ADVERSARIAL LEARNING FRAMEWORK FOR CEREBROVASCULAR LANDMARK DETECTION USING CROSS-MODALITY INFORMATION

Zimeng Tan[1,2], *Jianjiang Feng*[1,2(✉)], *Wangsheng Lu*[3], *Yin Yin*[3], *Guangming Yang*[3], *and Jie Zhou*[1,2]

[1] Department of Automation, Tsinghua University, Beijing, China

[2] Beijing National Research Center for Information Science and Technology, Beijing, China

[3] UnionStrong (Beijing) Technology Co.Ltd, Beijing, China

ABSTRACT

Anatomical landmark detection plays an important role in cerebrovascular analysis and clinical treatment. However, due to the complex structure and similar local appearance around landmarks, the popular heatmap regression based methods suffer from the landmark confusion problem. In this work, we propose an adversarial learning framework for cerebrovascular landmark detection in MRA images by leveraging cross-modality information. Specifically, we exploit an unpaired large-scale CTA dataset to complement the limited MRA training data. The generator is modified as a U-Net based heatmap regression network, and the discriminator is trained using both MRA and CTA datasets to distinguish between multi-channel heatmap groundtruth and prediction. A relative coordinate matrix and a distance map are introduced to enhance landmark location distribution. Extensive experiments demonstrate the superior and robust performance of our method, even with very limited MRA training data.

Index Terms— Cerebrovascular landmark detection, adversarial learning framework, cross-modality information

1. INTRODUCTION

Cerebrovascular disease is one of the leading causes of disability and fatality worldwide [1]. Anatomical landmark detection is a key technology in cerebrovascular analysis, which represents the vascular hierarchical structure explicitly and provides a prerequisite for subsequent medical image processing, such as vessel centerline extraction [2] and image registration [3]. In this paper, we focus on the Circle of Willis (CoW) in cerebral magnetic resonance angiography (MRA) images, which undertakes important physiological functions and can normally be divided into 20 arterial segments according to 19 bifurcation landmarks [4] (see Fig. 1(a) and (b)). Although there are many approaches dedicated to cerebrovascular research [5, 6, 7], accurate and robust cerebrovascular landmark localization remains a challenging problem.

A major challenge in cerebrovascular landmark detection is the confusion among multiple landmarks due to similar local appearance. That is, in popular heatmap regression based methods [8, 9, 10], the heatmap prediction of a certain landmark may have responses at several bifurcation positions (see Fig. 1(c)). These false positive responses may lead to misdetection of the target landmark to other bifurcation positions and result in anatomically impossible landmark configurations. Furthermore, widespread physiological variations in the population [11] and disease-related effects may change the local appearance around landmarks, which are difficult to determine. Even expert annotators rely on professional expertise and clinical experience, such as statistical length of arterial segment and bilateral symmetry, to deal with complex cases. Therefore, it is essential to incorporate global structural information and anatomical prior knowledge into cerebrovascular landmark detection framework.

On the other hand, one of the fundamental challenges in medical image analysis is data scarcity [12]. Limited training

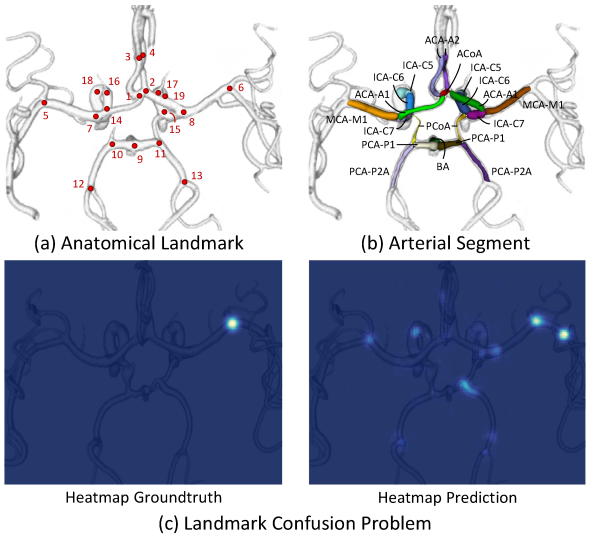

Fig. 1. Illustration of (a) anatomical landmarks, (b) arterial segments, and (c) landmark confusion problem exemplified by one landmark. The heatmap groundtruth and prediction are shown in 2D form using maximum intensity projection.

data may not cover sufficient variation patterns, leading to biased network training. In this paper, we propose to exploit a large-scale CTA dataset to assist landmark detection on a small-scale MRA dataset. Despite the large differences in appearance, these two modalities share many anatomical characteristics, such as vascular structure and landmark layout. The key insight is to leverage unpaired multi-modal images and capture the cross-modality information.

Recently, generative adversarial networks (GANs) [13] have gained considerable attention and achieved promising performance in medical image segmentation [14], synthesis [15] and reconstruction [16]. Taking inspiration from Kanazawa et al. [17], we propose an adversarial learning framework to accomplish cerebrovascular landmark detection. Specifically, as shown in Fig. 2, we modify the generator as a U-Net [18] based multi-channel heatmap regression network. A discriminator is trained using both MRA and CTA datasets to learn the data-driven anatomical prior and global structural information implicitly. Note that only landmark heatmaps are sent to the discriminator to avoid image domain differences between different modalities. In addition, we introduce a coordinate system and a distance map to emphasize the spatial statistic pattern of landmark distribution. The discriminator is expected to distinguish between the heatmap groundtruth and prediction, encouraging the generator to output anatomically plausible landmark predictions.

In summary, our main contributions are three-fold: (1) we propose an adversarial learning framework to accomplish cerebrovascular landmark detection in MRA images, which leverages an unpaired CTA dataset to learn cross-modality information. (2) Only the multi-channel heatmaps are fed into the discriminator to exclude modality-dependent effects. A relative coordinate matrix and a distance map are introduced to enhance the positional prior. The discriminator acts as a data-driven supervision, which learns the global structural information and guides the generator to output anatomically plausible predictions. (3) Extensive experiments demonstrate the effectiveness and robustness of the proposed method, even with very limited MRA training data.

2. METHOD

2.1. Data Preparation

Sufficient training data is fundamental in modeling data manifold for deep neural networks. However, data scarcity has been a long-standing challenge in medical image domain [12]. To tackle the challenge posed by small-scale MRA images (70 training scans), we exploit a separate large-scale CTA dataset (500 scans) to introduce much more complex structure variations. The image of a certain modality can be disentangled into modality-variation features such as imaging pattern and grayscale distribution, and patient-variation features such as vascular structure and landmark layout, with the latter shared among different modalities of the same patient. Therefore, after removing the modality-dependent features, the large-scale CTA dataset can complement the limited MRA training data to guide the network to learn the landmark statistical distribution. It is noteworthy to mention that we do not rely on paired MRA-CTA images, which are difficult to obtain clinically, and only the landmark annotations of CTA dataset are used. To keep the images of different modalities in the same coordinate system, we additionally exploit a pair of MRA-CTA scans acquired from a single patient, which have been aligned using manually annotated vascular binary segmentation. Then, the entire MRA and CTA datasets are registered to the single-person template of the corresponding modality respectively based on intensity.

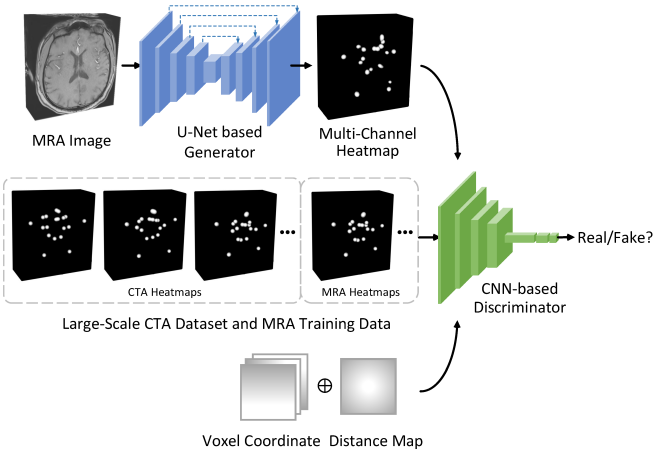

Fig. 2. The overview of the adversarial learning framework. The multi-channel heatmaps are projected into a single volume by channel-wise summation for better visualization.

2.2. Adversarial Framework

Standard generative adversarial network consists of a generator G and a discriminator D. As shown in Fig. 2, we replace the generator with a U-shape heatmap regression network. This voxel-to-voxel method [8] has been proven more effective than regressing landmark coordinates directly, which extracts location information and maintains similar data form as input. Specifically, the coordinates of each landmark are converted to a separate heatmap channel, where a Gaussian distribution is centered at the landmark position. The heatmap temperature $H_i(x)$ on voxel x ranging in $[0, 1]$ represents the probability of the ith landmark, which is determined according to the standard deviation δ and the Euclidean distance from voxel x to landmark x_i. Formally, the multi-channel heatmap for 19 landmarks is defined as follows:

$$H_i(x) = e^{-\frac{1}{2\delta^2}(x-x_i)^2}, i = 1, 2, ..., 19. \quad (1)$$

To solve the gradient vanishing problem, the plain convolutional layer in the original U-Net [18] is replaced with

residual block [19], which contains two convolutional operators and a shortcut connection. The long skip connections between encoder and decoder enable flexible feature fusion and preserve more local details for preciser detection. During inference, the voxel with the maximum probability is chosen as the predicted position of the corresponding landmark.

Given the similar local appearance, the generator network is prone to be confused among different landmarks. In the heatmap prediction of a certain landmark, highlighted areas may also appear on other landmark positions, which results in mislocalization of the target landmark and anatomically impossible landmark prediction. Thus, we exploit a discriminator to introduce adversarial regularization. The heatmap groundtruth and generator prediction are fed into the discriminator. Following the CNN architecture, the discriminator acts as a binary classifier and outputs a scalar indicating whether the input heatmap corresponds to real data or not. To overcome the modality differences in image domain, the original image is excluded from the discriminator input, such that both MRA training data and large-scale CTA dataset can be utilized for supervision, which cover sufficient morphological and structural variations. Furthermore, to enhance the position information, a normalized coordinate matrix and a distance map relative to the volume center are fed into the discriminator. In this way, a well-trained discriminator can capture the global structural prior implicitly, and the generator is discouraged from predicting heatmap according to anatomically implausible landmark configuration.

2.3. Training Strategy and Loss Functions

To accelerate convergence, we adopted a sequential training scheme composed of initial pretraining of the generator network and joint training of the entire framework. In the pretraining phase, only MRA training data is utilized. Given an input MRA image I, L2 loss between heatmap groundtruth H and prediction $G(I)$ is used for supervision. To tackle the class imbalance problem, i.e., the Gaussian spot occupies a small proportion of the output volume, we weight the L2 loss using the exponential form of H, which can be formulated as:

$$\mathcal{L}_{\text{heat}} = E_{I,H\sim P_{\text{MRA}}(I,H)}\left[w\left\|(H-G(I))\right\|_2^2\right],\ w=\alpha^H. \tag{2}$$

where $P_{\cdot}$ denotes data distribution, and the base α is set to 1000 empirically. In the joint training phase, both the MRA and CTA datasets are included. Following the standard practice [13], the generator G and discriminator D are trained alternatively. The loss function of generator G consists of heatmap loss and adversarial loss, with a weight λ_{adv} controlling the trade-off:

$$\mathcal{L}_G = \mathcal{L}_{\text{heat}} - \lambda_{\text{adv}} E_{I\sim P_{\text{MRA}}(I)}\left[\log D(G(I))\right]. \tag{3}$$

Note that the coordinate matrix and distance map in the input of D are omitted for simplicity. Given a fixed G, the discriminator D is updated by minimizing:

$$\begin{aligned}\mathcal{L}_D = &- E_{H\sim P_{\text{MRA}}\bigcup P_{\text{CTA}}(H)}\left[\log D(H)\right] \\ &- E_{I\sim P_{\text{MRA}}(I)}\left[\log(1-D(G(I)))\right].\end{aligned} \tag{4}$$

Table 1. Quantitative results evaluated by MRE, associated SD, and SDR, with the best performance shown in bold. "D-MRA" and "D-MRA-CTA" denote the adversarial learning framework trained with only MRA dataset and with both MRA and CTA images. The superscript "†" indicates the introduction of the coordinate matrix and distance map.

Method	MRE (SD) (mm)	SDR (%)				
		2mm	3mm	4mm	5mm	6mm
Payer et al. [8]	1.83 (0.66)	75.82	84.05	89.80	92.76	95.23
Noothout et al. [20]	2.10 (**0.39**)	57.24	79.61	92.11	95.39	**97.04**
Baseline	3.08 (1.80)	82.40	85.69	88.82	90.13	91.12
D-MRA	1.88 (1.19)	82.89	89.64	92.60	94.41	95.23
D-MRA†	1.73 (1.01)	85.53	90.46	93.59	94.24	95.23
D-MRA-CTA	1.53 (0.89)	86.02	90.79	93.75	95.07	95.56
D-MRA-CTA† (Ours)	**1.45** (0.72)	**86.18**	**91.61**	**93.75**	**95.56**	96.22
75% training data	1.83 (1.34)	83.22	89.31	92.93	94.41	95.07
50% training data	2.23 (1.22)	79.11	87.01	90.63	92.43	93.75

3. EXPERIMENTS

3.1. Datasets and Implement Details

We exploited a public MRA dataset and an in-house CTA dataset to train and evaluate our method. The public MRA dataset contains 102 healthy scans selected from the UNC dataset (https://public.kitware.com/Wiki/TubeTK/Data, with a total of 109 scans), where the images with incomplete ICA-C5 segments due to incomplete scanning range (2 scans), missing unilateral ACA-A segment (3 scans), and strong noise interference (2 scans) were excluded. We randomly selected 70 scans as the training set, with the remaining 32 scans for testing. The CTA dataset consists of 500 scans collected clinically with acute ischemic stroke. Note that these two datasets differ not only in imaging modalities, but also in population and health status. Nineteen predefined landmarks were annotated manually on both datasets by one of the authors and verified by an experienced neurosurgeon. All the scans were spatially normalized to $0.513 \times 0.513 \times 0.8$ mm^3 and automatically cropped to $192 \times 160 \times 96$ according to the mean distribution of landmark annotations.

Considering that image registration has been conducted, only tiny random translation was applied for data augmentation. We trained all the networks using an Adam optimizer with a learning rate of 0.0001. The method was implemented in PyTorch on a single NVIDIA GeForce RTX 3090 GPU.

3.2. Results

We utilized the mean radial error (MRE) and the successful detection rate (SDR) as metrics to evaluate our method.

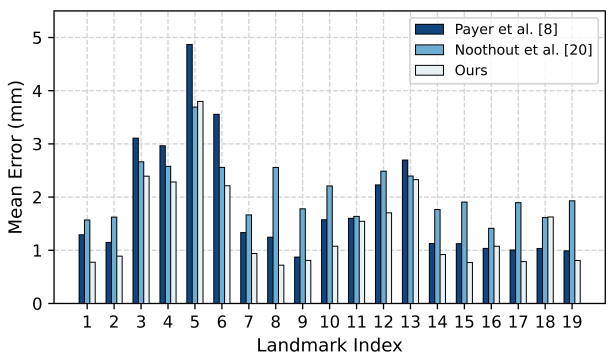

Fig. 3. Mean radial errors (MREs) of all landmarks in test set. See Fig. 1 for the meaning of landmark index.

The MRE calculates the mean Euclidean distance (in mm) between groundtruth and predicted landmark positions. The associated standard deviation (SD) is also reported. The SDR measures the successful detection percentage whithin predefined precision threshold. Five precision thresholds (2 mm, 3 mm, 4 mm, 5 mm, and 6 mm) were used in our experiments. The quantitative results are presented in Table 1.

We compared the proposed method with two state-of-the-art deep learning method for anatomical landmark detection [8, 20]. Payer et al. [8] proposed a spatial configuration component to learn the landmark distribution implicitly, while Noothout et al. [20] performed classification and regression in parallel and developed a global-to-local localization framework. Our method achieves the lowest detection error of 1.45±0.72 mm, showing significant improvements by 0.38 mm (21% reduction) and 0.65 mm (31% reduction) than [8] and [20], respectively. The method proposed in [20] prevents large outlier predictions through the classification branch and thus obtains the lowest SD of MRE and the highest SDR within 6 mm threshold. In contrast, our method maintains superior and robust performance in SDR within different precision thresholds. The MREs for each landmark are shown in Fig. 3. Some landmarks are inherently more difficult due to variable vessel shape and interference of branches, such as the bifurcation landmarks between MCA-M1 and M2 segments (i.e., landmark 5 and 6) and between PCA-P2A and P2P segments (i.e., landmark 12 and 13). Three typical samples are illustrated in Fig. 4 for qualitative comparison.

To verify the effectiveness of each component, we conducted ablation experiments. The heatmap regression network is indicated as the baseline, then the discriminator, coordinate matrix and distance map are gradually added to the framework. In particular, we compared the discriminator trained with only MRA training data and with both MRA and CTA datasets. Compared to the baseline, the proposed method provides a significant improvement by 1.2 mm in MRE when training using only MRA images. By leveraging the CTA dataset, the detection error is reduced by 0.35 mm,

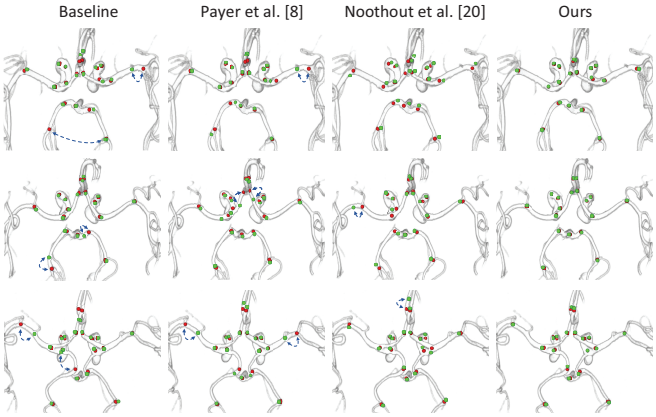

Fig. 4. Visual comparison of different methods. The landmark groundtruth and predicted positions are shown in red and green dots respectively, with the blue dashed arrows indicating large detection errors. Some landmarks are not drawn due to occlusion. Each row corresponds to one example.

with the SDRs improving accordingly. This indicates that the discriminator learns the structural information and guides the output of the generator to approach the real landmark distribution. Introducing the coordinate system and distance map further boosts the overall performance.

Furthermore, we compared the performance of models with different scales of MRA training data. In the proposed adversarial framework, the discriminator captures the cross-modality information to suppress erroneous responses of the generator. Therefore, only a small amount of MRA data is required for the heatmap regression network to learn local features. As shown in Table 1, even in challenging experimental setting (50% reduction of training data, 35 scans), the performance shows just a slight decrease by 0.78 mm in MRE.

4. CONCLUSION

In this paper, we presented an adversarial learning framework for cerebrovascular landmark detection in MRA images, which leverages datasets from different modalities and models cross-modality information. The discriminator is trained to distinguish between the heatmap groundtruth and prediction, which learns the global structural knowledge implicitly and encourages the generator to output anatomically plausible predictions. The relative coordinate matrix and distance map further enhance the validity of landmark distribution. The experimental results indicate that our method achieves state-of-the-art performance and provides a significant improvement. Although only tested using a large number of CTA samples to assist landmark detection in MRA image, we believe that the proposed method is also effective in the reverse scenario. It is interesting to extend the proposed framework to other multi-modal landmark detection problems.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study got ethical approval of Xuanwu Hospital of Capital Medical University (2020009) for using the clinically collected cerebral CTA dataset. The UNC dataset is publicly available.

6. REFERENCES

[1] Ruifang Dong, Zhiling Dong, Hongmei Liu, Fangkun Shi, and Junfeng Du, "Prevalence, risk factors, outcomes, and treatment of obstructive sleep apnea in patients with cerebrovascular disease: a systematic review," *Journal of Stroke and Cerebrovascular Diseases*, vol. 27, no. 6, pp. 1471–1480, 2018.

[2] Li Chen, Thomas Hatsukami, Jenq-Neng Hwang, and Chun Yuan, "Automated intracranial artery labeling using a graph neural network and hierarchical refinement," in *MICCAI*, 2020, pp. 76–85.

[3] Sepideh Almasi, Alexandra Lauric, Adel Malek, and Eric L Miller, "Cerebrovascular network registration via an efficient attributed graph matching technique," *MedIA*, vol. 46, pp. 118–129, 2018.

[4] Hrvoje Bogunović, José María Pozo, Rubén Cárdenes, Luis San Román, and Alejandro F Frangi, "Anatomical labeling of the circle of willis using maximum a posteriori probability estimation," *IEEE TMI*, vol. 32, no. 9, pp. 1587–1599, 2013.

[5] Hao Zhang, Likun Xia, Ran Song, Jianlong Yang, Huaying Hao, Jiang Liu, and Yitian Zhao, "Cerebrovascular segmentation in MRA via reverse edge attention network," in *MICCAI*, 2020, pp. 66–75.

[6] Lei Mou, Yitian Zhao, Huazhu Fu, Yonghuai Liu, Jun Cheng, Yalin Zheng, et al., "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *MedIA*, vol. 67, pp. 101874, 2021.

[7] Zimeng Tan, Jianjiang Feng, Wangsheng Lu, Yin Yin, Guangming Yang, and Jie Zhou, "Cerebrovascular landmark detection under anatomical variations," in *ISBI*, 2022, pp. 1–5.

[8] Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler, "Integrating spatial configuration into heatmap regression based CNNs for landmark localization," *MedIA*, vol. 54, pp. 207–219, 2019.

[9] Wei Liu, Yu Wang, Tao Jiang, Ying Chi, Lei Zhang, and Xian-Sheng Hua, "Landmarks detection with anatomical constraints for total hip arthroplasty preoperative measurements," in *MICCAI*, 2020, pp. 670–679.

[10] Jingyuan Xu, Hongtao Xie, Chuanbin Liu, Fang Yang, Sicheng Zhang, Xun Chen, and Yongdong Zhang, "Hip landmark detection with dependency mining in ultrasound image," *IEEE TMI*, vol. 40, no. 12, pp. 3762–3774, 2021.

[11] Fethi Emre Ustabaşıoğlu, Serdar Solak, et al., "Magnetic resonance angiographic evaluation of anatomic variations of the circle of willis," *The Medical Journal of Haydarpaşa Numune Training and Research Hospital*, vol. 59, no. 3, pp. 291–295, 2019.

[12] Daniel C Castro, Ian Walker, and Ben Glocker, "Causality matters in medical imaging," *Nature Communications*, vol. 11, no. 1, pp. 1–10, 2020.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[14] Huisi Wu, Xuheng Lu, Baiying Lei, and Zhenkun Wen, "Automated left ventricular segmentation from cardiac magnetic resonance images via adversarial learning with multi-stage pose estimation network and co-discriminator," *MedIA*, vol. 68, pp. 101891, 2021.

[15] Tianyang Zhang, Huazhu Fu, Yitian Zhao, Jun Cheng, Mengjie Guo, Zaiwang Gu, et al., "SkrGAN: Sketching-rendering unconditional generative adversarial networks for medical image synthesis," in *MICCAI*, 2019, pp. 777–785.

[16] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, et al., "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE TMI*, vol. 37, no. 6, pp. 1310–1321, 2017.

[17] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik, "End-to-end recovery of human shape and pose," in *CVPR*, 2018, pp. 7122–7131.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[20] Julia MH Noothout, Bob D De Vos, Jelmer M Wolterink, Elbrich M Postma, Paul AM Smeets, Richard AP Takx, et al., "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE TMI*, vol. 39, no. 12, pp. 4011–4022, 2020.