computer-aided diagnosis method, significantly reducing the entire diagnosis time. With the recent success of deep learning, researchers are

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_25.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 252–261, 2023.
https://doi.org/10.1007/978-3-031-43904-9_25

able to raise the reliability of CAD methods and assist clinicians in diagnosing more complex clinical tasks. However, a reliable machine learning-based CAD method usually relies on the supervision of abundant annotated training data. Yet diseased pathological data are rare and diverse, and acquiring reliable pathological annotations are labor-intensive and expertise-required. As a result, the difficulty of data collection restricts the development of the supervised CAD.

Due to the difficulty of acquiring the abundant annotated training data, the current SOTA method, *i.e.*, CSM [14], proposes a MIL-based WVAD manner to specifically tackle one specific disease detection task, *i.e.*, colorectal cancer diagnosis via colonoscopy. Considering the case of colonoscopy, the CSM's anomaly detection setting is used to handle the rare and diverse diseased pathological data by commonly assuming that only video-level annotations are available for training. Furthermore, its video setting concerns the temporal correlation within data. The setting of such MIL-based weakly supervision prevents the need for abundant annotated training data by assuming that merely the video-level annotations, including normal and diseased ones, are available for training.

Similar to the previous MIL-based WVAD methods [4,13,14], our model assumes all training snippets (consecutive video frames) within a non-diseased video are all normal snippets, yet each diseased video has at least one abnormal snippet. Furthermore, the proposed contrastive feature decoupling network treats disease detection as an out-of-distribution task. Precisely, our CFD learns a memory bank to learn normal features. A snippet that failed to be well reconstructed with these normal features is considered diseased. On the other hand, the residual of a snippet and its reconstructed one reveals the snippet's abnormal ingredients. Consequently, we are able to decouple each snippet as normal ingredients (reconstructed parts) and abnormal ingredients (residual parts) by leveraging the memory bank. With the decoupled snippet-level feature ingredients, our CFD employs both the normal and abnormal feature ingredients via a contrastive learning paradigm to concurrently optimize video-level and snippet-level disease scores for pursuing more accurate detection.

To assess the proposed contrastive feature decoupling network, we conduct experiments on two datasets, *i.e.*, Polyp and PANDA-MIL. The main contributions are summarized as follows.

- Our contrastive feature decoupling network learns a memory bank to learn normal atoms for decoupling each snippet as normal and diseased feature ingredients as opposite contrastive learning samples. Such a feature decoupling intrinsically fits the contrastive learning paradigm for optimizing MIL objectives on bags and instances. The ablation study shows the decoupled diseased ingredients enable accurate disease detection, and the accompanied contrastive learning paradigm provides further improvement.
- We introduce the new biomedical imaging dataset for prostate cancer detection, *i.e.*, PANDA-MIL. The dataset is organized to fit the MIL-based WVAD task, including video-level annotations and video-wise format data.
- We demonstrate the generalization of CFD by achieving new SOTA performance on two biomedical imaging datasets concerning different pathological modalities, *i.e.*, the colonoscopy videos in Polyp and prostate tissue biopsies in PANDA-MIL.

2 Related Work

2.1 Disease Detection

With the evolution of artificial intelligence techniques in the past decades, deep learning has shown its potential for computer-aided diagnosis of various symptoms [6,8,14,17,22]. For example, Li *et al.* [8] established a large-scale attention-based database and designed a specialized model using retinal fundus images for detecting glaucoma. Windsor *et al.* [17] constructed a transformer-based model to detect spinal cancer for MRI scans. More recently, Tian *et al.* [14] formulated polyp detection in a WVAD scheme while tackling polyp detection using colonoscopy videos to search colon polyps in the temporal sequence. Unlike previous methods of handling one specific pathological modality, we simultaneously address disease detection across pathological modalities of colonoscopy videos and prostate tissue biopsies using our contrastive feature decoupling network.

2.2 Contrastive Learning

The characteristics of self-supervised learning are defining the proxy objective or addressing pretext tasks using pseudo labels for the unlabeled instances. One popular branch is contrastive learning which shows a remarkable ability to obtain the desired semantic representation from various perspectives. For example, CoLA [21] tackled action localization by proposing snippet contrast loss to refine the feature representations of hard snippets according to the easily discriminative snippets. CSM [14] borrowed the concept from CoLA and defined the hard/easy snippets for representing normal/abnormal from colonoscopy videos. They empirically selected hard snippets based on the transitional edge and missed disease snippets, such as an occlusive polyp. In this work, we employ a similar contrastive learning strategy as CSM while preventing their rule-based contrastive training samples selection by intrinsically leveraging the decoupled features derived from our feature decoupling process via memory bank.

3 Method

We aim to design a MIL-based WVAD model for tackling disease detection across different pathological modalities. Our model contains an offline trained memory bank to store feature atoms before the CFD training procedure, which associates a contrastive loss to boost the model performance using decoupled features per instance. Winin the MIL scenario, our model employs two classifiers to enable reasoning of the disease scores at instance and bag levels. Figure 1 overviews the working flow of the proposed contrastive feature decoupling network.

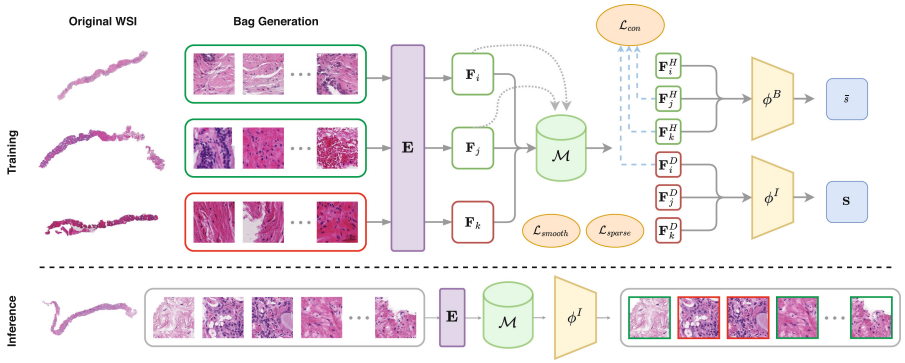

Fig. 1. Illustration of our contrastive feature decoupling (CFD) network. Given a **bag** composing multiple *instances* (also known as a **video** comprising *snippets* in Polyp or a **WSI** consisting of *patches* in PANDA-MIL), our CFD employs I3D as feature encoder $\mathbf{E}$ to generate bag-level feature $\mathbf{F}$, which subsequently decoupled as the normal proxy $\mathbf{F}^H$ and abnormal proxy $\mathbf{F}^D$ through the offline trained memory bank $\mathcal{M}$. The gray dotted arrows indicate the required normal features for constructing $\mathcal{M}$. Besides the regular MIL-based losses, *i.e.*, $\mathcal{L}_{sep}$ and $\mathcal{L}_{cls}$, we employ auxiliary losses of $\mathcal{L}_{smooth}$ and $\mathcal{L}_{sparse}$ to regularize the decoupled features within each bag and employ contrastive loss $\mathcal{L}_{con}$ to regularize the opposite decoupled features across bags. The notations ϕ^B and ϕ^I denote the bag-level and instance-level classifiers, respectively. The green box indicates a normal bag/instance, while the red box represents abnormal ones.

3.1 Memory Bank Construction

Given dataset $\mathcal{D}$ comprising normal sub-dataset $\mathcal{D}_0$ and abnormal sub-dataset $\mathcal{D}_1$, we first encode all instances per bag $\mathbf{B} \in \mathcal{D}$ into instance-level feature set $\mathbf{F} = \{f^t\}_{t=1}^T \in \mathbb{R}^{T\times C}$ via a pre-trained feature extractor $\mathbf{E}$. That is, $\mathbf{F} = \mathbf{E}(\mathbf{B})$, T is the number of instances, and C represents the instance-level feature dimension. We then collect all normal instance-level features $f \in \mathbb{R}^{1\times C}$ from $\mathcal{D}_0$ to learn the memory bank $\mathcal{M}$ by using the dictionary learning technique [7] via

$$\underset{\mathcal{M},\{\mathbf{w}_t\}}{\operatorname{argmin}} \sum_{B\in\mathcal{D}_0} \sum_{t=1}^{T} (\|f^t - \mathcal{M}\mathbf{w}_t\|^2 + \lambda\|\mathbf{w}_t\|_0)\,, \tag{1}$$

where $\mathcal{D}_0$ is the normal sub-dataset collected from the training split, $\mathbf{w}_t$ is the learned weights within the memory bank learning process, and λ is a hyperparameter to constrain the memory bank sparsity.

3.2 Contrastive Feature Decoupling

With the learned normal instance features stored in the memory bank $\mathcal{M}$, we are able to reconstruct a normal version for any given bag-level feature $\mathbf{F}$. Such a normal version is denoted as a normal-like feature $\mathbf{F}^H$. To this end, we reconstruct a given $\mathbf{F}$ concerning $\mathcal{M}$ via the following equation

$$\mathbf{F}^H = \sigma\left(\phi^q(\mathbf{F})\phi^k(\mathcal{M})^\intercal\right)\phi^v(\mathcal{M})\,, \tag{2}$$

where σ stands for the softmax function; ϕ^q, ϕ^k, and ϕ^v respectively represent the query, key, and value linear projections, as introduced in the self-attention framework [15]. To make a robust learning process by mining the hard and easy instances for the subsequent contrastive loss, CSM [14] designed a rule-based instances selection concerning the transitional edge and missed disease instances, such as occlusion or invisibility from the polyp. Different from the CSM model, we generate the disease-like feature $\mathbf{F}^D$ referring to $\mathbf{F}^H$ as follows:

$$\mathbf{F}^D = \omega \odot \mathbf{F}\,, \tag{3}$$

where $\omega \in \mathbb{R}^{1\times C}$ is the weight for reweighting $\mathbf{F}$ by channel-wise multiplication.

For depicting the degree of disease/abnormal to estimate the channel-wise weight ω, we consider attending to the distant features with respect to the normal ones, *i.e.*, $\mathbf{F}^H$. In practice, we estimate the degree of disease/abnormal based on the difference between $\mathbf{F}$ and $\mathbf{F}^H$ by

$$\omega = \text{sigmoid}(\phi^d(\mathcal{G}(\Psi(\mathbf{F}) - \Psi(\mathbf{F}^H))))\,, \tag{4}$$

where ϕ^d, $\mathcal{G}$, and Ψ are linear projections, global average pooling, and the multi-scale temporal network [13], respectively.

With the decoupled features $\mathbf{F}^H$ and $\mathbf{F}^D$, our MIL-based instance-level classifier aims to carry out the discriminating decision, and the bag-level classifier seeks the prediction as fitting annotation y as possible. Following recent MIL-based methods, we select top-K instances of normal and abnormal in each bag to form the separation loss as

$$\mathcal{L}_{sep} = \sum_{\substack{\mathbf{F}_i^D \in \mathcal{D}_0 \\ \mathbf{F}_j^D \in \mathcal{D}_1}} \sum_{K} \left(\left|\delta - \|\{\mathbf{F}_i^D\}_K\|_2\right| + \|\{\mathbf{F}_j^D\}_K\|_2\right), \tag{5}$$

where δ is the hyperparameter for constrained margin and $\{\cdot\}_K$ is the operator that selects top-K instances. The other common loss used in the recent MIL-based works is the classification loss building upon the binary cross entropy

$$\mathcal{L}_{cls} = BCE(\{\mathbf{S}\}_K, y) + BCE(\bar{s}, y)\,, \tag{6}$$

where $\mathbf{S} = \phi^I(\mathbf{F}^D)$ represents the instance-level prediction inferred by an instance-level classifier ϕ^I and $\bar{s} = \phi^B(\mathcal{G}((1-\omega)\odot\mathbf{F}^H))$ means the bag-level prediction resulting from a bag-level classifier ϕ^B.

3.3 Regularization

Motivated by the recent WVAD methods [4,13,14,18], which adopt auxiliary losses to regularize the learning procedure, we consider the conventional regularization losses, such as temporal smoothness and sparsity, as follows

$$\mathcal{L}_{smooth} = \sum_{i=1}^{\|\mathcal{D}\|} \frac{1}{T-1} \sum\nolimits_{t=1}^{T-1} \|f_i^{t+1} - f_i^t\|^2\,, \quad \mathcal{L}_{sparse} = \sum_{i=1}^{\|\mathcal{D}\|} \frac{1}{T} \sum\nolimits_{t=1}^{T} \|f_i^t\|\,. \tag{7}$$

By using the decoupled features $\mathbf{F}^H$ and $\mathbf{F}^D$, we are ready to regularize the opposite decoupled features across bags with the aid of a contrastive loss. An expected contrastive loss aims to make our model attract features from the same category while distracting the features from distinct classes. In practice, we formulate such a contrastive loss by

$$\mathcal{L}_{con} = \sum_{\mathbf{F}_i^D, \mathbf{F}_j^H \in \mathcal{D}_0} \log \frac{\exp\left[\frac{1}{\tau}(\mathbf{F}_i^D)^\intercal \mathbf{F}_j^H\right]}{\exp\left[\frac{1}{\tau}(\mathbf{F}_i^D)^\intercal (\mathbf{F}_j^H)\right] + \sum_{\mathbf{F}_k^H \in \mathcal{D}_1} \exp\left[\frac{1}{\tau}(\mathbf{F}_i^D)^\intercal (\mathbf{F}_k^H)\right]}, \quad (8)$$

where τ denotes the temperature parameter in the normalized temperature-scaled loss. Notice that (8) is simplified for the sake of clarity. A complete objective should consider the symmetric form by switching $\mathcal{D}_1$ and $\mathcal{D}_0$ in (8).

4 Experiments

4.1 Dataset and Metric

We evaluate our model against SOTAs on the existing Polyp [14] dataset and the PANDA-MIL dataset introduced in this work. We employ the same evaluation criteria as the previous work for a fair comparison. Please refer to the supplementary material for the statistics of the two datasets.

Polyp. This dataset collects colonoscopy videos from Hyper-Kvasir [1] and LDPolypVideo [10]. Its training split contains 163 videos of video-level annotations, and the testing split includes 90 videos of frame-level annotations.

PANDA-MIL. The Prostate cANcer graDe Assessment (PANDA) challenge [2] comprises over 10K whole-slide images (WSIs) of digitized hematoxylin and eosin-stained biopsies originating from Radboud University Medical Center and Karolinska Institute. PANDA-MIL collects the eosin-stained biopsies with region-based masks indicating the benign (normal) and cancerous (abnormal) tissue, combined by stroma and epithelium. To fit the MIL-based WVAD task, we non-overlapped partition each WSI (bag) into patches (instances) and only keep those patches comprising tissue over the 50% patch size. Each kept patch gets its patch-level annotations from PANDA, and a WSI comprising any abnormal patch is treated as an abnormal WSI. In sum, PANDA-MIL's training split contains 3,925 bags of bag-level annotations, and the testing split includes 975 bags of instance-level annotations.

Metric. We follow the previous methods [4,13,14] to employ the instant-level Area Under Curve (AUC) and the Average Precision (AP) for a fair comparison. The larger values of both metrics mean better disease detection performance.

Table 1. Comparison with AUC and AP metrics on Polyp and PANDA-MIL datasets.

Method	Publication	Polyp		PANDA-MIL	
		AUC	AP	AUC	AP
MIST [4]	CVPR'21	94.53	72.85	82.84	75.45
RTFM [13]	ICCV'21	96.30	77.96	85.12	78.17
S3R [18]	ECCV'22	98.32	86.21	86.19	78.33
CSM [14]	MICCAI'22	98.41	86.63	76.52	73.12
CFD		**99.51**	**88.13**	**87.28**	**80.78**

4.2 Implementation Details

All the evaluated methods in the experiment used the same feature encoder, *i.e.*, I3D [3] pre-trained on Kinetics-400 [5], for a fair comparison. Our method is trained using Adam optimizer with the learning rate of 0.001, batch size 32, and 200 epochs. Each bag/video is encoded into $T = 32$ snippets among both datasets via linear interpolation.

4.3 Comparison Results

Table 1 shows the compared results of our CFD model against recent WVAD methods [4,13,14,18] for tackling the disease detection task. The results in Table 1 demonstrate that our CFD consistently outperforms all the other methods on two datasets. Precisely, our model achieves the new SOTA by 1.1% AUC and 1.5% AP improvements on the Polyp dataset and 1.09% AUC and 2.45% AP improvements on the PANDA-MIL dataset. Please refer to the supplementary material for the completed results, including more WVAD methods [12,16,20,24].

Figure 2 visualizes one disease detection result of our CFD model on the PANDA-MIL dataset. The disease score per instance/patch predicted by our method is close to the ground-truth annotations, in which the clear margin between cancerous and benign validates the robust prediction of the proposed CFD model.

4.4 Ablation Study

We analysis on why the CFD network performs better than other methods listed in Table 1 by ablating the contributed components in CFD. The ablation study in Table 2 is conducted on the PANDA-MIL dataset to evaluate the effectiveness of the memory bank and loss functions in our model. Row one in Table 2 indicates our CFD model without regularization, yet it has shown better AUC values than other methods besides the S3R. While employing the three loss functions for regularization, as described in Sect. 3.3, each loss function shows its improvement in our model performance. The contrastive loss contributes the most to AUC improvement, enabling our model to achieve the SOTA performance.

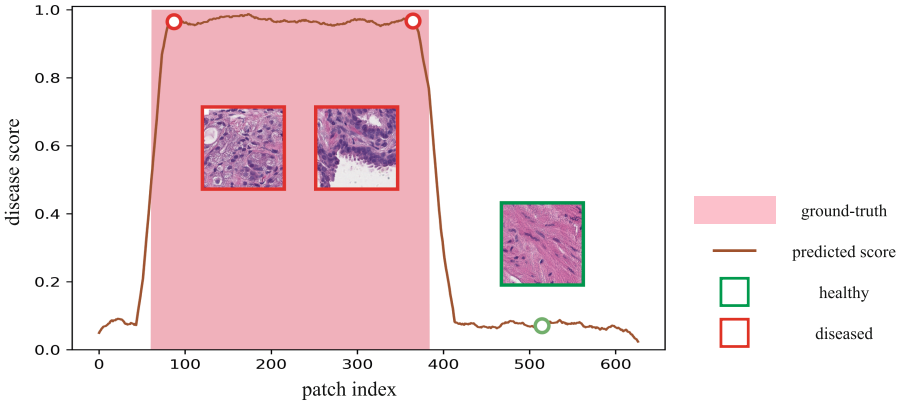

Fig. 2. Qualitative results of our CFD model on one testing prostate tissue biopsy of the PANDA-MIL dataset. Our disease detection results are close to the ground-truths.

Table 2. Ablation study on model components with the AUC metric on PANDA-MIL.

$\mathcal{M}$	$\mathcal{L}_{sparse}$	$\mathcal{L}_{smooth}$	$\mathcal{L}_{con}$	AUC
✓				85.36
✓	✓			85.79
✓	✓	✓		86.15
✓	✓	✓	✓	**87.28**

5 Conclusion

This paper casts disease detection as a MIL-based WVAD task and introduces Contrastive Feature Decoupling (CFD) network to learn a memory bank boosted with contrastive learning. With the learned feature atoms stored in the memory bank, our contrastive feature decoupling is able to decouple each snippet as normal and abnormal proxies. Further, the decoupled abnormal proxies highlight the abnormal feature ingredients for better reasoning the disease score. Our feature decoupling intrinsically fits the contrastive learning paradigm to define opposite training samples for model optimization. Besides, we introduce a new dataset of prostate cancer detection, i.e., PANDA-MIL, to provide a biomedical imaging dataset concerning a different pathological modality. Experiments demonstrate that our CFD network achieves new SOTA performance on the Polyp and PANDA-MIL datasets, indicating that our method effectively addresses the disease detection task across different pathological modalities.

Acknowledgement. This research was supported by National Science and Technology Council of Taiwan, R.O.C., under Grants NSTC 112-2221-E-002-189-MY2 and MOST 111-2221-E-002-174.

References

1. Borgli, H., et al.: Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Sci. Data **7**(1), 283 (2020)
2. Bulten, W., et al.: The PANDA challenge: prostate cANcer graDe assessment using the Gleason grading system (2020). https://doi.org/10.5281/zenodo.3715938
3. Carreira, J., Zisserman, A.: Quo vadis, action recognition? A new model and the kinetics dataset. In: CVPR, pp. 6299–6308 (2017)
4. Feng, J.C., Hong, F.T., Zheng, W.S.: MIST: multiple instance self-training framework for video anomaly detection. In: CVPR, pp. 14009–14018 (2021)
5. Kay, W., et al.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017)
6. Kazeminia, S., Sadafi, A., Makhro, A., Bogdanova, A., Albarqouni, S., Marr, C.: Anomaly-aware multiple instance learning for rare anemia disorder classification. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13438, pp. 341–350. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16452-1_33
7. Kreutz-Delgado, K., Murray, J.F., Rao, B.D., Engan, K., Lee, T.W., Sejnowski, T.J.: Dictionary learning algorithms for sparse representation. Neural Comput. **15**(2), 349–396 (2003)
8. Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: Attention based glaucoma detection: a large-scale database and CNN model. In: CVPR, pp. 10571–10580 (2019)
9. Ma, X.Y., et al.: DSP-NET: deeply-supervised pseudo-siamese network for dynamic angiographic image matching. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 44–53. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_5
10. Ma, Y., Chen, X., Cheng, K., Li, Y., Sun, B.: LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 387–396. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_37
11. Shao, W., et al.: Weakly supervised registration of prostate MRI and histopathology images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 98–107. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_10
12. Sultani, W., Chen, C., Shah, M.: Real-world anomaly detection in surveillance videos. In: CVPR, pp. 6479–6488 (2018)
13. Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J.W., Carneiro, G.: Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In: ICCV, pp. 4975–4986 (2021)
14. Tian, Y., et al.: Contrastive transformer-based multiple instance learning for weakly supervised polyp frame detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 88–98. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_9
15. Vaswani, A., et al.: Attention is all you need. In: NIPS, pp. 5998–6008 (2017)
16. Wan, B., Fang, Y., Xia, X., Mei, J.: Weakly supervised video anomaly detection via center-guided discriminative learning. In: ICME, pp. 1–6 (2020)
17. Windsor, R., Jamaludin, A., Kadir, T., Zisserman, A.: Context-aware transformers for spinal cancer detection and radiological grading. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 271–281. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_26

18. Wu, J.C., Hsieh, H.Y., Chen, D.J., Fuh, C.S., Liu, T.L.: Self-supervised sparse representation for video anomaly detection. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13673, pp. 729–745. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19778-9_42
19. Xia, Y., et al.: Effective pancreatic cancer screening on non-contrast CT scans via anatomy-aware transformers. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 259–269. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_25
20. Zaheer, M.Z., Mahmood, A., Astrid, M., Lee, S.-I.: CLAWS: clustering assisted weakly supervised learning with normalcy suppression for anomalous event detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12367, pp. 358–376. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58542-6_22
21. Zhang, C., Cao, M., Yang, D., Chen, J., Zou, Y.: Cola: weakly-supervised temporal action localization with snippet contrastive learning. In: CVPR, pp. 16010–16019 (2021)
22. Zhang, W., et al.: A multi-task network with weight decay skip connection training for anomaly detection in retinal fundus images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13432, pp. 656–666. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16434-7_63
23. Zhao, H., et al.: Towards unsupervised ultrasound video clinical quality assessment with multi-modality data. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13434, pp. 228–237. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_22
24. Zhong, J.X., Li, N., Kong, W., Liu, S., Li, T.H., Li, G.: Graph convolutional label noise cleaner: train a plug-and-play action classifier for anomaly detection. In: CVPR, pp. 1237–1246 (2019)