2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-2923-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISBI52829.2022.9761512

A META-LEARNING APPROACH FOR MEDICAL IMAGE REGISTRATION

Heejung Park[1*] *Gyeong Min Lee*[1*] *Soopil Kim*[1] *Ga Hyung Ryu*[2,3] *Areum Jeong*[2,3]
Min Sagong[2,3 †] *Sang Hyun Park*[1†]

[1]Department of Robotics Engineering, DGIST, Daegu, South Korea
[2]Department of Ophthalmology, Yeungnam University College of Medicine, Daegu, South Korea
[3]Yeungnam Eye Center, Yeungnam University Hospital, Daegu, South Korea

ABSTRACT

Non-rigid registration is a necessary but challenging task in medical imaging studies. Recently, unsupervised registration models have shown good performance, but they often require a large-scale training dataset and long training time. Therefore, in real world application where only dozens to hundreds of image pairs are available, existing models cannot be practically used. To address these limitations, we propose a novel unsupervised registration model which is integrated with a gradient-based meta learning framework. In particular, we train a meta learner which finds an optimal initialization point of parameters by utilizing various registration datasets. To quickly adapt to diverse tasks, the meta learner was updated to get close to the center of parameters which are fine-tuned for each registration task. Thereby, our model can adapt to unseen domain tasks via a short fine-tuning process and perform accurate registration. To verify the superiority of our model, we train the model using various types of medical data sets such as retinal Optical Coherence Tomography Angiography (OCTA) for choroidal vasculature, body CT scans, and brain MRI scans and then test it on registration of unseen retinal OCTA for Superficial Capillary Plexus (SCP). In our experiments, the proposed model obtained significantly improved performance in terms of accuracy and training time compared to other registration models.

Index Terms— Registration, Unsupervised learning, Meta learning, Fine-tuning, VoxelMorph

1. INTRODUCTION

Non-rigid registration is necessary to quantify image changes over time or between different patients. Recently, unsupervised deep learning-based registration methods [1, 2, 3, 4] that do not require ground truth have been proposed and achieved state-of-the-art performance in many applications. VoxelMorph [4] predicts a deformation map and transform a moving image using a spatial transformer [5]. The spatial transformer enables the backpropagation of the network to maximize the similarity between deformed moving image and target image. Based on this model, several follow-up works have been proposed. Zhao et al. [6] proposed a cascaded framework to address large displacements by recursively deforming the images. Mahapatra et al. [7] utilize a GAN architecture, while Lee et al. [8] utilize additional features to perform the registration between images with different characteristics.

Though successful, most methods have addressed the image registration problem in a single domain with a large number of training data. Thus, the performance is often limited when there are few image pairs available for training. Guan et al. [9] proposed a transfer learning method for 2D/3D medical image registration to transfer knowledge from multiple patients to the target, but the model was also trained with in a single domain data. Recently, meta learning has been proposed to address the limitation of supervised learning that required a large amount of data [10, 11, 12]. In particular, MAML [13] finds an initialization point that can quickly adapt to various tasks with only a few data samples. Nichol et al. [14] proposed Reptile to reduce the computation of MAML by reformulating the second-order framework to the first-order. They have been applied to various tasks such as few-shot segmentation [15, 16] and zero-shot super resolution [17]. Wang et al. [18] proposed a meta-learning method that addressed 3D point cloud registration. The weights of registration model are estimated by external meta modules. However, the model is trained in supervised manner using ground truth, which is difficult to be applied to the medical image registration since it is expensive to make ground truth of the deformation map.

To address this problem, we propose a meta-learning based unsupervised registration method that is quickly adaptable to various domains. To achieve this, we optimize the parameters of VoxelMorpth registration model [4] via Reptile [14] that is one of gradient-based meta learning methods. Specifically, the registration model is first trained using all multi-domain datasets and then fine-tuned by meta-training to find an optimal initialization point which can quickly adapt

[*]The authors contribute equally to this paper.
[†]Corresponding authors

to various registration problems. The model is updated to get close to the center of parameters fine-tuned for each registration task. To evaluate our proposed method, we trained our model with multi-domain datasets including Optical Coherence Tomography Angiography (OCTA) chroid, abdomen CT and brain MRI scans, and then tested to an OCTA Superficial Caillary Plexus(SCP) dataset.

The main contributions of this work are summarized as follows: (1) we propose a novel registration framework which quickly adapts to new tasks on unseen domains via meta-learning framework. (2) our framework can effectively utilize existing registration data to learn a new registration task from unseen domains. Lastly, (3) we empirically demonstrate the benefit of our method with comparisons with conventional deep learning based registration methods and transfer learning based methods.

2. METHOD

An overview of our proposed model is shown in Fig. 1. We utilize data from multiple domains D_{source} to learn robust registration model. First, an unsupervised registration model G_{θ} is pretrained with D_{source} to learn initial parameters θ_{pt} to predict a displacement map ϕ for moving and fixed image pairs (M, F). Here, any deep-learning based registration model can be used. In our experiment, we used VoxelMorph [4] as the registration model that contains a U-Net [19] style encoder-decoder with a spatial transformer [5]. M and F are concatenated in the channel direction and used as input for G_{θ}. The deformation field ϕ is obtained as a result of $G_{\theta}(M, F)$ and used to make a deformed moving image $M(\phi)$ using the spatial transformer. For optimization, we employ a loss defined by the similarity between $M(\phi)$ and F and the smoothness of ϕ as:

$$L(F, M, \phi) = -CC(F, M(\phi)) + \lambda \sum \| \bigtriangledown \phi \|^2, \qquad (1)$$

where $CC(F, M(\phi))$ is the cross-correlation (CC) in 9×9 window between F and $M(\phi)$, $\sum \| \bigtriangledown \phi \|^2$ penalizes the magnitude of ϕ gradients, and λ is a parameter to weight two terms. At this initial training stage, all data from multiple domains are used as a single domain problem to find representative features from image pairs in all domain. By learning the representative features, we can boost the meta learning process which is sometimes unstable to learn meta knowledge when trained from scratch. After loss convergence, the pretrained model θ_{pt} is saved.

Then, $G_{\theta_{meta}}$ is trained using the same dataset following the manner of Reptile [14] which is a gradient-based meta learning framework. Specifically, we sample several tasks $T_1, T_2, ..., T_m$ from D_{source} where m is the number of tasks and then randomly select (M, F) pairs from each task for the task-level training. For each task, we find fine-tuned parameter θ_t by updating θ_{pt} k times via gradient descent using the selected (M, F) pairs with the loss following Eq. (1). After we compute m fine-tuned parameters, we initially set θ_{pt} as θ_{meta} and then repeatedly update θ_{meta} using $(\theta_t - \theta_{meta})$ as a gradient as:

$$\theta_{meta} = \theta_{meta} + \alpha \frac{1}{m} \sum_{t=1}^{m} (\theta_t - \theta_{meta}), \qquad (2)$$

where α is the learning rate. We repeat this procedure until θ_{meta} converges. Through the meta learning stage, the model learns transferable knowledge that is adaptable to various tasks and thus it can be quickly updated when a new task is given.

Finally, if a new task is given, $G_{\theta_{meta}}$ is fine-tuned to $\hat{\theta}$ using a target domain data D_{target}. When performing the fine-tuning at this stage, all pairs in D_{target} were used to update the parameters using the loss in Eq. (1). After the adaptation, $\hat{\theta}$ is finally used to perform the registration of test data. Note that the test data can be included in D_{target} for model update since the proposed model is an unsupervised learning based model. However, if the model parameters are updated to overfit the testing images, it may be difficult to show the effect of fine tuning properly. Thus, in our experiments, D_{target} was divided into a training set for fine-tuning of $\hat{\theta}$ and a separate test set for testing.

We implemented the proposed method using pytorch [20] on Intel i9-9900K CPU, NVIDIA GeForce RTX 2080 Ti with 64 GB RAM. To update θ_t and θ_{meta}, Adam [21] optimizer was used with the learning rate α of $1e^{-4}$. In each meta-learning step, a batch consisted of randomly sampled 3 tasks ($m = 3$) and k was set as 10. We used the same gradient regularization parameter λ set to 1 for θ_t and θ_{meta}.

3. EXPERIMENTAL RESULTS

In our experiment, we used four datasets including retinal OCTA SCP, retinal OCTA choroid, abdomen CT, and Brain MRI. Both OCTA SCP and choroid datasets contained 368 moving and fixed image pairs collected from local university hospital, some of which were taken from same subjects at different times. Note that the color of the blood vessels and the texture of background of OCTA SCP and OCTA choroid images look completely different from each other. The abdomen CT and brain MRI images were obtained from public Decathlon dataset [22]. Here, we define three tasks according to modality (T1w, T1Gd, and T2w) from the brain MRI dataset and two tasks in the abdomen CT dataset according to different image characteristics. Each 3D volume was divided into multiple axial slices and adjacent two slices were defined as a (M, F) pair. All images were resized to a size of 400×400 and histogram equalization was applied. Also, the range of intensity was rescaled to [0,1]. For training, we defined a set of five tasks as the source data $D_{source} = \{T_{brainT1}, T_{brainT1Gd}, T_{brainT2}, T_{abdomen}, T_{Choroid}\}$. Retinal OCTA

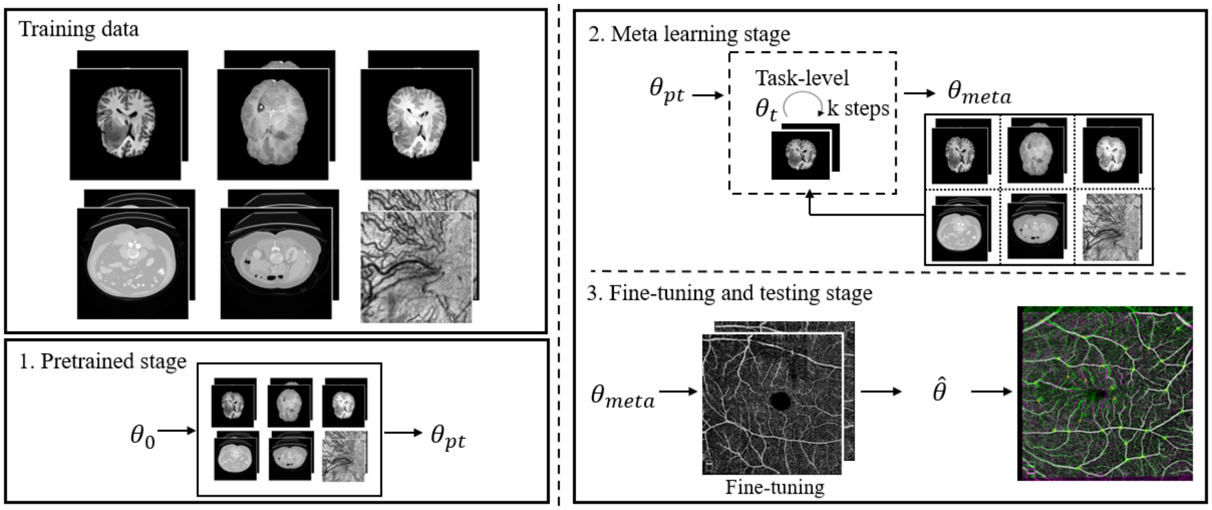


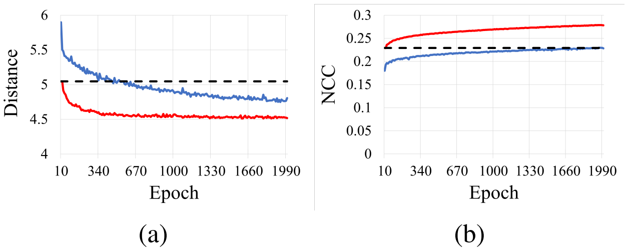
Fig. 1. Overview of the proposed method.

Table 1. Performance comparison of the proposed model against the other baselines on OCTA SCP dataset. (Pixel distance ± std and NCC ± std)

Method	Distance	NCC
Not deformed	10.07 ± 3.99	0.06 ± 0.02
Affine	7.82 ± 4.43	0.10 ± 0.04
B-spline	5.20 ± 3.50	0.24 ± 0.07
VoxelMorph-Seen	5.55 ± 3.99	0.24 ± 0.09
VoxelMorph-Unseen	5.85 ± 3.99	0.22 ± 0.08
Transfer	6.19 ± 3.59	0.16 ± 0.08
Fine-tune - 10 epochs	5.90 ± 3.66	0.18 ± 0.08
Fine-tune - 2000 epochs	4.81 ± 2.93	0.24 ± 0.08
Ours - 10 epochs	5.05 ± 3.27	0.23 ± 0.09
Ours - 2000 epochs.	4.52 ± 2.93	0.28 ± 0.10

SCP was used as target domain D_{target}. OCTA SCP dataset was divided into a training set T_{train} (294 pairs) for fine-tuning and a test set T_{test} (74 pairs). To evaluate the registration performance of unsupervised models, we manually labeled 20∼30 bifurcation points on image pairs in T_{test}. As evaluation metrics, we measured average pixel distances between deformed bifurcation points on $M(\phi)$ and correspondences on F and the normalized cross-correlation (NCC) between $M(\phi)$ and F.

To evaluate the registration performance, we compared our method against six models such as Affine, B-spline, *VoxelMorph-Seen, VoxelMorph-Unseen, Transfer*, and *Fine-tune* with different training strategy. As conventional methods, Affine and B-spline were implemented using SimpleITK library [23, 24] with NCC similarity measure. *VoxelMorph-Seen, VoxelMorph-Unseen, Transfer* are deep-learning based methods, which use the same registration network as ours. *VoxelMorph-Seen* and *VoxelMorph-Unseen* were both trained using only the target domain data. *VoxelMorph-Seen* was trained using both T_{train} and T_{test}, while *VoxelMorph-Unseen* was trained using T_{train}. *Transfer* and *Fine-tune* models were transfer learning-based models. *Transfer* was trained with D_{source} and then tested on T_{test} without additional update. Meanwhile, *Fine-tune* was additionally trained from the parameters of *Transfer* using T_{train} and then applied to T_{test}. Our proposed model performed the task-level training from the parameters of *Transfer*. It was trained with D_{source} in task-level training, and then fine-tuned using T_{train}. We tested these models after their loss converges, i.e., 25k, 38k, 6k, and 23k epochs for *VoxelMorph-Seen*, *VoxelMorph-Unseen*, pre-training and meta-iterations, respectively.



(a) (b)

Fig. 2. Comparison of fine-tuning phase between *Fine-tune* (solid blue) and Ours (solid red) with Ours-10 epoch performance (dotted black). (a) pixel distance and (b) NCC were recorded in every 10 epochs from 10 to 2000 epochs.

3.1. Results

We present average scores and standard deviations of pixel distance between matching points and NCC between deformed moving and target images of comparison methods in

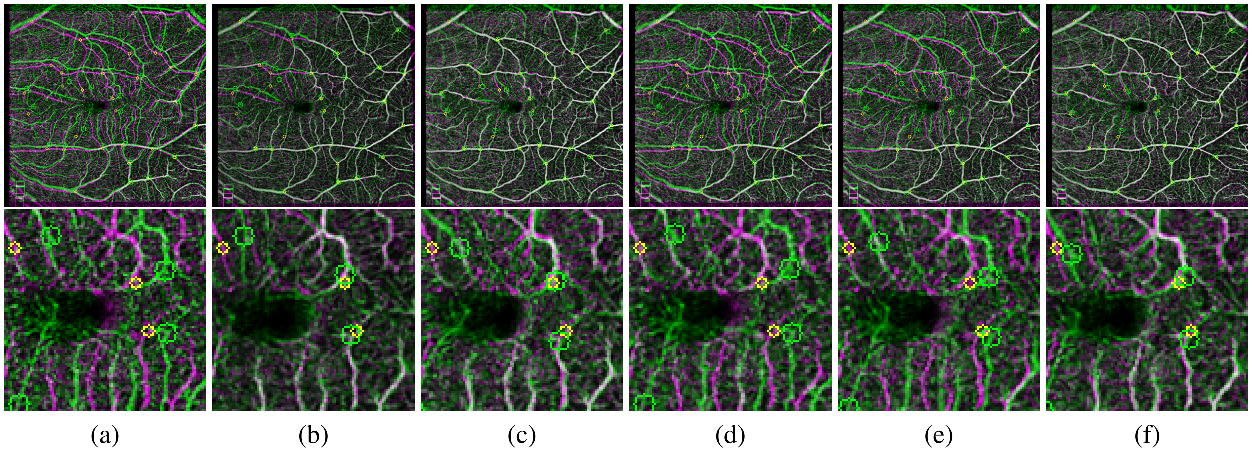


Fig. 3. Qualitative results. Each image shows an overlap of $M(\phi)$ (green) and F (purple) with corresponding points in deformed points (green) and fixed points (yellow). Overlapping area appears in white. (a) Not deformed, (b) B-spliine, (c) *VoxelMorph-seen*, (d) *Transfer*, (e) *Fine-tune* - 10 epochs, (f) Ours - 10 epochs

Table 1. The affine registration method showed the lowest performance since it cannot predict pixel-level deformation map. On the other hand, B-spline showed better performance than *VoxelMorph-Seen* and *VoxelMorph-Unseen*. However, several model parameters need to be set manually according to the data and the performance improvement may be limited since it is not able to utilize the features of images in various domains.

VoxelMorph-Seen and *VoxelMorph-Unseen* show better performance than *Transfer* since *Transfer* is not trained with the data from D_{target}. Meanwhile, when the model is fine-tuned using the data from D_{target} (i.e., *Fine-tune*), it shows better performance than *VoxelMorph-Seen* and *VoxelMorph-Unseen*. On the other hand, our proposed model showed significantly improved performance over *Fine-tune*, i.e., -0.85 and -0.29 pixel distance in 10 and 2000 epochs, respectively. Moreover, ours with only 10 epochs update produced a better performance compared to *VoxelMorph-Seen* which uses T_{test} for training. This shows that our model can produce a better registration results even with less number of training data for the target task.

Fig. 2 shows the comparison of the convergence speed between our proposed method and *Fine-tune*. As shown in (a) pixel distance and (b) NCC graphs, our model achieved much better performance than *Fine-tune*. *Fine-tune* required a large number of model updates to improve the performance and saturation (i.e. more than 500 in pixel distance graph and 1600 epochs in NCC graph), and the results of the model trained for a long time were comparable to the initial results of our method. On the other hand, the proposed method requires only about 600 epochs to converge in NCC graph and achieved better performance even after saturation. These results show that our proposed method can be quickly optimized to the unseen task with better performance.

Fig. 3 shows qualitative results of comparison methods. Our qualitative observations are consistent with quantitative comparisons. *VoxelMorph-Seen* and *Transfer* achieved relatively poor registration results. Although B-spline achieved better performance than these methods, it cannot properly register the area with large changes. In addtion unnatural deformations were often observed in some regions that do not include distinct features. When the model is fine-tuned, there was improvement compared to *Transfer* as shown in *Fine-tune*-10 epochs. However, we still could observe gaps between the corresponding points and vessels. On the other hand, our model obtained closer distances between corresponding points and well-aligned vessels. Compared to other methods, the proposed model works robustly even in regions with large changes.

4. CONCLUSIONS

In this paper, we propose a meta registration model which is adaptable to unseen tasks by utilizing existing various registration tasks. We integrated VoxelMorph model and Reptile meta-learning framework. The model finds optimal parameters that can be quickly applied to various tasks since it utilized various dataset on meta-learning. We expect that our proposed model can be extended to more challenging tasks such as cross-domain image registration.

Acknowledgement

This research was supported by the NRF grant funded by the Korean Government(MSIT)(No. 2019R1C1C1008727) and IITP grant funded by the MSIT (No. 2021-0-02068, Artificial Intelligence Innovation Hub)

5. REFERENCES

[1] Hongming Li and Yong Fan, “Non-rigid image registration using fully convolutional networks with deep self-supervision,” *arXiv preprint:1709.00799*, 2017.

[2] Bob D de Vos, Floris F Berendsen, Max A Viergever, et al., “End-to-end unsupervised deformable image registration with a convolutional neural network,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 204–212. Springer, 2017.

[3] Bob D de Vos, Floris F Berendsen, Max A Viergever, et al., “A deep learning framework for unsupervised affine and deformable image registration,” *Medical image analysis*, vol. 52, pp. 128–143, 2019.

[4] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, et al., “Voxelmorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[5] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al., “Spatial transformer networks,” *Advances in neural information processing systems*, vol. 28, pp. 2017–2025, 2015.

[6] Shengyu Zhao, Yue Dong, Eric I Chang, et al., “Recursive cascaded networks for unsupervised medical image registration,” in *ICCV*, 2019, pp. 10600–10610.

[7] Dwarikanath Mahapatra, Bhavna Antony, Suman Sedai, and Rahil Garnavi, “Deformable medical image registration using generative adversarial networks,” in *ISBI*, 2018, pp. 1449–1453.

[8] Gyoeng Min Lee, Kwang Deok Seo, Hye Ju Song, et al., “Unsupervised learning model for registration of multi-phase ultra-widefield fluorescein angiography,” in *MICCAI*, 2020, pp. 201–210.

[9] Shaoya Guan, Tianmiao Wang, Kai Sun, and Cai Meng, “Transfer learning for nonrigid 2d/3d cardiovascular images registration,” *IEEE Journal of Biomedical and Health Informatics*, 2020.

[10] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales, “Learning to compare: Relation network for few-shot learning,” in *CVPR*, 2018, pp. 1199–1208.

[11] Soopil Kim, Sion An, Philip Chikontwe, and Sang Hyun Park, “Bidirectional rnn-based few shot learning for 3d medical image segmentation,” *AAAI*, 2020.

[12] Sion An, Soopil Kim, Philip Chikontwe, and Sang Hyun Park, “Few-shot relation learning with attention for eeg-based motor imagery classification,” in *IROS*, 2020, pp. 10933–10938.

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in *International Conference on Machine Learning*, 2017, pp. 1126–1135.

[14] Alex Nichol, Joshua Achiam, and John Schulman, “On first-order meta-learning algorithms,” *arXiv preprint:1803.02999*, 2018.

[15] Quande Liu, Qi Dou, and Pheng-Ann Heng, “Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains,” in *MICCAI*, 2020, pp. 475–485.

[16] Sean M Hendryx, Andrew B Leach, Paul D Hein, and Clayton T Morrison, “Meta-learning initializations for image segmentation,” *arXiv preprint:1912.06290*, 2019.

[17] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho, “Meta-transfer learning for zero-shot super-resolution,” in *CVPR*, 2020, pp. 3516–3525.

[18] Lingjing Wang, Yu Hao, Xiang Li, and Yi Fang, “3d meta-registration: Learning to learn registration of 3d point clouds,” 2020.

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

[20] Adam Paszke, Sam Gross, Francisco Massa, et al., “PyTorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

[21] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint:1412.6980*, 2014.

[22] Amber L Simpson, Michela Antonelli, Spyridon Bakas, et al., “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” *arXiv preprint:1902.09063*, 2019.

[23] Bradley Christopher Lowekamp, David T Chen, Luis Ibáñez, and Daniel Blezek, “The design of simpleitk,” *Frontiers in neuroinformatics*, vol. 7, pp. 45, 2013.

[24] Ziv Yaniv, Bradley C Lowekamp, Hans J Johnson, and Richard Beare, “Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research,” *Journal of digital imaging*, vol. 31, no. 3, pp. 290–303, 2018.