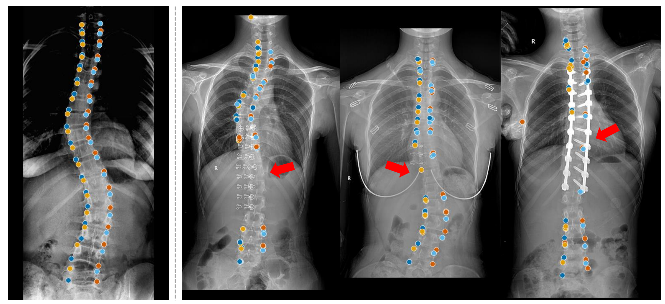
Style- and Structure-Aware Vertebral Landmark Detection under Implant-Induced Domain Shift

Zhihao Yao[1], Jingwei Qu[1], Congyu Wang[2], Ronggang Chen[1], Zhigang Cai[1], Jianwei Liao[1], Xinyu Liu[2]

***Abstract*— Public scoliosis landmark benchmarks are dominated by preoperative radiographs, whereas clinical deployment often involves postoperative follow-up images with substantial implant-induced appearance shifts. We study source-only vertebral landmark detection, where the target domain is unseen during training and model selection. Built on a heatmap-based detector, our style-structure framework inserts random *MixStyle* into early ResNet34 stages to perturb source feature statistics, and introduces *Slot-Wise Structure Prior Decoding* (*SWSPD*) to enforce vertically coherent vertebral ordering. Under a public-to-private protocol, the model is trained and validated on a public benchmark and tested directly on a held-out private clinical cohort. Experiments show that our method preserves competitive in-domain performance while improving source-only cross-domain robustness. On the implant-bearing subset, it reduces SMAPE from 18.99 to 13.69 and CMAE from 4.85 to 3.80 compared with the strongest competing source-only baseline.**



(a) (b)

Fig. 1. (a) A representative image from the public-domain dataset, where vertebral landmarks are relatively easy to identify. (b) Representative images from our private-domain dataset. Pedicle screws and other structured occlusions introduce domain-specific interference, often causing missing points and false predictions in heatmap-based methods. Different colors encode the four corner points of each vertebra.

I. INTRODUCTION

Adolescent idiopathic scoliosis (AIS) is a common structural spinal deformity, and its radiographic evaluation spans screening, treatment planning, and postoperative follow-up [12], [9], [2], [19]. Accurate vertebral landmark localization is central to Cobb angle measurement, vertebral morphology analysis, and other structured assessments, making reliable landmark detection essential across patient-care stages.

A standard standing anteroposterior (AP) whole-spine radiograph contains 12 thoracic and 5 lumbar vertebrae, each represented by four corner landmarks, yielding 68 landmarks in total. Existing detectors perform well on public preoperative benchmarks, as shown in Figure 1(a), but these datasets are usually free of internal fixation hardware. In contrast, postoperative follow-up images often contain pedicle screws, rods, and other high-density occlusions that alter vertebral appearance and boundary patterns.

Under implant-induced occlusions, existing methods are more prone to missed landmarks, false positives, and anatomically inconsistent predictions, as seen in Figure 1(b). Thus, strong preoperative benchmark performance may not directly translate to postoperative use.

Existing methods are typically coordinate-regression or heatmap-based. Coordinate-regression methods can model global spatial relationships, but their geometric plausibility often depends on network design or additional constraints. We retain the heatmap paradigm because of its robustness to pose variation, but address its dependence on local appearance evidence, which is vulnerable to metallic artifacts, occlusions, and repetitive structures. Our goal is therefore to improve cross-domain appearance robustness while adding structural constraints that suppress anatomically inconsistent predictions.

In clinical deployment, target-hospital postoperative data are often limited, private, or expensive to annotate. Rather than relying on target-domain adaptation, a practical strategy is to synthesize richer style variation from source samples alone [20], improving generalization to unseen domains without target exposure.

Building on this motivation, we propose a source-only style-structure framework for robust vertebral landmark detection under postoperative implant-induced domain shift. The main contributions are threefold:

- We formulate postoperative vertebral landmark detection as a source-only domain generalization problem with no target-domain access during training or model selection.
- We propose a style-structure framework that combines *MixStyle*-based appearance perturbation with *Slot-Wise Structure Prior Decoding* (*SWSPD*) for anatomically ordered decoding.
- We validate the framework on the public benchmark and by direct transfer to a held-out private cohort with and without implants, with ablations showing complementary style and structure gains.

Zhihao Yao and Jingwei Qu contributed equally to this work. Jingwei Qu and Xinyu Liu are the corresponding authors.

[1]College of Computer and Information Science, Southwest University.
[2]Department of Orthopedics, Qilu Hospital of Shandong University.

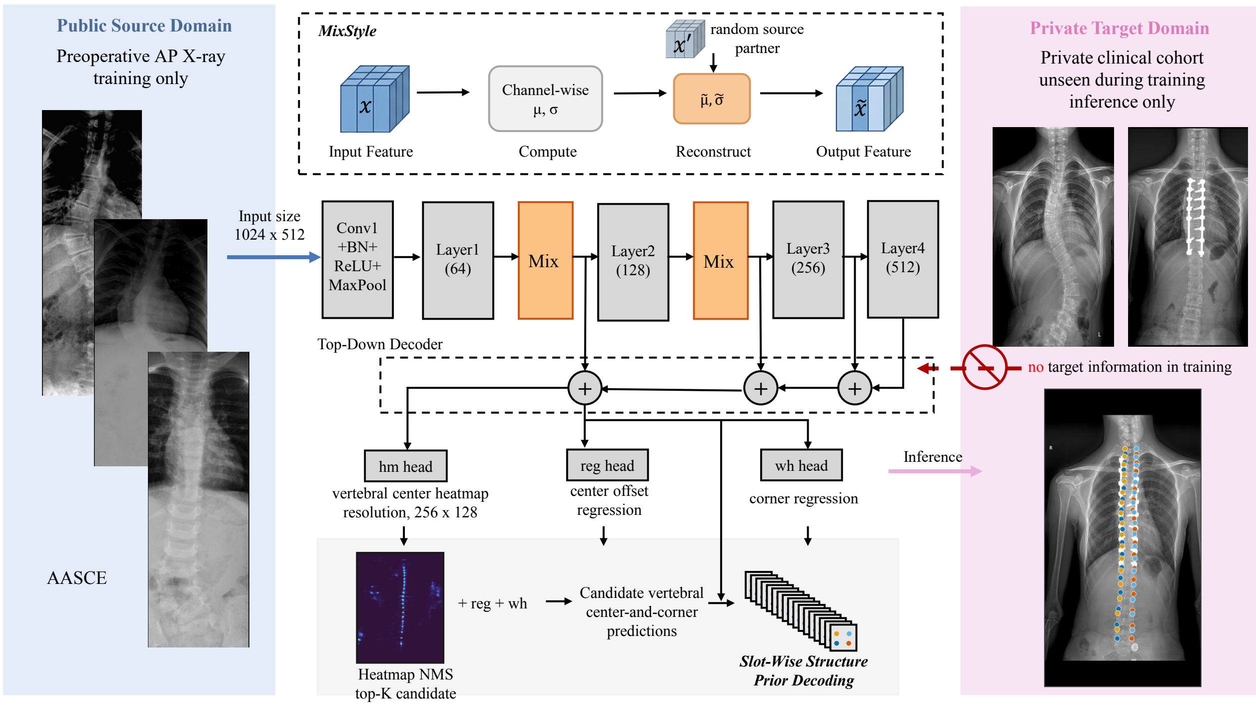


Fig. 2. Overview of the proposed source-only framework for cross-domain vertebral landmark detection.

II. RELATED WORK

A. Vertebral Landmark Detection for Scoliosis Analysis

The AASCE 2019 challenge established a widely used benchmark for automatic scoliosis assessment from AP radiographs [12]. Existing methods mainly use heatmap-based localization or coordinate regression. Representative heatmap-based approaches include SCN [10], Yi *et al.* [16], H3R [17], SLSN [18], and FARNet [1], which improve localization through multi-resolution features, vertebra-aware representations, or refined supervision. Wang *et al.* further incorporate global shape regularity into coordinate regression [14]. These studies highlight the value of vertebra-aware representation and structural modeling, but are mostly evaluated under benchmark-aligned settings. Our work instead studies public-to-private transfer to postoperative radiographs with metallic implants, where the appearance shift is substantially stronger and conventional detector robustness remains insufficiently explored.

B. Adaptation and Source-Only Domain Generalization

Cross-domain learning often relies on domain adaptation, where target-domain data are available during training. Domain-adversarial learning, for example, learns task-discriminative and domain-invariant features through adversarial supervision [4]. In many medical settings, however, target data are limited or inaccessible because of privacy, collection cost, or protocol constraints. This motivates source-only domain generalization, where no target images, labels, statistics, or model-selection signals are used.

Recent source-only methods show that perturbing style-related feature statistics can improve robustness to appearance variation [20]. *MixStyle* mixes instance-level feature statistics across source samples in intermediate representations, simulating latent domain variation without pixel synthesis or target supervision. This is consistent with style-transfer studies showing that shallow feature statistics capture appearance while largely preserving semantic content [5], [7], [3]. For implant-induced shift, this source-only formulation better matches our clinical constraint than target-aware adaptation.

III. METHODOLOGY

A. Framework Overview

To address postoperative domain shift, we propose the source-only framework in Figure 2. A style branch applies random *MixStyle* to source-domain features, while *SWSPD* uses a slot-wise vertical prior and dynamic programming to select a globally consistent vertebral sequence.

B. Problem Setup

The source dataset is defined as:

$$\mathscr{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}, \tag{1}$$

where x_i^s is a public preoperative radiograph and y_i^s denotes the corresponding vertebral landmarks.

The target dataset is defined as:

$$\mathscr{D}_t = \{x_j^t\}_{j=1}^{N_t}, \tag{2}$$

where x_j^t denotes private clinical radiographs drawn from a different clinical distribution.

In our main setting, $\mathscr{D}_t$ is unavailable during training and checkpoint selection. We learn f_θ using only $\mathscr{D}_s$ and evaluate

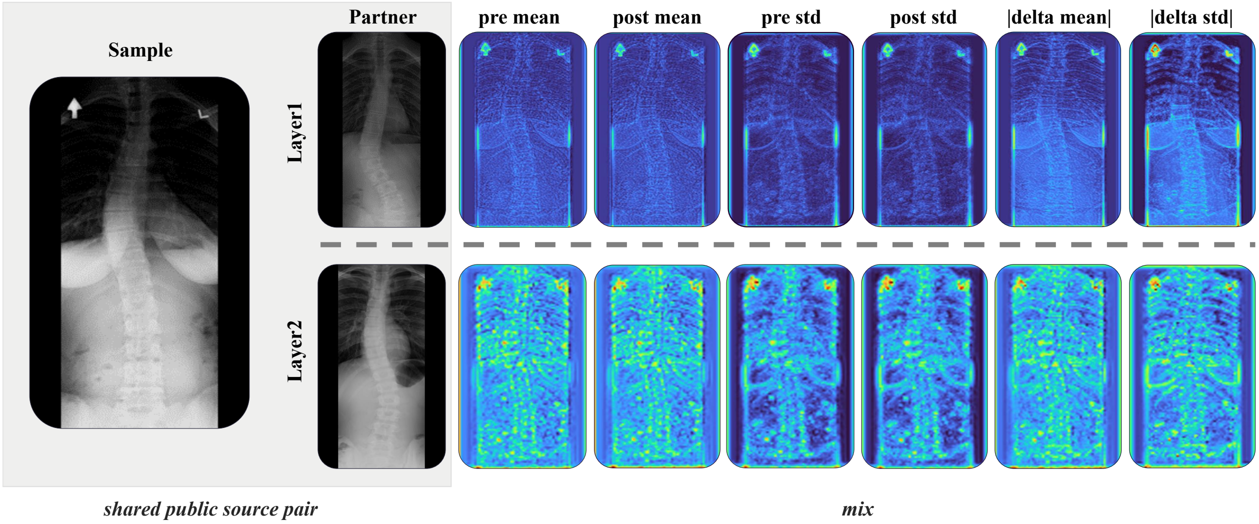


Fig. 3. Visualization of the source-only style branch implemented with random *MixStyle* on public training images.

it directly on the unseen target domain; mixed-source and target-aware training are excluded from the strict protocol.

The style branch of the proposed framework inserts random *MixStyle* into the early backbone stages. Let $x \in \mathbb{R}^{C\times H\times W}$ denote an intermediate feature map. Following *MixStyle* [20], we compute its channel-wise mean and standard deviation as:

$$\mu(x) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{:,h,w}, \tag{3}$$

$$\sigma(x) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(x_{:,h,w}-\mu(x)\right)^2+\varepsilon}. \tag{4}$$

For another source sample x' drawn by a random permutation within the same mini-batch, we sample $\lambda \sim \mathrm{Beta}(\alpha,\alpha)$ and define mixed statistics:

$$\tilde{\mu} = \lambda\mu(x)+(1-\lambda)\mu(x'), \tag{5}$$

$$\tilde{\sigma} = \lambda\sigma(x)+(1-\lambda)\sigma(x'). \tag{6}$$

The transformed feature is then defined as:

$$\tilde{x} = \tilde{\sigma}\odot\frac{x-\mu(x)}{\sigma(x)}+\tilde{\mu}. \tag{7}$$

Figure 3 visualizes this source-only perturbation by showing channel-aggregated feature statistics before and after *MixStyle*. Random source pairs exchange feature statistics within a mini-batch, producing appearance variation without changing annotations or using target images.

This feature-space design perturbs appearance without label adjustment, image synthesis, or target-domain batches. The `random` mixing mode is especially suitable for our protocol because it uses only source-domain samples and does not require paired source-target batches. Since implants and acquisition differences alter low-level statistics in postoperative radiographs, we insert *MixStyle* after `layer1` and `layer2` to reduce over-reliance on source-domain style cues and encourage more stable structural representations.

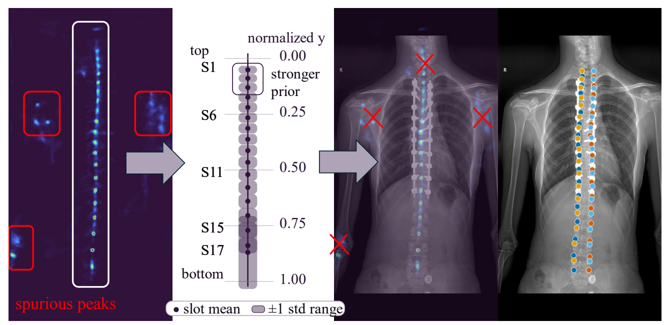


Fig. 4. Visualization of the slot-wise vertical position prior used in *SWSPD*.

C. Slot-Wise Structure Prior Decoding (SWSPD)

Given an input radiograph I, the network prediction is defined as:

$$f_\theta(I) = \{\hat{Y}^{hm},\hat{Y}^{reg},\hat{Y}^{wh}\}, \tag{8}$$

where $\hat{Y}^{hm}$ is the vertebral center heatmap, $\hat{Y}^{reg}$ denotes subpixel center offsets, and $\hat{Y}^{wh}$ parameterizes the four corner vectors of each vertebral box. The decoder applies non-maximum suppression to $\hat{Y}^{hm}$, extracts a top-K center pool, refines each center using $\hat{Y}^{reg}$, and expands it into a vertebral box using $\hat{Y}^{wh}$.

Rather than selecting 17 vertebrae independently from local peaks, the decoder searches for a fixed-length anatomically ordered sequence. Candidates are sorted vertically, and dynamic programming identifies the best superior-to-inferior path by favoring high heatmap confidence and penalizing implausible neighbor transitions.

The transition cost contains two geometric regularizers: adjacent candidates are discouraged from having unrealistically small vertical gaps relative to vertebral height, and consecutive vertebrae are encouraged to maintain consistent box scale in both height and width. A gap-smoothness term further penalizes abrupt changes in inter-vertebral spacing across three consecutive vertebrae. Together, these terms bias decoding toward a coherent spine layout rather than isolated high-response distractors.

To further suppress off-spine peaks, we incorporate a slot-wise vertical position prior over the normalized y-coordinates of the 17 vertebrae. For slot s, the decoder maintains a mean position μ_s and standard deviation σ_s, estimated from source-domain training annotations. During decoding, a candidate assigned to slot s receives a quadratic penalty when its normalized vertical location deviates from μ_s:

$$\mathscr{L}_{pos}(s,i) = \lambda_{pos}\alpha_s\left(\frac{\tilde{y}_i-\mu_s}{\sigma_s}\right)^2, \tag{9}$$

where $\tilde{y}_i$ is the normalized vertical coordinate of candidate i, and α_s is a slot-dependent scaling factor. We set $\alpha_s = 1$ for most slots and use larger weights for the first few superior slots to anchor upper vertebrae against off-spine distractions. This prior remains fully source-only because it is estimated from source annotations.

Figure 4 visualizes the slot-wise prior: the middle panel summarizes the per-slot mean and ± 1 standard-deviation range, while the side panels show how the prior suppresses off-spine peaks and yields an ordered landmark sequence.

IV. DATASETS AND EVALUATION METRICS

A. Datasets

The AASCE dataset contains 609 AP spinal radiographs from London Health Sciences Centre, Canada[1], split into 481 training/validation images and 128 test images. Each image has 68 landmarks for the four corners of 17 vertebrae. We denote these splits as $AASCE_{train}$ and $AASCE_{test}$.

Our private dataset contains 216 AP radiographs from scoliosis patients acquired with a GE Healthcare X-ray system. It includes 147 images reserved as target-domain data for DANN or relaxed fine-tuning, 24 validation images for those relaxed protocols, and 45 test images. The test set has 18 preoperative images without implants and 27 postoperative images with implants, reported as *Without Implants* and *With Implants*. In the main source-only setting, private training and validation splits are never used for training or checkpoint selection. Splits are patient-level, and two experts with more than 6 years of experience independently annotated 68 landmarks per image. DANN and relaxed-protocol baselines are therefore reported separately from the strict source-only table.

B. Metrics

We evaluate landmark localization accuracy using the Normalized Mean Squared Error (NMSE), defined as:

$$NMSE = \frac{1}{M}\sum_{i=1}^{M}\left[\left(\frac{d_{x,i}-g_{x,i}}{W}\right)^2+\left(\frac{d_{y,i}-g_{y,i}}{H}\right)^2\right], \tag{10}$$

where M is the number of landmarks, $(d_{x,i}, d_{y,i})$ and $(g_{x,i}, g_{y,i})$ are predicted and ground-truth coordinates, and W, H are image size.

[1]https://aasce19.github.io/#challenge-dataset

To assess clinical relevance, we evaluate Cobb angles derived from predicted landmarks using SMAPE and CMAE, following the AASCE protocol. Let a_{ji} and b_{ji} denote predicted and ground-truth Cobb angles of image j and curve type $i \in \{PT, MT, TL\}$, corresponding to proximal thoracic, main thoracic, and thoracolumbar/lumbar curves. SMAPE is:

$$SMAPE = \frac{100}{N}\sum_{j=1}^{N}\frac{\sum_{i=1}^{3}|a_{ji}-b_{ji}|}{\sum_{i=1}^{3}(a_{ji}+b_{ji})}, \tag{11}$$

where N is the number of test images.

Since Cobb angles are angular quantities, CMAE measures the mean circular discrepancy. We first define:

$$s_j = \sum_{i=1}^{3}\sin\left(\frac{\pi}{180}|a_{ji}-b_{ji}|\right), \tag{12}$$

$$c_j = \sum_{i=1}^{3}\cos\left(\frac{\pi}{180}|a_{ji}-b_{ji}|\right). \tag{13}$$

Then CMAE is computed as:

$$CMAE = \frac{1}{N}\sum_{j=1}^{N}\operatorname{atan2}(s_j, c_j)\frac{180}{\pi}. \tag{14}$$

V. EXPERIMENTS

A. Experimental Protocol

We compare with six vertebral landmark localization methods, namely Yi *et al.* [16], H3R [17], SLSN [18], W-Transformer [15], FARNet [1], and Cascaded CNNs [14], as well as the domain adaptation baseline DANN [4]. Cascaded CNNs is coordinate-regression-based, while the others are heatmap-based. The six landmark baselines and our method follow the same source-only protocol: training on $AASCE_{train}$ and testing on $AASCE_{test}$ and the private test set. DANN is target-aware and uses the 147-image private training split during training.

B. Source-Only Cross-Domain Evaluation

Table I reports strict source-only transfer to the private cohort. Our method achieves the best NMSE, SMAPE, and CMAE for Overall, Without Implants, and With Implants. On the implant-bearing subset, it attains 13.69 SMAPE and 3.80 CMAE, indicating stronger structural consistency for Cobb-angle estimation under severe implant-induced shift. Relative to other source-only direct-transfer methods, the consistent gains in SMAPE and CMAE suggest improved global geometry and Cobb-angle stability, not merely point-wise coordinate refinement.

Figure 5 shows that competing methods often miss or misplace landmarks, especially in the upper spine and highly curved regions. In implant-bearing cases, bright hardware and partial occlusion can attract heatmap responses, causing false detections and locally disordered vertebral configurations. Our method produces more complete and anatomically consistent sequences, indicating better suppression of implant-induced distractions.

TABLE I

STRICT SOURCE-ONLY CROSS-DOMAIN EVALUATION ON THE PRIVATE CLINICAL DATASET.

Method	Overall			Without Implants			With Implants		
	NMSE↓	SMAPE↓	CMAE↓	NMSE↓	SMAPE↓	CMAE↓	NMSE↓	SMAPE↓	CMAE↓
SLSN	3.53E-2	54.52	36.75	1.36E-2	39.16	30.61	4.97E-2	64.76	40.85
Yi *et al.*	7.13E-3	17.45	7.05	4.66E-3	15.13	10.34	8.78E-3	18.99	4.85
W-Transformer	8.81E-3	47.15	32.29	6.18E-3	28.30	23.19	1.06E-2	59.72	38.36
H3R	1.13E-2	29.28	11.54	1.16E-2	18.45	9.93	1.11E-2	36.50	12.62
FARNet	6.63E-2	60.93	47.39	2.79E-2	45.19	38.48	9.18E-2	71.43	53.33
Cascaded CNNs	1.12E-2	25.34	10.80	1.37E-2	23.55	13.96	9.58E-3	26.53	8.70
Ours	**5.69E-3**	**13.66**	**5.75**	**3.07E-3**	**13.61**	**8.68**	**7.44E-3**	**13.69**	**3.80**

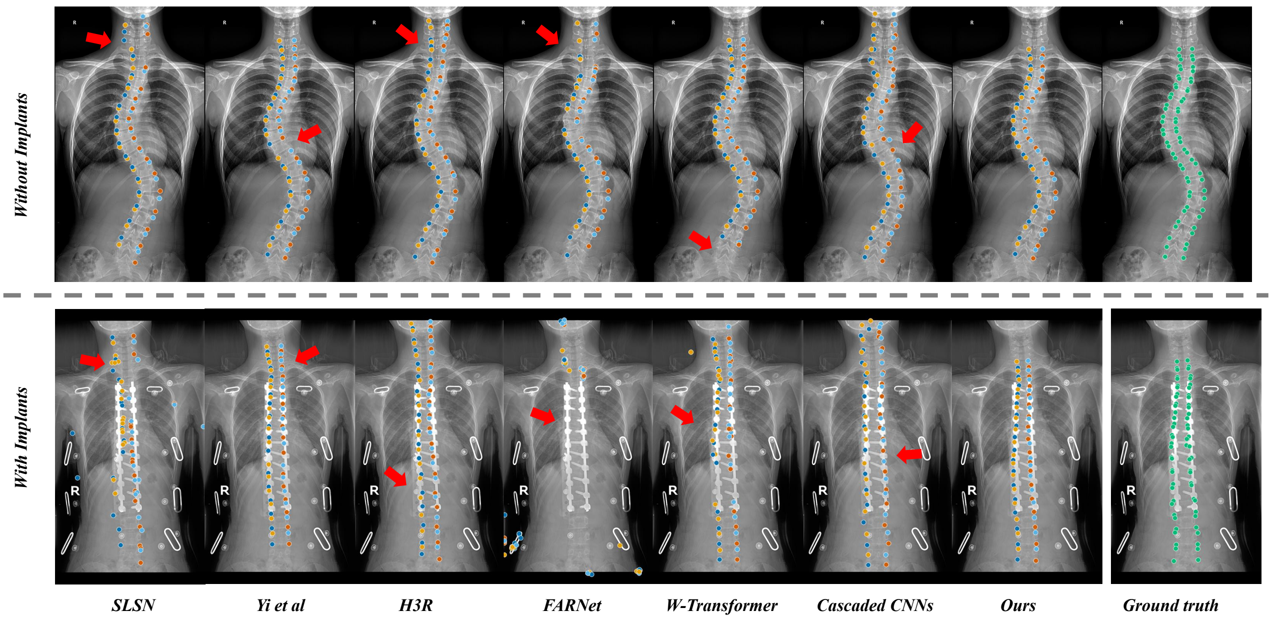


Fig. 5. Qualitative results on target-domain images with and without implants. Different colors denote the four corner points of each vertebra.

TABLE II

PERFORMANCE COMPARISON ON THE PUBLIC TEST SET.

Method	NMSE↓	SMAPE↓	CMAE↓
Wang *et al.* (2019a)	–	23.43	–
Huo *et al.*	–	20.51	–
Horng *et al.*	–	16.48	–
Wang *et al.* (2019b)	–	12.97	–
W-Transformer	–	8.26	–
SLSN	2.41E-3	17.81	8.79
Yi *et al.*	**1.24E-3**	8.67	3.73
H3R	3.34E-3	18.35	8.87
FARNet	4.68E-3	24.99	14.34
Cascaded CNNs	3.44E-3	19.89	9.32
Ours	1.54E-3	**7.90**	**3.40**

C. Public-Benchmark Performance

Table II reports $AASCE_{test}$ performance. Our method achieves the best SMAPE and CMAE among the reproduced methods while maintaining competitive NMSE. Although Yi *et al.* obtains the lowest NMSE, our method yields more accurate Cobb-angle estimation, showing that the style-structure framework preserves a competitive in-domain operating point rather than trading it away for robustness.

For completeness, the upper block lists previously reported benchmark results from Huo *et al.* [8], W-Transformer [15], and results quoted by W-Transformer [13], [11], [6]; only SMAPE is shown where NMSE or CMAE is unavailable.

TABLE III

IMPLANT-BEARING SUBSET RESULTS UNDER A RELAXED PROTOCOL.

Method	NMSE↓	SMAPE↓	CMAE↓
SLSN	1.35E-3	30.47	11.36
Yi *et al.*	6.61E-4	17.08	4.16
W-Transformer	**5.54E-4**	21.62	7.58
H3R	7.92E-4	29.83	7.35
FARNet	1.68E-3	32.19	14.41
Cascaded CNNs	2.27E-3	43.46	10.30
DANN	6.46E-3	18.82	4.71
Ours (no fine-tuning)	7.44E-3	**13.69**	**3.80**

D. Comparison Under a Looser Protocol

Table III gives a relaxed comparison on implant-bearing images: DANN uses private target data, other baselines are

TABLE IV

ABLATION STUDY ON THE IMPLANT-BEARING SUBSET UNDER THE CROSS-DOMAIN SETTING.

Method	NMSE↓	SMAPE↓	CMAE↓
Baseline	8.20E-3	19.29	5.23
+ *MixStyle*	9.22E-3	17.24	4.96
+ *MixStyle* + *SWSPD*	7.44E-3	13.69	3.80

fine-tuned on the private domain, and our method remains without private-domain fine-tuning. Although our method does not achieve the best NMSE, it still obtains the lowest SMAPE and CMAE. This suggests that target-domain supervision can improve point-level localization, while source-only style-structure regularization remains competitive for clinically relevant angle consistency.

E. Ablation Study

Table IV reports ablations on the implant-bearing subset. *MixStyle* improves Cobb-angle metrics over the baseline, reducing SMAPE from 19.29 to 17.24 and CMAE from 5.23 to 4.96. Adding *SWSPD* further improves all metrics to 7.44E-3 NMSE, 13.69 SMAPE, and 3.80 CMAE, showing complementary gains from appearance perturbation and structure-aware decoding.

VI. DISCUSSION

The results show that the proposed framework remains competitive on the public source benchmark and improves direct transfer to the private cohort, especially in implant-bearing cases. This supports combining appearance-level perturbation with structure-aware decoding when local image evidence becomes unreliable under domain shift. Importantly, the strongest improvements appear in angle-based metrics, which are more directly tied to scoliosis assessment than isolated coordinate errors.

The two components are complementary. *MixStyle* improves tolerance to contrast, texture, and artifact shifts by perturbing feature statistics during training. However, appearance robustness alone is insufficient because landmark failures are often structured rather than purely local. *SWSPD* therefore suppresses anatomically implausible predictions through slot-wise ordering and stable inter-vertebral transitions.

The framework does not explicitly model implant geometry, severe occlusions, or institution-specific imaging variation, so postoperative domain shift remains challenging in complex cases. Future work may incorporate uncertainty-aware structural reasoning, explicit implant-artifact modeling, or target-aware adaptation when allowed.

VII. CONCLUSIONS

This paper studies source-only vertebral landmark detection from public preoperative radiographs to unseen private clinical images. The proposed style-structure framework combines *MixStyle* for appearance robustness with *SWSPD* for anatomically ordered decoding. Experiments on a public benchmark and a private clinical cohort show competitive source-domain performance and stronger postoperative cross-domain robustness.

REFERENCES

[1] Y. Ao and H. Wu. Feature aggregation and refinement network for 2D anatomical landmark detection. *Journal of Digital Imaging*, 36(2):547–561, 2023.

[2] J. Cool, A. M. Post, B. J. van Royen, M. Maas, G. J. Streekstra, F. S. Jamaludin et al. Radiological follow-up strategies in adolescent idiopathic scoliosis patients. a best evidence synthesis by systematic review. *Brain and Spine*, p. 105865, 2025.

[3] V. Dumoulin, J. Shlens, and M. Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016.

[4] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette et al. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

[5] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423, 2016.

[6] M.-H. Horng, C.-P. Kuok, M.-J. Fu, C.-J. Lin, and Y.-N. Sun. Cobb angle measurement of spine from X-ray images using convolutional neural network. *Computational and mathematical methods in medicine*, 2019(1):6357171, 2019.

[7] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pp. 1501–1510, 2017.

[8] L. Huo, B. Cai, P. Liang, Z. Sun, C. Xiong, C. Niu et al. Joint spinal centerline extraction and curvature estimation with row-wise classification and curve graph network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 377–386. Springer, 2021.

[9] A. H. Jinnah, K. A. Lynch, T. R. Wood, and M. S. Hughes. Adolescent idiopathic scoliosis: advances in diagnosis and management. *Current reviews in musculoskeletal medicine*, 18(2):54–60, 2025.

[10] C. Payer, D. Štern, H. Bischof, and M. Urschler. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Medical image analysis*, 54:207–219, 2019.

[11] J. Wang, L. Wang, and C. Liu. A multi-task learning method for direct estimation of spinal curvature. In *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, pp. 113–118. Springer, 2019.

[12] L. Wang, C. Xie, Y. Lin, H.-Y. Zhou, K. Chen, D. Cheng et al. Evaluation and comparison of accurate automated spinal curvature estimation algorithms with spinal anterior-posterior X-ray images: The AASCE2019 challenge. *Medical image analysis*, 72:102115, 2021.

[13] L. Wang, Q. Xu, S. Leung, J. Chung, B. Chen, and S. Li. Accurate automated Cobb angles estimation using multi-view extrapolation net. *Medical image analysis*, 58:101542, 2019.

[14] Z. Wang, J. Lv, Y. Yang, Y. Lin, Q. Li, X. Li et al. Accurate scoliosis vertebral landmark localization on x-ray images via shape-constrained multi-stage cascaded CNNs. *Fundamental Research*, 4(6):1657–1665, 2024.

[15] Y. Yao, W. Yu, Y. Gao, J. Dong, Q. Xiao, B. Huang et al. W-transformer: accurate Cobb angles estimation by using a transformer-based hybrid structure. *Medical Physics*, 49(5):3246–3262, 2022.

[16] J. Yi, P. Wu, Q. Huang, H. Qu, and D. N. Metaxas. Vertebra-focused landmark detection for scoliosis assessment. In *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*, pp. 736–740. IEEE, 2020.

[17] B. Yu and D. Tao. Heatmap regression via randomized rounding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8276–8289, 2021.

[18] C. Zhang, J. Wang, J. He, P. Gao, and G. Xie. Automated vertebral landmarks and spinal curvature estimation using non-directional part affinity fields. *Neurocomputing*, 438:280–289, 2021.

[19] J. L. Zheng, Y. Li, G. Hogue, M. Johnson, J. B. Anari, M. D. Regan et al. What imaging does my AIS patient need? a multi-group survey of provider preferences. *Spine deformity*, 13(2):351–359, 2025.

[20] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021.