ProtoReg: Prototype Guidance and Dual-level Uncertainty-weighted Consistency for Semi-supervised Pulmonary Perfusion Regression

Yitao Zhuang[1†], Yao Pu[1†], Zhenxi Zhang[1], Peixin Yu[1], Zihan Li[1], Mayang Zhao[1], Wai Yin Ho[3], Jocelyn Chu[4], Gary Ka Wai Chan[3], Jing Cai[1*], and Ge Ren[1,2*]

[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR
[2] The Hong Kong Polytechnic University Shenzhen Research Institute, The Hong Kong Polytechnic University, China
[3] Department of Radiology, Queen Mary Hospital, Hong Kong SAR
[4] Department of Radiology & Nuclear Medicine, Tuen Mun Hospital, Hong Kong SAR

Abstract. Pulmonary perfusion imaging is clinically important, but paired 3D CT–perfusion annotations are difficult to obtain. Semi-supervised learning (SSL) can exploit abundant unlabeled CT volumes, yet existing SSL methods are less effective for voxel-wise dense regression because continuous pseudo-labels lack reliable confidence estimates and strong consistency may amplify teacher noise. To address this, we propose ProtoReg, a semi-supervised framework for 3D non-contrast CT-to-perfusion regression with prototype-guided uncertainty modeling and sub-unit-aware learning. ProtoReg adopts a teacher–student framework with weak-to-strong consistency and maintains a learnable functional prototype bank to estimate uncertainty via prototype-induced entropy. Based on this uncertainty, we introduce a Dual-level Uncertainty-weighted Consistency Loss (DUW-CL), which re-weights consistency at both voxel and lung sub-unit levels to suppress noisy pseudo-label propagation. We further incorporate 1,000 external unlabeled CT volumes from public datasets to improve generalization without additional perfusion annotations. Experiments on two independent CT-perfusion cohorts show consistent improvements over strong SSL baselines, and ablations validate the effectiveness of prototype guidance, DUW-CL, and supervoxel-aware modeling. Code is publicly available at `https://github.com/johnhamtom/ProtoReg_main`.

Keywords: Lung imaging · Pulmonary perfusion · Semi-supervised learning · CT-to-SPECT regression · Consistency regularization · Uncertainty estimation.

* Corresponding authors. E-mail: gary-ge.ren@polyu.edu.hk, jing.cai@polyu.edu.hk.
† Equal contribution.

1 Introduction

Pulmonary perfusion maps reflect regional pulmonary blood flow and are essential for regional lung-function assessment, supporting the detection of pulmonary abnormalities and informing function-aware radiotherapy planning [13]. However, current acquisition techniques often require additional dedicated scans and rely on radioactive tracers or contrast agents, leading to increased cost, operational burden, and limited accessibility [7]. These practical limitations have motivated the exploration of deep learning-based surrogates that infer perfusion maps from routinely acquired chest CT.

Recent years have seen growing evidence that deep networks can infer lung SPECT-like perfusion from non-contrast chest CT, achieving moderate-to-high voxel-wise agreement with reference perfusion and suggesting that pulmonary parenchymal textures encode function-related cues for CT-to-perfusion translation [17, 16, 13]. Nevertheless, most existing methods still depend on sufficiently paired CT–perfusion data for supervised training or fine-tuning. In practice, paired acquisitions remain costly and scarce, whereas large-scale unlabeled CT scans are far more accessible, motivating semi-supervised learning under limited paired supervision.

Semi-supervised learning (SSL) has achieved remarkable success in image classification and segmentation [28, 15, 26]. Recent prototype- and uncertainty-based methods, including CPCL [23], UPCoL [14], and DyCON [4], further improve SSL by leveraging class-level prototypes, uncertainty estimates, or probability maps. However, these methods mainly rely on discrete category-level signals, whereas CT-to-perfusion regression requires reliability estimation for continuous voxel-wise pseudo-labels without direct confidence measures. Noisy pseudo-labels may then be reinforced by self-training or strong consistency regularization, leading to confirmation bias and unstable optimization [9, 2, 21].

This problem is especially challenging in lung functional imaging, where 3D perfusion patterns are heterogeneous and susceptible to motion-related inconsistencies [10]. Accordingly, SSL for pulmonary functional mapping is still in its infancy, with only limited recent work on adjacent directions [24]. Furthermore, relying solely on voxel-wise constraints may reinforce spatially fragmented pseudo-label noise, motivating the incorporation of sub-unit–level regional consistency [27, 29].

In this paper, we formulate 3D CT-to-perfusion synthesis as voxel-wise dense regression and propose ProtoReg, a teacher–student semi-supervised framework that learns from limited paired CT–perfusion scans together with large-scale unlabeled CT volumes. ProtoReg uses an EMA teacher to generate continuous pseudo-targets, and trains the student with a consistency objective under plausible CT perturbations, enabling effective utilization of unlabeled data for dense regression.

A central challenge in semi-supervised dense regression is pseudo-label confidence estimation: continuous voxel-wise targets lack a natural confidence score, and naive consistency can amplify noise. Motivated by prototype-based representation learning [5], we maintain a learnable functional prototype bank and propose DUW-CL, a Dual-level Uncertainty-weighted Consistency loss. DUW-CL derives soft voxel-wise weights from prototype-induced entropy at voxel and lung sub-unit levels. Without hard filtering, all voxels are retained but assigned uncertainty-adaptive penalties: Early in training, the model focuses more on low-uncertainty regions at both voxel and sub-unit levels and tends to reduce the overall entropy; later, it shifts to refinement with milder entropy-guided weighting.

Overall, our contributions are fourfold: (1) We propose ProtoReg, a semi-supervised framework that integrates functional prototype guidance with uncertainty-aware optimization for accurate 3D lung perfusion estimation. (2) We introduce prototype-induced entropy as a reliability signal for continuous pseudo-targets, mitigating erroneous pseudo-label propagation in dense regression. (3) We propose DUW-CL, a Dual-level Uncertainty-weighted Consistency Loss that stabilizes training and promotes coherent functional patterns. (4) We demonstrate consistent gains on two independent CT-perfusion cohorts, with further improvements by incorporating external unlabeled CT volumes.

2 Method

2.1 Problem Setup

We study semi-supervised pulmonary perfusion synthesis from 3D non-contrast CT. The training set consists of a labeled set $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ and an unlabeled set $\mathcal{D}_u = \{x_j\}_{j=1}^{N_u}$, where x is a CT volume and y is the paired perfusion map (continuous-valued). We learn a voxel-wise regressor f_θ that outputs $\hat{y} = f_\theta(x)$. All voxel-wise losses are computed within the lung ROI Ω.

2.2 Overview of ProtoReg

An overview of ProtoReg is shown in Fig. 1. ProtoReg follows an EMA teacher–student paradigm with two networks of identical architecture: a student network f_θ and a teacher network $f_{\bar{\theta}}$. The teacher is updated by the exponential moving average (EMA) of the student at each iteration.

Labeled training. For labeled samples $(x_l, y_l) \in \mathcal{D}_l$, the student produces $\hat{y}_l = f_\theta(x_l)$ and is optimized by a supervised regression loss $\mathcal{L}_{sup} = \mathcal{L}_{sup}(\hat{y}_l, y_l)$ within Ω. Meanwhile, labeled perfusion maps provide reliable functional cues to update the prototype space: we quantize y_l into K functional bins and apply a prototype classification loss $\mathcal{L}_{proto}$ (Sec. 2.3) to train the prototype head and update the prototype bank.

Unlabeled training. For unlabeled samples $x_u \in \mathcal{D}_u$, we apply two augmentation strengths to obtain a weak view x_u^w and a strong view x_u^s. Then the student and teacher generate predictions:

$$\hat{y}_u^s = f_\theta(x_u^s), \quad \hat{y}_u^w = f_{\bar{\theta}}(x_u^w). \tag{1}$$

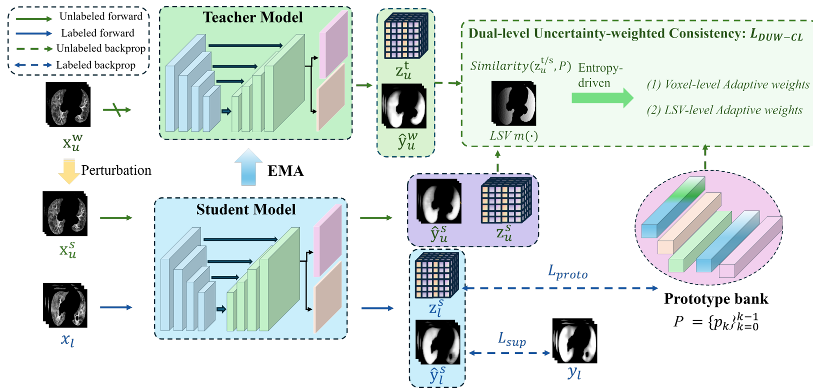

Fig. 1. Overview of the proposed ProtoReg framework.

We treat the teacher prediction $\hat{y}_u^w$ as the continuous pseudo target and enforce weak-to-strong consistency. Instead of confidence thresholding, we propose a Dual-level Uncertainty-weighted Consistency Loss (DUW-CL, Sec. 2.4), which estimates pseudo-label reliability via prototype-induced entropy and yields a voxel-wise reliability weight map $w(v)$ to suppress unreliable regions. The unlabeled objective is $\mathcal{L}_u = \mathcal{L}_{\text{DUW-CL}}$.

Overall objective and EMA update. The overall training objective is

$$\mathcal{L} = \underbrace{\mathcal{L}_{sup} + \lambda_{proto}\mathcal{L}_{proto}}_{\text{labeled}} + \underbrace{\mathcal{L}_{\text{DUW-CL}}}_{\text{unlabeled}}, \tag{2}$$

and the EMA update is $\bar{\theta} \leftarrow \alpha\bar{\theta} + (1-\alpha)\theta$.

2.3 Prototype Update Strategy

We maintain a bank of K functional prototypes $\mathcal{P} = \{p_k\}_{k=0}^{K-1}$ to structure the voxel embedding space. Given a perfusion value $y(v)$, we first normalize it into a fixed range and then quantize it into K discrete bins. Each voxel v is thus assigned a bin index $c(v) \in \{0, \ldots, K-1\}$ (from ground truth on labeled data).

For labeled samples $(x, y) \in \mathcal{D}_l$, the student network produces a voxel-wise embedding $z(v) \in \mathbb{R}^C$. We update the prototype space by encouraging $z(v)$ to be close to its assigned prototype $p_{c(v)}$. Specifically, we ℓ_2-normalize both embeddings and prototypes and compute cosine-similarity logits

$$s_k(v) = \frac{\langle \tilde{z}(v), \tilde{p}_k \rangle}{\tau}, \quad k = 0, \ldots, K-1, \tag{3}$$

where τ is a temperature scalar and $s(v) = [s_0(v), \ldots, s_{K-1}(v)]$. The prototype update loss is defined as a cross-entropy classification objective over the K

prototypes:

$$\mathcal{L}_{proto} = \frac{1}{|\Omega|} \sum_{v \in \Omega} \mathrm{CE}\big(\mathrm{softmax}(s(v)),\, c(v)\big). \tag{4}$$

In practice, prototypes $\{p_k\}$ are implemented as learnable parameters and updated jointly with the student network via back-propagation. This loss refines $\mathcal{P}$ so that voxels with similar perfusion levels form compact, prototype-aligned clusters in the embedding space.

2.4 Dual-level Uncertainty-weighted Consistency Loss

Inspired by DyCON [4], we propose Dual-level Uncertainty-weighted Consistency Loss (DUW-CL) for unlabeled data, consisting of (i) a Dual-level Uncertainty-weighted Consistency term and (ii) a voxel-entropy regularizer. In this section, s and t denote student (on the strong view) and teacher (on the weak view), respectively.

Prototype-induced entropy. Let $z_s(v), z_t(v) \in \mathbb{R}^C$ be voxel-wise prototype embeddings and $\mathcal{P} = \{p_k\}_{k=0}^{K-1}$ the prototype bank (Sec. 2.3). We ℓ_2-normalize embeddings/prototypes and compute cosine-similarity logits

$$s^k_{(\cdot)}(v) = \frac{\langle \tilde{z}_{(\cdot)}(v), \tilde{p}_k \rangle}{\tau}, \quad k = 0, \ldots, K-1, \tag{5}$$

followed by assignment probabilities $q_{(\cdot)}(v) = \mathrm{softmax}(s_{(\cdot)}(v))$ and entropy

$$H_{(\cdot)}(v) = -\sum_{k=0}^{K-1} q^k_{(\cdot)}(v) \log\big(q^k_{(\cdot)}(v) + \epsilon\big), \quad (\cdot) \in \{s, t\}. \tag{6}$$

(I) Dual-level Uncertainty-weighted Consistency. We partition the lung ROI Ω into M disjoint lung sub-units, implemented as lung supervoxels (LSV), and denote the sub-unit index of voxel v by $m(v) \in \{1, \ldots, M\}$. We compute LSV mean-entropy maps by averaging entropy within each LSV and broadcasting the values back to voxels:

$$\bar{H}_{(\cdot)}(v) = \frac{1}{|\{u : m(u) = m(v)\}|} \sum_{u: m(u) = m(v)} H_{(\cdot)}(u). \tag{7}$$

Here, β_{vox} and β_{lsv} control entropy-weighting sharpness: larger values emphasize low-entropy predictions, while smaller values yield flatter weights. We then define voxel-/LSV-level denominators

$$\begin{aligned} d_{\mathrm{vox}}(v) &= \exp\big(\beta_{\mathrm{vox}} H_s(v)\big) + \exp\big(\beta_{\mathrm{vox}} H_t(v)\big), \\ d_{\mathrm{lsv}}(v) &= \exp\big(\beta_{\mathrm{lsv}} \bar{H}_s(v)\big) + \exp\big(\beta_{\mathrm{lsv}} \bar{H}_t(v)\big), \end{aligned} \tag{8}$$

and obtain the reliability weight

$$w(v) = \mathbf{1}_{\Omega}(v) \cdot \Big((d_{\mathrm{vox}}(v) + \epsilon)(d_{\mathrm{lsv}}(v) + \epsilon)\Big)^{-1}. \tag{9}$$

The normalized weighted consistency loss is

$$\mathcal{L}_{\text{cons}} = \frac{\sum_{v\in\Omega} w(v)\,\ell_{\text{reg}}(\hat{y}_s(v),\, \text{sg}(\hat{y}_w(v)))}{\sum_{v\in\Omega} w(v) + \epsilon}, \tag{10}$$

where $\text{sg}(\cdot)$ denotes stop-gradient, ℓ_{reg} is SmoothL1 loss. Following DyCON, we apply a geometric decay schedule $\beta(t) = \max\left(\beta_{\min}, \beta_{\max}\left(\frac{\beta_{\min}}{\beta_{\max}}\right)^{t/T}\right)$, and set $\beta_{\text{vox}} = \beta_{\text{lsv}} = \beta(t)$ unless stated otherwise.

(II) Voxel entropy regularization and final loss. We regularize prototype assignments by

$$\mathcal{L}_{\text{ent}} = \frac{1}{|\Omega|}\sum_{v\in\Omega}\frac{H_s(v) + H_t(v)}{2}, \tag{11}$$

which discourages overly diffuse assignments and mitigates a degenerate solution where the student increases entropy to down-weight $w(v)$, and define the final unlabeled loss

$$\mathcal{L}_{\text{DUW-CL}} = \lambda_{\text{cons}}\mathcal{L}_{\text{cons}} + \lambda_{\text{ent}}\mathcal{L}_{\text{ent}}. \tag{12}$$

3 Experiments

3.1 Experimental Settings

Datasets and pre-processing. We conduct experiments on two cohorts. Cohort A is a private hospital dataset of 233 PE-suspected patients from Queen Mary Hospital (2019–2023), approved by the Institutional Review Board (Approval No. UW 20-105; age: 65.5 ± 15.8 years; male: 39.4%). Cohort B is an access-controlled dataset of 100 thoracic CT studies with rigidly co-registered lung perfusion SPECT and lobe segmentations [1]. Each case contains paired 3D non-contrast CT and $^{99\text{m}}$Tc-MAA SPECT. We also sample 1000 unlabeled CT volumes with axial slices $Z \geq 100$ from COVID19-CT-dataset [18], LIDC-IDRI [3], and MSD Task06_Lung [19]. For pre-processing, following [6], lung lobe masks were obtained using a pre-trained lobe segmentation model and merged as the lung ROI. Lung sub-units were generated by MaskSLIC [12] within each lobe at a target scale of $1\,\text{cm}^3$. The original CT and SPECT volumes were approximately $512 \times 512 \times 200$ and 128^3, respectively. Only the pre-segmented lung ROI was resampled/cropped to 128^3 to preserve CT–SPECT correspondence. CT intensities were clipped to $[-1000, 0]$ HU and linearly normalized, while SPECT was normalized by 1st/99th percentile stretching.

Training and evaluation. ProtoReg is implemented in PyTorch with a mean-teacher setup, where the student and EMA teacher ($\alpha = 0.996$) share a 3D U-Net [8] backbone and a lightweight prototype head with two $1\times1\times1$ convolutions. We use weak-to-strong consistency: both views share weak spatial transforms, while the student additionally receives intensity-only RandAugment

Table 1. Comparison on Cohort A.

Setting (L/U)	Method	Spearman $\rho \uparrow$	MAE $\downarrow$	SSIM $\uparrow$	FL20 $\uparrow$	FL40 $\uparrow$
100%/0%	Sup-only (3D U-Net)	0.780±0.115	0.132±0.032	0.758±0.117	0.908±0.072	0.817±0.109
100%/1000CT	MT-Reg	0.786±0.121	0.128±0.035	0.768±0.126	0.912±0.071	0.823±0.120
100%/1000CT	FixMatch-Reg	0.779±0.128	0.133±0.034	0.757±0.129	0.905±0.081	0.822±0.115
100%/1000CT	UniMatch-Reg	0.772±0.132	0.137±0.032	0.741±0.130	0.903±0.084	0.818±0.114
100%/1000CT	UniMatchV2-Reg	<u>0.797±0.100</u>	<u>0.126±0.030</u>	<u>0.777±0.099</u>	<u>0.912±0.071</u>	<u>0.831±0.105</u>
100%/1000CT	ProtoReg	**0.800±0.090**	**0.122±0.028***	**0.781±0.092**	**0.913±0.072**	**0.834±0.102**
20%/0%	Sup-only (3D U-Net)	0.765±0.120	0.134±0.032	0.750±0.121	0.901±0.071	0.797±0.104
20%/80%	MT-Reg	<u>0.776±0.109</u>	<u>0.132±0.032</u>	<u>0.758±0.108</u>	<u>0.908±0.070</u>	0.816±0.102
20%/80%	FixMatch-Reg	0.773±0.125	0.136±0.032	0.754±0.122	0.901±0.069	0.801±0.098
20%/80%	UniMatch-Reg	0.776±0.115	0.139±0.030	0.752±0.111	0.899±0.071	0.803±0.095
20%/80%	UniMatchV2-Reg	0.776±0.110	0.133±0.033	0.745±0.115	0.903±0.080	<u>0.818±0.110</u>
20%/80%	ProtoReg	**0.778±0.110**	**0.129±0.036***	**0.764±0.118***	**0.912±0.068***	**0.822±0.105**

* indicates significant improvement over UniMatchV2-Reg.

without geometric transforms (N=2 operations from noise/smoothing and intensity/histogram shifts). Models are trained for up to 500 epochs with Adam (batch size 1, lr 1×10^{-4}, wd 1×10^{-4}) and early stopping on the validation set. The supervised loss combines voxel-/LSV-level SmoothL1 and prototype update loss ($\lambda_{\text{proto}} = 0.01$). DUW-CL uses $\tau = 0.1$, $\lambda_{\text{cons}} = 1$, $\lambda_{\text{ent}} = 0.001$, and DyCON-style geometric decay with $\beta_{\max} = 5$, $\beta_{\min} = 0.5$, and $T = 10{,}000$.

All metrics are computed within the lung ROI Ω and reported as per-case mean ± standard deviation on the test set. We evaluate MAE, SSIM, Spearman's ρ, and functional-lung overlap indices FL20/FL40 [11]. For $x \in \{0.2, 0.4\}$, $R_x = \{v \in \Omega \mid y(v) \geq x \max_{u\in\Omega} y(u)\}$ and $\hat{R}_x = \{v \in \Omega \mid \hat{y}(v) \geq x \max_{u\in\Omega} \hat{y}(u)\}$; $\text{FL}x = \text{Dice}(\hat{R}_x, R_x)$.

3.2 Results and Visualizations

Comparison with Baselines Since no prior SSL method is tailored to 3D CT-to-perfusion regression, we adapt representative, transferable SSL baselines for dense prediction—Mean Teacher [22] and FixMatch [20]/UniMatch-style [25, 26] consistency learning—resulting in MT-Reg, FixMatch-Reg, UniMatch-Reg, and UniMatchV2-Reg. All methods share the same 3D U-Net backbone, splits, pre-processing, augmentations, evaluation protocol, and major hyperparameters for fair comparison. As shown in Table 1 and Table 2, ProtoReg consistently achieves best or comparable performance on both Cohort A and Cohort B under two SSL settings (fixed 80/20 train–test split): in-domain SSL (20% labeled + in-domain unlabeled CT) and external-unlabeled SSL (100% labeled + 1000 external unlabeled CT). On Cohort B with full labels, generic SSL baselines can underperform Sup-only, likely due to external-domain mismatch or noisy pseudo-targets; ProtoReg remains more robust. Representative visualizations are shown in Fig. 2.

Ablation Studies We conduct ablations on Cohort A under the 20% labeled setting (Table 3). A moderate prototype granularity performs best. Replacing our uncertainty re-weighting with fixed-threshold filtering based on the teacher's prototype-assignment probability (max prototype similarity probability > 0.90 /0.95) consistently underperforms, suggesting hard filtering is suboptimal for

Table 2. Comparison on Cohort B.

Setting (L/U)	Method	Spearman $\rho \uparrow$	MAE $\downarrow$	SSIM $\uparrow$	FL20 $\uparrow$	FL40 $\uparrow$
100%/0%	Sup-only (3D U-Net)	0.810±0.091	**0.119±0.026**	0.792±0.091	0.872±0.053	0.796±0.066
100%/1000CT	MT-Reg	0.802±0.087	0.122±0.025	0.786±0.089	0.868±0.053	0.790±0.068
100%/1000CT	FixMatch-Reg	0.797±0.078	0.126±0.024	0.767±0.081	0.865±0.056	0.787±0.069
100%/1000CT	UniMatch-Reg	0.801±0.089	0.126±0.026	0.774±0.088	0.864±0.055	0.773±0.071
100%/1000CT	UniMatchV2-Reg	0.806±0.082	0.123±0.025	0.780±0.091	0.871±0.051	0.794±0.065
100%/1000CT	ProtoReg	**0.814±0.077***	0.119±0.026*	**0.799±0.085***	**0.873±0.050**	**0.798±0.065**
20%/0%	Sup-only (3D U-Net)	0.740±0.085	0.147±0.023	0.702±0.091	0.846±0.050	0.735±0.064
20%/80%	MT-Reg	0.759±0.083	0.146±0.029	0.723±0.098	0.857±0.053	0.757±0.064
20%/80%	FixMatch-Reg	0.767±0.075	0.139±0.026	0.730±0.087	**0.862±0.053**	0.764±0.065
20%/80%	UniMatch-Reg	0.749±0.091	0.145±0.032	0.718±0.114	0.854±0.050	0.758±0.077
20%/80%	UniMatchV2-Reg	0.758±0.087	0.145±0.029	0.722±0.105	0.856±0.049	0.753±0.072
20%/80%	ProtoReg	**0.769±0.102***	**0.138±0.027***	**0.739±0.109***	0.861±0.048	**0.767±0.066***

* indicates significant improvement over UniMatchV2-Reg.

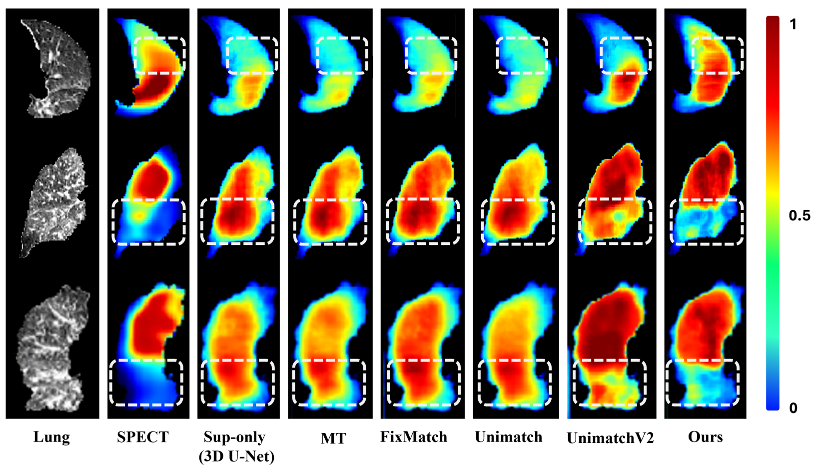

Fig. 2. Qualitative comparisons on CT-to-perfusion SPECT regression.

Table 3. Ablation studies on Cohort A under the 20% labeled setting.

Variant	Spearman $\rho \uparrow$	MAE $\downarrow$	SSIM $\uparrow$	FL20 $\uparrow$	FL40 $\uparrow$
(A) Number of prototype bins K					
$K = 2$	0.773±0.110	0.132±0.037	0.757±0.112	0.910±0.072	0.821±0.107
$K = 5$	**0.778±0.110**	**0.129±0.036**	**0.764±0.118**	**0.912±0.068**	**0.822±0.105**
$K = 10$	0.766±0.105	0.133±0.034	0.753±0.107	0.908±0.068	0.819±0.102
(B) DUW-CL design choices					
th_consist=0.90	0.767±0.113	0.135±0.031	0.747±0.108	0.902±0.069	0.816±0.099
th_consist=0.95	0.773±0.119	0.133±0.033	0.751±0.117	0.904±0.071	0.820±0.110
only voxel (w/o LSV)	0.774±0.105	0.132±0.037	0.758±0.112	0.909±0.072	**0.823±0.106**
ours (voxel+LSV)	**0.778±0.110**	**0.129±0.036***	**0.764±0.118***	**0.912±0.068***	0.822±0.105

* indicates significant improvement over voxel-only (w/o LSV).

dense regression with continuous pseudo targets. Finally, removing the LSV-level weighting ("only voxel") degrades MAE, SSIM, FL20, validating the benefit of regional uncertainty in the full voxel+LSV design.

4 Conclusion

In this paper, we propose ProtoReg, a teacher–student semi-supervised framework for 3D non-contrast CT-to-perfusion regression with limited paired data. It leverages a learnable functional prototype bank to compute prototype-induced entropy for voxel-wise uncertainty, and proposes DUW-CL, a Dual-level Uncertainty-weighted Consistency loss that suppresses noisy pseudo-labels and promotes anatomically coherent patterns. Experiments on two independent cohorts show consistent gains in correlation, reconstruction quality, and high-functional indicators, and remain effective when incorporating 1,000 external unlabeled CTs. Future work will extend to multi-center generalization, additional functional targets, and more robust sub-unit modeling and uncertainty calibration.

Acknowledgments. This research was partly supported by the Shenzhen Medical Research Fund (A2503079), Health and Medical Research Fund (11222456, 10240147) of the Health Bureau, Guangdong Basic and Applied Basic Research Foundation (2025A15 15012926), and Shenzhen Science and Technology Program (JCYJ20230807140403007).

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

References

1. Amini, E., Klein, R.: Lung lobe segmentation: performance of open-source moose, totalsegmentator, and lungmask models compared to a local in-house model. European Radiology Experimental **9**(1), 86 (2025)
2. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International joint conference on neural networks (IJCNN). pp. 1–8. IEEE (2020)
3. Armato, S.G.I., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., et al.: Data from lidc-idri (2015). https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX, accessed via TCIA
4. Assefa, M., Naseer, M., Ganapathi, I.I., Ali, S.S., Seghier, M.L., Werghi, N.: Dycon: Dynamic uncertainty-aware consistency and contrastive learning for semi-supervised medical image segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30850–30860 (2025)
5. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in neural information processing systems **33**, 9912–9924 (2020)
6. Chen, Z., Huang, Y.H., Kong, F.M., Ho, W.Y., Ren, G., Cai, J.: A super-voxel-based method for generating surrogate lung ventilation images from ct. Frontiers in physiology **14**, 1085158 (2023)
7. Chen, Z., Li, Z., Huang, Y.H., Xiong, T., Teng, X., Li, B., Kipritidis, J., Keall, P.J., Reinhardt, J.M., Ge, H., et al.: A review of advances in lung function imaging and its applications for functional lung avoidance in radiation therapy. In: Seminars in nuclear medicine. Elsevier (2025)

8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
9. Dai, W., Li, X., Cheng, K.T.: Semi-supervised deep regression with uncertainty consistency and variational model ensembling via bayesian neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 7304–7313 (2023)
10. Dupuis, J., Harel, F., Nguyen, Q.T.: Molecular imaging of the pulmonary circulation in health and disease. Clinical and Translational Imaging **2**(5), 415–426 (2014)
11. Farr, K.P., West, K., Yeghiaian-Alvandi, R., Farlow, D., Stensmyr, R., Chicco, A., Hau, E.: Functional perfusion image guided radiation treatment planning for locally advanced lung cancer. Physics and Imaging in Radiation Oncology **11**, 76–81 (2019)
12. Irving, B.: maskslic: regional superpixel generation with application to local pathology characterisation in medical images. arXiv preprint arXiv:1606.09518 (2016)
13. Liu, Y.K., Cisneros, J., Nair, G., Stevens, C., Castillo, R., Vinogradskiy, Y., Castillo, E.: Perfusion estimation from dynamic non-contrast computed tomography using self-supervised learning and a physics-inspired u-net transformer architecture. International Journal of Computer Assisted Radiology and Surgery **20**(5), 959–970 (2025)
14. Lu, W., Lei, J., Qiu, P., Sheng, R., Zhou, J., Lu, X., Yang, Y.: Upcol: Uncertainty-informed prototype consistency learning for semi-supervised medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 662–672. Springer (2023)
15. Pu, Y., Zhang, Q., Qian, C., Zeng, Q., Li, N., Zhang, L., Zhou, S., Zhao, G.: Semi-supervised segmentation of coronary dsa using mixed networks and multi-strategies. Computers in Biology and Medicine **156**, 106493 (2023)
16. Ren, G., Li, B., Lam, S.k., Xiao, H., Huang, Y.H., Cheung, A.L.y., Lu, Y., Mao, R., Ge, H., Kong, F.M., et al.: A transfer learning framework for deep learning-based ct-to-perfusion mapping on lung cancer patients. Frontiers in oncology **12**, 883516 (2022)
17. Ren, G., Zhang, J., Li, T., Xiao, H., Cheung, L.Y., Ho, W.Y., Qin, J., Cai, J.: Deep learning-based computed tomography perfusion mapping (dl-ctpm) for pulmonary ct-to-perfusion translation. International Journal of Radiation Oncology* Biology* Physics **110**(5), 1508–1518 (2021)
18. Shakouri, S., Bakhshali, M.A., Layegh, P., Kiani, B., Masoumi, F., Ataei Nakhaei, S., Mostafavi, S.M.: Covid19-ct-dataset: an open-access chest ct image repository of 1000+ patients with confirmed covid-19 diagnosis. BMC research notes **14**(1), 178 (2021)
19. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
20. 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)
21. Sun, X., Wang, R., Wang, Q., Wu, Y., Jia, X., Meng, D.: Semi-supervised regression with heteroscedastic pseudo-labels. arXiv preprint arXiv:2510.15266 (2025)

22. 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)
23. Xu, Z., Wang, Y., Lu, D., Yu, L., Yan, J., Luo, J., Ma, K., Zheng, Y., Tong, R.K.y.: All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation. IEEE Journal of Biomedical and Health Informatics **26**(7), 3174–3184 (2022)
24. Xue, P., Zhang, J., Ma, L., Li, Y., Ji, H., Ren, T., Hu, Z., Ren, M., Zhang, Z., Dong, E.: Semi-supervised temporal attention network for lung 4d ct ventilation estimation. Computerized Medical Imaging and Graphics **123**, 102551 (2025)
25. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7236–7246 (2023)
26. Yang, L., Zhao, Z., Zhao, H.: Unimatch v2: Pushing the limit of semi-supervised semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 3031–3048 (2025)
27. Zhang, J., Wu, T., Ding, C., Zhao, H., Guo, G.: Region-level contrastive and consistency learning for semi-supervised semantic segmentation. In: Raedt, L.D. (ed.) Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22. pp. 1622–1628. International Joint Conferences on Artificial Intelligence Organization (7 2022). https://doi.org/10.24963/ijcai.2022/226, main Track
28. Zhang, Z., Yin, G., Zhang, B., Liu, W., Zhou, X., Wang, W.: A semantic knowledge complementarity based decoupling framework for semi-supervised class-imbalanced medical image segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 25940–25949 (2025)
29. Zhang, Z., Zheng, F., Iltaf, A., Han, Y., Cheng, Z., Du, Y., Li, B., Liu, T., Zhou, S.: Adaptive morph-patch transformer for aortic vessel segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 13015–13024 (2026)