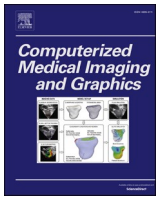
Computerized Medical Imaging and Graphics 123 (2025) 102522

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag



Domain adaptive detection framework for multi-center bone tumor detection on radiographs

Bing Li [a,b,1], Danyang Xu [c,1], Hongxin Lin [a], Ruodai Wu [d], Songxiong Wu [d], Jingjing Shao [c], Jinxiang Zhang [c], Haiyang Dai [e], Dan Wei [f], Bingsheng Huang [a,*], Zhenhua Gao [c,f,**], Xianfen Diao [a,g,***]

[a] Medical AI Lab, School of Biomedical Engineering, Medical School, Shenzhen University, Shenzhen, China
[b] Medical Imaging Department, The First Affiliated Hospital of Guangdong Pharmaceutical University, China
[c] Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China
[d] Radiology Department, Shenzhen University General Hospital and Shenzhen University Clinical Medical Academy, Shenzhen, China
[e] Department of Radiology, People's Hospital of Huizhou City Center, Huizhou, Guangdong, China
[f] Department of Radiology, Huiya Hospital of The First Affiliated Hospital, Sun Yat-Sen University, Huizhou, Guangdong, China
[g] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Medical School, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:
Adversarial learning
Bone tumor detection
Domain adaptation
Radiography

ABSTRACT

Automatic bone tumor detection on radiographs is crucial for reducing mortality from bone cancer. However, the performance of the detection methods may be considerably affected when deployed to bone tumor data in a distinct domain, which could be attributed to the differences in the imaging process and can be solved by training with a large amount of annotated data. However, these data are difficult to obtain in clinical practice. To address this challenge, we propose a domain-adaptive (DA) detection framework to effectively bridge the domain gap of bone tumor radiographs across centers, consisting of four parts: a multilevel feature alignment module (MFAM) for image-level alignment, Wasserstein distance critic (WDC) for quantization of feature distance, instance feature alignment module (IFAM) for instance-level alignment, and consistency regularization module (CRM), which maintains the consistency between the domain predictions of MFAM and IFAM. The experimental results indicated that our framework can improve average precision (AP) with an intersection over union threshold of 0.2 (AP@20) on the source and target domain test sets by 1 % and 8.9 %, respectively. Moreover, we designed a domain discriminator with an attention mechanism to improve the efficiency and performance of the domain-adaptative bone tumor detection model, which further improved the AP@20 on the source and target domain test sets by 2 % and 10.7 %, respectively. The proposed DA model is expected to bridge the domain gap and address the generalization problem across multiple centers.

1. Introduction

Bone tumors encompass primary and secondary neoplastic bone lesions with diverse biological behaviors (Choi and Ro, 2021). Primary bone tumors, such as osteosarcomas and giant cell tumors, are common around the knee joint and include the distal femur, proximal tibia, and proximal fibula (Arora et al., 2012; Beird et al., 2022; Cole et al., 2022). Bone tumor detection is crucial for bone tumor classification and follow-up therapy (He et al., 2020; von Schacky et al., 2022; Liu et al., 2022). Digital radiography (DR) has been the first-line imaging modality for assessing bone lesions owing to its ability to evaluate lesion location, internal matrix, and borders of bone lesions, with fast acquisition and cost-efficiency compared with computed tomography (CT) and magnetic resonance imaging (MRI) (Li et al., 2023; Bestic et al., 2020). However, plain radiography is a two-dimensional imaging modality, and radiographs have a relatively low contrast resolution compared with

* Corresponding author.
** Corresponding author at: Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China.
*** Corresponding author at: Medical AI Lab, School of Biomedical Engineering, Medical School, Shenzhen University, Shenzhen, China.
E-mail addresses: huangb@szu.edu.cn (B. Huang), gaozhh@mail.sysu.edu.cn (Z. Gao), laodiao@szu.edu.cn (X. Diao).
[1] Bing Li and Danyang Xu contributed equally to this work.

https://doi.org/10.1016/j.compmedimag.2025.102522
Received 7 December 2023; Received in revised form 14 January 2025; Accepted 2 March 2025
Available online 7 March 2025

CT images (Gould et al., 2007). Approximately 30–50 % of the trabecular bones may have been destroyed before the lesions are visible to the naked eye on plain radiographs (Krych et al., 2014). Consequently, some bone lesions may be obscured and misdiagnosed by visual observation in clinical practice (Liu et al., 2022). Many radiologists, especially inexperienced radiologists, lack sufficient training to identify and assess bone tumors on radiographs. Moreover, different device conditions and bone tumor types may result in variant visualizations with high domain discrepancies and further increase the risk of missed diagnosis by radiologists (Liu et al., 2022; Rosenberg, 2013; Errani et al., 2020). Therefore, automatic computer-aided diagnostic systems are necessary to assist in the diagnosis.

Artificial intelligence algorithms have been used to automatically detect tumors and demonstrate significant clinical value. Zhao et al. (Zhao et al., 2022) designs a novel fully automatic magnetic resonance imaging (MRI) vertebrae tumor diagnosis network, which firstly predict vertebrae labels (e.g., T12, L1, L2, etc.) and bounding boxes and then predict the vertebrae diagnostic labels (e.g., tumor/non-tumor). Zhao et al. (Zhao et al., 2021) proposes a united adversarial learning framework to simultaneously segment and detect liver tumors by using multi-modality non-contrast MRI. For automatic bone tumor detection, the YOLO algorithm has been used for detecting and classifying bone lesions on radiographs (Li et al., 2023); the multitask learning method has been used for simultaneous detection, segmentation, and classification of primary bone tumors (von Schacky et al., 2021). Despite these efforts lay a solid foundation for bone tumor detection, these methods rely excessively on large amounts of well-annotated data, which can be extremely tedious and time-consuming for medical image analysis. Moreover, they did not consider the significant differences in bone tumor data among different centers, and may suffer from significant performance deterioration in a new scenario with a large domain discrepancy caused by different hospitals, scanner vendors, and patient populations (Fig. 1).

Domain adaptive (DA) methods can narrow the distribution difference between the source domain and the target domain, and improve the performance of the model in cross domain tasks. Several studies use supervised DA methods to perform cross domain tasks. Wang et al. (Wang et al., 2019) designs a novel domain attention module to allow the detection network to leverage shared knowledge and adapt to the distribution characteristics across domains. Liu et al. (Liu et al., 2020) designs a feature decoupling module to separate features into shared features and domain-specific features, improving the performance of cross domain tasks by reducing the interference of domain-specific features on segmentation tasks. Zhao et al. (Zhao et al., 2023) designs a contour attraction term and a comprehensive contour quality loss for helping to overcome the challenges of data heterogeneity and image characteristics complexity. These works can effectively solve the problem of domain differences, however, a significant number of annotations are still required. Unsupervised domain adaptation (UDA) is an effective method for simultaneously addressing limited annotated data and bridging domain gaps, which includes techniques such as adversarial learning, cross-modal matching and meta-learning to reduce distribution differences between labeled source domain data and unlabeled target domain data through knowledge transfer to improve the generalization ability of the model (Oza et al., 2023). Cross-modal matching typically requires completely paired data from the source and target domains. Meta-learning typically requires training with data from multiple domains, and each domain requires large datasets (Liang et al., 2021). Adversarial learning is considered more suitable for this study because of the single-modal and unmatched data and the lack of positive data in many centers. UDA methods based on adversarial learning (Chen et al., 2018) exploit a domain discriminator with a gradient reversal layer (GRL) (Ganin and Lempitsky, 2015) to distinguish data from the source and target domains. When the domain discriminator cannot distinguish between the source and target domain data, the model can align the features between the domains and extract domain-invariant information for adaptation to the target domain. Therefore, this method can achieve satisfactory performance on target domain data with significant differences in distribution from the training set, even without annotations.

To address domain discrepancies in bone tumor data from different centers, we proposed a novel unsupervised adversarial domain adaptation framework that integrates an attention mechanism for bone tumor detection based on the YOLOv5m algorithm (Jocher et al., 2021). This approach enhances the detection performance of the target domain without incurring high labeling costs.

The UDA bone tumor detection framework comprises a Muti-level Feature Adaptation Module (MFAM), Wasserstein Distance Critic (WDC), Instance Feature Adaptation Module (IFAM), Consistency Regularization Module (CRM), and an attention mechanism. Considering the diverse manifestations of bone tumors, a certain layer of the feature extraction network may not be sufficient to cover all features of bone tumors. Therefore, we designed an MFAM that performs domain adaptation on features at three different depth layers of the feature extraction

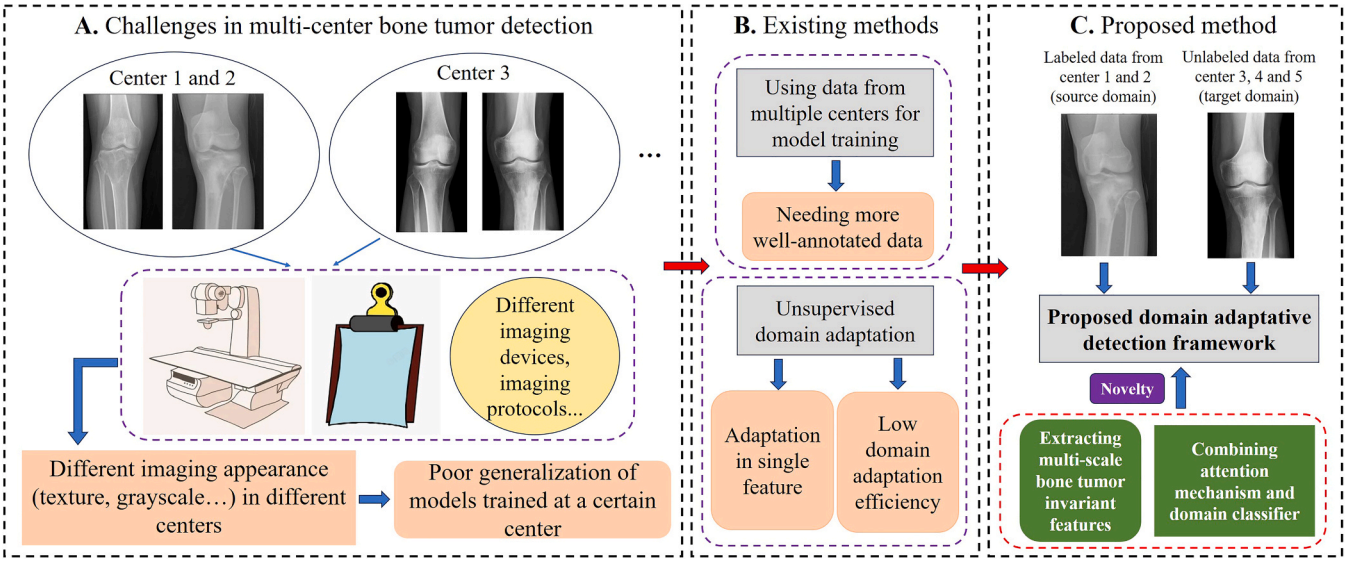


Fig. 1. Difference between our proposed method and existing methods to overcome the challenges in multi-center bone tumor detection. Note: Center 2 is an affiliated branch of Center 1.

network, thereby helping the network extract multiscale domain-invariant bone tumor features. MFAM can align the global domain shift, such as signal strength, image style, and texture; however, it cannot handle the shift of the local instance, such as tumor shapes, sizes, and viewpoints (Chen et al., 2018). Therefore, we designed an IFAM that can align instance-level features and reduce the differences between instances in the domain to generate domain-invariant detection results. For the same image, the domain classification labels of the global image and local instance features were identical. If the predicted labels of the two are inconsistent, they will have a negative impact on model training. The designed CRM minimizes the domain discrepancy between global image-level and instance-level predictions, thereby guiding the network to generate more domain-independent proposals. An inherent challenge in adversarial learning is ensuring the balance and stability of the training (Arora et al., 2017). Inspired by the Wasserstein GAN (Arjovsky et al., 2017), we combined the WDC (Villani, 2009) with a domain discriminator, which can quantify the distance between distributions and alleviate the equilibrium challenge. Considering that most of the images have redundant background areas, we combined the attention mechanism with a domain discriminator to refine the features and improve the feature alignment efficiency.

In general, our contributions are as follows:

- We proposed a domain-adaptive detection and localization framework to address the issue of bone tumor data distribution differences, and comparative experiments showed that the framework proposed

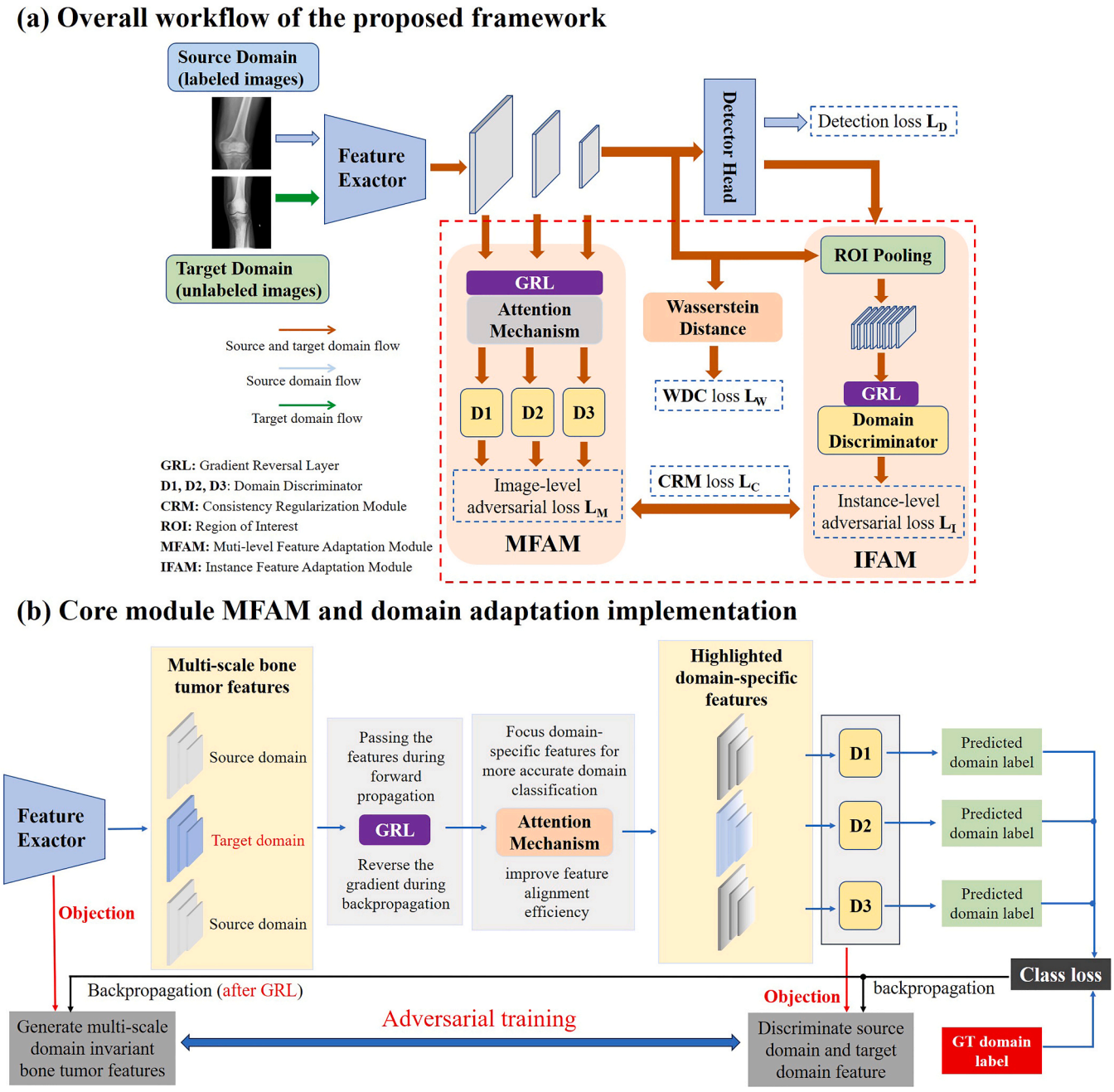


Fig. 2. Overview of the proposed framework for bone tumor detection. Fig. 2(a) shows overall structure of the domain adaptation framework that includes MFAM, WDC, IFAM and CRM. MFAM and IFAM use domain discriminators to enable the original detector to learn domain-invariant features on image and instance levels, respectively. L_C regularizes these extracted domain-invariant features at these two levels. WDC quantifies the feature differences between domains. Fig. 2(b) shows how the domain adaptation is implemented.

in our study is superior to the other two currently state-of-the-art bone tumor detection methods.
- We proposed multilevel feature adaptation module (MFAM) to align global features of different depth layers, which showed better domain adaptation performance than single-layer feature alignment.
- We integrated attention mechanism into domain discriminator, which enhanced domain adaptation capability and reduced inter domain differences.

2. Methods

An overview of the proposed bone tumor detection and localization framework is shown in Fig. 2, which consists of a YOLOv5m network and a domain adaptation network (red dashed part). The adaptive part includes MFAM for image-level feature alignment, WDC (Villani, 2009) for stabilizing the training process of the adversarial network and evaluating the difference between two distributions, IFAM for local feature alignment, and CRM for maintaining domain prediction consistency for global and local feature alignment. The labeled source domain images and unlabeled target domain images are simultaneously fed into the feature extractor, and the domain adaptation network aligns the extracted features of the source and target domains.

2.1. MFAM

Owing to the fixed kernel convolutional layer in the network, it is difficult to capture the accurate features of objects with various proportions and aspect ratios at a certain level of the feature layer (Liu et al., 2021). Small objects are usually detected in shallow feature layers with small receptive fields, whereas large objects are usually detected in deep feature layers with large receptive fields (Cui et al., 2019). Bone tumors have various imaging manifestations (such as texture and size) on plain radiographs (Errani et al., 2020). To enable the detection network to extract multiscale domain-invariant features of multiscale bone tumors, we proposed MFAM, which is composed of three adversarial training branches with the same structure but with different input scales (Fig. 2). The detailed structure of a branch in MFAM is illustrated in Fig. 3. Each domain discriminator consisted of a GRL layer, two CGR structures (convolutional layer, group normalization layer, and ReLU activation function layer), and a classification layer using a 1×1 convolution. Given the input image of the source or target domain, the feature extractor of YOLOv5m can generate three multiscale features of different sizes, denoted as P_1, P_2 and P_3 (the scales are 1/8, 1/16, and 1/32 times the original input size of the image, respectively).

The multiscale features are fed into the GRL layers that reverse the sign of the gradient during backpropagation to optimize the backbone feature generator. Subsequently, the feature P_l after the GRL layer is fed into the domain discriminator to generate the domain prediction $d(u,v,l)$ which is a one-dimensional vector with only one value for each location (u,v). The binary cross-entropy with logits (BCEWithLogits) loss was used to calculate the loss of MFAM. With both the source image i and its corresponding target image j as inputs to the network, the loss function of MFAM is calculated as follows:

$$L_G = -\sum_{(i,j,l,u,v)} \lambda_l [y \log \sigma(\boldsymbol{d}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{l})) + (1-y)\log(1-\sigma(\boldsymbol{d}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{l})))], \quad (1)$$

where y is the ground-truth domain label. $y = 0$ refers to the source domain and $y = 1$ refers to the target domain. λ_l is the trade-off factor in each branch in MFAM. σ is the sigmoid function.

Given that most radiographic images contain extensive background regions, the extracted image-level features typically contain several redundant features. As shown in Fig. 3, we incorporated the attention mechanism into the domain discriminator to extract radiographic image features and improve the feature alignment efficiency. Specifically, we add an attention mechanism to the GRL layer of the discriminator. The refined features are fed into the domain discriminator to generate the domain prediction. In this study, we have tried six classic attention mechanisms, namely criss-cross attention (CCNet) (Huang et al., 2019), squeeze-and-excitation networks (SENet) (Hu et al., 2018), frequency channel attention networks (FCANet) (Qin et al., 2021), convolutional block attention module (CBAM) (Woo et al., 2018), coordinate attention (CoordAtt) (Hou et al., 2021) and spatial pyramid attention (SPA) (Guo et al., 2020).

2.2. WDC

The binary loss-based domain discriminator (which can be approximated to the JS-divergence under most conditions (Cuturi and Peyré, 2016)) is unstable in the adversarial training process and may not effectively align domain shifts (Long et al., 2018). Inspired by the Wasserstein GAN (Arjovsky et al., 2017), WDC addresses these issues by quantifying feature distance and measuring the similarity between distributions, even when they do not overlap. We combined WDC with domain discriminators to provide more stable gradients, minimize domain discrepancies, and simultaneously confuse the domain classifier. Given the feature distributions P_s of source image i and P_t of target image j, the loss function of WDC can be calculated as follows:

$$L_W = \sum_{(i,j)} \inf_{\gamma \in \pi(P_s,P_t)} E_{(x,y)\in\gamma}[\|x-y\|], \quad (2)$$

where $\pi(P_s,P_t)$ is the set consisting of all joint distribution of P_s and P_t, γ is one of the possible joint distribution, x and y are two samples belonging to γ, $\|x-y\|$ and $E_{(x,y)\in\gamma}[\|x-y\|]$ represent the distance between two samples and the expectation of distance between samples, respectively, inf means to select a joint distribution γ to minimize the expectation. The Sinkhorn iteration method (Cuturi, 2013) is used to determine the optimal joint distribution.

2.3. IFAM

The instance-level feature refers to the feature vector of the area related to the instance before the detection head. Based on the image-level adaptation, we designed IFAM to reduce the instance-level domain shift. Specifically, the detection results from YOLOv5's three different-scale detection layers were used to extract instance-level features from their corresponding feature maps by ROI pooling. The

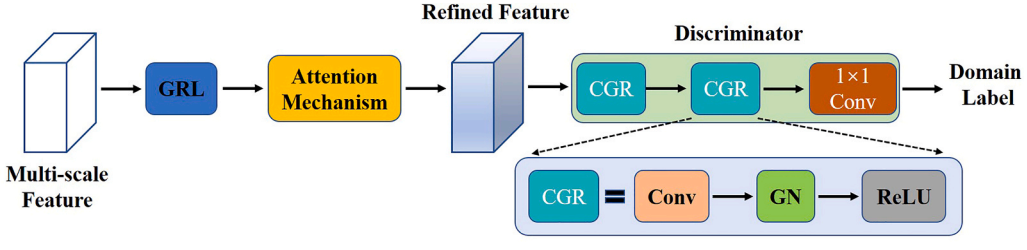


Fig. 3. Single branch in the MFAM. GN: group normalization.

BCEWithLogits loss was used to calculate the IFAM loss. Given an input image x and its k-th proposal, the instance feature f_k is extracted and fed into the domain discriminator with a GRL to generate a domain prediction $d(f_k)$. The loss function of IFAM was calculated as follows:

$$L_I = -\sum_{(x,k)}[y\log\sigma(d(f_k)) + (1-y)\log(1-\sigma(d(f_k)))] \quad (3)$$

2.4. CRM

Although image- and instance-level alignments enable the network to produce domain-invariant features, it may not produce domain-invariant detections, which are critical for object detection. Because the classifiers should make consistent predictions for features at both the image and instance levels, CRM is proposed for regularizing the extracted domain-invariant features at these two levels. CRM facilitates optimal cooperation between MFAM and IFAM, minimizing the domain discrepancy between image-level and instance-level predictions, thereby guiding the network to generate more domain-independent proposals. The loss function of CRM is calculated as follows:

$$L_C = \sum_{(i,j,k)}\left\|\frac{1}{S}\sum_{(u,v)}\boldsymbol{d}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{l}) - d(f_k)\right\|_2, \quad (4)$$

where $\|.\|_2$ is the L2 distance, and S is the total number of pixels in all image-level features.

2.5. Domain adaptative framework optimization

The proposed domain-adaptative framework was jointly optimized in an end-to-end fashion by detection loss L_{det}, MFAM loss L_G, WDC loss L_W, IFAM loss L_I and CRM loss L_C. The basic detection framework was implemented based on YOLOv5m (Jocher et al., 2021), and the loss function comprised the following three components:

$$L_{\text{det}} = \alpha L_{cls} + \beta L_{obj} + \gamma L_{loc} \quad , \quad (5)$$

Specifically, L_{cls}, L_{obj} and L_{loc} represent classification, confidence, and localization losses, respectively. α, β and γ are hyper-parameters to balance the optimization. Based on the original loss of the YOLOv5m detector L_{det}, the overall objective loss of our domain-adaptative detection framework can be expressed as:

$$L = L_{\text{det}} + L_G + \lambda_W L_W + \lambda_I L_I + \lambda_C L_C, \quad (6)$$

where λ_I, λ_C, and λ_W are hyper-parameters to balance the optimization.

3. Experiments

This section presents detailed information regarding the bone tumor radiographic image dataset. The implementation details (including training environment, hyperparameter setting, and experimental arrangement) were outlined. Finally, the performance evaluation metrics for the bone tumor detection tasks in this study were explained.

3.1. Dataset

This study was approved by the Institutional Review Board, which waived the requirement for written informed consent. We collected data on 685 patients diagnosed with bone tumors using radiography combined with CT or MRI/clinical follow-up or histopathology. The tumors were classified according to the fifth World Health Organization classification of bone tumors, published in 2020 (Soft tissue and bone tumours, WHO Classification of Tumours Editorial Board, 5th edn., 2020. IARC Press, Lyon, p.338.). In addition, 927 non-tumor patients with normal knee joint were recruited from five centers. The details of the dataset are listed in Table 1. Center 2 is an affiliated branch of Center 1, and both hospitals use the same type of digital radiography equipments with automatic exposure control technology, and the same filming position for same bone tumors. Thus, there is a high degree of homogeneity in the domain information between the two hospitals. Therefore, the labeled data from centers 1 and 2 were combined as the source domain dataset (596 patients with tumors), which is randomly split into the source domain training set (70 %), validation set (10 %), and test set (20 %). Data from the other three centers were used as the target domain training dataset (unlabeled data), which included 89 tumor patients and 927 non-tumor patients. The target domain test set consists of 89 tumor patients and 100 non-tumor patients randomly selected from centers 3 and 4. The source domain and target domain test sets were used to evaluate the model's performance.

Table 1
Characteristics of the study participants.

	Center 1	Center 2	Center 3	Center 4	Center 5
Number of patients	544	52	437	525	54
Tumor patients	544 (100.0)	52 (100.0)	26 (5.9)	9 (1.7)	54 (100.0)
Non-tumor patients	0 (0.0)	0 (0.0)	411 (94.1)	516 (98.3)	0 (0.0)
Number of lesions	1088	111	55	20	112
Age (mean ± SD) in years	33.4 ± 11.5	30.6 ± 17.2	25.3 ± 14.0	24.5 ± 13.0	31.4 ± 12.4
Sex					
Males (%)	321 (59.0)	42 (80.8)	305 (69.8)	286 (54.5)	32 (46.3)
Females (%)	223 (41.0)	10 (19.2)	132 (30.2)	239 (45.5)	22 (53.7)
Scan parameters (AEC)*					
Tube voltage (kV)	56–77		50–85	52–90	55–96
Exposure dose (mAs)	1–44		1–54	1–57	1–17
Exposure time (ms)	2–176		3–214	1–165	2–64

SD, standard deviation. AEC, automatic exposure control. The numbers in parentheses represent the percentage of cases in a particular class. *Center 2 is an affiliated branch of Center 1.

3.2. Data annotation

By referring to the corresponding CT or MR images, a senior radiologist (G.Z.H., with 16 years of experience in interpreting musculoskeletal radiographs) used LabelImg software (version 1.8.1, *an open-source image labeling program based on Python,* https://github.com/heartexlabs/labelImg) to annotate all bone tumor lesions with bounding boxes. These annotations serve as the reference standard for automatic bone tumor detection.

3.3. Evaluation metrics

The average precision value (AP) (Everingham et al., 2010) with an intersection over union (IoU) threshold of 0.2 was used as the main evaluation metric, denoted as AP@20. All bounding boxes localized by the DL model were matched with a reference standard. Each reference standard matched only one box. If multiple boxes have IoUs exceeding the given threshold, the box with the highest confidence score was selected. The boxes that matched the reference standard were counted as true positives (TPs). The boxes unmatched by the reference standard were false positives (FPs), whereas the reference standards unmatched by any box were false negatives (FNs). By setting a specific confidence score threshold, boxes with lower confidence scores were discarded, and TP, FP, and FN counts were calculated for the remaining boxes. By systematically changing the confidence threshold, we computed both precision and recall according to Equations (7) and (8) to generate a

precision-recall (PR) curve. Given the need for high sensitivity in bone tumor screening, precision and recall are specifically calculated and presented with a confidence threshold of 0.1 (Wang et al., 2020).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

3.4. Implementation details

All experiments were performed using Python 3.8.5 and PyTorch 1.7.1, using an NVIDIA TITAN RTX 24 GB GPU. An SGD optimizer was used to optimize the parameters of the model. The initial learning rate was set to 0.01. In the warm-up stage, the learning rate adjustment strategy of one-dimensional linear interpolation was adopted, after which cosine annealing was applied. The training batch size, input size of the images, and number of epochs were set to 8, 640 × 640, and 300, respectively.

The hyper-parameters in the framework include $\alpha = 0.5$, $\beta = 1$, $\gamma = 0.05$, $\lambda_l = 0.1, 0.15, 0.2$ for l in ranges from 1 to 3 in Eq. 1, $\lambda_I = 0.03$, $\lambda_C = 0.1$ and $\lambda_W = 0.005$ in the overall objective in Eq. 5. For all experiments, model weights are initialized with those pretrained with data from the common objects in context database (Lin et al., 2014).

The original YOLOv5m, trained on the source domain data without domain adaptation was used as the baseline. We first compared the impact of each domain discriminator in MFAM on the model's performance to assess the effectiveness of MFAM compared to a single discriminator. To verify the effectiveness of the proposed components in the DA detection frame, including MFAM, WDC, IFAM, and CRM, we performed an ablation study by incrementally incorporating these components. Notably, the MFAM used in the above experiments did not incorporate attention mechanisms. Finally, to determine whether attention mechanisms enhanced feature alignment, we compared the effectiveness of several classic attention mechanisms in domain-adaptive detection frameworks. The adjustable parameters for each attention mechanism were set according to the optimal parameters recommended in their respective studies.

To compare the performances of different components in the proposed detection network, we plotted the PR curves for each model. In addition, the feature distribution maps of the source and target test set data were plotted to assess whether the proposed method effectively bridged the gap in the feature distribution. Specifically, we extracted 1024 dimensional features of the source and target data from the last layer of the backbone of the trained YOLOv5m model and then used t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) to visualize the feature distribution.

To demonstrate the superiority of the proposed method in the field of bone tumor detection, comparative experiments with two state-of-the-art bone tumor detection methods (Li et al., 2023; von Schacky et al., 2021) were conducted.

4. Results

4.1. Comparison of single discriminator (D1, D2 or D3) and MFAM (D1 +D2 +D3)

We first compared the impact of the three domain discriminators (D1, D2, and D3) in the MFAM on the model performance (Table 2). The AP@20 of the baseline detection model YOLOv5m for the source and target domain test sets were 0.955 and 0.707, respectively. The model achieves an optimal AP@20 of 0.969 on the source domain test set when using only discriminator D2 in MFAM and an optimal AP@20 of 0.757 on the target domain test set when using three discriminators simultaneously (MFAM). Each discriminator in the MFAM exhibited varying degrees of improvement in the performance of the source domain and target test set. The results indicate that the model exhibited the optimal performance when using MFAM compared to only using single discriminator (achieving a 5.6 % improvement in AP@20 on two test sets). Aligning features at different scales enhances the extraction of richer multiscale domain-invariant features from bone tumor datasets. The PR curves of each model after adding the domain discriminators to the different feature layers are shown in Fig. 4.

Table 2
Detection performance of three discriminators (D1, D2, D3) in MFAM. The best results on two datasets are highlighted in bold. Baseline means the original YOLOv5m, trained on the source domain data without domain adaptation.

	Domain discriminator (Input feature size)	AP@20	Recall	Precision	F1
Source domain test set (119 patients from centers 1 and 2)	Baseline	0.955	0.920	**0.982**	0.950
	MFAM-D1 (256*80*80)	0.967	0.933	0.961	0.947
	MFAM-D2 (512*40*40)	**0.969**	**0.958**	0.950	**0.954**
	MFAM-D3 (512*20*20)	0.959	0.920	0.965	0.942
	D1 + D2 + D3 (MFAM)	0.961	0.950	0.890	0.919
Target domain test set (189 patients from centers 3, 4 and 5)	Baseline	0.707	0.605	**0.889**	0.720
	MFAM-D1 (256*80*80)	0.746	0.686	0.765	**0.723**
	MFAM-D2 (512*40*40)	0.714	0.714	0.567	0.632
	MFAM-D3 (512*20*20)	0.723	0.612	0.776	0.684
	D1 + D2 + D3 (MFAM)	**0.757**	**0.795**	0.457	0.580

4.2. Ablation experiments

The effectiveness of the proposed components in the DA detection frame, including MFAM, WDC, IFAM, and CRM is described in this section. The results of the ablation experiments are summarized in Table 3.

As shown in Table 3, AP@20 increased as the components were added, with the final domain adaptation framework showing a 9.9 % improvement in AP@20 on the source and target domain test sets compared with the conventional object detector. Compared with the baseline object detector on the target domain test set, MFAM can increase AP@20 by 5.0 %, which indicates that minimizing the global image feature discrepancy of bone tumor radiographs can lead to more precise box predictions. The WDC further quantifies the feature distance between domains and alleviates the impact of domain discrepancies. The addition of WDC improved AP@20 by 1.9 % on the target domain test set. Moreover, performing alignment at the instance level resulted in an AP@20 gain of 0.6 % on the target domain test set. Incorporating CRM improved AP@20 by 1.4 % on the target domain test set, which facilitates the detection performance by making consistent predictions for MFAM and IFAM. The entire domain adaptation framework with all the components achieved optimal performance on both the source and target domain test sets. Fig. 5 shows the PR curves of the model for the two test sets during the ablation experiment.

4.3. Exploration experiment on attention mechanism

The results of the bone tumor detection model incorporating the attention mechanism into the discriminator branch in the domain-adaptive framework are shown in Table 4. Based on the DA detection framework (adding MFAM, WDC, IFAM, and CRM) in Section 4.2, the model incorporating FCANet demonstrated optimal performance on the

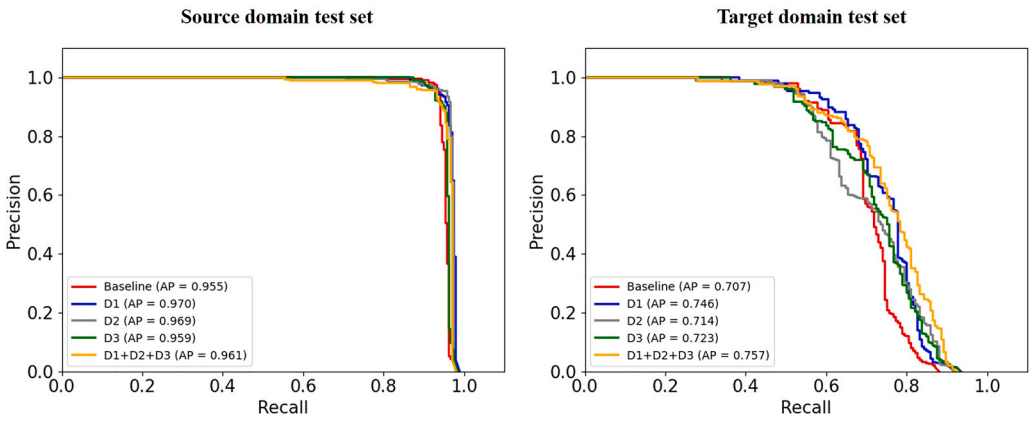


Fig. 4. PR curves of the model on source (left) and target (right) domain test sets when using domain discriminators on different feature layers.

Table 3
Ablation study on key components of the bone tumor detection framework. The best results on two datasets are highlighted in bold.

Dataset	Method	AP@20	Recall	Precision	F1
Source domain test set (119 patients from centers 1 and 2)	Baseline	0.955	0.920	**0.982**	**0.950**
	MFAM	0.961	0.950	0.890	0.919
	MFAM+WDC	0.953	0.950	0.900	0.924
	MFAM+WDC+IFAM	0.952	**0.954**	0.904	0.928
	MFAM+WDC+IFAM+CRM	**0.965**	0.950	0.890	0.919
Target domain test set (189 patients from centers 3, 4, and 5)	Baseline	0.707	0.605	**0.889**	**0.720**
	MFAM	0.757	0.795	0.457	0.580
	MFAM+WDC	0.776	0.805	0.548	0.652
	MFAM+WDC+IFAM	0.782	**0.832**	0.427	0.564
	MFAM+WDC+IFAM+CRM	**0.796**	0.778	0.629	0.696

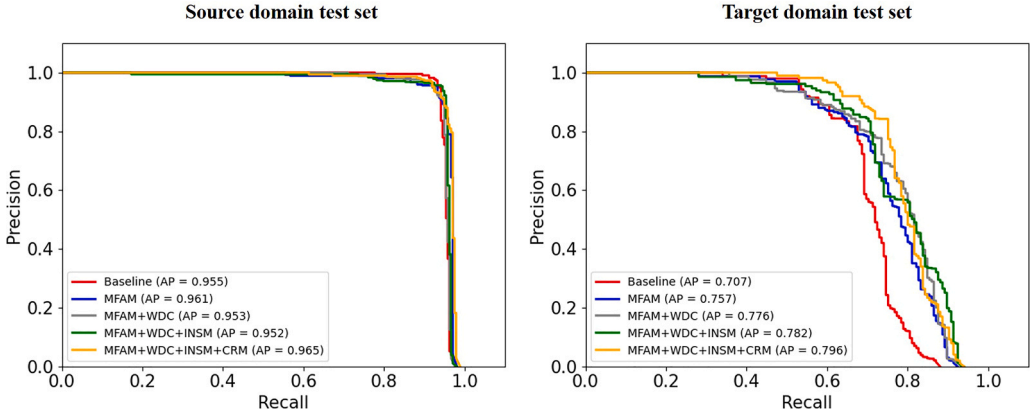


Fig. 5. PR curves of the model on the source (left) and target (right) domain test sets during the ablation experiment of each component in the domain adaptive framework.

source domain test set, with AP@20 increasing from 0.965 to 0.975. The model incorporating SPA demonstrated optimal performance on the target domain test set, and AP@20 increased from 0.796 to 0814. Based on these results, the performance of the model incorporating the SPA was the best, achieving a 1.4 % AP@20 improvement. Compared to the baseline, the domain-adaptative model incorporating SPA achieved a total of 15.3 % AP@20 improvement on the two test sets. Fig. 6 shows the PR curves of the domain-adaptative model incorporating the attention mechanism on the two test sets.

4.4. Visualization of feature distribution

To examine whether our proposed domain adaptative detection frame could bridge the gap in the feature distribution extracted from the source and target samples, the t-SNE method was used. Fig. 7 shows the t-SNE results of the feature distribution in the source- and target-domain test sets. Red and blue dots denote source and target domains, respectively. Fig. 7 (A and B) show the visualization results of the baseline model (original YOLOv5m trained on source domain data) and the adaptative model, respectively. Fig. 7 (C-H) show the visualization results of the domain adaptative model combining various attention mechanisms. The results illustrated in Fig. 7(A) indicate that the baseline model has a significant gap between the features generated from the source and target samples; however, the feature distribution interval after adaptation was small, as shown in Fig. 7 (B-H). Fig. 7 demonstrates similar feature distributions extracted by the proposed domain-adaptative model between the source and target samples, which suggests that the DA model can effectively address the domain shift problem

Table 4
Performance of the detection model combining a domain-adaptive framework and attention mechanisms. The best results on two datasets are highlighted in bold.

Dataset	Method	AP@20	Recall	Precision	F1
Source domain test set (119 patients from centers 1 and 2)	Baseline	0.955	0.920	**0.982**	0.950
	DA (MFAM+WDC+IFAM+CRM)	0.965	0.950	0.890	0.919
	DA + CCNet	0.960	0.950	0.919	0.934
	DA + SENet	0.957	0.950	0.897	0.923
	DA + FCANet	**0.975**	**0.962**	0.946	0.954
	DA + CBAM	0.965	0.954	0.966	**0.960**
	DA + CoordAtt	0.960	0.954	0.904	0.928
	DA + SPA	0.961	0.950	0.926	0.938
Target domain test set (189 patients from centers 3, 4, and 5)	Baseline	0.707	0.605	**0.889**	0.720
	DA (MFAM+WDC+IFAM+CRM)	0.796	0.778	0.629	0.696
	DA + CCNet	0.806	0.822	0.613	0.702
	DA + SENet	0.813	**0.832**	0.570	0.677
	DA + FCANet	0.797	0.784	0.597	0.678
	DA + CBAM	0.799	0.795	0.631	0.704
	DA + CoordAtt	0.804	0.816	0.619	0.704
	DA + SPA	**0.814**	0.805	0.696	**0.747**

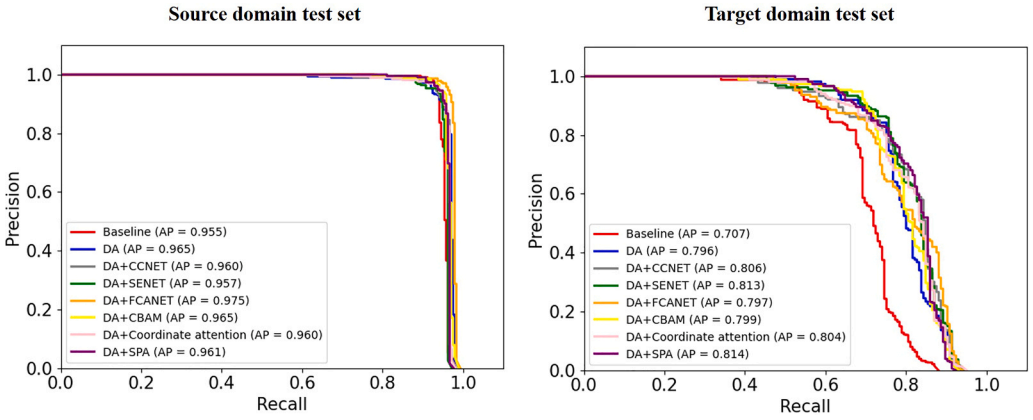


Fig. 6. PR curves of the domain-adaptative model incorporating attention mechanisms on the source (left) and target (right) domain test sets.

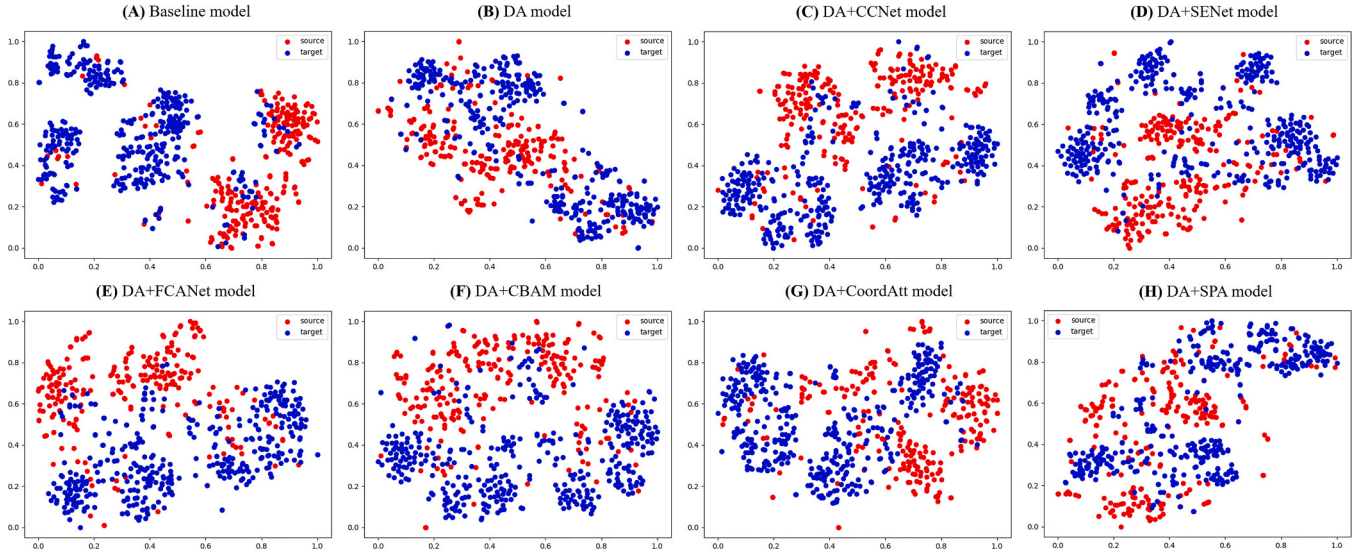


Fig. 7. Visualization of feature distribution using the t-SNE algorithm. (A): baseline model (YOLOv5m trained on the source domain). (B): domain-adaptative model. (C)-(H): domain-adaptative model combining attention mechanisms. The red dots and blue dots represent the feature distributions learned from the source and target samples, respectively. The model is expected to eliminate the gap in features between the source and target domains.

in multicenter bone tumor data.

4.5. Comparison of other bone tumor detection methods

We compared our method with two state-of-the-art bone-tumor detection methods by Li et al. and Von et al. (Li et al., 2023; von Schacky et al., 2021). As shown in Table 5, the method proposed by Li et al. achieved AP@20 of 0.944 and 0.757 on the source and target test sets, respectively, while, Von et al.'s method achieved AP@20 of 0.937 and 0.741 on the source and target test sets, respectively. Our method

outperformed both of these methods, with AP@20 scores of 0.961 and 0.814 on the source and target test sets, respectively, showing significant improvement, especially on the target domain.

5. Discussion

In this study, we proposed a domain-adaptive bone-tumor detection framework based on the YOLOv5 model. Our results demonstrated that the framework effectively addresses distribution differences in multi-center bone tumor data and enhances bone tumor lesion detection performance in the target domain. Moreover, integrating attention mechanisms into the domain-adaptive framework further improved bone tumor detection.

Significant differences were observed in the performance of the baseline model in the source domain (AP@20 =0.955) and target domain (AP@20 =0.707) test sets. Furthermore, t-SNE visualization revealed substantial differences in the feature distributions between the source and target domains. Data distribution differences can be caused by different hospitals, scanner vendors, and patient populations, and may also be related to subtypes of bone tumors as bone tumor subtypes have distinct imaging manifestations on plain radiographs (Errani et al., 2020).

MFAM, which simultaneously used three discriminators on different feature layers, exhibited better performance than that of a single-domain discriminator. Shallow features usually contain only low-level information, such as contours and edges, whereas deep features represent richer target information (Zeiler and Fergus, 2014). Bone tumors exhibit various imaging features and sizes. The alignment of single-scale features may not be sufficient to extract multiscale features for bone tumor detection, resulting in a less significant performance improvement. Using three domain discriminators simultaneously on three different-scale feature layers of the detection network, multiple domain discriminators can promote the extraction of richer multiscale domain-invariant features from bone tumor datasets. Moreover, we observed that aligning shallow features yielded the best results, likely because shallow features (such as tumor edges, textures, and basic visual details) are more effective for bone tumor detection tasks.

In the ablation experiment, model performance improved with the addition of each component, demonstrating their effectiveness. However, after incorporating WDC, we observed a slight decrease in AP@20 on the source domain test set. This may be because of the inherent differences between non-tumor and bone tumor patients, where WDC's quantification of inter-domain features might reduce distribution differences between domains while increasing variations within the source domain. We observe that the addition of IFAM slightly improved the model's performance on the target domain test set, but the performance was slightly decreased on the source domain test set. This may be because most of the instances output by the network are in the background region, and only a few instances contain target feature information (Liu et al., 2021). Therefore, aligning some redundant instance features may result in a slight improvement (Zhang et al., 2021; Wang et al., 2021). However, the addition of CRM enhanced the performance of the model for both test sets. For the same image, the predictions of MFAM and IFMA should be consistent because inconsistency may lead to adverse effects. CRM ensures consistent predictions between MFAM and IFAM, which can avoid many erroneous feature alignments and improve model alignment efficiency. The precision of the domain-adaptive model with all components was lower than that of the baseline model. The recall and precision depend on the confidence threshold of the PR curve. Compared with precision, high recall is usually required for bone tumor screening in clinical practice. Precision and recall were specifically calculated and presented with a confidence threshold of 0.1 (Wang et al., 2020), rather than the optimal threshold.

Table 4 shows that adding attention mechanisms to domain adaptive frameworks can improve alignment efficiency and further improve model performance. However, the addition of the attention mechanism resulted in only slight performance improvements. A possible explanation is that the hyperparameters used in the attention mechanism are not applicable for bone tumor detection. In addition, similar results across different attention mechanisms suggest that model performance may have reached a bottleneck due to the limited tumor data (89 cases) in the target domain.

Our proposed model outperformed the other two bone tumor detection models (Li et al., 2023; von Schacky et al., 2021) on both the source domain and target domain test sets. Compared with the source domain test set, the proposed model outperforms the other two models on the target domain test set, which highlights its effectiveness in addressing large distribution differences across different centers.

Although the proposed domain-adaptive detection framework narrows the gap between domains and improves performance in the target domain, there remains a gap between the performance of the current model on the target domain (AP@20 =0.814) and the source domain (AP@20 =0.961). This discrepancy may be due to the limited amount of bone tumor data (87 cases) in the target domain training set. Insufficient training data during feature alignment resulted in a limited adaptive ability of the model to reduce the distribution differences between domains. Future research with more bone tumor data is warranted to effectively train the adaptative model.

Table 5
Quantitative comparison of two state-of-the-art bone tumor detection methods and our proposed framework. The best results on two datasets are highlighted in bold.

Dataset	Method	AP@20	Recall	Precision	F1
Source domain test set (119 patients from centers 1 and 2)	Li et al. (2023)	0.944	0.941	**0.945**	**0.943**
	von et al. (2022)	0.937	0.933	0.895	0.914
	ours	**0.961**	**0.950**	0.926	0.938
Target domain test set (189 patients from centers 3, 4, and 5)	Li et al. (2023)	0.757	0.746	0.600	0.665
	von et al. (2022)	0.741	0.762	0.520	0.618
	ours	**0.814**	**0.805**	**0.696**	**0.747**

6. Conclusion

We proposed an unsupervised adaptive object detection framework as an effective one-stage detector for bone tumor detection across centers. Our framework incorporated MFAM and IFAM into the original detection network to perform multiscale image-level and instance-level feature alignment, respectively. The two modules are also connected by the proposed CRM to enhance the compatibility and efficiency of the feature-level adaptation. Moreover, the WDC was used to quantify the feature distance between domains and stabilize the adversarial training process. This method allows for a well-trained detection model to be obtained in new scenarios without annotated labels. The detection model trained by our framework can remove the modules for domain adaptation when the model is deployed to avoid increasing the time consumed in the inference phase. The experimental results for the source and target domain test sets verified the effectiveness of the adaptive model for bone tumor detection.

CRediT authorship contribution statement

Xu Danyang: Resources, Data curation. **Li Bing:** Writing – original draft, Methodology, Conceptualization. **Wu Ruodai:** Methodology. **Lin Hongxin:** Methodology. **Shao Jingjing:** Data curation. **Wu Songxiong:** Methodology. **Dai Haiyang:** Data curation. **Zhang Jinxiang:** Data curation. **Huang Bingsheng:** Writing – review & editing, Supervision, Conceptualization. **Wei Dan:** Data curation. **Diao Xianfen:** Supervision, Project administration, Conceptualization. **Gao Zhenhua:** Resources,

Funding acquisition, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was jointly funded by Basic and Applied Basic Research Foundation of Guangdong Province, China (Grant No.2025A1515011777; No.2025A1515011661), the Natural Scientific Foundation of China (No. 62371303), the Shenzhen-Hong Kong Institute of Brain Science—Shenzhen Fundamental Research Institutions of China (No. 2024SHIBS0003).

Data availability

The authors do not have permission to share data.

References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. : Proc. Int. Conf. Mach. Learn. Res. 214–223.
Soft tissue and bone tumours, WHO Classification of Tumours Editorial Board, 5th edn., 2020. IARC Press, Lyon, p. 338.
Arora, R.S., Alston, R.D., Eden, T.O., Geraci, M., Birch, J.M., 2012. The contrasting age-incidence patterns of bone tumours in teenagers and young adults: implications for aetiology. Int. J. Cancer 131 (7), 1678–1685.
Arora, S., Ge, R., Liang, Y., et al., 2017. Generalization and equilibrium in generative adversarial nets (gans). : Proc. Int. Conf. Mach. Learn. Res. 224–232.
Beird, H.C., Bielack, S.S., Flanagan, A.M., et al., 2022. Osteosarcoma. Nat. Rev. Dis. Prim. 8 (1), 77.
Bestic, J.M., Wessell, D.E., Beaman, F.D., et al., 2020. ACR Appropriateness Criteria® primary bone tumors. J. Am. Coll. Radiol. 17 (5), S226–S238.
Chen, Y., Li, W., Sakaridis, C., et al., 2018. Domain adaptive faster r-cnn for object detection in the wild. : Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 3339–3348.
Choi, J.H., Ro, J.Y., 2021. The 2020 WHO classification of tumors of bone: an updated review. Adv. Anat. Pathol. 28 (3), 119–138, 2021.
Cole, S., Gianferante, D.M., Zhu, B., Mirabello, L., 2022. Osteosarcoma: a surveillance, epidemiology, and end results program-based analysis from 1975 to 2017. Cancer 128 (11), 2107–2118.
Cui, Z., Li, Q., Cao, Z., et al., 2019. Dense attention pyramid networks for multi-scale ship detection in SAR images. IEEE Trans. Geosci. Remote Sens. 57 (11), 8983–8997.
Cuturi, M., 2013. Sinkhorn distances: lightspeed computation of optimal transport. Adv. Neural Inf. Process. Syst. 26.
Cuturi, M., Peyré, G., 2016. A smoothed dual approach for variational Wasserstein problems. SIAM J. Imaging Sci. 9 (1), 320–343.
Errani, C., Tsukamoto, S., Mavrogenis, A.F., 2020. Imaging analyses of bone tumors. JBJS Rev. 8 (3), e0077.
Everingham, M., Van Gool, L., Williams, C.K.I., et al., 2010. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. 88, 303–338.
Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. : Proc. Int. Conf. Mach. Learn. Res. 1180–1189.
Gould, C.F., Ly, J.Q., Lattin, Jr.G.E., et al., 2007. Bone tumor mimics: avoiding misdiagnosis, 124-41 Curr. Probl. Diagn. Radiol. 36 (3), 124-41.
Guo, J., Ma, X., Sansom, A., et al., 2020. Spanet: Spatial pyramid attention network for enhanced image recognition. : IEEE Int. Conf. Multimed. Expo. (ICME) 1–6.
He, Y., Pan, I., Bao, B., et al., 2020. Deep learning-based classification of primary bone tumors on radiographs: A preliminary study. EBioMedicine 62, 103121.
Hou, Q., Zhou, D., Feng, J., 2021. Coordinate attention for efficient mobile network design. : Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 13713–13722.
Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the. IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.
Huang, Z., Wang, X., Huang, L., et al., 2019. Ccnet: criss-cross attention for semantic segmentation. : Proc. IEEE/CVF Int. Conf. Comput. Vis. 603–612.
Jocher, G., Stoken, A., Borovec, J., et al.: ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models (2021). https://doi.org/10.5281/zenodo.4679653.
Krych, A., Odland, A., Rose, P., et al., 2014. Oncologic conditions that simulate common sports injuries, 223-34 J. Am. Acad. Orthop. Surg. 22, 223-34.
Li, J., Li, S., Li, X., et al., 2023. Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model. Eur. Radiol. 33 (6), 4237–4248.
Liang, P., Wu, P., Liu, Z., et al., 2021. Cross-modal generalization: learning in low resource modalities via meta-alignment. : Proc. 29th ACM Int. Conf. Multimed. 2680–2689.
Lin, T.Y., Maire, M., Belongie, S., et al., 2014. Microsoft COCO: common objects in context. : Proc. Eur. Conf. Comput. Vis. 740–755.
Liu, Q., Dou, Q., Yu, L., et al., 2020. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans. Med. Imaging 39 (9), 2713–2724.
Liu, X., Guo, X., Liu, Y., et al., 2021. Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images. Med. Image Anal. 71, 102052.
Liu, R., Pan, D., Xu, Y., et al., 2022. A deep learning-machine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors. Eur. Radiol. 32 (2), 1371–1383.
Long, M., Cao, Z., Wang, J., et al., 2018. Conditional adversarial domain adaptation. Adv. Neural Inf. Process. Syst. 31.
Oza, P., Sindagi, V.A., Sharmini, V.V., et al., 2023. Unsupervised domain adaptation of object detectors: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 46 (6), 4018–4040.
Qin, Z., Zhang, P., Wu, F., et al., 2021. Fcanet: frequency channel attention networks. : Proc. IEEE/CVF Int. Conf. Comput. Vis. 783–792.
Rosenberg, A.E., 2013. WHO classification of soft tissue and bone: summary and commentary. Curr. Opin. Oncol. 25 (5), 571–573.
von Schacky, C.E., Wilhelm, N.J., Schäfer, V.S., et al., 2021. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. Radiology 301 (2), 398–406.
von Schacky, C.E., Wilhelm, N.J., Schäfer, V.S., et al., 2022. Development and evaluation of machine learning models based on X-ray radiomics for the classification and differentiation of malignant and benign bone tumors. Eur. Radiol. 32 (9), 6247–6257.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).
Villani, C., 2009. Optimal transport: old and new. springer, Berlin. https://doi.org/10.1007/978-3-540-71050-9.
Wang, X., Cai, Z., Gao, D., Vasconcelos, N., 2019. Towards universal object detection by domain attention. : Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 7289–7298.
Wang, J., He, Y., Fang, W., et al., 2021. Unsupervised domain adaptation model for lesion detection in retinal OCT images. Phys. Med. Biol. 66 (21), 215006.
Wang, X., Zhang, R., Kong, T., Li, L., Shen, C., 2020. SOLOv2: dynamic and fast instance segmentation. Adv. Neural Inf. Process. Syst. 33, 17721–17732.
Woo, S., Park, J., Lee, J.Y., et al., 2018. Cbam: convolutional block attention module. Proc. Eur. Conf. Comput. Vis. 3–19.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. Proc. Eur. Conf. Comput. Vis. 818–833.
Zhang, S., Tuo, H., Hu, J., et al., 2021. Domain adaptive YOLO for one-stage cross-domain detection. In: Asian Conference on Machine Learning. PMLR, pp. 785–797.
Zhao, S., Chen, B., Chang, H., Chen, B., Li, S., 2022. Reasoning discriminative dictionary-embedded network for fully automatic vertebrae tumor diagnosis. Med. Image Anal. 79, 102456.
Zhao, J., Li, D., Xiao, X., et al., 2021. United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI. Med Image Anal. 73, 102154.
Zhao, S., Wang, J., Wang, X., et al., 2023. Attractive deep morphology-aware active contour network for vertebral body contour extraction with extensions to heterogeneous and semi-supervised scenarios. Med Image Anal. 89, 102906.