Progressive Knowledge Distillation for Automatic Perfusion Parameter Maps Generation from Low Temporal Resolution CT Perfusion Images

Moo Hyun Son[1], Juyoung Bae[1], Elizabeth Tong[4], and Hao Chen[1,2,3(✉)]

[1] Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong
{mhson, jbaeaa, jhc}@cse.ust.hk
[2] Department of Chemical and Biological Engineering,
The Hong Kong University of Science and Technology, Hong Kong
[3] Division of Life Science,
The Hong Kong University of Science and Technology, Hong Kong
[4] Department of Pediatric Radiology, Stanford University School of Medicine,
Stanford, CA, USA
etong@stanford.edu

Abstract. Perfusion Parameter Maps (PPMs), generated from Computer Tomography Perfusion (CTP) scans, deliver detailed measurements of cerebral blood flow and volume, crucial for the early identification and strategic treatment of cerebrovascular diseases. However, the acquisition of PPMs involves significant challenges. Firstly, the accuracy of these maps heavily relies on the manual selection of Arterial Input Function (AIF) information. Secondly, patients are subjected to considerable radiation exposure during the scanning process. In response, previous researches have attempted to automate AIF selection and reduce radiation exposure of CTP by lowering temporal resolution, utilizing deep learning to predict PPMs from automated AIF selection and temporal resolutions as low as $\frac{1}{3}$. However, the effectiveness of these approaches remains marginally significant. In this paper, we push the limits and propose a novel framework, Progressive Knowledge Distillation (PKD), to generate accurate PPMs from $\frac{1}{16}$ standard temporal resolution CTP scans. PKD uses a series of teacher networks, each trained on different temporal resolutions, for knowledge distillation. Initially, the student network learns from a teacher with low temporal resolution; as the student is trained, the teacher is scaled to a higher temporal resolution. This progressive approach aims to reduce the large initial knowledge gap between the teacher and the student. Experimental results demonstrate that PKD can generate PPMs comparable to full-resolution ground truth, outperforming current deep learning frameworks. Our code is available at https://github.com/mhson-kyle/progressive-kd.

Keywords: Knowledge Distillation · Radiation Reduction · CT Perfusion · Multi-task learning.

1 Introduction

Computed Tomography Perfusion (CTP) imaging has established itself as a highly efficient and essential protocol in diagnosing cerebral pathologies, including stroke, tumors, and vascular disorders. CTP imaging entails the acquisition of sequential 3D CT scans post-contrast agent injection, yielding 4D images that visualize the contrast agent's temporal distribution. However, the high sensitivity to noise and artifacts limits the direct use of CTP scans [10]. Clinicians therefore derive PPMs from CTP data to address these issues [11]. This involves deconvolving the time-concentration curve (TCC) of each voxel against the arterial input function (AIF), representing the TCC of the main feeding arteries. From CTP analyses, PPMs, including Time-to-Maximum (Tmax), Cerebral Blood Volume (CBV), and Cerebral Blood Flow (CBF), are produced. PPMs are crucial for radiologists in making informed treatment decisions. PPMs not only provide these crucial hemodynamic measurements, but their derivation from CTP data also aims to overcome some of the clinical limitations associated with traditional CTP analysis. However, the current process of deriving PPMs is full of challenges: dependency on manual selection of AIF and high radiation exposure during the scanning stage [2, 22, 23]. These issues are particularly detrimental in clinical diagnostics and treatment, underscoring the urgent need for a more efficient and safer approach.

Existing approaches to automate AIF selection and reduce radiation exposure of CTP scans face significant challenges that require a two-step process—predicting AIF and generating PPMs. From this two-step method, utilizing the predicted AIF information to predict the PPMs inherently introduces errors [3, 13, 16, 21, 28]. However, little work has been explored to fully automate the PPM generation without the need for AIF information [7, 25]. A straightforward solution to reduce the radiation exposure of CTP scans is lowering the temporal resolution. Although low temporal resolution CTP scan can significantly reduce radiation exposure, it fails to accurately capture the dynamics of cerebral blood flow and generate reliable PPMs. Thus, existing works have attempted to reconstruct high temporal resolution scans from low temporal resolution and generate PPMs from the reconstructed frames. Utilizing the generated CTP scans to generate PPMs inadvertently introduces defects and degrades the performance [15, 26]. Despite limited performance, these works offer a starting point for solutions to manual AIF selection and radiation concerns. In response, our research proposes a deep learning model to directly derive PPMs from low temporal resolution CTP scans without AIF information specifically leveraging multi-task learning and knowledge distillation.

Recent advancements in knowledge distillation have highlighted the issue of the knowledge gap between large teacher models and small student models, which can lead to ineffective learning [4, 8]. Thus, the largest teacher is not the best teacher for knowledge distillation [24]. Some approaches have been explored to address this gap: employing smaller teacher assistant models [19], multiple teacher models [27], and a combination of teacher and teacher assistant models [24]. However, existing works heavily rely on the same teacher model throughout

training, which fails to account for the student's increasing knowledge. Minimal research has explored adjusting the teacher's complexity with student's training [20, 29]. The significance is that as a student's knowledge expands with ongoing training, the teacher's knowledge must increase correspondingly. However, no relevant research attempted to leverage multiple teachers of various knowledge, each corresponding to a training stage of the student.

Here, we propose a novel framework called Progressive Knowledge Distillation (PKD) to address the large knowledge gap between high and low temporal resolution models. This framework is distinctive in its progressive adaptation of the teacher model during the training process. The teachers are systematically and progressively changed during training corresponding to the student's training stage to maintain a small knowledge gap between the teacher and the student. Remarkably, our model is capable of generating PPMs from both AIF-Free and low temporal resolution CTP scans within a single framework

In this paper, our contributions are as follows. 1) We achieve state-of-the-art performance in automatic PPM generation without AIF selection by using a multi-task learning framework. 2) We propose a novel framework, Progressive Knowledge Distillation (PKD), which generates accurate PPMs from CTP scans while reducing radiation exposure to less than 7% of conventional levels. This reduction is achieved by effectively bridging the large knowledge gap in the student model using a progressive teacher approach. 3) Extensive experimental results confirm that our Progressive Knowledge Distillation outperforms competing Knowledge Distillation methods.

2 Methodology

2.1 Multi-Task Learning

Our approach utilizes a single 4D sequence of raw CTP scans, comprising T time points and S axial slices. We create S input volumes, each combining T slices of identical axial height, represented as $x \in \mathbb{R}^{H \times W \times T}$. Input x passes through a shared encoder E, yielding an encoded bottleneck feature f. Three task-specific decoders then process this feature to generate the PPMs, $y_{\text{map}} = \{y_{\text{Tmax}}, y_{\text{CBV}}, y_{\text{CBF}}\} \subset \mathbb{R}^{H \times W}$. By utilizing a shared encoder, we enhance the correlation between the PPMs and foster a generalized and refined feature set for synergistic task learning. PPMs are inherently intercorrelated since all PPMs derive from the same underlying cerebral blood flow dynamics [6]. As a result, utilizing a single encoder is well justified due to the same underlying principle of the PPMs. We incorporate 3D V-Net [18] as our backbone to leverage its robust capabilities in analyzing volumetric data and take advantage of the spatio-temporal information of CTP scans. Optimization of the generated PPMs is conducted using a combined loss function of mean absolute error (MAE), structural similarity index measure (SSIM), and perceptual loss [9] for specific purposes: MAE ensures pixel-level accuracy for numerical integrity, SSIM maintains image structural and textural fidelity, and perceptual loss preserves essential features and style for overall accuracy. Let $\mathcal{L}_{\text{GT}}$ denote the total loss and $\mathcal{L}_{\text{map}}$ denote

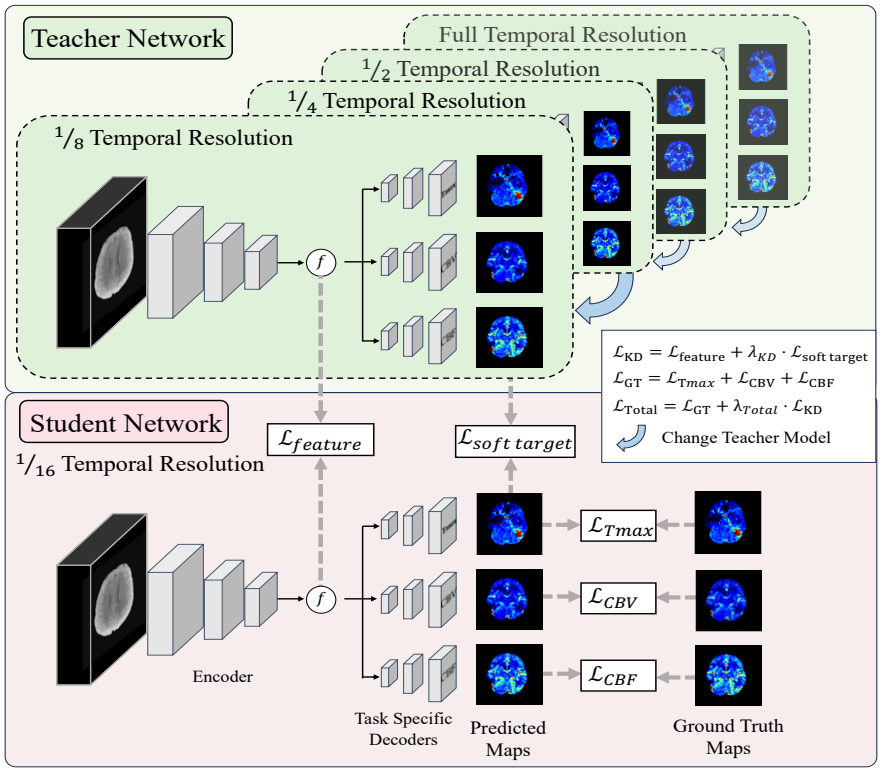

Fig. 1. Progressive Knowledge Distillation framework consists of two main parts: the teacher networks and the student network of the same 3D V-Net [18] architecture (encoder and 3 decoders). In addition to loss computation between the generated maps and the ground truth, the framework aligns the bottleneck feature (f) and output of the student network with the teacher's bottleneck feature and soft target output. See the appendix for the detailed architecture of the components.

the loss associated with PPMs, with map $\in \{\text{Tmax}, \text{CBV}, \text{CBF}\}$. Consequently, the loss equation can be formulated as follows

$$\mathcal{L}_{\text{map}} = \|\hat{y}_{\text{map}} - y_{\text{map}}\| + 1 - \text{SSIM}(\hat{y}_{\text{map}}, y_{\text{map}}) + \text{Perceptual}(\hat{y}_{\text{map}}, y_{\text{map}}) \quad (1)$$

$$\mathcal{L}_{\text{GT}} = \mathcal{L}_{\text{Tmax}} + \mathcal{L}_{\text{CBV}} + \mathcal{L}_{\text{CBF}}. \quad (2)$$

2.2 Progressive Knowledge Distillation

We introduce a simple yet effective approach-the Progressive Knowledge Distillation framework that leverages multiple teacher models of different knowledge dynamically to guide the student model. The student model initially distills knowledge from a teacher model that has a small knowledge gap with the

student. As training continues, the student model's knowledge increases and correspondingly needs a larger teacher. Here, the teacher model is substituted with a teacher model of a higher level of knowledge, trained on higher temporal resolution. Thus, the student model gradually distills the knowledge of full resolution and addresses the large knowledge gap. The core of this methodology is the implementation of a progressive scaling of the teacher model. Initially, the student model is trained using $\frac{1}{16}$ temporal resolution CTP scan and distills knowledge from the teacher model trained on $\frac{1}{8}$ temporal resolution. To maintain progressive and gradual distillation, the teacher model's knowledge is incrementally increased in stages ($\frac{1}{4}$, $\frac{1}{2}$, and finally full temporal resolution) as training progresses and when specific criteria are satisfied, such as no reduction in distillation loss for 5 consecutive epochs. By gradually enhancing the complexity of the knowledge from a low temporal resolution to a full temporal resolution, PKD smooths the learning curve and simplifies optimization. This structured progression prevents overwhelming the student model with a large knowledge gap. The knowledge is precisely aligned between the bottleneck feature (f_t) and the soft-targeted output ($\hat{y}_t$) of the teacher model and the bottleneck feature (f_s) and the output ($\hat{y}_s$) of the student model. The importance of guiding the bottleneck feature arises from training the encoder to represent the same latent space vector for low and high temporal resolution CTP scans. The overall loss is defined as

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{GT}} + \lambda_{\text{Total}} \cdot \mathcal{L}_{\text{KD}} \tag{3}$$

$$\text{where } \mathcal{L}_{KD} = \|f_s - f_t\| + \lambda_{\text{KD}} \cdot \mathcal{L}_{map} \tag{4}$$

where λ is used to determine the strength of the distillation loss.

Figure 1 provides a high-level schematic representation of this progressive learning approach, illustrating the transitional role of the teacher models in bridging the knowledge gap between low and high temporal resolution CTP scans. Please refer to the supplementary material for the pseudocode representation.

3 Experiments

3.1 Datasets and Implementation Details

We utilize a composite of public and private datasets featuring 3D Brain CTP scans and corresponding PPMs (Tmax, CBF, CBV). The public dataset, from the ISLES 2018 challenge [12, 17], includes 156 patient-level data (677 H $\times$ W $\times$ T Brain CTP scans and PPMs). Our private in-house dataset consists of 468 patient-level data (3593 H $\times$ W $\times$ T Brain CTP scans and PPMs). We employ two distinct hold-out experimental settings to validate our models' effectiveness. Our first setting utilizes a subset of our in-house dataset, manually chosen by clinicians for performance validation, as our testing set. The second setting utilizes the testing set from the public ISLES 2018 challenge dataset as a testing benchmark. We present the details in Table 1. Experimenting with datasets from different centers with varying acquisition protocols ensure that our results

Table 1. Details of experiment settings. ✓ indicates training set and ✗ indicates hold-out testing set.

Experimental	In-House		ISLES2018		Total	
Settings	Train	Test	Train	Test	Train	Test
#1	✓	✗	✓	✓	573 Cases (3860 H × W × T CT Scans)	50 Cases (410 H × W × T CT Scans)
#2	✓	✓	✓	✗	560 Cases (4025 H × W × T CT Scans)	63 Cases (245 H × W × T CT Scans)

are reliable and robust across different situations, fostering openness and reproducibility. The CTP scans are preprocessed: normalized and resized to match 256 × 256 spatial resolution and 40 frames temporal resolution. From the processed temporal CTP volume of size 256 × 256 × 40, we randomly extract a volumetric patch of size 128 × 128 × 32 for input. All PPMs used in our experiments are produced using commercially available software (RAPID, IschemaView, Melnao Park, CA) [14] and manually cross-checked with clinicians.

We simulate $\frac{1}{n}$ low temporal resolution by selecting every n^{th} frame of T frame CTP scan, resulting in $\lfloor \frac{T}{n} \rfloor$ number of frames. We further augment the data, which can be readily applied in clinical settings, by reconstructing the reduced temporal resolution data to its original frame count by duplicating frames according to their reduced temporal rate ($\frac{1}{n}$). By filling the gaps between frames, it bridges the knowledge gap between different temporal resolutions.

All experiments are conducted on a single NVIDIA RTX 3090 GPU for 80,000 iterations using Adam optimizer with the momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.99$. We initialize the learning rate to 0.001 and gradually decreased linearly. The evaluation metrics employed in this analysis include root mean squared error (RMSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR) on images scaled to [0, 24] seconds for Time-to-Maximum (Tmax), [0, 200] ml/100g for Cerebral Blood Volume (CBV), and [0, 1000] ml/100g/min for Cerebral Blood Flow (CBF).

3.2 Experimental Results

Perfusion Parameter Map Estimation We compare our methods with AIF-Free models, particularly UNET-GAN-Assisted [1] (hybrid of U-Net architectural frameworks and adversarial training method) and UniToBrain [7] (modification of U-Net architectural method). Tables 2 and 3 each present the testing results on settings #1 and #2. As shown in Table 2, our model outperforms all competing methods in both full and $\frac{1}{16}$ temporal resolution. We confirm the significance of this result by conducting a one-tailed paired t-test at a significance level $\alpha < 0.05$. We further investigate the effectiveness of PKD by comparing the results of $\frac{1}{8}$ temporal resolution. Similarly, we confirm that our model under the PKD setting can accurately generate perfusion parameter maps indistinguishable from the ground truth maps. PKD even outperforms the teacher model in Tmax map generation. Building on the concept from [5] that student models can occasionally surpass teacher models by distilling distinct levels of knowledge, we give credit to our PKD approach for learning from critical but neglected information from various temporal resolution teacher models. The proposed Progressive

Table 2. Results of perfusion parameter map generation setting #1. **Bolded** values indicate the best performance in each temporal resolution. Underlined values indicate the best performance across all rows. RMSE values are given in their original units (see section 3.1)

	Temporal Resolution	Tmax			CBV			CBF		
		RMSE ↓	SSIM ↑	PSNR ↑	RMSE ↓	SSIM ↑	PSNR ↑	RMSE ↓	SSIM ↑	PSNR ↑
Ours	Full	**1.201**	**0.913**	**27.965**	**8.892**	**0.918**	**28.167**	**48.163**	**0.932**	**27.693**
UNET-GAN-Assisted [1]	Full	1.355	0.903	26.646	9.144	0.917	27.893	49.680	0.929	27.413
UniToBrain [7]	Full	1.314	0.900	26.834	9.117	0.916	27.798	50.196	0.926	27.051
Ours w/ PKD	1/8	**1.033**	**0.923**	**29.709**	**9.000**	**0.915**	**28.015**	**48.207**	**0.928**	**27.729**
Ours w/o PKD	1/8	1.516	0.869	25.023	9.536	0.912	27.467	52.536	0.921	26.909
UNET-GAN-Assisted [1]	1/8	1.769	0.852	23.662	9.662	0.908	27.252	58.508	0.909	25.676
UniToBrain [7]	1/8	1.460	0.885	25.728	9.778	0.895	27.040	51.565	0.922	26.829
Ours w/ PKD	1/16	**1.436**	**0.879**	**25.790**	**9.004**	**0.916**	**28.061**	**49.639**	**0.926**	**27.430**
Ours w/o PKD	1/16	1.516	0.869	25.023	9.536	0.912	27.467	52.536	0.921	26.909
Ours w/ DGKD [24]	1/16	1.721	0.859	24.279	10.661	0.893	26.806	62.742	0.900	25.593
Ours w/ KD [8]	1/16	1.511	0.870	25.463	9.594	0.910	27.506	52.298	0.922	26.820
Ours w/ MTKD [27]	1/16	1.464	0.875	25.661	9.518	0.910	27.403	51.502	0.922	26.897
Ours w/ TAKD [19]	1/16	1.492	0.871	25.356	9.273	0.913	27.842	52.217	0.921	27.095
UNET-GAN-Assisted [1]	1/16	2.171	0.842	21.782	10.049	0.898	26.889	57.849	0.907	25.691
UniToBrain [7]	1/16	1.726	0.838	24.044	10.321	0.893	26.483	59.901	0.884	25.310

Table 3. Results of perfusion parameter map generation setting #2. **Bolded** values indicate the best performance in each temporal resolution. RMSE values are given in their original units (see section 3.1)

	Temporal Resolution	Tmax			CBV			CBF		
		RMSE ↓	SSIM ↑	PSNR ↑	RMSE ↓	SSIM ↑	PSNR ↑	RMSE ↓	SSIM ↑	PSNR ↑
Ours	Full	**1.438**	**0.896**	**26.189**	**10.290**	**0.923**	**27.129**	**67.864**	**0.926**	**25.042**
UNET-GAN-Assisted [1]	Full	1.974	0.831	22.605	10.971	0.923	26.937	73.437	0.911	24.313
UniToBrain [7]	Full	1.673	0.885	24.807	11.733	0.907	26.091	77.697	0.909	23.741
Ours w/ PKD	1/16	1.666	**0.865**	**24.184**	**6.356**	**0.950**	**31.030**	**43.881**	**0.943**	**28.145**
Ours w/o PKD	1/16	2.233	0.838	21.918	13.053	0.893	25.445	102.771	0.864	21.585
Ours w/ DGKD [24]	1/16	1.962	0.846	22.927	13.409	0.890	25.222	106.195	0.859	21.162
Ours w/ KD [8]	1/16	**1.632**	0.863	24.062	11.307	0.912	26.375	74.510	0.906	24.018
Ours w/ MTKD [27]	1/16	1.716	0.863	24.152	6.958	0.944	30.220	45.319	0.940	27.824
Ours w/ TAKD [19]	1/16	1.987	0.838	22.845	11.431	0.916	26.906	81.931	0.899	23.470

Knowledge Distillation (PKD) method has significantly improved performance across all metric values, underscoring its critical role in bridging the gap between full and low temporal resolution datasets. This advancement stems from progressively changing the teacher model in parallel with the student's learning progress which avoids overwhelming the student model with a significant knowledge gap. In Table 3, we further present our model's performance on experiment setting #2. Here, the ISLES 2018 testing dataset is not specifically designed for PPM evaluation, and its limited number of 3D CTP data could have lead to relatively unstable results. Nonetheless, comparing its performance with the UNET-GAN-Assisted [1] model and UniToBrain [7] model, our model shows the best performance.

In Figure 2, we present a qualitative analysis comparing our model with alternative PPM estimation methods. Both our models trained on full temporal resolution and $\frac{1}{16}$ temporal resolution with PKD shows close agreement with the

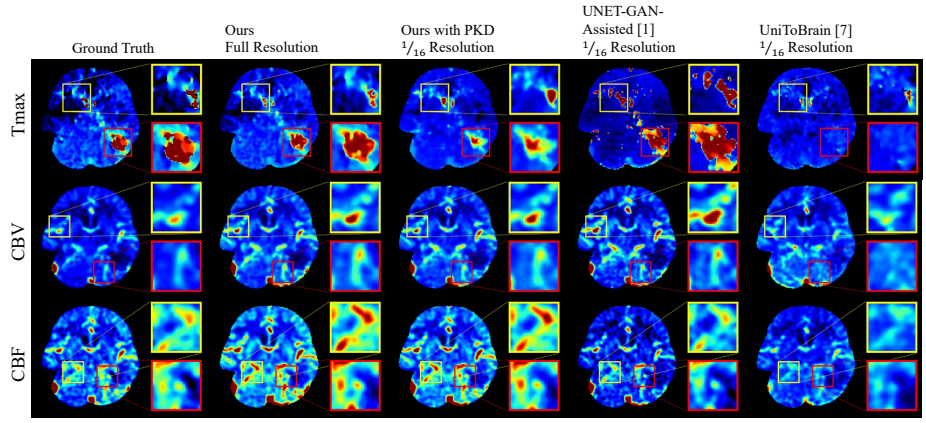

Fig. 2. Visual comparison of the generated perfusion parameter maps. Red and yellow boxed areas represent lesions with the first and second largest disagreement between maps.

ground truth across all PPMs. In contrast, the UNET-GAN-Assisted [1] model tends to over-highlight certain areas, including those of lower intensity (yellow boxed area in Figure 2 for all Tmax and CBV). UniToBrain [7] model captures a more limited scope of detail suppressing high-intensity areas (red boxed area in Figure 2 for all Tmax, CBV, and CBF). These results support our model's superior applicability in clinical settings.

Knowledge Distillation We further explore the effectiveness of our Progressive Knowledge Distillation by comparing with existing knowledge distillation methods that leverage different sizes of teacher models: vanilla Knowledge Distillation [8], TAKD [19], MTKD [27], and DGKD [24]. In our Progressive Knowledge Distillation, we progressively guide the student model, trained at $\frac{1}{16}$ temporal resolution, using teacher models trained at increasing resolutions: $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, and full. In the TAKD method, we select a single teacher assistant (TA) model trained at $\frac{1}{4}$ temporal resolution. For MTKD, we employ three different teacher models, each trained at full temporal resolution. Lastly, in the DGKD framework, we simultaneously utilize teacher models at all temporal resolutions: $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, and full. As reported in Tables 2 and 3, Progressive Knowledge Distillation outperforms all knowledge distillation methods compared. These experimental results confirm that PKD effectively reduces the knowledge gap and distills knowledge of the full temporal resolution CTP scan. While other methods use the same teacher model throughout the training, our PKD scales the teaching models during the training, thereby progressively enriching the knowledge.

4 Discussion and Conclusion

In this work, we introduce a novel framework, Progressive Knowledge Distillation, for the automatic generation of perfusion parameter maps from low-resolution CTP scans, aimed to reduce the radiation exposure. Extensive experiments confirm that our framework can generate accurate perfusion parameter maps from $\frac{1}{16}$ temporal resolution CTP scans. In contrast to existing approaches, our method scales the teacher model progressively during training to bridge the knowledge gap of the student model. We further confirm PKD's effectiveness by comparing it with existing knowledge distillation methods. While the proposed framework demonstrates promising results, future research could investigate its generalization to different modalities and its robustness to noise and artifacts in the input data. We acknowledge limitations such as extended training requirements and resource intensity due to the use of teacher models. Despite these challenges, our approach maintains rapid inference times of a few seconds, demonstrating significant clinical potential.

Acknowledgments. This work was supported by the Hong Kong Innovation and Technology Fund (Project No. MHP/002/22) and HKUST (Project No. FS111).

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

References

1. Asaduddin, M., Roh, H.G., Kim, H.J., Kim, E.Y., Park, S.: Perfusion Maps Acquired From Dynamic Angiography MRI Using Deep Learning Approaches. Journal of Magnetic Resonance Imaging **57**(2), 456–469 (2023)
2. Brenner, D.J., Hall, E.J.: Computed tomography–an increasing source of radiation exposure. The New England Journal of Medicine **357**(22), 2277–2284 (2007)
3. Chen, X., Tran, A.P., Elkin, R., Benveniste, H., Tannenbaum, A.R.: Visualizing Fluid Flows via Regularized Optimal Mass Transport with Applications to Neuroscience. Journal of Scientific Computing **97**(2), 26 (2023)
4. Cho, J.H., Hariharan, B.: On the Efficacy of Knowledge Distillation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 4793–4801 (2019)
5. Deng, X., Zheng, J., Zhang, Z.: Personalized Education: Blind Knowledge Distillation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022, vol. 13694, pp. 269–285. Springer Nature Switzerland, Cham (2022)
6. Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R.: Deconvolution-Based CT and MR Brain Perfusion Measurement: Theoretical Model Revisited and Practical Implementation Details. International Journal of Biomedical Imaging **2011**, 467563 (2011)
7. Gava, U.A., D'Agata, F., Tartaglione, E., Renzulli, R., Grangetto, M., Bertolino, F., Santonocito, A., Bennink, E., Vaudano, G., Boghi, A., Bergui, M.: Neural network-derived perfusion maps: A model-free approach to computed tomography perfusion in patients with acute ischemic stroke. Frontiers in Neuroinformatics **17** (2023)

8. Hinton, G., Vinyals, O., Dean, J.: Distilling the Knowledge in a Neural Network (2015)
9. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 694–711. Lecture Notes in Computer Science, Springer International Publishing, Cham (2016)
10. Kasasbeh, A.S., Christensen, S., Parsons, M.W., Campbell, B., Albers, G.W., Lansberg, M.G.: Artificial Neural Network Computer Tomography Perfusion Prediction of Ischemic Core. Stroke **50**(6), 1578–1581 (2019)
11. Kheradmand, A., Fisher, M., Paydarfar, D.: Ischemic stroke in evolution: predictive value of perfusion computed tomography. Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association **23**(5), 836–843 (2014)
12. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The Virtual Skeleton Database: An Open Access Repository for Biomedical Research and Collaboration. Journal of Medical Internet Research **15**(11), e245 (2013)
13. Koundal, S., Elkin, R., Nadeem, S., Xue, Y., Constantinou, S., Sanggaard, S., Liu, X., Monte, B., Xu, F., Van Nostrand, W., Nedergaard, M., Lee, H., Wardlaw, J., Benveniste, H., Tannenbaum, A.: Optimal Mass Transport with Lagrangian Workflow Reveals Advective and Diffusion Driven Solute Transport in the Glymphatic System. Scientific Reports **10**(1), 1990 (2020)
14. Laughlin, B., Chan, A., Tai, W., Moftakhar, P.: RAPID Automated CT Perfusion in Clinical Practice. In: Pract Neurol. vol. 2019, pp. 41–55 (2019)
15. Lin, C., Wu, T., Lin, C., Hung, S., Chiu, C., Liu, M.J., Teng, M., Chang, F., Guo, W., Chang, C.: Can Iterative Reconstruction Improve Imaging Quality for Lower Radiation CT Perfusion? Initial Experience. American Journal of Neuroradiology **34**(8), 1516–1521 (2013)
16. Liu, P., Lee, Y.Z., Aylward, S.R., Niethammer, M.: Perfusion Imaging: An Advection Diffusion Approach. IEEE Transactions on Medical Imaging **40**(12), 3424–3435 (2021)
17. Maier, O., Menze, B.H., von der Gablentz, J., Hani, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Götz, M., Haeck, T., Halme, H.L., Havaei, M., Reyes, M.: ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis **35**, 250–269 (2017)
18. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571 (2016)
19. Mirzadeh, S.I., Farajtabar, M., Li, A., Levine, N., Matsukawa, A., Ghasemzadeh, H.: Improved Knowledge Distillation via Teacher Assistant. Proceedings of the AAAI Conference on Artificial Intelligence **34**(04), 5191–5198 (2020)
20. Qian, B., Wang, Y., Yin, H., Hong, R., Wang, M.: Switchable Online Knowledge Distillation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 449–466. Lecture Notes in Computer Science, Springer Nature Switzerland, Cham (2022)
21. de la Rosa, E., Sima, D.M., Menze, B., Kirschke, J.S., Robben, D.: AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning. Medical Image Analysis **74**, 102211 (2021)
22. Shan, F., Xing, W., Qiu, J., Zhang, Z., Yang, S.: First-pass CT perfusion in small peripheral lung cancers: effect of the temporal interval between scan acquisitions on the radiation dose and quantitative vascular parameters. Academic Radiology **20**(8), 972–979 (2013)

23. Smith, A.B., Dillon, W.P., Gould, R., Wintermark, M.: Radiation Dose-Reduction Strategies for Neuroradiology CT Protocols. American Journal of Neuroradiology **28**(9), 1628–1632 (2007)
24. Son, W., Na, J., Choi, J., Hwang, W.: Densely Guided Knowledge Distillation using Multiple Teacher Assistants. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9375–9384. IEEE, Montreal, QC, Canada (2021)
25. Winder, A., d'Esterre, C.D., Menon, B.K., Fiehler, J., Forkert, N.D.: Automatic arterial input function selection in CT and MR perfusion datasets using deep convolutional neural networks. Medical Physics **47**(9), 4199–4211 (2020)
26. Xiao, Y., Gupta, A., Sanelli, P.C., Fang, R.: STAR: Spatio-Temporal Architecture for Super-Resolution in Low-Dose CT Perfusion. In: Wang, Q., Shi, Y., Suk, H.I., Suzuki, K. (eds.) Machine Learning in Medical Imaging, vol. 10541, pp. 97–105. Springer International Publishing, Cham (2017)
27. You, S., Xu, C., Xu, C., Tao, D.: Learning from Multiple Teacher Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1285–1294. ACM, Halifax NS Canada (2017)
28. Zhang, Q., Spincemaille, P., Drotman, M., Chen, C., Eskreis-Winkler, S., Huang, W., Zhou, L., Morgan, J., Nguyen, T.D., Prince, M.R., Wang, Y.: Quantitative transport mapping (QTM) for differentiating benign and malignant breast lesion: Comparison with traditional kinetics modeling and semi-quantitative enhancement curve characteristics. Magnetic Resonance Imaging **86**, 86–93 (2022)
29. Zhu, Y., Wang, Y.: Student Customized Knowledge Distillation: Bridging the Gap Between Student and Teacher. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5037–5046. IEEE, Montreal, QC, Canada (2021)