ITMatch: Arch-Guided Semi-Supervised Tooth Arrangement via Iterative Confidence Evaluation

Chengyuan Wang[1], Zhihui He[1], Li Chen[1]✉, Shidong Yang[1], Guiyu Sun[1], Fan Duan[1], Shuo Wang[2], and Yanheng Zhou[3]

[1] School of Software, Tsinghua University, Beijing, China
chenlee@tsinghua.edu.cn
[2] Peking University Hospital of Stomatology, Orthodontic, Beijing, China
[3] Tsinghua University Affiliated Beijing Tsinghua Changgung Hospital, Dental, Beijing, China

Abstract. Automated tooth arrangement is a crucial stage in digital orthodontic planning. Existing learning-based methods are based on large-scale expert-designed treatment plans, but high-quality arrangement results are difficult to obtain. Semi-supervised learning is commonly applied in scenarios with limited labeled data. However, due to the challenge of evaluating the confidence of pseudo-labels, previous works have not effectively explored semi-supervised tooth arrangement as a regression problem. To address this, we propose a semi-supervised tooth arrangement framework guided by dental arch priors and iterative confidence evaluation. We establish a teacher-student-based semi-supervised framework and introduce a weak-to-strong consistency regularization tailored for 3D point clouds. Inspired by optimization problems, we iteratively analyze errors to assess the confidence of pseudo-labels generated by the teacher network, mitigating the challenge of filtering low-quality pseudo-labels in regression. In addition, we predict the dental arch width to reduce the complexity of learning intricate transformations and leverage it as orthodontic prior information to improve arrangement accuracy. Our framework fills a critical gap in the field, and its core ideas can be generalized to other regression tasks. On a high-quality dataset, our method achieves competitive results with minimal labeled data. Code and typical data are available at `https://github.com/oblivionis-tgw/ITMatch`.

Keywords: Semi-Supervised Regression · Tooth Arrangement.

1 Introduction

Digital orthodontic planning, which leverages computational methods to design optimal treatment plans, is progressively reshaping modern dental practices [9]. Automated tooth arrangement is a key step in digital orthodontic planning, aiming to predict optimal tooth positions to achieve both functional occlusion and aesthetic outcomes. In recent years, advances in deep learning have facilitated the application of learning-based methods to the tooth arrangement task. Several methods [18,3,15,8] have explored point cloud-based predictions of final tooth positions, but the performance is heavily constrained by the availability of high-quality labeled data. Due to the complexity of orthodontic treatment and individual patient variability, manual annotation is expensive and difficult to obtain, limiting the scalability of supervised learning methods.

Semi-supervised learning (SSL) offers a promising alternative by leveraging a limited number of labeled samples alongside a large volume of unlabeled data. While SSL has achieved remarkable success in classification tasks [1,12,17], its application to regression problems remains an active research area. One key challenge is the reliable assessment of pseudo-labels for continuous target values [7,14]. Existing approaches [5,4,19,16] using image, textual, or point cloud inputs employ uncertainty-based filtering strategies for pseudo-labels in regression tasks such as age estimation and object detection, but struggle with teeth transformations due to complex geometric and spatial relationships among teeth.

Tooth arrangement presents additional challenges beyond standard regression: it requires both local and global structural perception, estimation of final positions, and the regression of rigid transformation parameters. Training a robust point cloud regressor under limited supervision is particularly difficult, often leading to high pseudo-label noise. Therefore, it is crucial to incorporate domain-specific priors to simplify learning and improve pseudo-label reliability.

In this paper, we formulate automated tooth arrangement as a 3D point cloud regression problem and propose a semi-supervised regression framework based on a teacher-student paradigm, where the student network learns from pseudo-labels provided by the teacher network of the same architecture. Consistency regularization is commonly used in SSL to enforce consistency in network predictions under different input perturbations. In our case, given dental cases with identical tooth morphology, the network should naturally yield similar predictions when provided with different perturbations of the initial tooth positions. Therefore, we design a weak-to-strong consistency regularization strategy tailored to this problem within our framework.

A key challenge in semi-supervised regression is pseudo-label confidence estimation, particularly in 3D geometric transformations. Since tooth arrangement operates in an inherently iterative manner, we draw inspiration from iterative optimization methods [6], where convergence is assessed based on error reduction between consecutive iterations. i.e., the algorithm is considered converged when $|f(f(x_i)) - f(x_i)| < \epsilon$. Inspired by this principle, we propose an iterative confidence evaluation mechanism, which quantifies pseudo-label reliability by measuring the discrepancy between consecutive predictions. This effectively filters out low-quality pseudo-labels, improving the stability of semi-supervised training. In addition, we introduce a regularization loss during supervised training, discouraging unnecessary re-arrangement of already well-aligned teeth.

To further enhance robustness under limited supervision, we employ a multi-task learning strategy to assist both regressor training and confidence evaluation. Specifically, we introduce an auxiliary regression head to predict the dental arch width, encoding the extracted features as orthodontic constraints, guiding the arrangement predictions. This auxiliary task not only simplifies the learning of intricate tooth transformations but also enhances the accuracy of pseudo-label confidence evaluation.

In summary, our major contributions are: 1) We propose the first semi-supervised regression framework for automated tooth arrangement, introducing a weak-to-strong consistency regularization method tailored for point cloud inputs. 2) We propose a novel iterative confidence evaluation mechanism to effectively filter pseudo-labels, addressing a key challenge in applying semi-supervised regression to tooth arrangement. 3) We incorporate dental arch priors into the framework, utilizing arch width prediction to simplify tooth transformation learning and enhance arrangement accuracy. 4) We validate our proposed approach on a high-quality tooth arrangement dataset and achieve competitive results with minimal labeled data.

2 Methods

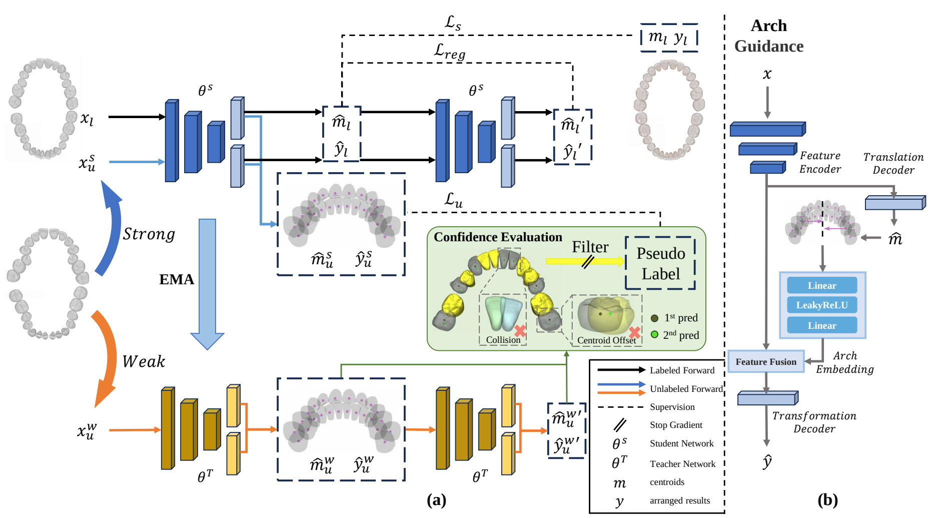

Fig. 1: The pipeline of our ITMatch framework.

2.1 Preliminary

In the automated tooth arrangement task, an input sample consists of a set of point cloud representations of teeth, denoted as $x = \{x^i \in \mathbb{R}^{3\times N} | i \in \boldsymbol{L}\}$, where $\boldsymbol{L}$ is the set of tooth labels, and N denotes the number of sampled points per tooth. Our objective is to train a network θ to regress the 6-DoF transformation parameters T^i for each tooth, such that the transformed teeth set $T * x$ closely aligns with the expert-designed arrangement, where * denotes the application of a 6-DoF transformation. This can be formulated as the following optimization problem:

$$\min_{\theta} \sum_{i \in \boldsymbol{L}} ||T^i * x^i - y^i||_2^2 \tag{1}$$

where $*$ denotes the application of a 6-DoF transformation. Following prior works, we parameterize the transformation T^i using a 3D translation vector and a unit quaternion to represent rotation. The network output is given by $T = \theta(x) = (\boldsymbol{t}, \boldsymbol{q})$.

2.2 Framework: ITMatch

Our semi-supervised tooth arrangement framework named ITMatch is illustrated in Figure 1. The framework consists of two networks with identical architectures: a student network θ_s and a teacher network θ_t. The teacher network is updated using the Exponential Moving Average (EMA) of the student network at each training iteration. A mini-batch D is divided into a labeled subset D_l and an unlabeled subset D_u. For labeled data x_l with ground truth annotations y_l, the student network θ_s is trained using supervised loss $\mathcal{L}_s = \mathcal{L}_s(x_l, y_l)$. For unlabeled data x_u, we first apply data augmentation with different intensity levels, generating weakly augmented x_u^w and strongly augmented x_u^s versions. To better accommodate the tooth arrangement task, we carefully design an augmentation library, detailed in Table 2. Next, the student network and teacher network generate predictions for the augmented samples:

$$\hat{y}_u^s = \theta_s(x_u^s) * x_u^s, \quad \hat{y}_u^w = \theta_t(x_u^w) * x_u^w \tag{2}$$

After confidence evaluation, the teacher network's reliable predictions are selected as pseudo-labels for training the student network. This encourages the network to learn consistent predictions under different perturbation levels. The unsupervised loss is formulated as $\mathcal{L}_u = \mathcal{L}_u(\hat{y}_u^s, h_{\text{conf}}(\hat{y}_u^w))$, where $h_{\text{conf}}(\cdot)$ represents the confidence-based filtering function.

2.3 Arch Perception and Guidance

Tooth arrangement requires the network to simultaneously capture both local tooth features and global dental arch structure, enabling it to determine target positions and regress transformation parameters. However, under limited supervision, directly adopting conventional semi-supervised learning paradigms often leads to low-quality pseudo-labels. In practice, aesthetically and functionally desirable dental arches exhibit a smooth arch-like curve, and orthodontists often rely on arch width as a guideline for arranging teeth.

To incorporate arch priors into the learning process, we introduce an additional auxiliary regressor to predict the control points of the dental arch. To simplify data annotation, we use tooth centroids as control points instead of manually annotated landmarks. Since the centroid naturally serves as the center of transformation for each tooth, we can supervise the network's translation predictions and use the estimated centroids to guide the final transformation regression. By predicting centroid displacement instead of directly regressing full 6-DoF transformations, we reduce the learning complexity, focusing the network on estimating a reasonable arch width rather than complex rotational transformations. Specifically, we introduce an auxiliary regressor ϕ', which shares a nearly identical architecture with the primary regressor ϕ, but is only for predicting the translation component $\boldsymbol{t}$. The predicted centroid positions for all teeth are given by $\hat{m} = \{\hat{m}_i \in \mathbb{R}^3 | i \in \boldsymbol{L}\}$, which can be computed as $\hat{m} = \boldsymbol{t} * x_m = \phi'(f) * x_m$, where f represents point cloud features encoded by the feature encoder.

To supervise the centroid prediction, we apply L2 loss to both labeled and unlabeled data:

$$\mathcal{L}_{m,s} = \frac{1}{|\boldsymbol{L}|} \sum_{i \in \boldsymbol{L}} ||\hat{m}_l^i - m_l^i||_2^2 \qquad \mathcal{L}_{m,u} = \frac{1}{|\boldsymbol{L}|} \sum_{i \in \boldsymbol{L}} W_{\text{conf}}^i ||\hat{m}_u^{s,i} - \hat{m}_u^{w,i}||_2^2 \tag{3}$$

Here, $W_{\text{conf}} \in \{0,1\}^{|\boldsymbol{L}|}$ is a confidence weighting vector, which will be detailed in the next section. Using the estimated centroid positions, as illustrated in Figure 1(b), we compute the directed distances x_r from the centroids to the median plane. These distances are then encoded into an arch width representation using a lightweight multi-layer perceptron (MLP). The arch features $f_r =$ `Linear(LeakyReLU(Linear(x_r)))` are fused with the original feature f and passed into ϕ to obtain the final transformation parameters, i,e, $T = \phi(f \oplus f_r)$, where $\oplus$ denotes element-wise addition, and the specific implementation of feature fusion can be substituted as needed.

2.4 Iterative Confidence Evaluation

Assessing the confidence of pseudo-labels is a fundamental challenge in semi-supervised regression. Inspired by iterative optimization methods, we propose a more principled approach: evaluating pseudo-label confidence based on the discrepancy between consecutive iterations. Given the network-predicted tooth arrangement $\hat{y}$, we feed it back into the same network to obtain a second-stage prediction $\hat{y}' = \theta(\hat{y}) * \hat{y}$. Ideally, $\hat{y}$ is expected to be close to the ground truth y, i.e., $||\hat{y} - y|| < \delta_1$. Additionally, if the network's rearrangement of an already well-aligned dentition is minimal, we have $||\theta(y) * y - y|| < \delta_2$. Since $\theta(\cdot)$ is a continuous function, it is easy to derive that $||\hat{y}' - \hat{y}|| < \delta$, where δ is also a small deviation. Thus, by evaluating the difference between $\hat{y}'$ and $\hat{y}$, we can assess how close $\hat{y}$ is to the target arrangement y, which serves as an indicator of pseudo-label confidence.

In practice, to reduce instability, we use the predicted centroids $\hat{m}_u^w$ instead of $\hat{y}_u^w$ for confidence estimation. Additionally, we incorporate a collision loss $\mathcal{L}_c$ [3], to evaluate whether the relative positions of adjacent teeth (x^i, x^j) are reasonable. This also serves as a complementary measure for assessing rotational transformations. The confidence weighting matrix is defined as:

$$W_{\text{conf}} = \mathbb{1}(||\hat{m}_u^w - \hat{m}_u^{w\prime}||_2 < \gamma) \cdot \mathbb{1}(\mathcal{L}_c(\hat{y}_u^w) < \gamma_c) \tag{4}$$

where W_{conf} is a binary vector of length $|\boldsymbol{L}|$, γ and γ_c are hyperparameters, and $\mathbb{1}(\cdot)$ denotes the indicator function.

To reinforce the network's understanding of well-aligned dentitions and prevent mode collapse, we apply the second-stage iterative process to labeled data as well. We introduce a regularization loss $\mathcal{L}_{\text{reg}}$ to minimize the discrepancy between first- and second-stage centroid predictions:

$$\mathcal{L}_{\text{reg}} = ||\hat{m}_l - \hat{m}_l'||_2^2 \tag{5}$$

The final objective function is formulated as $\mathcal{L} = \mathcal{L}_s + \lambda_u \mathcal{L}_u + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}}$, where λ_u and λ_{reg} are hyperparameters controlling the balance between different loss terms. Beyond the specific loss functions discussed above, our framework supports arbitrary point cloud regression losses. The details of the loss functions used in our implementation will be elaborated in Section 3.1.

3 Experiments

3.1 Experimental Settings

Loss Function and Training Details. We adopt the supervised loss functions proposed in [3] and strictly follow its hyperparameter settings. The loss components include: 1) **Reconstruction Loss** $\mathcal{L}_r^i = ||T^i * x^i - y^i||_2^2$. 2) **Transformation Parameter Loss** $\mathcal{L}_p^i = ||\boldsymbol{q}_{x^i} - \boldsymbol{q}_{x^{i*}}||$, where q_{x^i} and $q_{x^{i*}}$ indicate the predicted quaternion and ground truth quaternion of teeth x^i, respectively. 3) **Feature Consistency Loss** $\mathcal{L}_f$. 4) **Collision Loss** $\mathcal{L}_c$. The supervised and unsupervised losses share the same formulation and can be unified as:

$$\mathcal{L}_{\{s,u\}} = \frac{1}{2}\mathcal{L}_{r,\{s,u\}} + 20\mathcal{L}_{p,\{s,u\}} + \mathcal{L}_{f,\{s,u\}} + 2\mathcal{L}_{c,\{s,u\}} + \frac{1}{2}\mathcal{L}_{m,\{s,u\}} \tag{6}$$

For optimization, we use SGD [11] with an initial learning rate of 1×10^{-3} and weight decay set to 1×10^{-4}. Training is conducted on a single NVIDIA RTX 3090 GPU. The batch size is set to 16, with equal proportions of labeled and unlabeled data in each mini-batch. We sample $N = 512$ points from the point cloud data of each tooth using the Farthest Point Sampling method. Regarding the value of hyperparameters, we set $\gamma = \gamma_c = 0.5, \lambda_{\text{reg}} = 0.1$ by Grid-Search method. $\lambda_u(t_i) = \mathrm{e}^{-5(1-t_i/t_{\max})^2}$, where t_i represents the current training iteration, and $t_{\max} = 8000$ is the maximum number of iterations [10].

Datasets and Evaluation Metrics. We conduct experiments on the dataset proposed in [3], which consists of 909 cases from real-world orthodontic treatment plans, each annotated with transformation parameters provided by experts. We select its High-Quality dataset containing 544 cases, using 382, 54, and 108 cases for training, validation, and testing, with the same data preprocessing.

For evaluation, we adopt metrics similar to [15,3]. Specifically, we compute PCT@K, the percentage of predictions with errors below a threshold K, and use

Table 1: Quantitative comparison of different settings.

Labeled	Method	Backbone	$\mathrm{ME_{point}}\downarrow$	AUC↑	Backbone	$\mathrm{ME_{point}}\downarrow$	AUC↑
5%	LS	TANet [18]	1.2791	63.60	DTAN [3]	1.2775	64.06
	TA-MT		1.2187	66.45		1.1633	68.11
	TA-CPS		1.2013	66.99		1.1544	68.83
	ITMatch		**1.0773**	**71.26**		**1.0589**	**71.85**
10%	LS	TANet	1.1516	69.12	DTAN	1.1453	68.84
	TA-MT		1.0882	70.14		1.0723	71.37
	TA-CPS		1.1222	69.84		1.0547	71.99
	ITMatch		**0.9987**	**74.11**		**0.9775**	**74.82**
20%	LS	TANet	1.0423	73.23	DTAN	0.9880	74.48
	TA-MT		1.0049	73.69		0.9795	74.78
	TA-CPS		1.0178	73.40		0.9730	74.99
	ITMatch		**0.9386**	**76.35**		**0.9170**	**77.02**
100%	FS	TANet	0.8846	78.22	DTAN	0.7952	81.48

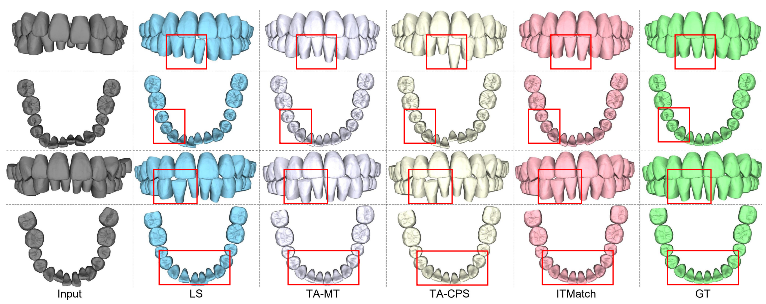

Fig. 2: Visualization results of different methods on 5% labeled setting.

it to construct a PCT curve with a maximum threshold of 3 mm and an interval of 0.01 mm. The AUC (area under the PCT curve) quantifies overall accuracy, while the mean point-wise distance error ($\mathrm{ME_{point}}$) (mm) measures the average Euclidean distance between predicted and ground truth point clouds.

3.2 Results and Analysis

Quantitative Comparison. To quantitatively evaluate our method, we performed experiments with proportions of data labeled 5% (19 cases), 10% (38 cases), and 20% (76 cases). Since no existing semi-supervised methods are specifically designed for tooth arrangement, we adapt two classic paradigms from semi-supervised learning and segmentation, Mean Teacher [13] and CPS [2], to suit this task, resulting in TA-MT and TA-CPS, respectively. For fair comparisons, we keep the loss functions, data augmentation, and hyperparameters identical across all methods. As shown in Table 1, thanks to the effective pseudo-label filtering strategy, our method consistently achieves the highest accuracy on different backbones and annotation ratios, significantly outperforming Limited Supervision (LS) and approaching the performance of Fully Supervision (FS). Visualization results are shown in Figure 2.

Table 2: Data augmentation strategies for weak-to-strong consistency

Method	**Weak**	**Strong**
Tooth Rotation	$[-30°, 30°]$	$[-36°, 36°]$
Translation(mm)	$\mathbb{N}(0, 1^2)$	$\mathbb{N}(0, 1.2^2)$
Resampling	-	$p = 50\%$
Global Flipping	-	$p = 25\%$

Table 3: Ablation study with 36 labeled cases(10%) on DTAN

TS	CR	ICE	AG	$\text{ME}_{\text{point}}\downarrow$	AUC↑
✓				1.0723	71.37
✓	✓			1.0303	72.90
✓	✓	✓		0.9948	74.20
✓	✓		✓	1.0078	73.70
✓	✓	✓	✓	**0.9775**	**74.82**
w/o $\mathcal{L}_{\text{reg}}$				1.0511	72.20
uncertainty filtering				1.0286	73.12

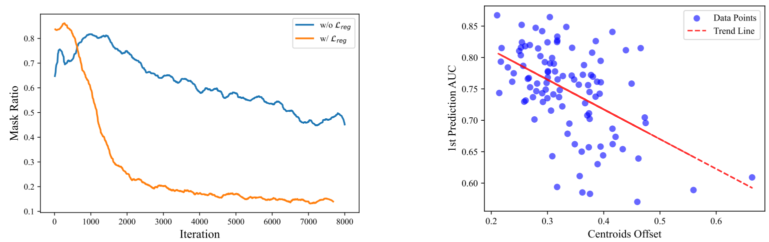

Fig. 3: Analysis of the effect of $\mathcal{L}_{\text{reg}}$ on pseudo-label utilization (left) and empirical evidence for the iterative confidence evaluation method (right).

Ablation studies. To validate the effectiveness of each component in our proposed framework, we conduct ablation studies under the 10% labeled setting. Starting from the Teacher-Student paradigm (TS), we progressively incorporate Weak-to-Strong Consistency Regularization (CR), Iterative Confidence Evaluation (ICE), and Arch Guidance (AG). Table 3 presents the contribution of each component. Notably, removing $\mathcal{L}_{\text{reg}}$, which plays a key role in ensuring the validity of confidence estimation via iteration error, results in a significant performance drop. Additionally, replacing ICE with an uncertainty-based pseudo-label filtering approach also leads to degraded performance.

Furthermore, we empirically verify the effectiveness of ICE. As shown in Figure 3, introducing $\mathcal{L}_{\text{reg}}$ enables the pseudo-label mask ratio to gradually decrease over training iterations, dynamically improving pseudo-label utilization. The right plot further illustrates a negative correlation between centroid displacement over iterations and prediction accuracy, providing strong empirical evidence for the validity of our approach.

4 Conclusion

In this paper, we propose ITMatch, the first semi-supervised regression framework tailored for tooth arrangement. Built upon a teacher-student paradigm, ITMatch enhances the learning of unlabeled data through weak-to-strong consistency regularization for point cloud inputs. To address the low confidence of pseudo-labels in 3D transformation regression, we introduce an iterative confidence evaluation mechanism inspired by optimization methods. Additionally, by incorporating arch-guided multi-task learning, we improve the robustness of the regressor under limited supervision. Experimental results on real-world datasets demonstrate that ITMatch achieves competitive performance. Future work will further explore pseudo-label confidence evaluation for iterative regression tasks and strategies to refine pseudo-labels and improve their utilization.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (Grant No. 61972221).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. Advances in neural information processing systems **32** (2019)
2. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
3. He, Z., Wang, C., Yang, S., Chen, L., Zhou, Y., Wang, S.: Differentiable collision-supervised tooth arrangement network with a decoupling perspective. arXiv preprint arXiv:2409.11937 (2024)
4. Jean, N., Xie, S.M., Ermon, S.: Semi-supervised deep kernel learning: Regression with unlabeled data by minimizing predictive variance. Advances in Neural Information Processing Systems **31** (2018)
5. Jo, Y., Kahng, H., Kim, S.B.: Deep semi-supervised regression via pseudo-label filtering and calibration. Applied Soft Computing **161**, 111670 (2024)
6. Kelley, C.T.: Iterative methods for optimization. SIAM (1999)
7. Kostopoulos, G., Karlos, S., Kotsiantis, S., Ragos, O.: Semi-supervised regression: A recent review. Journal of Intelligent & Fuzzy Systems **35**(2), 1483–1500 (2018)
8. Li, X., Bi, L., Kim, J., Li, T., Li, P., Tian, Y., Sheng, B., Feng, D.: Malocclusion treatment planning via pointnet based spatial transformation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 105–114. Springer (2020)
9. Littlewood, S.J., Mitchell, L.: An introduction to orthodontics. Oxford university press (2019)
10. Liu, J., Desrosiers, C., Zhou, Y.: Semi-supervised medical image segmentation using cross-model pseudo-supervision with shape awareness and local context constraints. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII. pp. 140–150. Springer (2022)
11. Robbins, H., Monro, S.: A stochastic approximation method. The annals of mathematical statistics pp. 400–407 (1951)
12. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)

13. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
14. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine learning **109**(2), 373–440 (2020)
15. Wang, C., Wei, G., Wei, G., Wang, W., Zhou, Y.: Tooth alignment network based on landmark constraints and hierarchical graph structure. IEEE Transactions on Visualization and Computer Graphics (2022)
16. Wang, H., Cong, Y., Litany, O., Gao, Y., Guibas, L.J.: 3dioumatch: Leveraging iou prediction for semi-supervised 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14615–14624 (2021)
17. Wang, Y., Chen, H., Heng, Q., Hou, W., Fan, Y., Wu, Z., Wang, J., Savvides, M., Shinozaki, T., Raj, B., et al.: Freematch: Self-adaptive thresholding for semi-supervised learning. arXiv preprint arXiv:2205.07246 (2022)
18. Wei, G., Cui, Z., Liu, Y., Chen, N., Chen, R., Li, G., Wang, W.: Tanet: towards fully automatic tooth arrangement. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16. pp. 481–497. Springer (2020)
19. Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., Liu, Z.: End-to-end semi-supervised object detection with soft teacher. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3060–3069 (2021)