A segmentation based deep learning framework for multimodal retinal image registration

Yiqian Wang[1], Junkang Zhang[1], Cheolhong An[1], Melina Cavichini[2], Mahima Jhingan[2], Manuel J. Amador-Patarroyo[2], Christopher P. Long[3], Dirk-Uwe G. Bartsch[2], **William R. Freeman[2]**, Truong Q. Nguyen[1]

1 UC San Diego JACOBS SCHOOL OF ENGINEERING Electrical and Computer Engineering

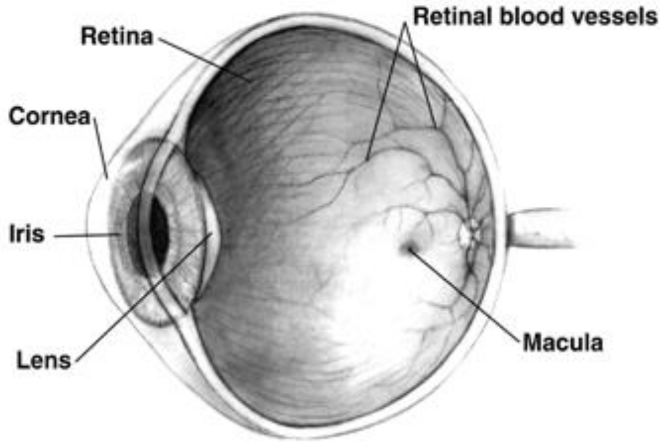
2 SHILEY EYE INSTITUTE UC SAN DIEGO

3 UC San Diego SCHOOL OF MEDICINE

Contents

Contents

The retina



Structure of the human eye (cross-sectional view)

- The retina is the innermost, light-sensitive layer of the eye
- Serve a function like the image sensors in a camera

Retinal diseases

- Age-related macular degeneration (AMD), diabetic retinopathy, glaucoma
- Severely damage the vision of patient

Normal vision

The same view with diabetic retinopathy

The same view with AMD

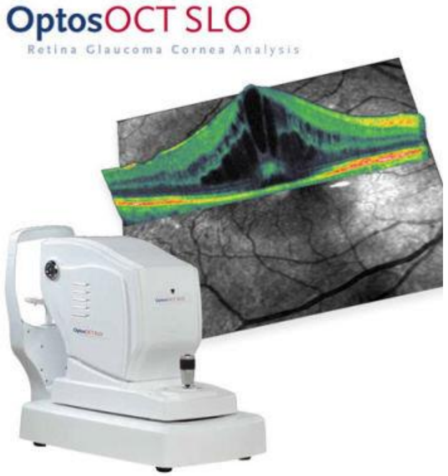
The same view with glaucoma

Images by National Eye Institute, National Institutes of Health

Retinal imaging

- The role of imaging in retinal diseases is critical
- Ophthalmologists face a large array of imaging devices
- Each device uses different methods, wavelengths, functional tests, angiographic dyes, optical systems, angles of view

TRC-NW7SF Mark II

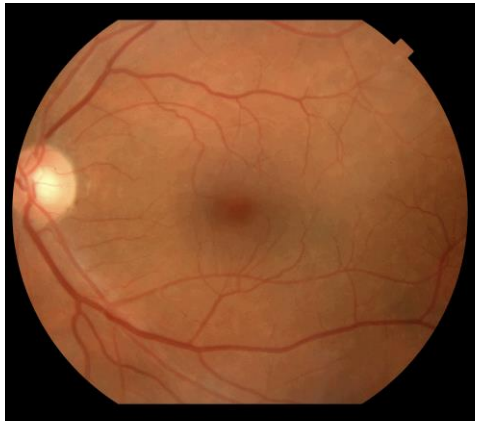


OptosOCT SLO

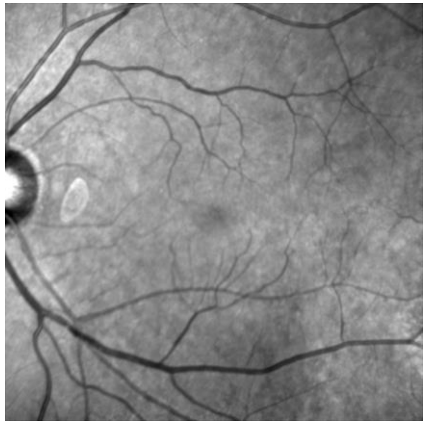
SPECTRALIS SPIRIT

Multimodal retinal image registration

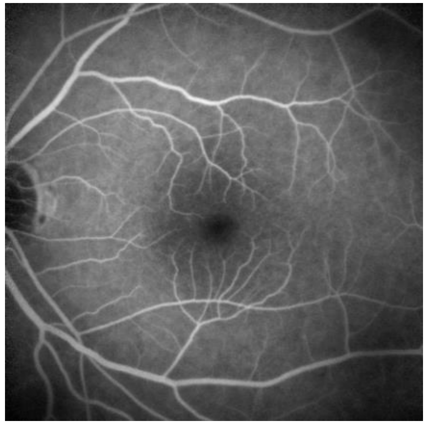
- **Motivation:** Integrate functional and structural evaluations into one co-localizable database
- **Challenge:** Different field of view, lens systems, light sources, manufactures …
- **Solution:** Retinal vessels are seen by all instruments, can be used to align different modalities



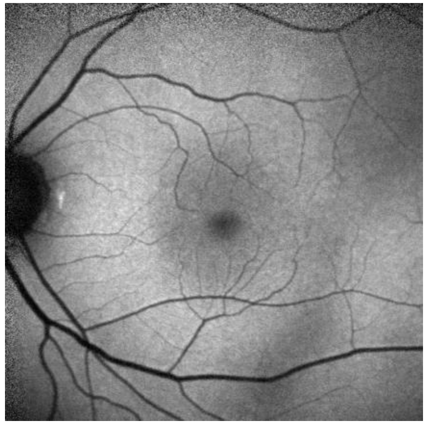
Color fundus (CF)



Infrared reflectance (IR)



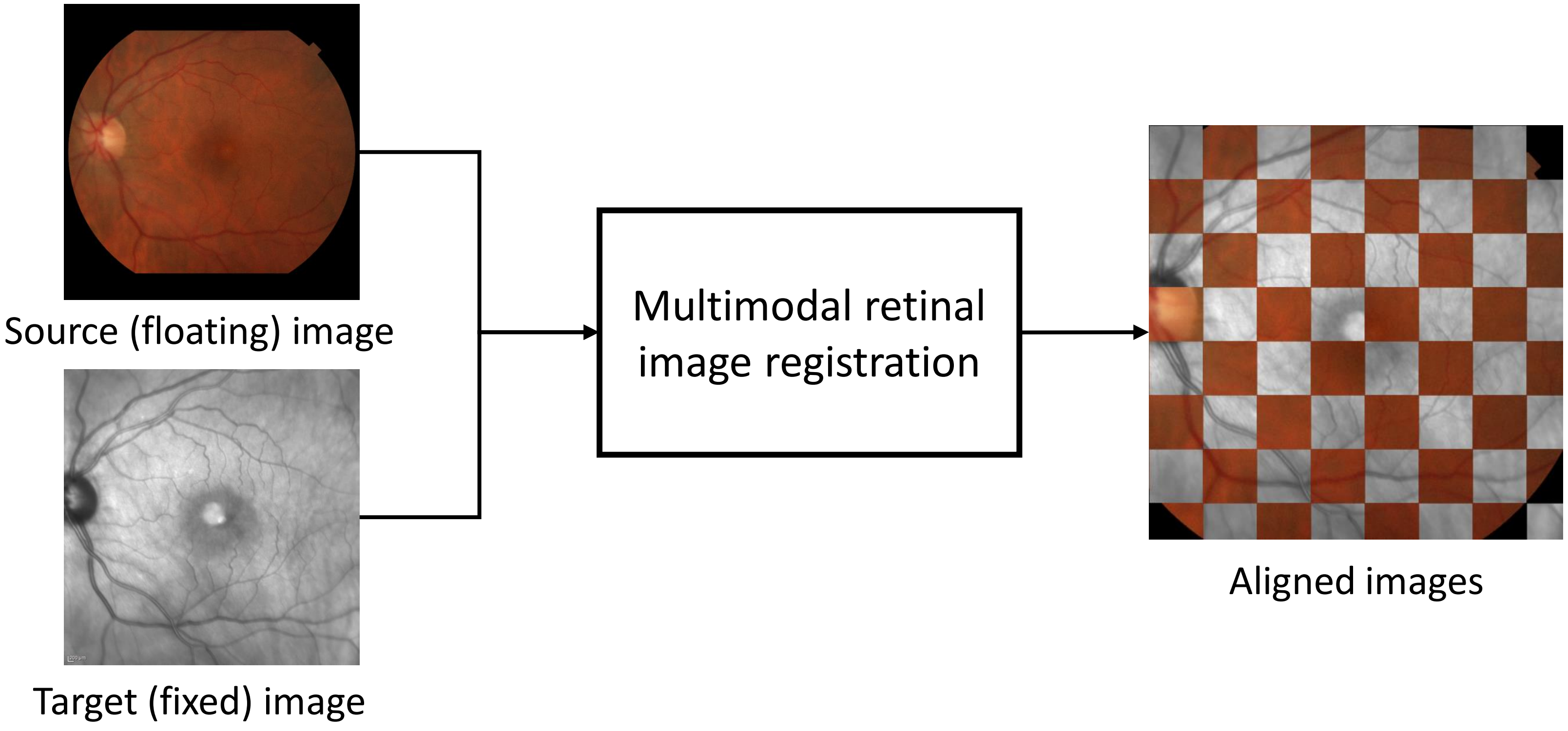
fluorescein angiography (FA)



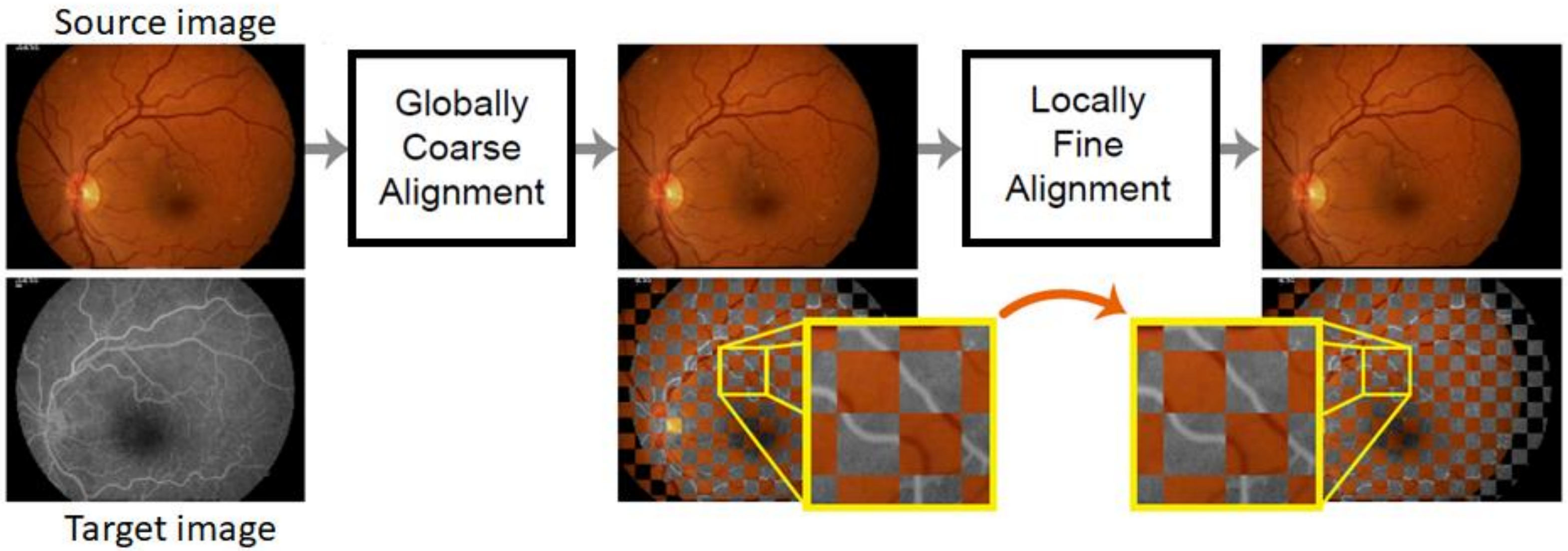
Blue-reflectance (BAF)

Multimodal retinal image registration

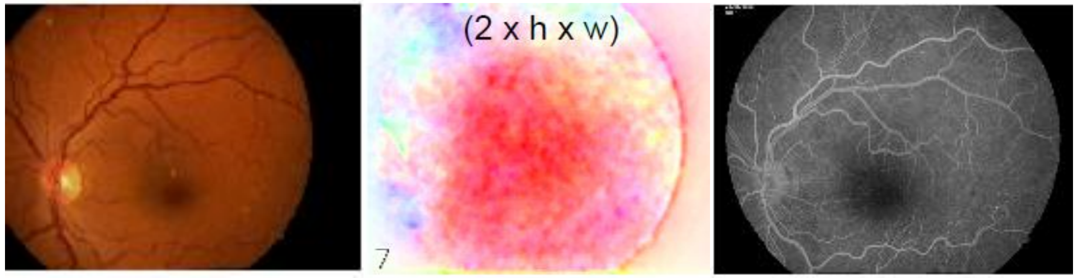
- Two images from different modalities
- Align (warp) source image to target image



General registration pipeline



The coarse-to-fine pipeline for multi-modal retinal image registration[13]



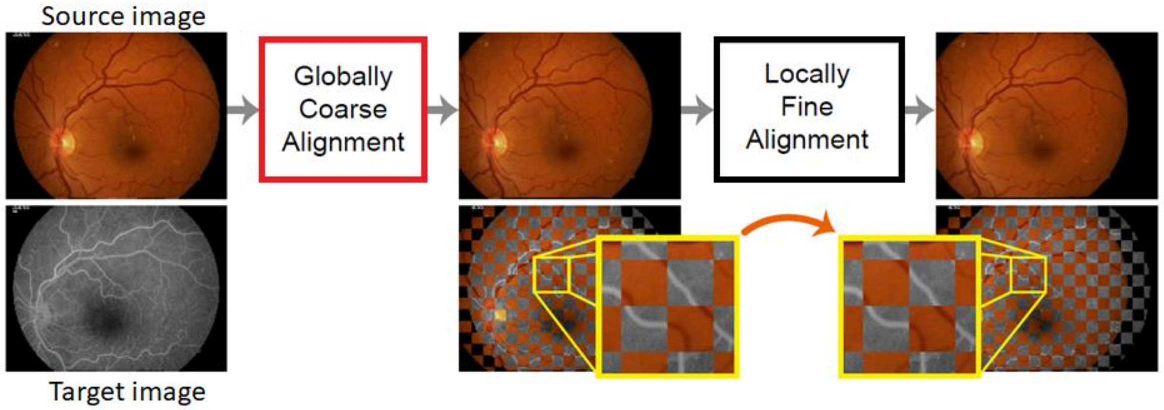
(a) source (b) Registration field (c) target

Deformable registration field [13]

[13] J. Zhang et al, "Joint vessel segmentation and deformable registration on multi-modal retinal images based on style transfer," in ICIP 2019

Research goal

- If the coarse alignment step is successful, the fine alignment step can improve accuracy
- However, if the coarse alignment completely fails / is too far away from ground-truth alignment, the fine alignment step cannot correct the previous result
- Therefore, improving the coarse alignment step is crucial to increase the success rate
- **Research Goal**: Design a **coarse alignment method** that is **robust** & **accurate**



Contents

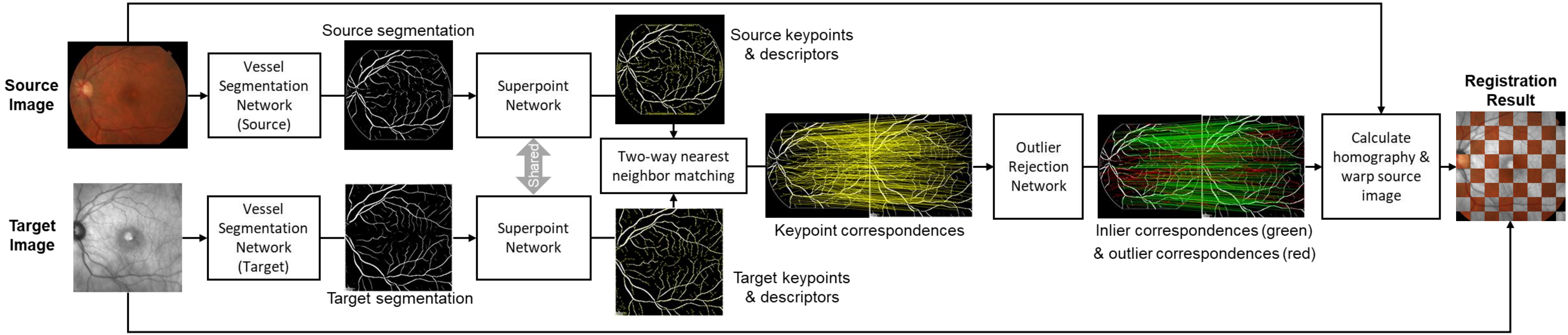
Related works

- **Area-based**: degrades for small overlap, rely on intensity, not suitable for multimodal
- **Feature-based**: detect feature points and find point correspondences
  - **Vessel extraction**: edge detection[5], mean phase image[2], vessel segmentation[6,12,13];
    - DRIU vessel segmentation network[12], unsupervised vessel segmentation network[13]
  - **Feature detection & description**: SIFT[7], Harris corner[14], HOG[15]; LIFT[17], UCN[18], SuperPoint[19]
  - **Outlier rejection**: LMEDS[20], RANSAC[9]; learned correspondences[24]
- **Learning-based**: using convolutional neural networks (CNN)
  - Multimodal retinal images[13], [26]: focus only on deformable, assume affinely aligned
  - VoxelMorph[25], DLIR[11]: CT/MRI images, single-modal only
  - CNNGeo[10]: multimodal natural image semantic alignment

Contents

Proposed method

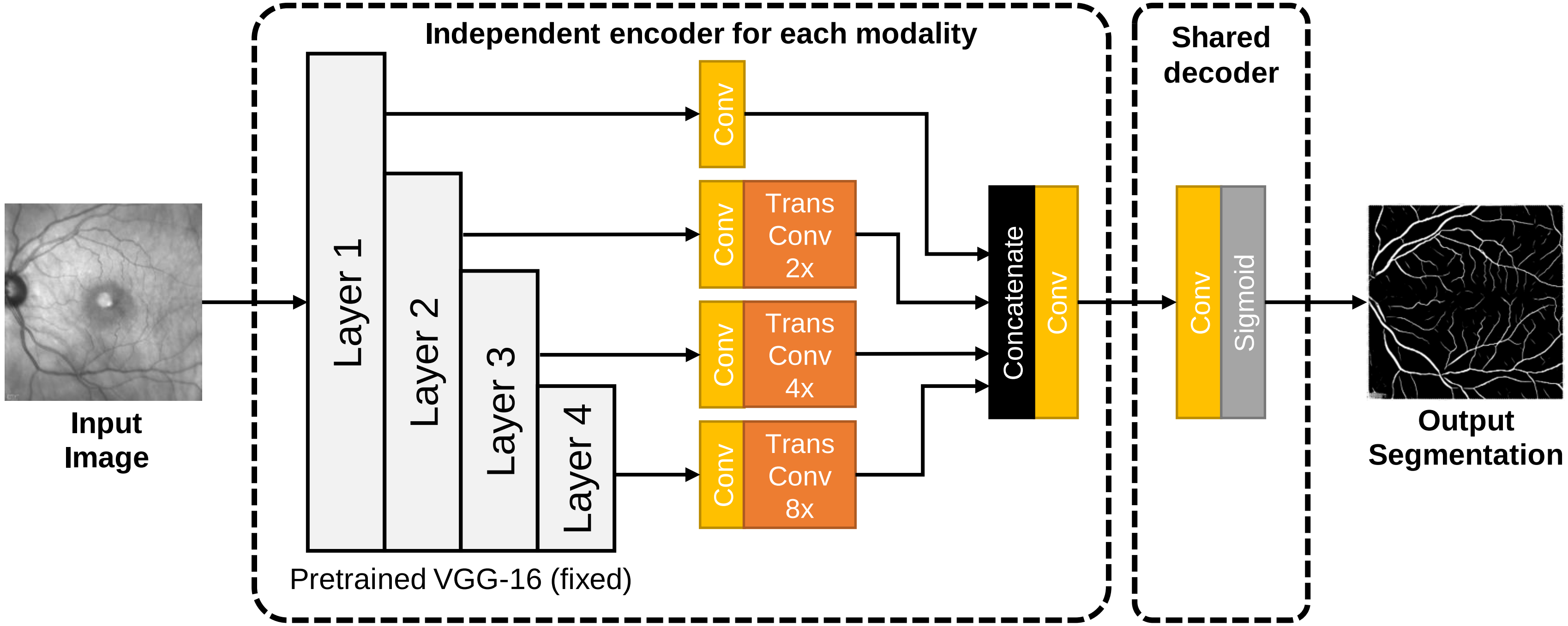
- Three neural networks for vessel segmentation, feature detection and description, and outlier rejection
- The first deep learning framework for multimodal retinal image coarse alignment



Proposed learning-based coarse alignment pipeline

Vessel segmentation network

[13] pretrained network is used

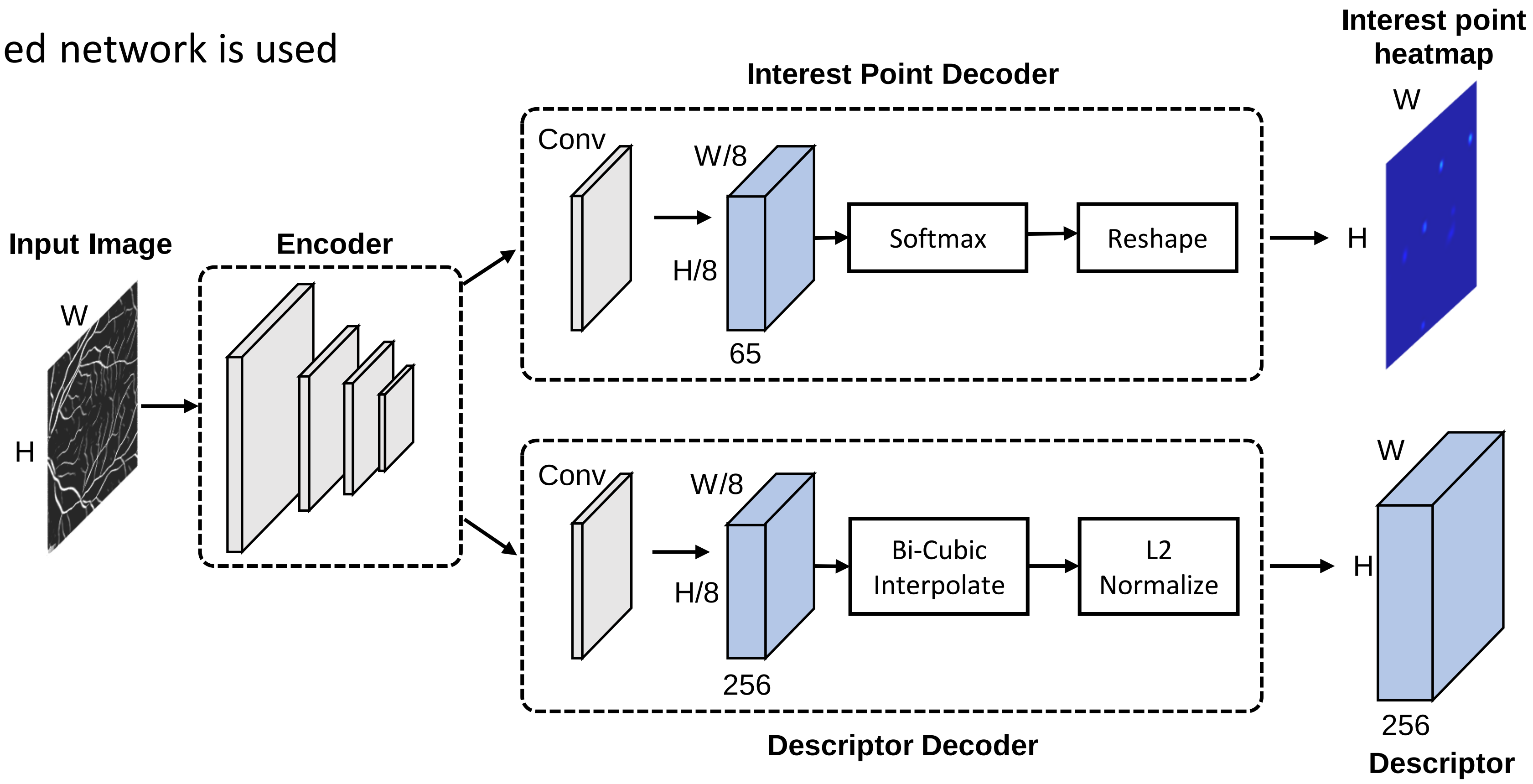


The vessel segmentation network

[13] J. Zhang et al, "Joint vessel segmentation and deformable registration on multi-modal retinal images based on style transfer," in ICIP 2019

SuperPoint network

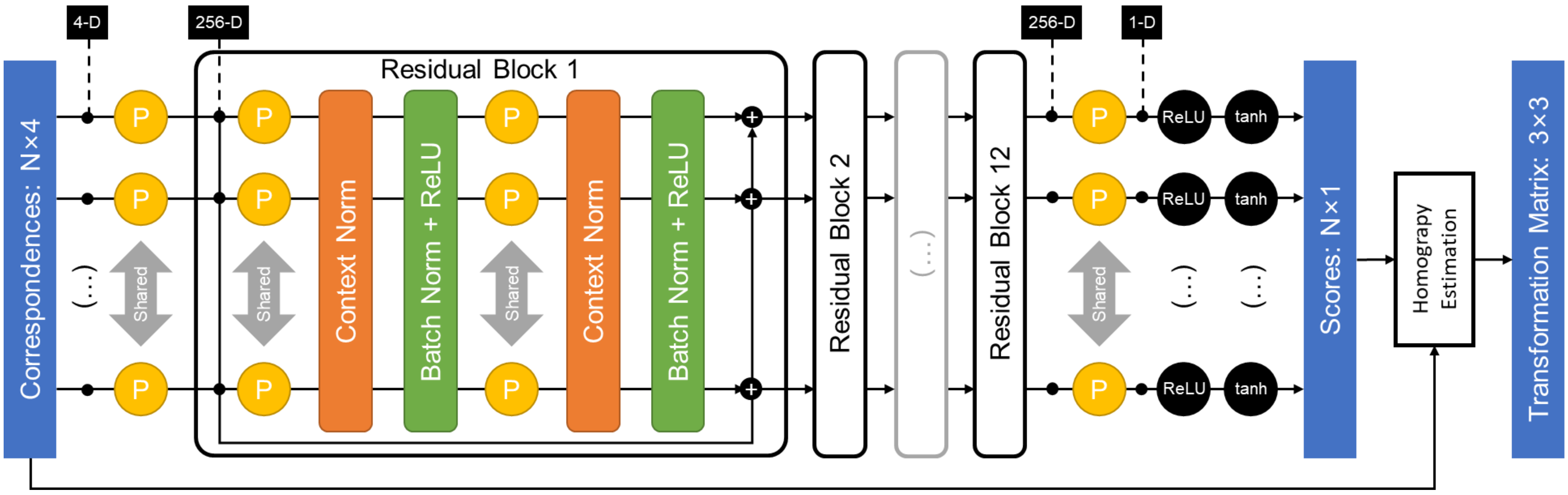
[19] pretrained network is used



The SuperPoint network

[19] D. DeTone et al, "SuperPoint: Self-supervised interest point detection and description," in CVPR 2018

Outlier rejection network



The outlier rejection network

[24] K. M. Yi et. al, "Learning to find good correspondences," in CVPR 2018

Training outlier rejection network

- Classification loss: $\mathcal{L}_{\text{class}}(\mathbf{x}, \mathbf{M}_{gt}) = \frac{1}{N}\sum_{i=1}^{N} \gamma_i H(y_i(\mathbf{M}_{gt}), \sigma(o_i(\mathbf{x})))$ $y_i(\mathbf{M}_{gt}) = \begin{cases} 1, & \text{if } \|T(\mathbf{p}_i, \mathbf{M}_{gt}) - \mathbf{p}'_i\| \leq 5 \text{ pixels} \\ 0, & \text{otherwise} \end{cases}$

- Matrix regression loss: $\mathcal{L}_{\text{matrix}}(\mathbf{x}, \mathbf{M}_{gt}) = \|\mathbf{M}_{gt} - \mathbf{M}(\mathbf{x})\|^2$

- Image registration loss: $\mathcal{L}_{\text{dice}}(\mathbf{x}, \mathcal{I}_{\text{src}}, \mathcal{I}_{\text{tgt}}) = 1 - \text{Dice}_s(\text{warp}(\mathcal{I}_{\text{src}}, \mathbf{M}(\mathbf{x})), \mathcal{I}_{\text{tgt}})$

(Binary) Dice coefficient: $\text{Dice}(\mathcal{I}_1, \mathcal{I}_2) = \frac{2 \times \sum(\mathcal{I}_1 \odot \mathcal{I}_2)}{\sum \mathcal{I}_1 + \sum \mathcal{I}_2}$ Soft Dice coefficient: $\text{Dice}_s(\mathcal{I}_1, \mathcal{I}_2) = \frac{2 \times \sum \text{ele\_min}(\mathcal{I}_1, \mathcal{I}_2)}{\sum \mathcal{I}_1 + \sum \mathcal{I}_2}$

- Total loss:

$$\mathcal{L}(\mathbf{x}, \mathcal{I}_{\text{src}}, \mathcal{I}_{\text{tgt}}, \mathbf{M}) = \lambda_{\text{class}}\mathcal{L}_{\text{class}}(\mathbf{x}, \mathbf{M}) + \lambda_{\text{matrix}}\mathcal{L}_{\text{matrix}}(\mathbf{x}, \mathbf{M}) + \lambda_{\text{dice}}\mathcal{L}_{\text{dice}}(\mathbf{x}, \mathcal{I}_{\text{src}}, \mathcal{I}_{\text{tgt}})$$

Contents

- Introduction
- Related Works
- Proposed Method
- Experiments
- Conclusion

Dataset

- Dataset collected from Jacobs Retina Center at Shiley Eye Institute
  - Source: CF image (RGB, 3000×2672)
  - Target: IR image (grayscale, 768×768 or 1536×1536)
- Training set: 530 pairs, validation set: 90 pairs, test set: 253 pairs
- Ground truth: transformation matrices
  - Manually labeled by selecting point correspondences in each image

Experiments

Dataset: Our test set (253 pairs of CF & IR)

Comparison:

- Conventional method [2]: mean phase image + dense HOG + RANSAC
- CNNGeo [10]: compare only affine registration step, pretrained and finetuned version

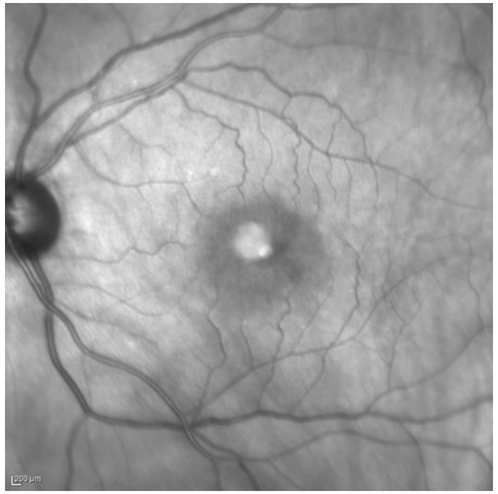
Criteria:

- **Robustness:** Success rate
  - Success registration is determined by the maximum error (MAE) on corresponding landmarks
  - Determine success registration by MAE < 20 pixels
- **Accuracy:** Dice coefficient
  - Our binary segmentation maps (threshold at 0.5)

$$\mathrm{Dice}(\mathcal{I}_1, \mathcal{I}_2) = \frac{2 \times \sum(\mathcal{I}_1 \odot \mathcal{I}_2)}{\sum \mathcal{I}_1 + \sum \mathcal{I}_2}$$

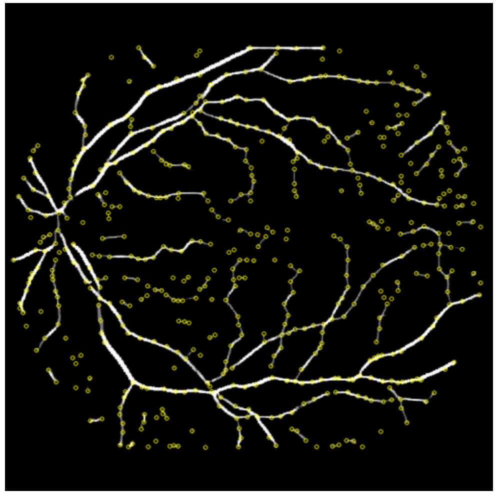
Example pair 1

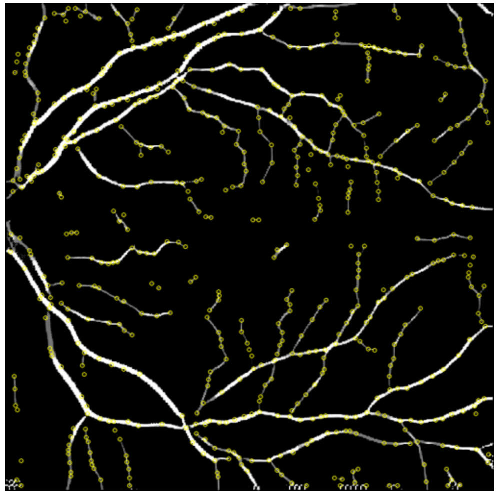

Source image

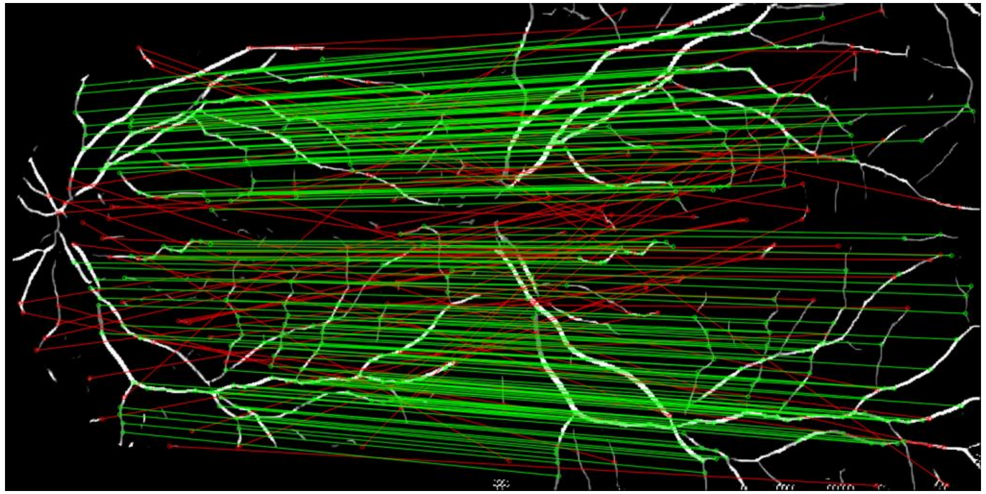

Target image


Source segmentation


Target segmentation

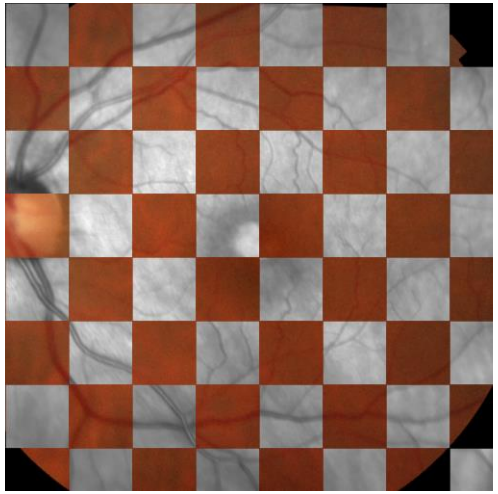

Source keypoints

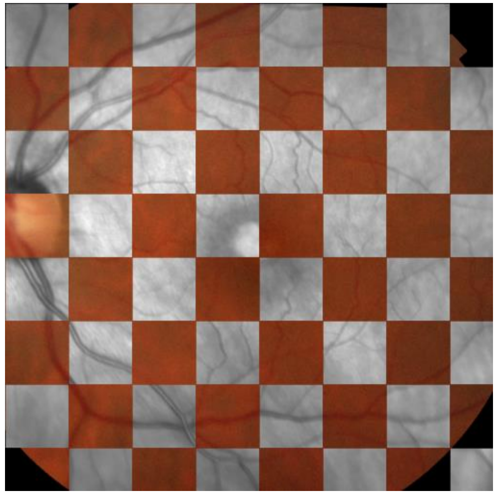

Target keypoints

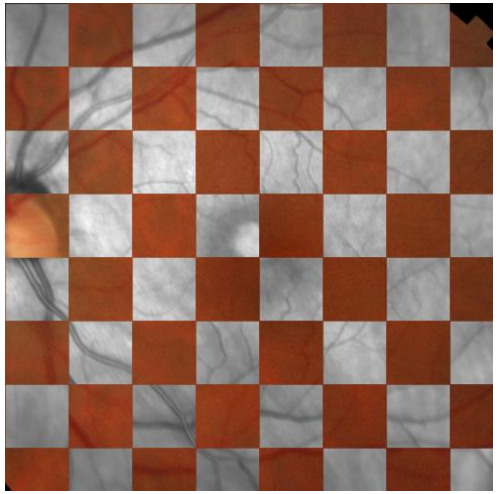

Inlier matches (green) & outlier matches (red)

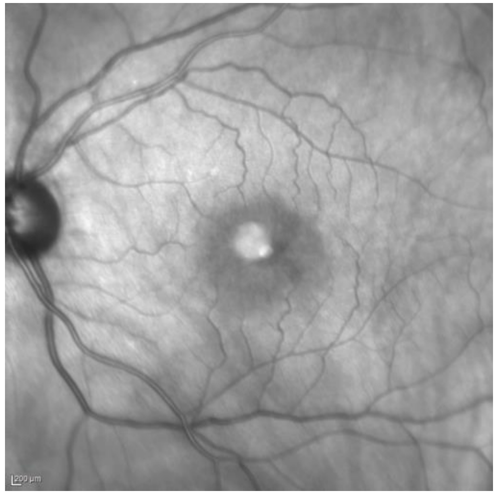
Registration result

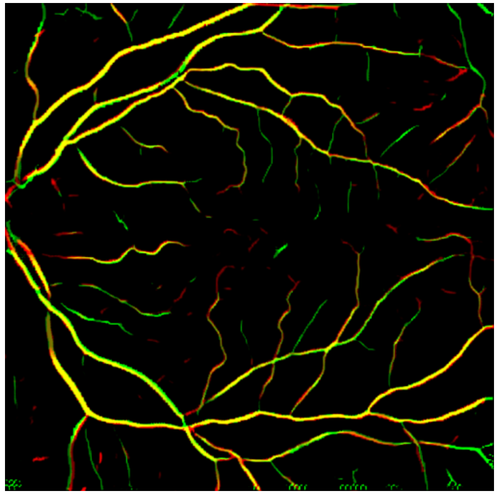

Source image

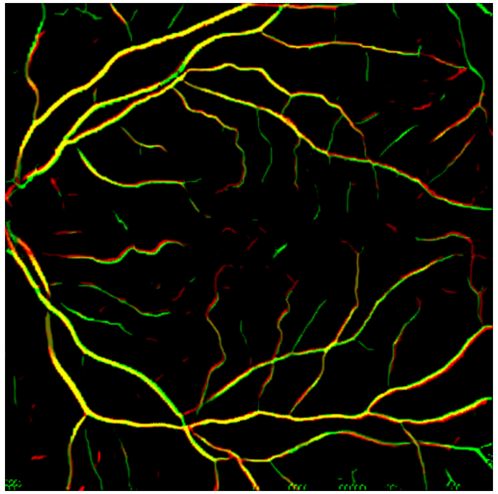

Proposed (MAE=2.2)

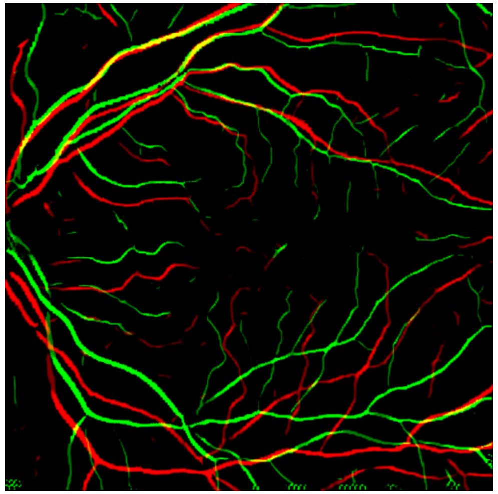

Conventional[2] (MAE=5.0)


CNNGeo[10] (MAE=95.9)


Target image

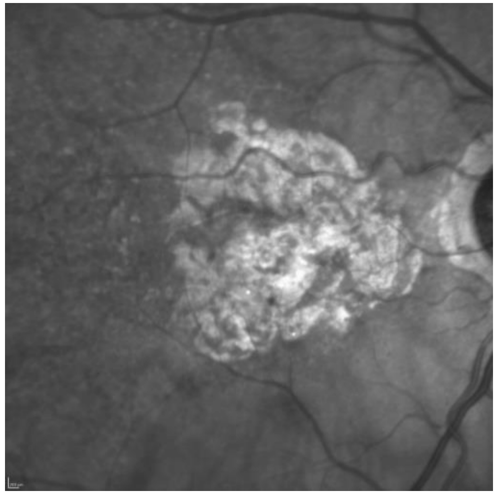

Proposed (Dice=0.7065)


Conventional[2] (Dice=0.6366)

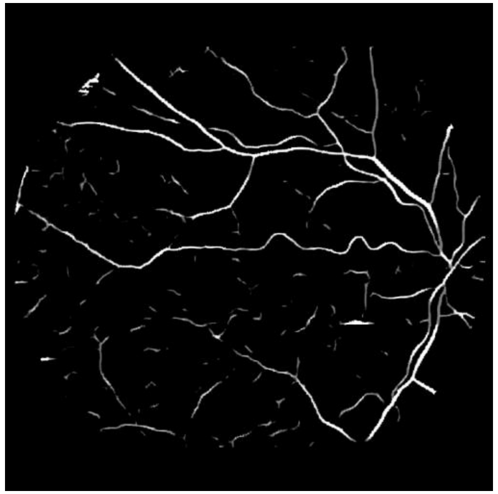

CNNGeo[10] (Dice=0.1295)

Example pair 2

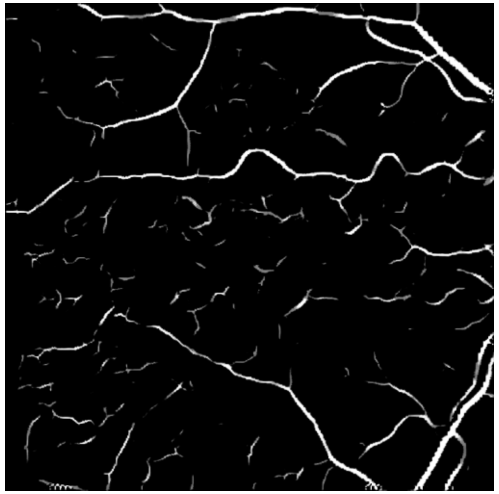
Source image



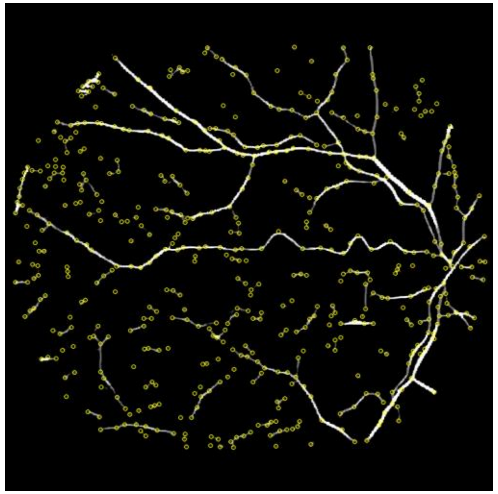
Target image



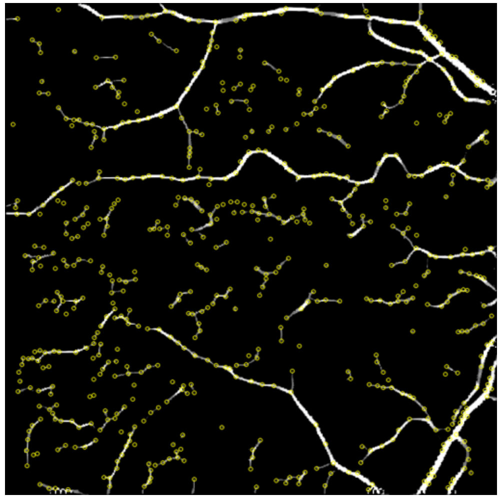
Source segmentation



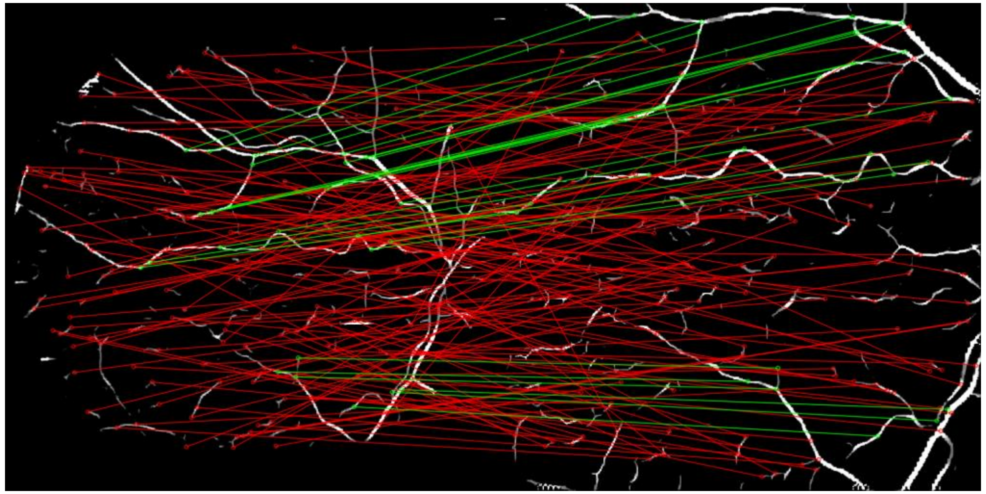
Target segmentation



Source keypoints



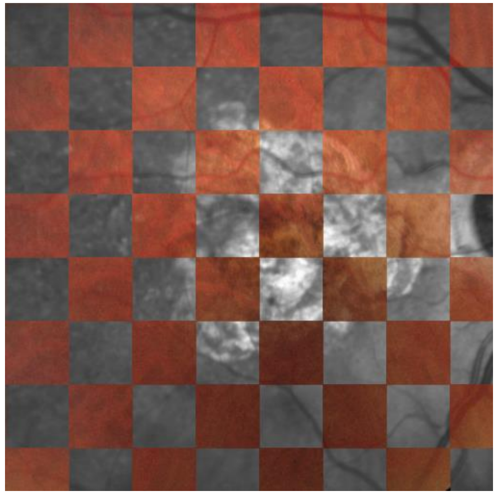
Target keypoints

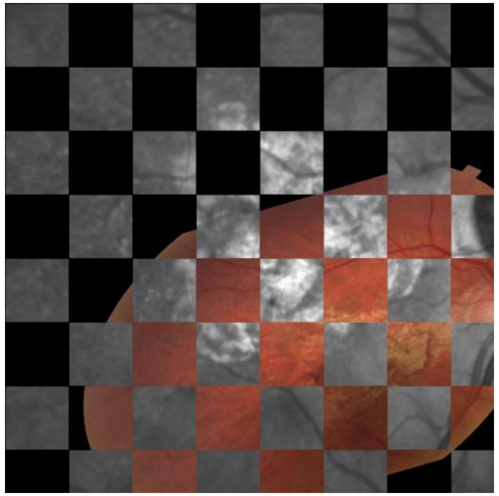


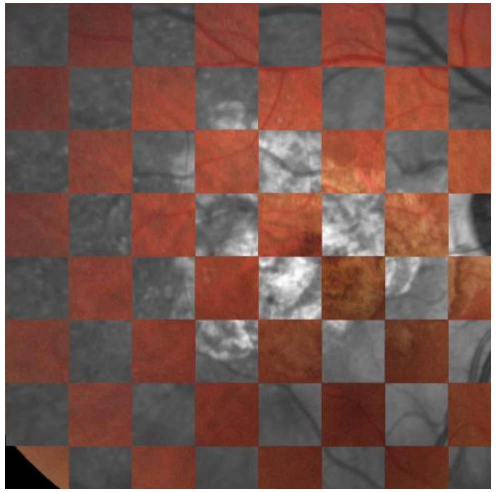
Inlier matches (green) & outlier matches (red)

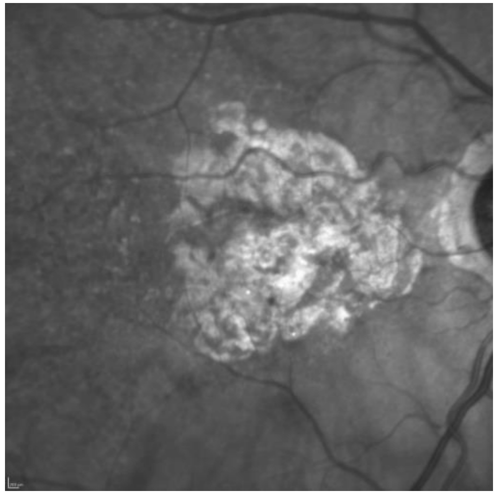
Registration result

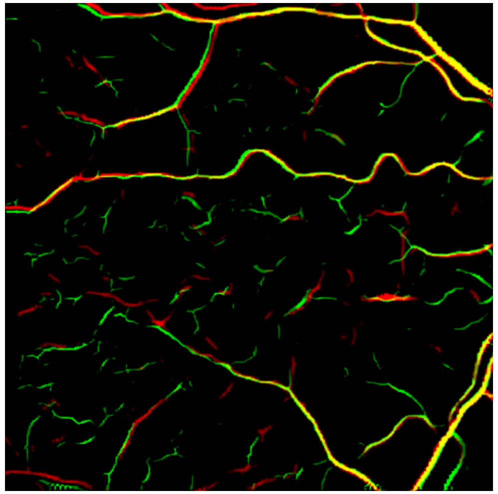

Source image

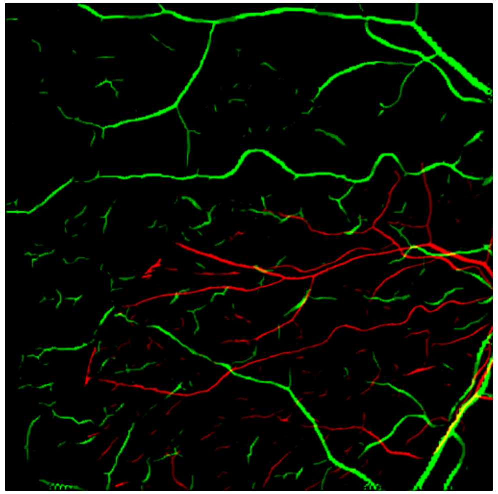

Proposed (MAE=7.1)

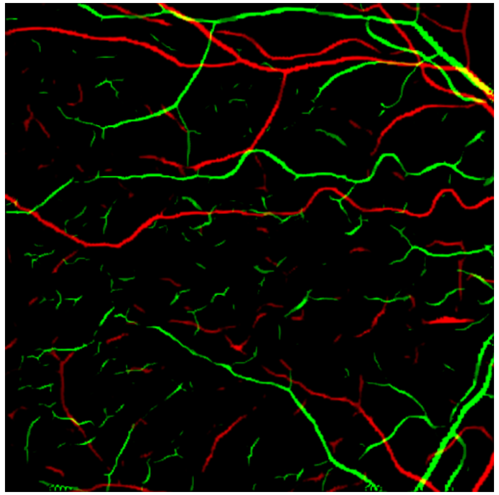

Conventional[2] (MAE=429.7)


CNNGeo[10] (MAE=150.3)


Target image


Proposed (Dice=0.5384)


Conventional[2] (Dice=0.0517)


CNNGeo[10] (Dice=0.0732)

Quantitative result

Table 1: Result using different combinations of algorithms on the test set

Method	Success Rate	Dice coefficient
(a) Phase + HOG + RANSAC (Method [2])	48.22% (122/253)	0.3084 ($\pm$0.2821)
(b) Phase + SuperPoint + RANSAC	79.84% (202/253)	0.4902 ($\pm$0.2304)
(c) Seg. + SuperPoint + RANSAC	85.37% (216/253)	0.4922 ($\pm$0.2162)
(d) Seg. + SuperPoint + OutlierNet (Proposed)	**94.07% (238/253)**	**0.5748 (±0.1796)**

Table 2: Result using different registration methods on the test set

Method	Success Rate	Dice coefficient
Method [2]	48.22% (122/253)	0.3084 ($\pm$0.2821)
CNNGeo [10] pretrained	0.79% (2/253)	0.0677 ($\pm$0.0281)
CNNGeo [10] finetuned	5.13% (13/253)	0.0734 ($\pm$0.0493)
Proposed Method	**94.07% (238/253)**	**0.5748 (±0.1796)**

*Dice coefficient before registration: 0.0399 ($\pm$0.0146)

[2] Z. Li et al, 2018, "Multi-modal and multi-vendor retina image registration," Biomedical optics express
[10] I. Rocco, et. al, "Convolutional neural network architecture for geometric matching," in CVPR 2017

Contents

- Introduction
- Literature Review
- Proposed Method
- Experiments
- Conclusion

Conclusion

- Proposed a deep learning framework for multimodal retinal image registration
- Focused on the globally coarse alignment step
- Vessel segmentation network + SuperPoint network + Outlier rejection network
- Significant improvement in both robustness and accuracy compared to previous conventional / learning-based registration methods in clinical dataset

ICASSP2020 Barcelona

Thank you!

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

SHILEY EYE INSTITUTE
UC SAN DIEGO

UC San Diego
SCHOOL OF MEDICINE

References

[1] T. J. MacGillivray, E. Trucco, J. R. Cameron, B. Dhillon, J. G. Houston, and E. J. R. Van Beek, “Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions,” The British journal of radiology, vol. 87, no. 1040, pp. 20130832, 2014.

[2] Z. Li, F. Huang, J. Zhang, B. Dashtbozorg, S. Abbasi Sureshjani, Y. Sun, X. Long, Q. Yu, B. t. Haar Romeny, and T. Tan, “Multi-modal and multi-vendor retina image registration,” Biomedical optics express, vol. 9, no. 2, pp. 410–422, 2018.

[3] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, “Registration of stereo and temporal images of the retina,” IEEE Transactions on medical imaging, vol. 18, no. 5, pp. 404–418, 1999.

[4] T. Chanwimaluang, G. Fan, and S. R. Fransen, “Hybrid retinal image registration,” IEEE transactions on information technology in biomedicine, vol. 10, no. 1, pp. 129–142, 2006.

[5] J. A. Lee, J. Cheng, B. H. Lee, E. P. Ong, G. Xu, D. W. K. Wong, J. Liu, A. Laude, and T. H. Lim, “A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1046–1053.

[6] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. t. Haar Romeny, “Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores,” IEEE transactions on medical imaging, vol. 35, no. 12, pp. 2631–2644, 2016.

[7] D. G. Lowe et al., “Object recognition from local scale invariant features.,” in ICCV, 1999, vol. 99, pp. 1150–1157.

[8] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in European conference on computer vision. Springer, 2006, pp. 404–417.

[9] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” Communications of the ACM, vol. 24, no. 6, pp. 381–395, 1981.

[10] I. Rocco, R. Arandjelovic, and J. Sivic, “Convolutional neural network architecture for geometric matching,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6148–6157.

References

[11] B. D. d. Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Iˇsgum, “A deep learning framework for unsupervised affine and deformable image registration,” Medical image analysis, vol. 52, pp. 128–143, 2019.

[12] K.-K. Maninis, J. Pont-Tuset, P. Arbel´aez, and L. Van Gool, “Deep retinal image understanding,” in International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 140–148.

[13] J. Zhang, C. An, J. Dai, M. Amador, D.-U. Bartsch, S. Borooah, W. R. Freeman, and T. Q. Nguyen, “Joint vessel segmentation and deformable registration on multi-modal retinal images based on style transfer,” in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 839–843.

[14] C. G. Harris, M. Stephens, et al., “A combined corner and edge detector.,” in Alvey vision conference. Citeseer, 1988, vol. 15, pp. 10–5244.

[15] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” 2005.

[16] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, “A partial intensity invariant feature descriptor for multimodal retinal image registration,” IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1707–1718, 2010.

[17] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, “Lift: Learned invariant feature transform,” in European Conference on Computer Vision. Springer, 2016, pp. 467–483.

[18] C. B. Choy, J. Gwak, S. Savarese, and M. Chandraker, “Universal correspondence network,” in Advances in Neural Information Processing Systems, 2016, pp. 2414–2422.

[19] D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-supervised interest point detection and description,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 224–236.

[20] P. J. Rousseeuw, “Least median of squares regression,” Journal of the American statistical association, vol. 79, no. 388, pp. 871–880, 1984.

References

[21] H. Zhang, X. Liu, G. Wang, Y. Chen, and W. Zhao, “An automated point set registration framework for multimodal retinal image,” in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 2857–2862.

[22] Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, “An efficient approach for robust multimodal retinal image registration based on ur-sift features and piifd descriptors,” EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, pp. 25, 2013.

[23] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother, “Dsac-differentiable ransac for camera localization,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6684–6692

[24] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, “Learning to find good correspondences,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2666–2674.

[25] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: a learning framework for deformable medical image registration,” IEEE transactions on medical imaging, 2019.

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

[27] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in European conference on computer vision. Springer, 2016, pp. 694–711.