Retinal Images Registration via Unsupervised Deep Learning

Giovana Augusta Benvenuto
Faculty of Science and Technology (FCT)
São Paulo State University (UNESP)
Presidente Prudente, Brazil
Email:giovana.a.benvenuto@unesp.br

Wallace Casaca
Institute of Biosciences, Letters and Exact Sciences (IBILCE)
São Paulo State University (UNESP)
São José do Rio Preto, Brazil
Email: wallace.casaca@unesp.br

***Abstract*—In ophthalmology and vision science applications, aligning a pair of retinal images is of paramount importance to support disease diagnosis and routine eye examinations. This paper introduces an end-to-end framework capable of learning the registration task in a fully unsupervised manner. The proposed approach combines Convolutional Neural Networks and Spatial Transformer Network into a unified pipeline that incorporates a similarity metric to gauge the difference between the images, enabling image alignment without requiring any ground-truth data. The validation study demonstrates that the model can successfully deal with several categories of *fundus* images, surpassing other recent techniques for retinal registration.**

I. Introduction

The Image Registration problem consists in finding a geometric or grid transformation that precisely aligns a given image with a reference one. This problem is of crucial importance in the field of Computer Vision, particularly in medical applications where digital imaging is frequently employed for diagnostic and disease monitoring purposes. This is particularly relevant for conditions such as ocular pathologies and disorders, such as *glaucoma* [1] and *diabetic retinopathy* [2].

In the context of ophthalmology, retinal (fundus) images are frequently captured and compared to other images taken at different times, scales, or even using different devices. Manual inspection of potential changes between two or more retinal images is a challenging, time-consuming, and error-prone task. Therefore, the utilization of specific computational techniques is necessary to automate this process. In this type of application, challenges related to eye fundus scanning, such as variations in lighting, scale, angulation, and positioning, are effectively addressed and corrected during the image registration process.

In recent years, the field of medical image registration has considerably benefited from the advancements achieved through the use of Deep Learning (DL). This progress has been highlighted in the works of Litjens et al. [3], Haskins et al. [4], and Fu et al. [5]. However, it is important to note that there are relatively few research proposals in this domain that specifically focus on dealing directly with *fundus* images, particularly in the context of unsupervised approaches.

Mahapatra et al. [6] employ a *Generative Adversarial Network* (GAN), which is a DL architecture consisting of two main components, a generator and a discriminator network, to align pairs of fundus images. Wang et al. [7] propose a framework that utilizes pre-trained networks for segmentation, detection, and feature description of retinal images to perform registration. In the work by Rivas-Villar et al. [8] a supervised network is implemented to address the alignment problem. Landmarks are transformed into heat maps and used by the network to learn and predict such maps during the inference step. However, it is noteworthy to mention that despite these methods demonstrate the capability to solve the image registration problem, they all require some form of reference data, such as a loss function, as part of their implementation.

In summary, most registration methods rely on supervised learning or the creation of synthetically generated data to be effective. While generating new labels can overcome the lack of reference data, it also introduces an additional complication in modeling the problem, raising the question of the reliability of artificially induced data in the field of medical imaging.

In this research, we propose an unsupervised deep learning strategy that combines a convolutional architecture, a spatial transformation module, and a loss function based on a similarity metric, with the goal of performing the registration task on a pair of fundus images without the need to use or acquire reference data (ground-truth) beforehand.

In summary, the main contributions introduced by the proposed approach are:

- The development of an integrated computational framework for performing end-to-end retina image registration using DL techniques, all while bypassing reliance on ground-truth data or artificially and/or manually generated reference resources.
- Establishing a functional and effective registration method capable of adapting to a large number of distinct classes of fundus image pairs.
- The combination of multiple DL networks with image analysis techniques such as the Isotropic Undecimated Wavelet and the Transformed and Linked Component Analysis enables the registration of fundus photographs even with segments of low quality and abrupt changes.

This work derives from Giovana's M.Sc. Dissertation from São Paulo State University

II. Methodology

The purpose of the our methodology is to achieve the unsupervised registration of a pair of fundus images, I_{Mov} and I_{Ref}. To accomplish this objective, we first extract blood veins, bifurcations, and other relevant components of the eye, producing the images B_{Mov} and B_{Ref}. These images serve as input to a Fully Convolutional Neural Network (FCNN) that implements a U-shaped architecture and outputs a correspondence grid between the images. In the subsequent learning step, a Spatial Transformer layer uses the matching grid to compute the transformation necessary to align the moving image to a reference one.

The integrated architecture employed in this project learns the task through a loss function that measures the similarity between the reference and transformed images. Finally, for refinement, we apply a mathematical morphology-based technique called Connected Component Analysis (CCA) [9] to remove noisy pixels that may appear during the learning process. This post-processing step helps to improve the overall quality of the registered images and enhance the accuracy of the registration results.

As a result, the model is capable to learn the registration task without the need for ground-truth annotations and any reference data. Figure 1 illustrates the proposed registration approach. Each stage of these pipeline is detailed in the next sections.

A. Network Input Preparation

The first phase of the proposed computational *framework* focuses on preprocessing the image pairs (I_{Ref} and I_{Mov}) to enhance the network's performance. Initially, the images are resized to 512×512 to reduce the total number of network parameters and then converted to grayscale.

The second step involves segmentation. This treatment aims to emphasize structures in the images that are relevant to solving the problem, specifically, the blood vessels and the optic disc. This process also addresses lighting issues and streamlines network conversion by removing unnecessary information. We utilized the Isotropic Undecimated Wavelet Transform (IUWT), a technique developed by [10] specifically for detecting and measuring retinal images. The resulting images, B_{Ref} and B_{Mov}, an be observed in the leftmost frame in Figure 1.

B. Learning a Deep Correspondence Grid

As previously stated, the initial learning mechanism incorporates a U-Net-like architecture with the objective of generating a deformation grid for the reference and moving images. The network takes the image pair B_{ref} and B_{Mov}, as input, which is then processed in the initial convolutional layer block. The first components of this architecture consists of two downsampling blocks, each composed of a max-pooling layer and two convolution layers. Within each block, the input size is halved according to the image resolution, while the number of analyzed features doubles.

During the second stage, two blocks are introduced as part of the network's upsampling process. These blocks consist of a deconvolution layer, which enlarges the input size while reducing the number of analyzed features, and two convolutional layers. Regarding the second step, the output data from the deconvolution are combined with the data obtained from the convolution block at the corresponding level in the previous step. This merging of data is achieved through concatenation, as indicated by the dashed arrows in Figure 2.

Except for the final convolutional layer, all convolutional layers in the current architecture employ the ReLU activation function and are accompanied by a Batch Normalization layer. The last convolutional layer utilizes a linear activation function and reduces the number of features (kernel) in order to produce a deformation field that matches the dimensions of the input data.

Figure 2 illustrates the network architecture implemented for generating a correspondence grid. Each layer is depicted as a colored block. The resolution of the data is specified beneath each block, while the number of kernels per layer is indicated in the upper right corner.

C. Spatial Transformer Network

Following our model, we incorporate a modified version of the Spatial Transformer Network (STN) architecture [11] to obtain a transformation model for mapping the moving image, B_{Mov}. The STN structure enables the network to dynamically apply scaling, rotation, and cropping, as well as non-rigid transformations to the moving image or feature map. Importantly, these transformations can be achieved without the need for additional training supervision or lateral optimization processes.

The STN incorporated as part of our integrated learning scheme consists of two core modules: grid generator and sampler. The grid generator module aims to align the correspondence positions in the target image B_{Mov} by iterating over the matching points previously determined by the network. Its primary objective is to generate a grid that facilitates the alignment process. Once the matches are properly found, the sampler module apply a bilinear interpolation to extracts the pixel values at each position, generating the definitive transformed image B_{Warp}. The middle frame of Figure 1 exemplifies the outputs of the STN modules implemented.

D. Loss Function

As our registration process does not rely on labeled data, we employ a loss function that utilizes an independent metric to assess the similarity between images. Specifically, we utilize the Normalized Cross-Correlation (NCC) as a mathematical measure for the loss function. NCC allows to quantitatively evaluate the degree of similarity between the images during training. Below is the equation for this measurement:

$$NCC(x,y) = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} T_{i,j}R_{i,j}}{\sqrt{\left(\sum_{i=0}^{m}\sum_{j=0}^{n} T_{i,j}^2\right)\left(\sum_{i=0}^{m}\sum_{j=0}^{n} R_{i,j}^2\right)}} \,. \quad (1)$$

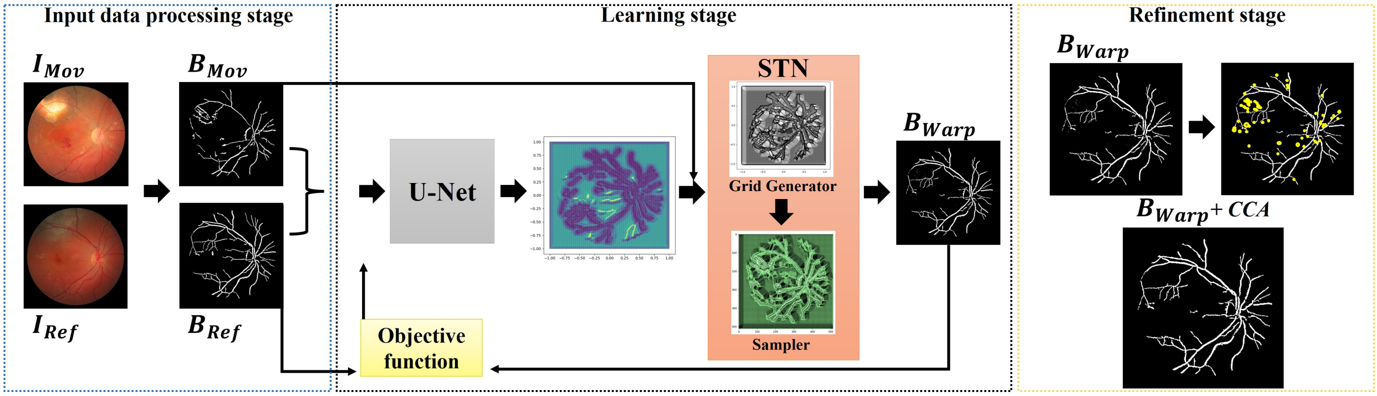

Fig. 1. Overview of the proposed *framework*. It includes a pre-processing step where the segmentation of the image pair occurs, followed by the learning step formed by the U-Net, the Spatial Transformation module and the Loss Function, and finally the post-processing is applied to refine the images.

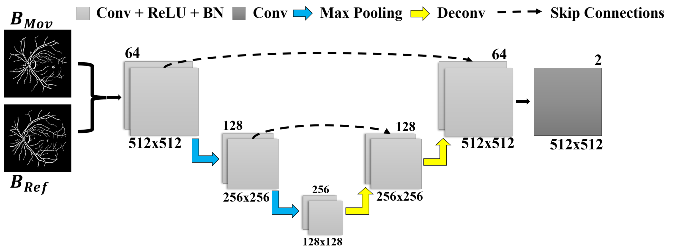

Fig. 2. The network architecture implemented to obtain a correspondence grid is represented by colored blocks, with each block denoting a layer. Below each block, the data resolution is specified, and the upper right corner shows the number of kernels per layer. The correspondence grid serves as the network's output, as displayed in the rightmost corner.

In Equation (1), $T_{i,j} = t(x+i, y+j) - \bar{t}_{x,y}$, $R_{i,j} = r(i,j) - \bar{r}$, and $t(i,j)$ and $r(i,j)$ are the pixel values at (i,j) w.r.t. the warped and reference images, B_{Warp} and B_{Ref}, respectively, while $\bar{r}$ and $\bar{t}$ give the average pixel values w.r.t. B_{Ref} and B_{Warp} [12].

The NCC metric is often selected as a similarity measure due to its robustness, as noted in [13], and its ability to provide high accuracy and adaptability, as mentioned in [14].

III. Experiments

In this section, we present a comprehensive description of the materials used, the conducted experiments, and the technical details of the implementation and evaluation processes, with the objective of providing a comprehensive understanding of the results and findings.

A. Datasets

To evaluate the performance of the proposed methodology for the registration task, we utilized three databases, comprising two publicly available datasets and one private dataset. Below, we provide the specifications of each database:

- **FIRE -** The High-resolution retinal database, available at [15], consists of 134 image pairs categorized into three distinct groups: A, S, and P. These categories are differentiated based on the estimated overlap between the pairs. Categories A and S exhibit an estimated overlap of more than 75%, while category P demonstrates a lower percentage of overlap.
- ***Image Quality Assessment Dataset* (Dataset 1) -** This public dataset, captured by [16], comprises 18 retinal image pairs, with each pair originating from a different individual. Each pair consists of a low-quality image, characterized by smokiness and blur, alongside a higher quality image of the same eye.
- **Private dataset (Dataset 2) -** This private dataset comprises 85 image pairs generously provided by an ophthalmologist who collaborated with our research. The dataset presents a real-world scenario encountered by medical experts during their routine examinations with real patients.

By employing diverse datasets, including both public and private sources, we aim to thoroughly evaluate the effectiveness and robustness of our proposed approach in different scenarios and settings.

B. Evaluation metrics

To quantitatively assess the registration results, we adopted well-known validation metrics that measure the alignment of the image pairs, including the Mean Squared Error (MSE) [6], [17], Structural Similarity Index Measure (SSIM) [17], Dice Coefficient (Dice) [7], [18]–[21], and Gain Coefficient (GC) [22], [23].

For the purpose of this article, we focused solely on the Dice coefficient metric to present the results. For a more comprehensive evaluation, we recommend referring to the accompanying dissertation [24] and the following citations: [25], [26]. These references provide extensive insights and discussions on the performance and effectiveness of our proposed methodology, considering various metrics and experimental setups.

The Dice coefficient is a widely used metric in the context of image registration, ranging between 0 and 1, where a value

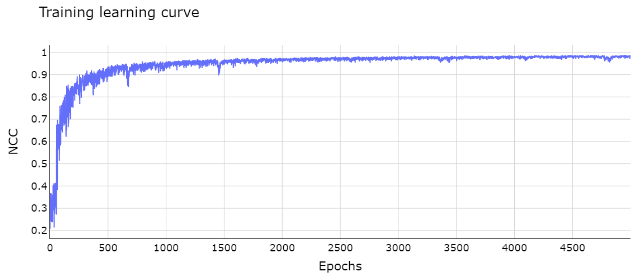

Fig. 3. Learning curve of the network after 5,000 epochs. The vertical axis represents the value of the loss (NCC) for each epoch.

of 1 indicates a perfect overlap. The mathematical calculation of this metric is governed by Equation (2):

$$Dice(B_{Ref}, B_{Warp}) = \frac{2 \times |B_{Ref} \cap B_{Warp}|}{|B_{Ref}| \cup |B_{Warp}|} \tag{2}$$

Where B_{Ref} represents the reference region of interest (ROI), and B_{Warp} represents the registered ROI.

C. Implementation details

The entire pipeline of the proposed approach was implemented in Python, making use of the OpenCV, Scikit-learn, and Tensorflow libraries. The learning process was trained using a routine consisting of eight batches of retinal image pairs for 5000 epochs. Optimization was achieved through the ADAM algorithm. The training was conducted on a cluster equipped with 32GB of RAM memory and two Intel(R) Xeon(R) E5-2690 processors.

The images used in the training were extracted from category S of the FIRE database, with a resolution of 512×512. Towards the end of the training process, we observed that the convergence of the network occurs approximately within the first two thousand iterations. After this point, the results remain stable with minimal fluctuations (see Figure 3).

IV. Results and Discussion

The results obtained from the developed methodology are discussed through both quantitative and qualitative evaluations. The qualitative observation was conducted through a visual inspection of the achieved registrations produced by our proposed approach, comparing them to registrations obtained using other methods from the literature. This qualitative analysis provides valuable insights into the visual quality and accuracy of the registrations, complementing the quantitative metrics and offering a comprehensive assessment of the proposed methodology's performance.

For the quantitative evaluation, we applied similarity metrics to quantify the percentage of overlap between the reference image B_{Ref} and the warped image B_{Warp}. In this article, we specifically focus on demonstrating the results achieved by the Dice Coefficient metric. For a more comprehensive evaluation, see the citations mentioned in Section III-B.

TABLE I
Comparison of registration methods using Dice Coefficient. The values in bold indicate the best performance.

Métodos	**FIRE** **A**	 **S**	 **P**	**Dataset 1**	**Dataset 2**
Before	0.2982	0.3418	0.1245	0.2922	0.3805
GFEMR	0.6023	0.8022	0.5919	0.6565	0.8009
VOTUS	0.6105	*0.8702*	*0.6149*	*0.8064*	*0.8188*
DIRNet	0.4982	0.6020	0.2630	0.5145	0.5744
Hu et al	*0.6303*	0.6948	0.5505	0.6155	0.6588
Proposed	**0.9505**	**0.9580**	**0.9109**	**0.9477**	**0.9467**

A. Comparison with State-of-the-Art and Deep Learning Techniques

To assess the competitiveness of our framework, we compared the achieved results with those obtained using other registration techniques. In the context of traditional methods, which employ optimization techniques for registration tasks, we considered the GFEMR [27] and VOTUS [23] algorithms.

To observe our results compared with other deep learning techniques, we also analyzed the DIRNet [28] network and the weakly supervised method proposed by Hu et al. [29]. For conducting these evaluations, we employed the same training process as utilized in our own framework. Following the specifications of each algorithm, we trained them using the same set of training images and the same number of epochs to ensure a consistent and fair comparison between our proposed approach and these state-of-the-art deep learning methods.

Table I presents a quantitative assessment of the results obtained by applying the Dice Coefficient metric to the compared registration methods. The rows represent the methods, while the columns demonstrate the average values achieved by each method for all applicable image pairs from each database. The row labeled 'Before' corresponds to the results without aligning the image pairs.

Observing the bold values in the table, which indicate superior performance, our proposed framework outperforms all the compared methods, regardless of the different databases. When compared to the second-best results (indicated by italicized values), our framework also demonstrates a significant improvement.

To analyze the distributional performance of the registration for the image pairs, we chose the box plot graph representation. In Figure 4, we can observe the variation of the metric results on the overlap of the reference and registered images for each method and database.

This graphical representation confirms our previous analysis and also reveals that among the compared methods, our proposed approach demonstrates the lowest variation in the registration values. This consistency highlights its ability to achieve high overlap registrations regardless of the pair of images being processed. This finding further emphasizes the robustness and reliability of our method across diverse image pairs and datasets.

In summary, it is noticeable that the traditional methods [23], [27] present better results when compared to the other

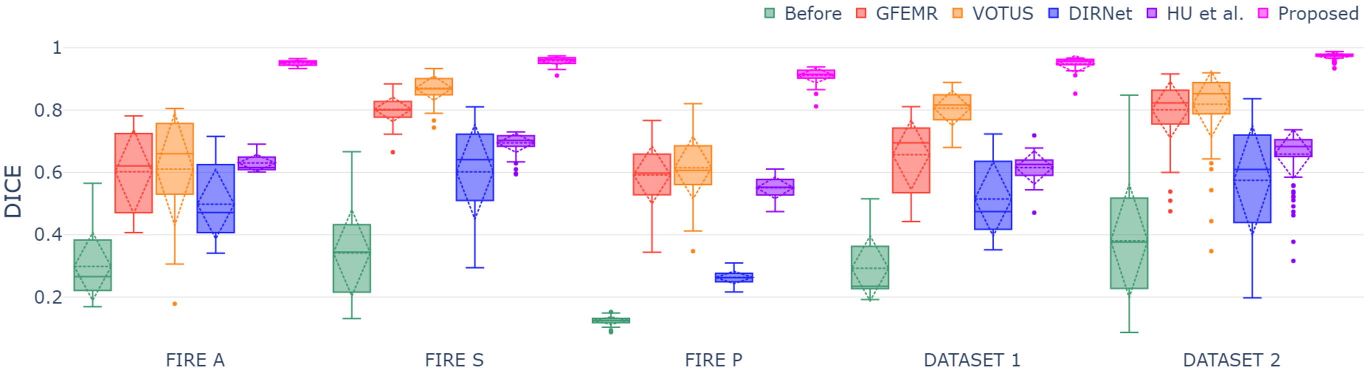

Fig. 4. Box plot representations of the results using the Dice Coefficient metric, with values closer to 1 indicating the best performance.

two deep learning methods [28], [29]. One plausible reason for this discrepancy in results is that deep learning approaches rely on the quantity of data and the number of training epochs to achieve the desired accuracy. On the other hand, such techniques are capable of generating transformed images independently of the input image pair, unlike optimization-based techniques which, in some cases, especially those that are initially more distinct, may fail to produce the registration. For instance, in the category P of the FIRE database, the GFEMR method failed to register 4 images, and VOTUS was unable to register 6 images.

Our framework not only generated a registered image for all tested pairs but also demonstrated good convergence with the volume of data on which it was trained, consistently achieving the best results in all tests conducted. It is noteworthy that the best results obtained by our architecture are in the category of the database on which it was trained (column FIRE S), but they are closely followed by the results from other databases. This indicates that the network has the ability to generalize the learned parameters to different image pairs, showcasing its versatility and adaptability to diverse datasets.

B. *Visual assessment of the comparison of results*

To conduct a visually qualitative analysis of the images generated by each method, we can refer to Figure 5. This figure, obtained from the approach adopted in [23], presents the segmented images B_{Ref} and B_{Warp} in green and magenta, respectively. Since these colors are complementary in the RGB spectrum, their combination results in the color white. Consequently, the white pixels in each image indicate the extent of overlap between them. This visual representation offers valuable insights into the accuracy and alignment achieved by the registration methods, allowing for a comprehensive assessment of their performance.

Analyzing the image registration methods via AP (DIRNet and Hu et al lines), both perform non-rigid registrations, which means that the transformations applied, in some cases, result in deformations in the images, causing them to lack the desired overlap.

The proposed framework, despite also employing non-rigid registration, considers a deformation field with the same dimensions as the input image pair in the network output. Consequently, when mapped for transformation, each point corresponds to a pixel, causing the applied deformation to distort the image B_{Mov} towards the reference image B_{Ref}. The structures in the transformed image B_{Warp} may exhibit differences from the original image B_{Mov} to make it visually closer to its reference, thus maximizing the overlap between both images as much as possible.

Another aspect that we can observe from Figure 5 is the role of segmentation in this framework. This process enables the registration of images acquired under diverse conditions. The column corresponding to *Dataset 1* presents an image pair with low illumination and a smoky occlusion. Through segmentation, it was possible to highlight the vascular structure of these images and perform the registration. In a practical sense, segmentation allows images in poor conditions to be registered, facilitating an initial observation in cases where there is a need for a new examination to replace the damaged image or even avoid a new procedure.

Despite the advantages of applying segmentation to the framework, it also represents a limitation of the proposed methodology since the network only generates segmented registrations.

The ultimate objective of image registration is to facilitate the comparison between pairs of fundus images, enabling ophthalmology professionals to quickly identify signals that indicate alterations, aiding in diagnosis and monitoring. The adopted visualization technique for this evaluation effectively highlights areas where structural changes exist between the reference and registered images.

Figure 6 illustrates the sequence of processing steps leading to the generation of the registration figure and the overlapped visualization. In this specific case, it is noticeable that certain vessel structures present in the reference image are absent in the second image, these are highlighted in green in the final image.

This visualization and comparison approach enhances the effectiveness of medical examinations, enabling practitioners

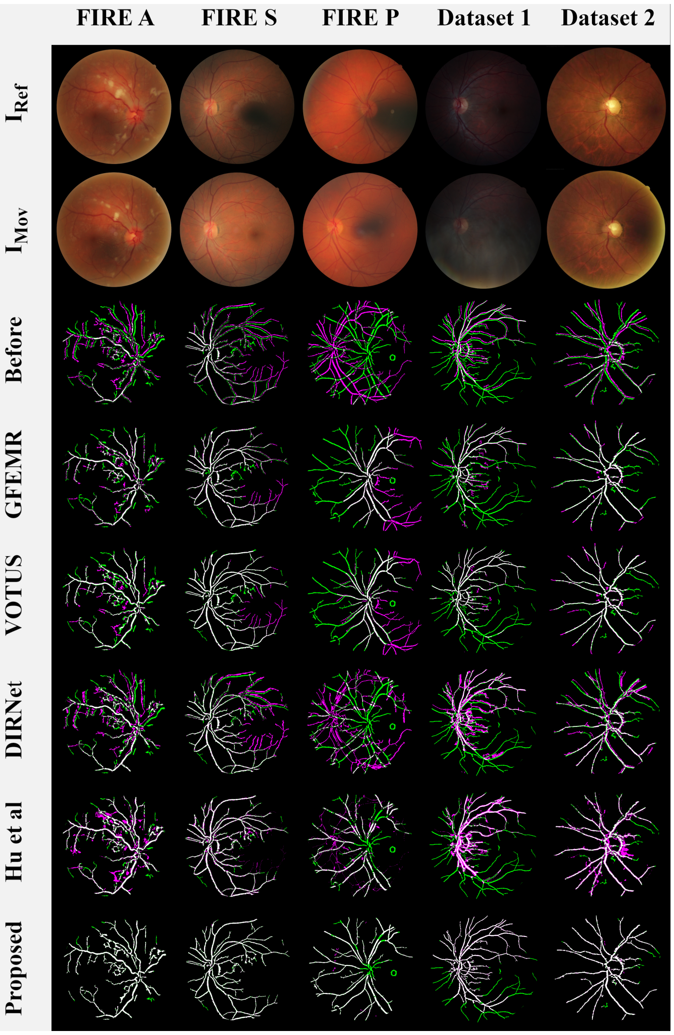

Fig. 5. Demonstration of the overlap between the reference and registered images. The original images from each database are displayed in the first two rows, while the subsequent rows show the overlaps between B_{Ref} in green and B_{Warp} in magenta for each compared method.

to identify and analyze relevant information efficiently, ultimately contributing to improved patient care and accurate diagnoses.

V. Conclusion

In this work, we addressed the problem of digital retinal image registration. We proposed our solution through an unsupervised computational framework that performs end-to-end registration of retinal images. This approach combines two neural networks that employ deep learning architectures.

The proposed framework is focused on the registration of segmented retinal images, as segmentation enables the accomplishment of this task even with low-quality images and makes it feasible for the technique to learn without the need for reference data, using a similarity metric instead.

By leveraging the benefits of segmentation, our framework becomes robust to image variations and allows for the registration of retinal images under diverse conditions, enhancing its practical applicability in real-world medical scenarios. The use of unsupervised learning and similarity metrics further simplifies the registration process and reduces the dependency on labeled data, facilitating the integration of our method into existing medical imaging workflows.

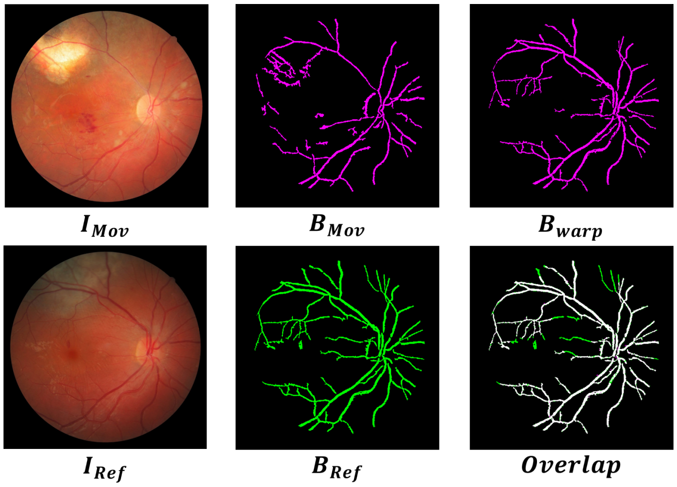

Fig. 6. Demonstration of the processing steps of the original image pair up to the superimposition of the final registration with the reference image.

The results obtained when comparing our proposed framework to other methods from the literature, including traditional optimization-based approaches using key-points and other strategies employing neural networks, demonstrated higher accuracy for all the employed databases and across all applied evaluation metrics.

Our proposed framework outperformed the compared methods in terms of registration accuracy, showcasing its superiority in handling diverse datasets and image variations. These results validate the effectiveness and robustness of our approach in achieving more accurate and reliable image registration outcomes.

Acknowledgment

The authors would like to thank the full support of the São Paulo Research Foundation (FAPESP), grant #2019/26288-7. The project is also associated to the Center for Mathematical Sciences Applied to Industry (CeMEAI), 2013/07375-0. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

VI. Publications and others contributions

The author's dissertation contains the following articles: [25] and [26]. Others contributions were also developed in parallel, related to the participation in an extension project called GECET - Girls in Engineering, Exact Sciences and Technology: [30] and the medical and social research involving the mapping of COVID-19 risk groups in Brazil: [31].

References

[1] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma: a review," *The Journal of the American Medical Association (JAMA)*, vol. 311, no. 18, pp. 1901–1911, 2014.

[2] K. M. Kim, T.-Y. Heo, A. Kim, J. Kim, K. J. Han, J. Yun, and J. K. Min, "Development of a fundus image-based deep learning diagnostic tool for various retinal diseases," *Journal of Personalized Medicine*, vol. 11, no. 5, p. 321, 2021.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841517301135

[4] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, no. 1, pp. 1–18, 2020.

[5] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.

[6] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 1449–1453.

[7] Y. Wang, J. Zhang, C. An, M. Cavichini, M. Jhingan, M. J. Amador-Patarroyo, C. P. Long, D.-U. G. Bartsch, W. R. Freeman, and T. Q. Nguyen, "A segmentation based robust deep learning framework for multimodal retinal image registration," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1369–1373.

[8] D. Rivas-Villar, Álvaro S. Hervella, J. Rouco, and J. Novo, "Color fundus image registration using a learning-based domain-specific landmark detection methodology," *Computers in Biology and Medicine*, vol. 140, p. 105101, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521008957

[9] L. He, X. Ren, Q. Gao, X. Zhao, B. Yao, and Y. Chao, "The connected-component labeling problem: A review of state-of-the-art algorithms," *Pattern Recognition*, vol. 70, pp. 25–43, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320317301693

[10] P. Bankhead, C. Scholfield, J. McGeown, and T. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PloS One*, vol. 7, no. 3, p. e32435, 2012.

[11] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *NIPS*, 2015.

[12] A. Kaso, "Computation of the normalized cross-correlation by fast fourier transform," *PLOS ONE*, vol. 13, no. 9, pp. 1–16, 09 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0203434

[13] M. Hisham, S. N. Yaakob, R. Raof, A. A. Nazren, and N. Wafi, "Template matching using sum of squared difference and normalized cross correlation," in *2015 IEEE Student Conference on Research and Development (SCOReD)*, 2015, pp. 100–104.

[14] Z. Cui, W. Qi, and Y. Liu, "A fast image template matching algorithm based on normalized cross correlation," *Journal of Physics: Conference Series*, vol. 1693, no. 1, p. 012163, dec 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1693/1/012163

[15] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. Argyros, "Fire: Fundus image registration dataset," *Journal for Modeling in Ophthalmology*, vol. 1, no. 4, pp. 16–28, 2017, source code: http://www.ics.forth.gr/cvrl/fire/.

[16] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 95–100.

[17] A. Kori and G. Krishnamurthi, "Zero Shot Learning for Multi-Modal Real Time Image Registration," *arXiv e-prints*, p. arXiv:1908.06213, 08 2019.

[18] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "Birnet: Brain image registration using dual-supervised fully convolutional networks," *Medical Image Analysis*, vol. 54, pp. 193–206, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841519300283

[19] B. D. De Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis*, vol. 52, pp. 128–143, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841518300495

[20] C. WANG, G. Papanastasiou, A. Chartsias, G. Jacenkow, S. A. Tsaftaris, and H. Zhang, "FIRE: Unsupervised bi-directional inter-modality registration using deep networks," *arXiv e-prints*, p. arXiv:1907.05062, Jul. 2019.

[21] T. Che, Y. Zheng, J. Cong, Y. Jiang, Y. Niu, W. Jiao, B. Zhao, and Y. Ding, "Deep group-wise registration for multi-spectral images from fundus images," *IEEE Access*, vol. 7, pp. 27 650–27 661, 2019.

[22] D. Motta, W. Casaca, and A. Paiva, "Fundus image transformation revisited: Towards determining more accurate registrations," in *IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 227–232.

[23] ——, "Vessel optimal transport for automated alignment of retinal fundus images," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6154–6168, 2019.

[24] G. A. Benvenuto, "Registro de imagens de retina via aprendizado profundo não-supervisionado," Master's thesis, Universidade Estadual Paulista (Unesp), 2022, disponível em: https://repositorio.unesp.br/handle/11449/217619.

[25] G. A. Benvenuto, M. Colnago, and W. Casaca, "Unsupervised deep learning network for deformable fundus image registration," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1281–1285.

[26] G. A. Benvenuto, M. Colnago, M. A. Dias, R. G. Negri, E. A. Silva, and W. Casaca, "A fully unsupervised deep learning framework for non-rigid fundus image registration," *Bioengineering*, vol. 9, no. 8, 2022. [Online]. Available: https://www.mdpi.com/2306-5354/9/8/369

[27] J. Wang, J. Chen, H. Xu, S. Zhang, X. Mei, J. Huang, and J. Ma, "Gaussian field estimator with manifold regularization for retinal image registration," *Signal Processing*, vol. 157, pp. 225–235, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168418303955

[28] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "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*, M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu, Eds. Cham: Springer International Publishing, 2017, pp. 204–212.

[29] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical Image Analysis*, vol. 49, pp. 1–13, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841518301051

[30] M. Colnago, C. F. A. Lages, H. S. Picoli, G. A. Benvenuto, T. B. Ghetti, and W. C. d. O. Casaca, "Um estudo de gênero a partir da distribuição de bolsas do programa universidade para todos," 2022.

[31] M. Colnago, G. A. Benvenuto, W. Casaca, R. G. Negri, E. G. Fernandes, and J. A. Cuminato, "Risk factors associated with mortality in hospitalized patients with covid-19 during the omicron wave in brazil," *Bioengineering*, vol. 9, no. 10, 2022. [Online]. Available: https://www.mdpi.com/2306-5354/9/10/584