DiffULD: Diffusive Universal Lesion Detection

Peiang Zhao[1,2], Han Li[1,2], Ruiyang Jin[1,2], and S. Kevin Zhou[1,2,3(✉)]

[1] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China
{pazhao,hanli21,ryjin}@mail.ustc.edu.cn, skevinzhou@ustc.edu.cn
[2] Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu 215123, People's Republic of China
[3] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, People's Republic of China

Abstract. Universal Lesion Detection (ULD) in computed tomography (CT) plays an essential role in computer-aided diagnosis. Promising ULD results have been reported by anchor-based detection designs, but they have inherent drawbacks due to the use of anchors: **(i) Insufficient training target** and **(ii) Difficulties in anchor design.** Diffusion probability models (DPM) have demonstrated outstanding capabilities in many vision tasks. Many DPM-based approaches achieve great success in natural image object detection without using anchors. But they are still ineffective for ULD due to the insufficient training targets. In this paper, we propose a novel ULD method, DiffULD, which utilizes DPM for lesion detection. To tackle the negative effect triggered by insufficient targets, we introduce a novel Center-aligned bounding box (BBox) padding strategy that provides additional high-quality training targets yet avoids significant performance deterioration. DiffULD is inherently advanced in locating lesions with diverse sizes and shapes since it can predict with arbitrary boxes. Experiments on the benchmark dataset DeepLesion [32] show the superiority of DiffULD when compared to state-of-the-art ULD approaches.

**Keywords:** Universal Lesion Detection · Diffusion Model

1 Introduction

Universal Lesion Detection (ULD) in computed tomography (CT) [3,13,14, 16,20,28–30,33–37,40,44] plays an important role in computer-aided diagnosis (CAD) [42,43]. The design of detection-only instead of identifying the lesion

types in ULD [1,6,17,19,22,24,39,41] prominently decreases the difficulty of this task for a specific organ (e.g., lung, liver), but it is still challenging for lesions vary in shapes and sizes among whole human body.

Previous arts in ULD are mainly motivated by the anchor-based detection framework, *e.g.*, Faster-RCNN [21]. These studies focus on adapting the detection backbone to universally locate lesions in CT scans. For instance, Li *et al.* [16] propose the so-called MVP-Net, a multi-view FPN with a position-aware attention mechanism to assist ULD training. Yang *et al.* [36–38] propose a series of 3D feature fusion operators to incorporate context information from several adjacent CT slices for better performance. Li *et al.* [15] introduce a plug-and-play transformer block to form hybrid backbones which can better model long-distance feature dependency.
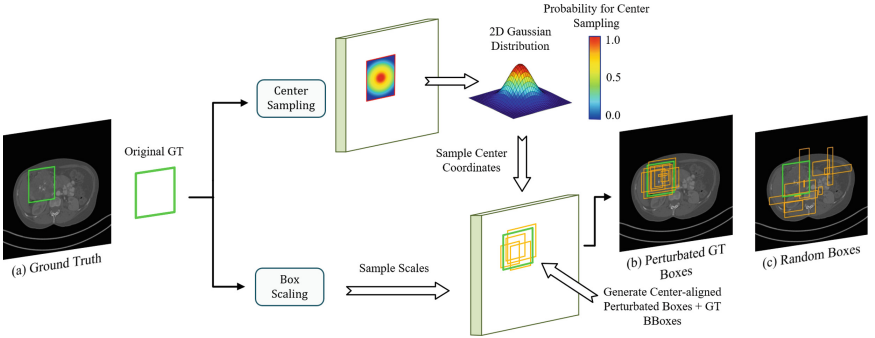
Fig. 1. Overview of the proposed center-aligned BBox padding strategy.

While achieving success, these anchor-based methods have inherent drawbacks: **(i) Insufficient training target.** In stage-1, anchor-based methods identify the positive (lesion) anchors and label them as the region of interest (RoI) based on the IoU between anchors and ground-truth (GT) bounding boxes (BBoxes). An anchor is considered positive if its IoU with any GT BBox is greater than the IoU threshold and negative otherwise [21]. The positive anchors are sufficient in natural images as they usually have many targets per image [13]. However, the number of lesions per CT scan is limited, most CT slices only contain one or two lesions (*i.e.*, detection targets in ULD) per CT slice [14]. Still applying the IoU-based anchor matching mechanism with such limited targets can lead to severe data imbalance and further hinders network convergence. Simply lowering the positive IoU threshold in the anchor-selecting mechanism can alleviate the shortage of positive anchors to some degree, but it leads to a higher false positive (FP) rate by labeling more low-IoU anchors as positive. **(ii) Difficulties in anchor design.** In anchor-based methods, the size, ratio and number of anchors are pre-defined hyper-parameters that significantly influence the detection performance [26]. Thus a proper design of anchor hyper-parameters

is of great importance. However, tuning anchor hyper-parameters is a challenging task in ULD because of the variety of lesions (target) with diverse diameters (from 0.21 to 342.5 mm). Even with a careful design, the fixed rectangle anchor boxes can be a kind of obstruction in capturing heterogeneous lesions, which have irregular shapes and vary in size.

To get rid of the drawbacks caused by the anchor mechanism, researchers resort to anchor-free detection, *e.g.*, FCOS [31] and DETR [5]. But these methods experience difficulties in achieving state-of-the-art results in ULD, as they lack the initialization of position prior provided by anchors.

Recently, the diffusion probabilistic model (DPM) [2,8,10,11,23] has demonstrated its outstanding capabilities in various vision tasks. Chen *et al.* follow the key idea of DPM and propose a noise-to-box pipeline, DiffusionDet [7], for natural image object detection. They achieved success on natural images with sufficient training targets, but still experience difficulties in dealing with tasks with insufficient training targets like ULD. This is because the DPM's denoising is a dense distribution-to-distribution forecasting procedure that heavily relied on a large number of high-quality training targets to learn targets' distribution accurately.

To address this issue, we hereby introduce a novel center-aligned BBox padding strategy in DPM detection to form a diffusion-based detector for Universal Lesion Detection, termed DiffULD.

As shown in Fig. 1, DiffULD also formulates lesion detection as a denoising diffusion process from noisy boxes to prediction boxes similar to [7], but we further introduce the center-aligned BBox padding before DiffULD's forward diffusion process to generate perturbated GT. Specifically, we add random perturbations to both scales and center coordinates of the original GT BBox, resulting in *perturbated boxes* whose centers remain aligned with the corresponding original GT BBox. Next, original GT boxes paired with these perturbated boxes are called *perturbated GT boxes* for simplicity. Finally, we feed the perturbated GT boxes to the model as the training objective during training. Compared with other training target padding methods (*e.g.*, padding with random boxes), our strategy can provide additional targets of higher quality, *i.e.*, center aligned with the original GT BBox. This approach effectively expands the insufficient training targets on CT scans, enhancing DPM's detection performance and avoiding deterioration triggered by adding random targets.

The following DPM training procedure contains two diffusion processes. i) In the forward training process, DiffULD corrupts the perturbated GT with Gaussian noise gradually to generate noisy boxes step by step. Then the model is trained to remove the noise and reconstruct the original perturbated GT boxes. ii) In the reverse inference process, the trained DiffULD can refine a set of randomly generated boxes iteratively to obtain the final detect predictions.

Our method gets rid of the drawbacks of pre-defined anchors and the deterioration of training DPM with insufficient GT targets. Besides, DiffULD is inherently advanced in locating targets with diverse sizes since it can predict with arbitrary boxes, which is an advantageous feature for detecting lesions of

irregular shapes and various sizes. To validate the effectiveness of our method, we conduct experiments against seven state-of-the-art ULD methods on the public dataset DeepLesion [32]. The results demonstrate that our method achieves competitive performance compared to state-of-the-art ULD approaches.

2 Method

In this section, we first formulate our overall detection process for DiffULD and then specify the training manner, inference process and backbone design.

2.1 Diffusion-Based Detector for Lesion Detection

Universal Lesion Detection can be formulated as locating lesions in input CT scan I_{ct} with a set of boxes predictions z_0'. For a particular box z', it can be denoted as $z' = [x_1, y_1, x_2, y_2]$, where x_1, y_1 and x_2, y_2 are the coordinates of the top-left and bottom-right corners, respectively.

We design our model based on a diffusion model mentioned in [7]. As shown in Fig. 2, our method consists of two stages, a forward diffusion (or training) process and a reverse refinement (or inference) process. In the forward process, We denote GT BBoxes as z_0 and generate corrupted training samples $z_1, z_2, ..., z_T$ for latter DiffULD training by adding Gaussian noise iteratively, which can be defined as:

$$(z_t \mid z_0) = \mathcal{N}\left(z_t \mid \sqrt{\bar{\alpha}_t} z_0, (1 - \bar{\alpha}_t) I\right) \tag{1}$$

where $\bar{\alpha}_t$ represents the noise variance schedule and $t \in \{0, 1, ..., T\}$. Subsequently, a neural network $f_\theta(z_t, t, I_{ct})$ conditioned on the corresponding CT scan I_{ct} is trained to predict z_0 from a noisy box z_T by reversing the noising process step by step. During inference, for an input CT scan I_{ct} with a set of random boxes, the model is able to refine the random boxes to get a lesion detection prediction box z_0', iteratively.

2.2 Training

In this section, we specify the training process with our novelty introduced 'Center-aligned BBox padding'.

Center-Aligned BBox Padding. As shown in Fig. 1, we utilize Center-aligned BBox padding to generate *perturbated boxes*. Then, the perturbated boxes are paired with original GT BBoxes, forming *perturbated GT boxes* which are used to generate corrupted training samples $z_1, z_2, ..., z_T$ for the latter DiffULD training by adding Gaussian noise iteratively.

For clarity, we denote $[x_c^i, y_c^i]$ as center coordinates of original GT BBox, where z^i, $[w^i, h^i]$ are the width and height of z^i.

We consider the generation in two parts: box scaling and center sampling. (i) *Box scaling:* We set a hyper-parameter $\lambda_{scale} \in (0, 1)$ for scaling. For z^i,
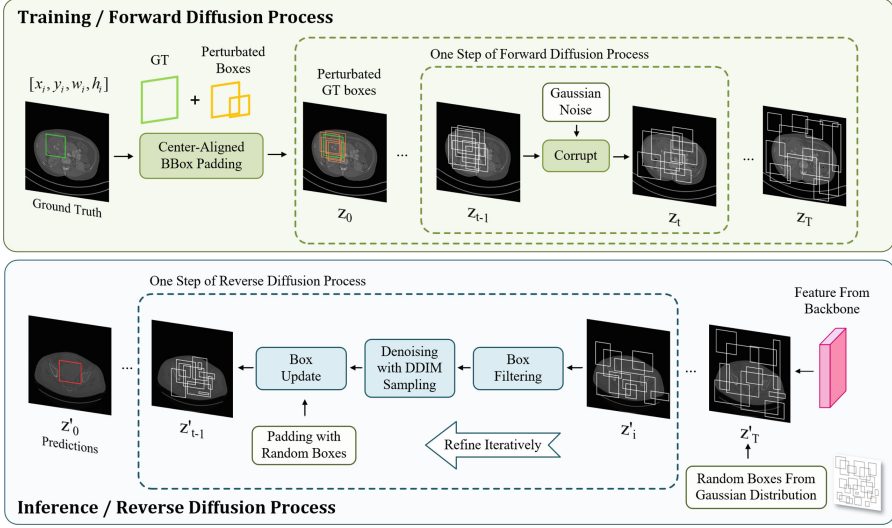
Fig. 2. Overview of DiffULD. The backbone extracts feature representation from an input CT scan. Then, the model takes the feature map and a set of noisy boxes as input and makes predictions.

The width and height of the corresponding perturbated boxes are randomly sampled in uniform distributions on $U_w \sim [(1 - \lambda_{scale})w^i, (1 + \lambda_{scale})w^i]$ and $U_h \sim [(1 - \lambda_{scale})h^i, (1 + \lambda_{scale})h^i]$. (ii) *Center sampling:* We sample the center coordinates $[x_c^{new}, y_c^{new}]$ of perturbated boxes from a 2D Gaussian distribution $\mathcal{N}$ whose probability density function can be denoted as:

$$f(x, y) = \exp\left(-\frac{\left(x - x_c^i\right)^2 + \left(y - y_c^i\right)^2}{2\sigma^2}\right), \qquad (x, y) \in z_i \qquad (2)$$

where σ is a size-adaptive parameter, which can be calculated according to the z^i's width and height:

$$\sigma = \frac{1}{6}(w^i + h^i). \qquad (3)$$

Besides, for each input CT scan I_{ct}, we collect all GT BBoxes in z and add random perturbations to them and generate multiple *perturbated boxes* for each of them. Thus the number of perturbated boxes in an image varies with its number of GT BBoxes. For better training, we fix the number of perturbated boxes as N for all training images.

As shown in Fig. 1., the perturbated boxes cluster together and their centers are still aligned with the corresponding original GT BBox. Subsequently, perturbated GT boxes z_0 are sent for corruption as the training objective.

Box Corruption. As shown in Fig. 1, we corrupt the parameters of z_0 with Gaussian noises. The noise scale is controlled by $\bar{\alpha}_t$ (in Eq. 1), which adopts the decreasing cosine scheduler in the different time step t.

Loss function. As the model generates the same number of (N) predictions for the input image, termed as a prediction set, the loss function should be set-wised [5]. Specifically, each GT is matched with the prediction by the least matching cost, and the overall training loss [5] can be represented as:

$$\mathcal{L} = \lambda_{L1box} \cdot \mathcal{L}_{L1box} + \lambda_{giou} \cdot \mathcal{L}_{giou} \tag{4}$$

where $\mathcal{L}_{L1box}$ and $\mathcal{L}_{giou}$ are the pairwise box loss. We adopt $\lambda_{L1box} = 2.0$ and $\lambda_{giou} = 5.0$.

2.3 Inference

At the inference stage, with a set of random boxes sampled from Gaussian distribution, the model does refinement step by step to obtain the final predictions $z_0^{'}$. For better performance, two key components are used:

Box Filtering. In each refinement step, the model receives a set of box proposals from the last step. As the prediction starts from arbitrary boxes and the lack of GT (lesion), most of these proposals are very far from lesions. Keeping refining them in the following steps will hinder network training. Toward efficient detection, we send the proposals to the detection head and remove the boxes whose confidential scores are lower than a particular threshold λ_{conf}. The remaining high-quality proposals are sent for followed DDIM sampling.

Box Update with DDIM Sampling. DDIM [27] is utilized to further refine the received box proposals by denoising. Next, these refined boxes are sent to the next step and start a new round of refinement. After multiple steps, final predictions are obtained.

However, we observe that if we just filter out boxes with low scores during iterative refinement, the model runs out of usable box proposals rapidly, which also leads to a deterioration in performance. Therefore, after the box updating, we add new boxes sampled from a Gaussian Distribution to the set of remaining boxes with. The number of box proposals per image is padded to the fixed number N before they are sent to the next refinement step.

2.4 Backbone Design

Our overall backbone design is identical to [16]. The input CT scan is rendered with different window widths and window levels. Then, multi-window features are extracted with a ConvNeXt-T [18] shared network and sent to 3D context feature fusion module. Subsequently, the fused feature is sent to the detector.

Multi-window Input. Most prior arts in ULD use a single and fixed window (*e.g.*, a wide window of [1024, 4096]) to render the input CT scan, which suppresses organ-specific information and makes it hard for the network to focus on the various lesions. Therefore, taking cues from [16], we introduce 3 organ-specific HU windows to highlight multiple organs of interest. Their window widths and window levels are: $W_1 = [1200, -600]$ for chest organs, $W_2 = [400, 50]$ for soft tissues and $W_3 = [200, 30]$ for abdomenal organs.

3D Context Feature Fusion. We modify the original A3D [37] DenseNet backbone for context fusion. We remove the first Conv3D Block and use the truncated network as our 3D context fusion module, which fuses the multi-window features from the last module.

3   Experiments

3.1   Settings

Our experiments are conducted on the standard ULD dataset DeepLesion [32]. The dataset contains 32,735 lesions on 32,120 axial slices from 10,594 CT studies of 4,427 unique patients. We use the official data split of DeepLesion which consists of 70%, 15%, 15% for training, validation, and test, respectively. Besides, we also evaluate the performance of 3 methods based on a revised test set from [4].

Training Details. DiffULD is trained on CT scans of size 512×512 with a batch size of 4 on 4 NVIDIA RTX Titan GPUs with 24GB memory. For hyperparameters, the threshold N for box padding is set to 300. λ_{scale} for box scaling is set to 0.4. λ_{conf} for box filtering is set to 0.5. We use the Adam optimizer with an initial learning rate of $2e-4$ and the weight decay as $1e-4$. The default training schedule is 120K iterations, with the learning rate divided by 10 at 60K and 100K iterations. Data augmentation strategies contain random horizontal flipping, rotation, and random brightness adjustment.

Evaluation Metrics. The lesion detection is classified as true positive (TP) when the IoU between the predicted and the GT BBox is larger than 0.5. Average sensitivities computed at 0.5, 1, 2, and 4 false-positives (FP) per image are reported as the evaluation metrics on the test set for a fair comparison (Table 2).

Table 1. Sensitivity (%) at various FPPI on the standard test set of DeepLesion. DKA-ULD [25] and SATr [15] are up-to-date SOTA ULD methods under the settings of 3 slices and 7 slices, respectively. The numbers in brackets indicate the performance gains of our method comparing to the previous SOTA methods under the same settings.

Methods	Slices	@0.5	@1	@2	@4	Avg.[0.5,1,2,4]
3DCE [33]	27	62.48	73.37	80.70	85.65	75.55
RetinaNet [44]	3	72.18	80.07	86.40	90.77	82.36
MVP-Net [16]	9	73.83	81.82	87.60	91.30	83.64
MULAN [34]	27	76.10	82.50	87.50	90.90	84.33
AlignShift [36]	3	73.00	81.17	87.05	91.78	83.25
A3D [37]	3	74.10	81.81	87.87	92.13	83.98
DiffusionDet [7]	3	76.71	83.49	88.21	91.70	85.03
DKA-ULD [25]	3	77.38	84.06	89.28	92.04	85.79
SATr [15]	7	81.02	86.64	90.69	93.30	87.91
DiffULD (Ours)	3	**77.84** (0.46↑)	**84.57** (0.51↑)	**89.41** (0.13↑)	**92.31** (0.40↑)	**86.03** (0.58↑)
DiffULD (Ours)	7	80.43	**87.16** (0.52↑)	**91.20** (0.51↑)	93.21	**88.00** (0.08↑)
FCOS [31]	3	56.12	67.31	73.75	81.44	69.66
CenterNet++ [9]	3	67.34	75.95	82.73	87.72	78.43
DN-DETR [12]	3	69.27	77.90	83.97	88.59	81.02

Table 2. Sensitivity (%) at various FPPI on the revised test set introduced by Lesion-Harvester [4].

Methods	Slices	@0.5	@1	@2	@4	Avg.[0.5,1,2,4]
A3D [37]	7	88.73	91.23	93.89	95.91	92.44
SATr [15]	7	91.04	93.75	95.58	96.73	94.28
DiffULD (Ours)	7	**91.65** (0.61↑)	**94.33** (0.58↑)	**95.97** (0.39↑)	**96.86** (0.13↑)	**94.70** (0.42↑)

Table 3. Ablation study of padding strategies at various FPs per image (FPPI).

Baseline	Duplicate	Gaussian	Uniform	Center-Aligned	FPPI = 0.5	FPPI = 1
✓					76.71	83.49
✓	✓				77.01	83.61
✓		✓			77.68	83.98
✓			✓		77.22	83.75
✓				✓	**77.84** (1.13↑)	**84.57** (1.08↑)

3.2   Lesion Detection Performance

We evaluate the effectiveness of DiffULD against anchor-based ULD approaches such as 3DCE [33], MVP-Net [16], A3D [37] and SATr [15] on DeepLesion. Several anchor-free natural image detection methods such as FCOS [31] and DN-DETR [12] are also introduced for comparison. We report the performance of DiffusionDet [7] trained with our proposed backbone in 2.4 as well. In addition,

we conduct an extensive experiment to explore DiffULD's potential on an revised test set of completely annotated DeepLesion volumes, introduced by Lesion-Harvester [4].

Table 1 demonstrates that our proposed DiffULD achieves favorable performances when compared to recent state-of-the-art anchor-based ULD approaches such as SATr on both 3 slices and 7 slices. It outperforms prior well-established methods such as A3D and MULAN by a non-trivial margin. This validates that with our padding strategy, the concise DPM can be utilized in general medical object detection tasks such as ULD and attain impressive performance.

3.3   Ablation Study

We provide an ablation study about our proposed approach: center-aligned BBox padding. As shown in Table 3., we compared it with various other padding strategies, including: (i) duplicating original GT boxes; (ii) padding random boxes sampled from a uniform distribution; (iii) padding random boxes sampled from a Gaussian distribution; (iv) padding with center-aligned strategy.

Our baseline method is training the diffusion model [7] directly with no box padding, using our proposed backbone in 2.4. The performance is increased by 0.30% by simply duplicating the original GT boxes. Padding random boxes following uniform and Gaussian distributions brings 0.51% and 0.91% improvement respectively. Our center-aligned padding strategy accounts for 1.13% improvement from the baseline. We attribute this performance boost to center-aligned padding's ability to provide high-quality additional training targets. It effectively expands the insufficient training targets on CT scans and enhances DPM's detection performance while avoiding deterioration triggered by adding random targets. This property is favorable for utilizing DPMs on a limited amount of GT like ULD.

4   Conclusion

In this paper, we propose a novel ULD method termed DiffULD by introducing the diffusion probability model (DPM) to Universal Lesion Detection. We present a novel center-aligned BBox padding strategy to tackle the performance deterioration caused by directly utilizing DPM on CT scans with sparse lesion BBoxes. Compared with other training target padding methods (e.g., padding with random boxes), our strategy can provide additional training targets of higher quality and boost detection performance while avoiding significant deterioration. DiffULD is inherently advanced in locating lesions with diverse sizes and shapes since it can predict with arbitrary boxes, making it a promising method for ULD. Experiments on both standard and revised DeepLesion datasets show that our proposed method can achieve competitive performance compared to state-of-the-art ULD approaches.

References

1. Baumgartner, M., Jäger, P.F., Isensee, F., Maier-Hein, K.H.: nnDetection: a self-configuring method for medical object detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 530–539. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_51

2. Boah, K., et al.: DiffuseMorph: unsupervised deformable image registration using diffusion model. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV, pp. 347–364. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19821-2_20

3. Cai, J., et al.: Deep volumetric universal lesion detection using light-weight pseudo 3D convolution and surface point regression. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 3–13. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_1

4. Cai, J., et al.: Lesion-harvester: iteratively mining unlabeled lesions and hard-negative examples at scale. IEEE Trans. Med. Imaging **40**(1), 59–70 (2020)

5. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 213–229. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_13

6. Chen, J., Zhang, Y., Wang, J., Zhou, X., He, Y., Zhang, T.: EllipseNet: anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12907, pp. 218–227. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_21

7. Chen, S., et al.: DiffusionDet: diffusion model for object detection. arXiv preprint arXiv:2211.09788 (2022)

8. Chen, T., et al.: A generalist framework for panoptic segmentation of images and videos. arXiv preprint arXiv:2210.06366 (2022)

9. Duan, K., et al.: CenterNet++ for object detection. arXiv preprint arXiv:2204.08394 (2022)

10. Ho, J., et al.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)

11. Holmquist, K., et al.: DiffPose: multi-hypothesis human pose estimation using diffusion models. arXiv preprint arXiv:2211.16487 (2022)

12. Li, F., et al.: DN-DETR: accelerate DETR training by introducing query DeNoising. In: IEEE CVPR, pp. 13619–13627 (2022)

13. Li, H., Han, H., Zhou, S.K.: Bounding maps for universal lesion detection. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 417–428. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_41

14. Li, H., Chen, L., Han, H., Chi, Y., Zhou, S.K.: Conditional training with bounding map for universal lesion detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 141–152. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_14

15. Li, H., et al.: SATr: slice attention with transformer for universal lesion detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI, pp. 163–174. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_16

16. Li, Z., Zhang, S., Zhang, J., Huang, K., Wang, Y., Yu, Y.: MVP-net: multi-view FPN with position-aware attention for deep universal lesion detection. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 13–21. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_2

17. Lin, C., Wu, H., Wen, Z., Qin, J.: Automated malaria cells detection from blood smears under severe class imbalance via importance-aware balanced group softmax. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 455–465. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_44

18. Liu, Z., et al.: A ConvNet for the 2020s. In: IEEE CVPR, pp. 11976–11986 (2022)

19. Luo, L., Chen, H., Zhou, Y., Lin, H., Heng, P.-A.: OXnet: deep omni-supervised thoracic disease detection from chest X-Rays. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 537–548. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_50

20. Lyu, F., Yang, B., Ma, A.J., Yuen, P.C.: A segmentation-assisted model for universal lesion detection with partial labels. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 117–127. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_12

21. Ren, S., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)

22. Ren, Y., et al.: Retina-match: ipsilateral mammography lesion matching in a single shot detection pipeline. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 345–354. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_33

23. Saharia, C., et al.: Photorealistic text-to-image diffusion models with deep language understanding. In: Advances in Neural Information Processing Systems, vol. 35, pp. 36479–36494 (2022)

24. Shahroudnejad, A., et al.: TUN-Det: a novel network for thyroid ultrasound nodule detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 656–667. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_62

25. Sheoran, M., et al.: An efficient anchor-free universal lesion detection in CT-scans. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2022)

26. Sheoran, M., et al.: DKMA-ULD: domain knowledge augmented multi-head attention based robust universal lesion detection. arXiv preprint arXiv:2203.06886 (2022)

27. Song, J., et al.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)

28. Tang, Y., et al.: ULDor: a universal lesion detector for CT scans with pseudo masks and hard negative example mining. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 833–836. IEEE (2019)

29. Tang, Y., et al.: Weakly-supervised universal lesion segmentation with regional level set loss. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 515–525. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_48

30. Tao, Q., Ge, Z., Cai, J., Yin, J., See, S.: Improving deep lesion detection using 3D contextual and spatial attention. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 185–193. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_21

31. Tian, Z., et al.: FCOS: fully convolutional one-stage object detection. In: IEEE ICCV, pp. 9627–9636 (2019)

32. Yan, K., Wang, X., Lu, L., Summers, R.M.: DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging **5**(3), 036501–036501 (2018)

33. Yan, K., Bagheri, M., Summers, R.M.: 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 511–519. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_58

34. Yan, K., et al.: MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 194–202. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_22

35. Yan, K., et al.: Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT. IEEE Trans. Med. Imaging **40**(10), 2759–2770 (2020)

36. Yang, J., et al.: *AlignShift*: bridging the gap of imaging thickness in 3D anisotropic volumes. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 562–572. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_55

37. Yang, J., He, Y., Kuang, K., Lin, Z., Pfister, H., Ni, B.: Asymmetric 3D context fusion for universal lesion detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 571–580. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_55

38. Yang, J., et al.: Reinventing 2D convolutions for 3D images. IEEE J. Biomed. Health Inform. **25**(8), 3009–3018 (2021)

39. Yu, X., et al.: Deep attentive panoptic model for prostate cancer detection using biparametric MRI scans. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 594–604. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_58

40. Zhang, S., et al.: Revisiting 3D context modeling with supervised pre-training for universal lesion detection in CT slices. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 542–551. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_53

41. Zhao, Z., Pang, F., Liu, Z., Ye, C.: Positive-unlabeled learning for cell detection in histopathology images with incomplete annotations. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 509–518. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_49

42. Zhou, S.K., et al.: Handbook of Medical Image Computing and Computer Assisted Intervention. Academic Press (2019)

43. Zhou, S.K., et al.: A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proc. IEEE **109**(5), 820–838 (2021)

44. Zlocha, M., Dou, Q., Glocker, B.: Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 402–410. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_45