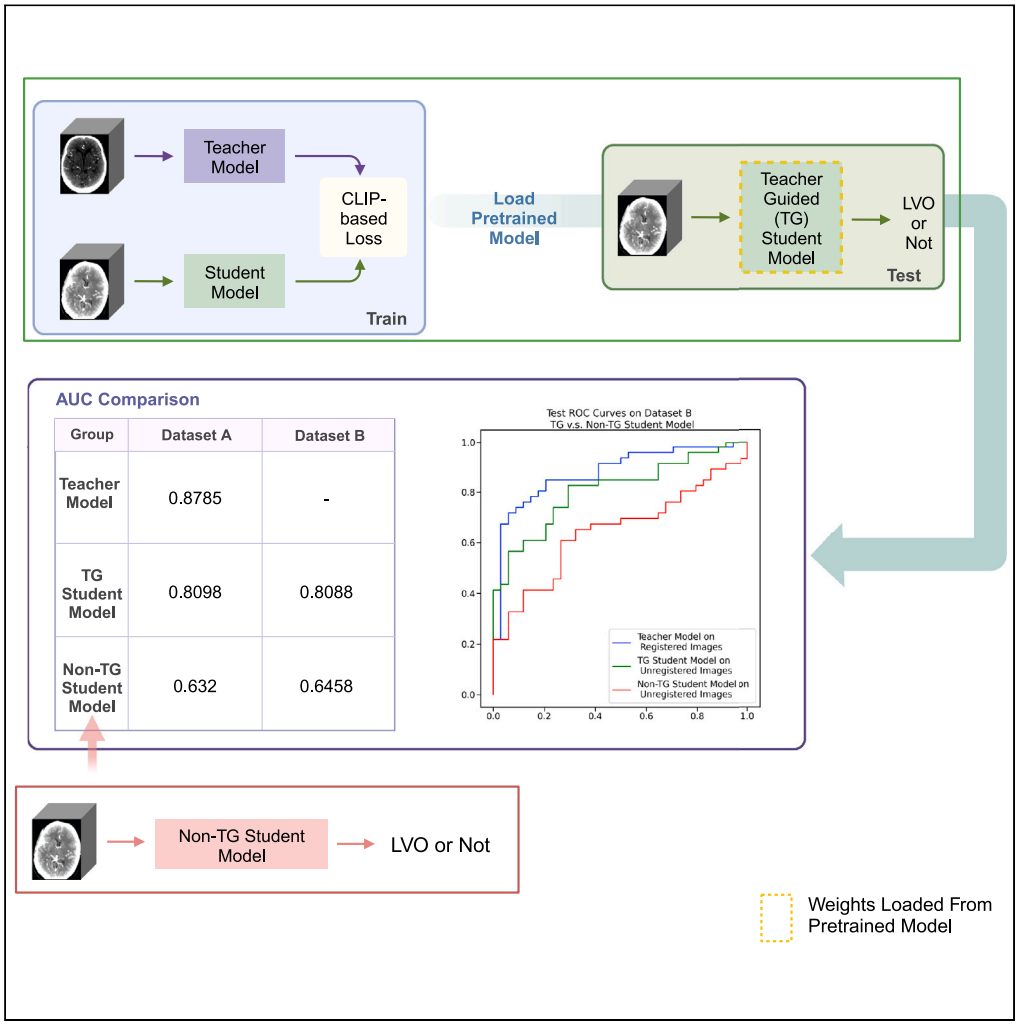
iScience

CellPress
OPEN ACCESS

Article

A self-supervised learning approach for registration agnostic imaging models with 3D brain CTA



Yingjun Dong, Samiksha Pachade, Xiaomin Liang, Sunil A. Sheth, Luca Giancardo

luca.giancardo@uth.tmc.edu

Highlights

Learning from a pre-trained teacher model and adapt knowledge to unregistered images

Knowledge distillation process does not require use of image labels

CLIP-based paradigm outperforms other contrastive learning strategies

We did not detect instances of "shortcut learning"

Dong et al., iScience *27*, 109004
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109004

iScience

CellPress
OPEN ACCESS

Article

A self-supervised learning approach for registration agnostic imaging models with 3D brain CTA

Yingjun Dong,[1] Samiksha Pachade,[1] Xiaomin Liang,[1] Sunil A. Sheth,[2,4] and Luca Giancardo[1,3,4,5,*]

SUMMARY

Deep learning-based neuroimaging pipelines for acute stroke typically rely on image registration, which not only increases computation but also introduces a point of failure. In this paper, we propose a general-purpose contrastive self-supervised learning method that converts a convolutional deep neural network designed for registered images to work on a different input domain, i.e., with unregistered images. This is accomplished by using a self-supervised strategy that does not rely on labels, where the original model acts as a teacher and a new network as a student. Large vessel occlusion (LVO) detection experiments using computed tomographic angiography (CTA) data from 402 CTA patients show the student model achieving competitive LVO detection performance (area under the receiver operating characteristic curve [AUC] = 0.88 vs. AUC = 0.81) compared to the teacher model, even with unregistered images. The student model trained directly on unregistered images using standard supervised learning achieves an AUC = 0.63, highlighting the proposed method's efficacy in adapting models to different pipelines and domains.

INTRODUCTION

Deep learning approaches have become the go-to method for automatically classifying medical conditions in brain images. As opposed to natural images, brain images, regardless of their image modality have to go through multiple pre-processing steps before using them as input to a deep learning approach. One of the most critical ones is image registration. This involves aligning the brain images to a common space, such as the Montreal Neurological Institute (MNI) template or a subject-specific template. This step is important to account for variations in brain size and shape across different subjects and to ensure that these changes are not used as "shortcuts" by the deep learning algorithm to perform the classification. Apart from adding additional computation, image registration is one of the most common points of failure which is typically addressed by adding manual or automatic quality assessment requirements to the pipeline. These drawbacks are particularly significant for acute stroke applications, where all interventions need to be extremely fast, and any minute lost will negatively affect the outcomes of patients.

Large vessel occlusion (LVO) is defined as vessel blockages of intracranial internal carotid artery (ICA), anterior cerebral arteries (ACA) A1, ACA A2, middle cerebral arteries (MCA) M1, MCA M2, and posterior cerebral arteries (PCA), which amounts up to 46% of acute ischemic strokes. 3D computed tomographic angiography (CTA) has been proven as an efficient and more precise way of medical imaging analysis for LVO detection.[1] Note that other imaging modalities, such as diffusion-weighted magnetic resonance imaging or CT-perfusion are used for other aspects of acute stroke care, however, for LVO detection CTA is the main diagnostic modality. In this work, we introduce a method based on a teacher-student model structure using unregistered 3D CTA without extra image pre-processing steps on LVO detection.

In recent years, contrastive learning has played a pivotal role in self-supervised learning approaches.[2–4] Contrastive learning methods attempt to maximize the similarities between positive pairs and minimize the differences between negative pairs, and by doing so learn valid semantic representations without the need for labels. Contrastive learning is a major contributor to the development of medical imaging, due to the high cost and insufficient of labeling medical images.

Automatic LVO detection with the deep learning method has been studied before. Olive-Gadea et al.[5] applied deep learning-based software named MethinksLVO to detect LVO on brain CTA, as their results showed, Methinks software works well in LVO detection in patients who are suspected to be diagnosed with acute ischemic stroke. Barman et al.[6] demonstrated a deep symmetry-sensitive CNN on brain CTA to investigate the changes between brain hemispheres for acute ischemic stroke detection. Czap et al.[7] investigated a deep learning method to detect LVO on a mobile stroke unit CTA, their work showed an efficient and accurate approach to LVO detection for prehospital patients, which shows practical advantage in clinical studies. Other works for LVO detection with deep convolutional neural networks (DCNN), include

[1]McWilliams School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin Street, Houston, TX, USA
[2]Department of Neurology, McGovern Medical School, University of Texas Health Science Center at Houston, 6431 Fannin Street, Houston, TX, USA
[3]Institute for Stroke and Cerebrovascular Diseases, University of Texas Health Science Center at Houston, 7000 Fannin Street, Houston, TX, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: luca.giancardo@uth.tmc.edu
https://doi.org/10.1016/j.isci.2024.109004

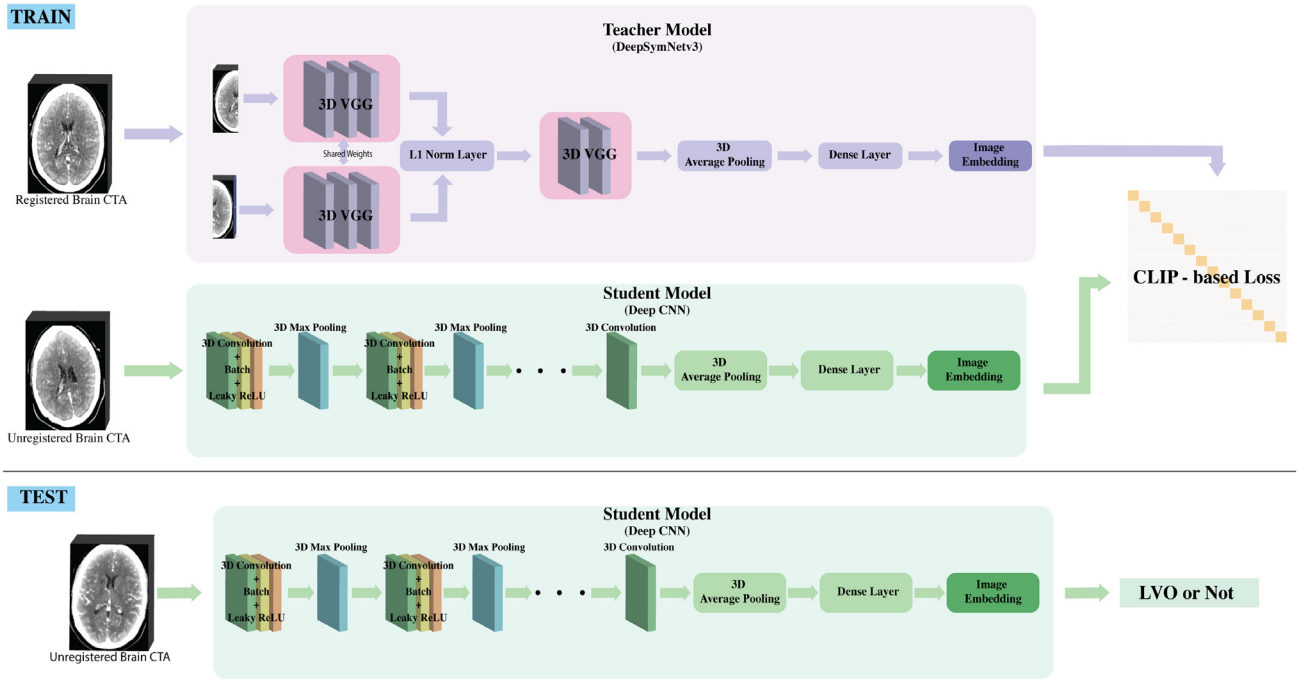


Figure 1. A pipeline of the proposed method

Structure of teacher-student model experiments, we utilized a CLIP-based loss strategy on pairs of registered and unregistered 3D CTA.

the work of Stib et al.,[8] Luijten et al.,[9] and Czap et al.[10] These works showed competitive performance in LVO detection using deep learning methods; however, they still required labeled, registered, and quality-checked data. In these works, imaging registration is one of the key preprocessing steps and potential source of error. In our study, we propose a teacher-student model based on self-supervised contrastive learning that utilizes unregistered and no quality checked images to break through the bottleneck.

Self-supervised contrastive learning has been proven for a useful method in medical imaging studies. Azizi et al.[11] introduced a self-supervised model of medical imaging classification. Their work has 3 steps, firstly, they conducted self-supervised learning on unlabeled natural images, then they applied a self-supervised model on unlabeled medical images, and they conducted fine-tuning on labeled medical images in the last step. Their work showed the efficiency and reliability of a self-supervised model on medical images. Taleb et al.[12] conducted a self-supervised multimodal contrastive learning on retinal fundus images and genetic data which is named ContIG. Their work showed considering genetic data in imaging models could improve the performance of image models.

After the contrastive language-image pretraining (CLIP) model was published, some works applied CLIP-based strategies to medical data. Wang et al.[13] developed a CLIP-based model using prior medical knowledge of unpaired images and reports named MedCLIP. Their proposed method could improve the classification performance with fewer datasets than state-of-the-art methods. Tiu et al.[14] also applied CLIP-based strategies on unlabeled chest X-ray images for self-supervised classification tasks. They pre-trained the image-report model and then utilized prompts and images for zero-shot classification. However, those works focus on 2D images. Our proposed study utilized 3D images with small batch sizes.

In this paper, we propose a general-purpose contrastive self-supervised learning method that converts a convolutional deep neural network designed for registered images to work on a different input domain, i.e., with unregistered images. The proposed pipeline is shown in Figure 1. During training, a teacher model pre-trained to classify LVO on registered images is used to guide a student model (with a different architecture) to learn a similar feature representation on unregistered brain CTA. The teacher model guides the student model by the modified CLIP loss without any explicit use of the image labels. In the testing phase, we only utilize unregistered CTA for LVO detection with the student model without any further fine-tuning. One example slice from the registered CTA and unregistered CTA of two patients is shown in Figure 2. In the rest of the paper, we refer to the student model trained in this manner as the teacher-guided (TG) student model. The model trained with a classic supervised loss on the LVO labels is referred to as the non-teacher-guided (non-TG) student model. The demographic information for subjects we used in this work is listed in Table 1.

RESULTS

To demonstrate the performance of our proposed method, we conducted a comparative analysis of the area under the receiver operating characteristic curve (AUC) scores for three different models: our proposed TG student model, the non-TG student model, and the teacher model itself.

While the TG student model utilizing a teacher-student structure, is designed to learn key features for LVO detection from registered 3D CTA images and apply the acquired knowledge to detect LVO using unregistered 3D CTA images without explicitly relying on the labels,

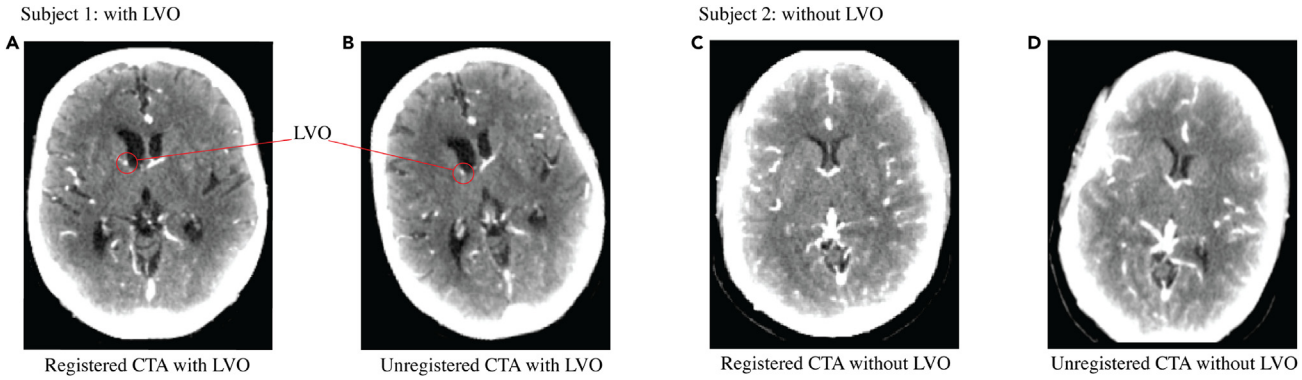


Figure 2. Axial slices from 3D CTA volumes

(A) shows a registered CTA image for a subject with LVO; (B) shows an unregistered CTA image for the same subject as a; (C) shows a registered CTA image for a subject without LVO; (D) shows an unregistered CTA image for the same subject as c.

the non-TG student model employs a standard the supervised learning method, and labels, on unregistered 3D CTA images for LVO detection.

To clarify, the non-TG student model is trained using supervised learning with unregistered images. We conducted testing experiments on the same dataset with all of the models. We implemented the testing experiments on the two datasets previously described. For Dataset A, each subject has an individual transformation, and for Dataset B, we performed 50 different transformations, but for each iteration, all of the subjects have the same transformation.

The main metric used in the experiments is AUC, which is the most common evaluation metric for LVO detection classifiers.[1,7,15] A higher AUC score indicates better performance.

The performance comparisons for different methods are shown in Table 2 and Figure 3. From Table 2 the AUC on the test dataset from the TG student model which utilized DCNN as an unregistered image encoder is 0.8098, the AUC from the teacher model which applied DeepSymNetv3[16] as a registered image encoder is 0.8785, and the AUC from the non-TG student model using DCNN for unregistered image encoder is 0.632. The TG student model shows better results compared with the non-TG student model, which means our proposed teacher-student structure works better than the common supervised method. As expected, the TG student model does not outperform the teacher model. Based on Table 2, no matter whether TG or non-TG, DCNN as an unregistered image encoder shows better results than Vision Transformer (ViT).

In Table 2, we compared three different contrastive strategies to perform the TG training of the student model. In addition to CLIP, we tested two other recent strategies: SimSiam[17] and SimCLR.[18] The results indicate that CLIP outperforms both SimSiam and SimCLR.

In addition, Table 3 shows the statistical significance of the changes in probabilities between the experiments. p values reported are computed with a Mann-Whitney U test to reject the null hypothesis that the output of the two models compared are part of the same distribution.

Computational time

Furthermore, we conducted a comparison of computational time between using registered and unregistered images, considering the efficiency requirements for clinical implementations. Utilizing registered images requires additional processing steps for image registration compared to unregistered images. We performed registration on 11 subjects, with an average registration time of 16.9 ($\pm$ 5.1) seconds per subject. This registration time can be avoided by using unregistered images, enabling quicker response and early intervention in acute stroke diagnosis cases.

DISCUSSION

In this study, we propose a knowledge distillation approach to enable a student model not only to learn from a pre-trained teacher model but also to adapt this knowledge using unregistered images. Registration is an essential pre-training step for the teacher model.

Utilizing unregistered images as input is a significant advantage for acute stroke application and LVO detection, as these systems are typically used to prioritize urgent cases and send alerts to the stroke team in case there is a need for urgent intervention. As such any additional

Table 1. Subjects demographic information

N	79.85%(321) LVO=yes, 20.15%(81) LVO=no
Age, mean(STD)	65.57 (14.94)
Gender	50.25%(202) male, 49.75%(200) female
Race	36.57%(147) White, 28.61%(115) Black or African American, 25.12%(101) Asian, 9.7%(39) other

Table 2. AUC on different methods

Group	Image Encoder	Loss	Dataset A AUC	Dataset B AUC
Teacher model on registered images	DeepSymNetv3	CLIP	0.8785	–
TG student model on unregistered images	DCNN	CLIP	0.8098	0.8088
Non-TG student model on unregistered images	DCNN	CLIP	0.632	0.6458
TG student model on unregistered images	ViT	CLIP	0.5729	0.4386
Non-TG student model on unregistered images	DCNN	CLIP	0.632	0.6458
Non-TG student model on unregistered images	ViT	CLIP	0.4687	0.4789
TG student model on unregistered images	DCNN	SimSiam	0.6477	0.6311
TG student model on unregistered images	DCNN	SimCLR	0.5632	0.5921

minute spent in pre-processing is detrimental to the patient's health. In addition, the need to use registered images as input for clinical applications would require an additional quality assurance step to make sure that no processing errors were made.

While our work takes inspiration from the knowledge distillation paradigm, there are significant differences from the typical knowledge distillation studies. We did not distill a complex model into a simple one to make it more efficient or use fewer parameters, but rather having a general-purpose 3D deep learning network without pre-training (the student) distill the knowledge from a pre-trained model with a custom architecture specialized for acute stroke neuroimaging (the teacher) and at the same time changing the image domain, i.e., going from registered to unregistered images.

As we discussed in Section *2*, the TG student model proposed performs significantly better on unregistered images compared with the non-TG student model, no matter which unregistered image encoder was used. The largest improvement was achieved with a relatively simple DCNN as a student model which led to improvements of around 0.17 points in AUC.

We compared the loss used in our strategy, i.e., CLIP, with other popular contrastive learning losses: SimSiam and SimCLR. CLIP significantly outperformed both of them. The most likely reason is that both SimSiam and SimCLR do not take do not attempt to minimize the cosine similarity (or maximize the distance) between registered/unregistered pairs coming from different subjects, as opposed to the CLIP loss.

While our strategy allows, in principle, to use any network architecture as a student model, larger student models will still be much more data hungry. This is apparent in our experiments using ViT, which is known for having excellent generalizability performance but only if pre-trained with a larger amount of data than DCNN models. In fact, our teacher-student approach improved the performance of the baseline ViT model (non-TG Student), but the AUC improvement obtained (0.57 vs. 0.46) is significantly less than the one obtained from on the model with the DCNN architecture (0.81 vs. 0.63).

In the development of this approach, we were cognizant of the risks of the student model learning shortcuts for the LVO classification by using the type of "misalignments." This is why we generated the datasets A and B. A model that uses transformation as a shortcut for the LVO prediction, would work well on dataset A, but perform poorly in dataset B. Our best-performing model, the student model based on the DCNN architecture, did not exhibit this behavior, giving us confidence of the generalizability of the LVO prediction.

In conclusion, this paper introduces a strategy that enables the conversion of a convolutional deep neural network originally designed for registered images to function effectively with unregistered images. By employing a self-supervised strategy that eliminates the need for labeled LVO data, our method employs the original model as a teacher and a new network as a student, facilitating the transfer of knowledge without relying on explicit labels.

Our results reveal that the trained model exhibits competitive LVO detection performance compared to the teacher model, even when handling unregistered images. In contrast, training the same student model directly on unregistered images using a standard supervised cross-entropy loss yields a significantly lower AUC. These findings underscore the potential of our training strategy to enhance the adaptation of existing models to different pre-processing pipelines and potentially other domains, surpassing the performance achieved by training models from scratch using standard supervised approaches.

Table 3. Methods performance comparison evaluated with Mann-Whitney U test

Methods	p value
TG student model on unregistered images DCNN vs. non-TG student model on unregistered images DCNN	9e-28
TG student model on unregistered images ViT vs. non-TG student model on unregistered images ViT	0.3e-28
Non-TG student model on unregistered images ViT vs. non-TG student model on unregistered images DCNN	9e-29
TG student model on unregistered images CLIP loss vs. TG student model on unregistered images SimSiam	9e-28
TG student model on unregistered images CLIP loss vs. TG student model on unregistered images SimCLR	9e-28
TG student model on unregistered images SimSiam vs. TG student model on unregistered images SimCLR	9e-28

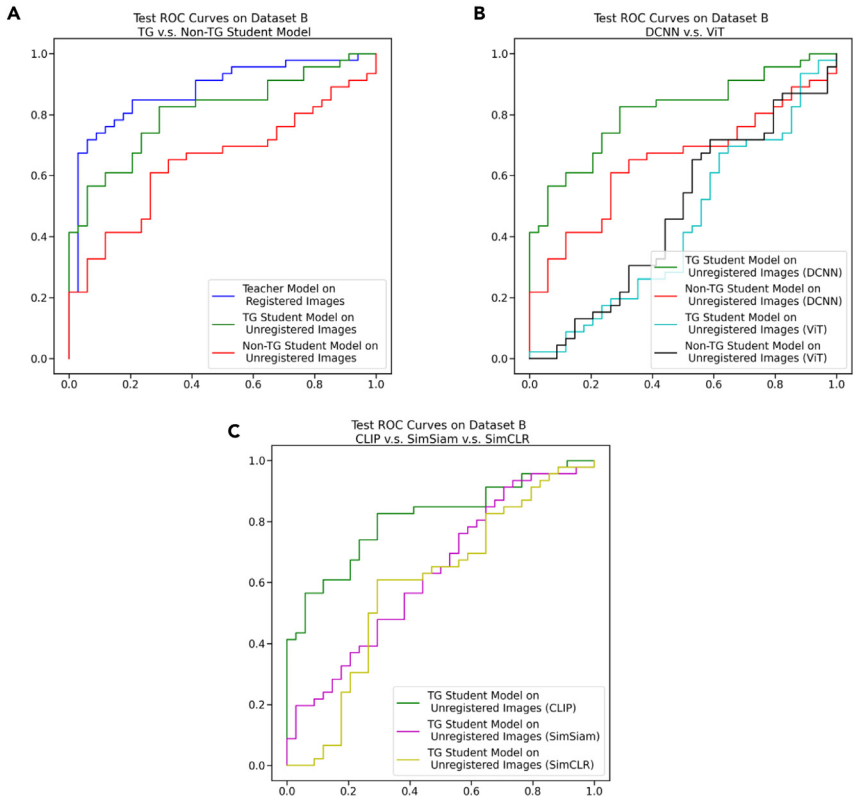


Figure 3. Receiver operating characteristic (ROC) curves comparisons

(A) shows ROC curves comparison on teacher model on registered images, teacher model guided student model on unregistered images, and non-teacher model guided student model on unregistered images; CLIP loss was used and DCNN worked as an unregistered image encoder; (B) shows ROC curves comparison on different unregistered image encoders. We utilized DCNN and ViT as unregistered image encoders. CLIP loss was used; (C) shows ROC curves comparison on different contrastive learning loss, we compared CLIP loss, SimSam loss, and SimCLR loss. DCNN was used as unregistered image encoders.

For future work, it would be valuable to explore the applicability of our proposed method to additional medical domains and datasets. Additionally, investigating the extension of this approach to other types of neural network architectures and evaluating its performance on larger and more diverse datasets could provide further insights and improvements.

Limitations of the study

The main limitation of our work is that the performance of the original teacher model is still superior to the student model; however, it should be noted that in these experiments, the teacher is the upper bound achievable, as the student has no access to the training labels. In addition, we only tested this approach with a very specific task, LVO detection, and registration type, linear registration.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability statement
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Image encoders
  - Student model - Convolutional neural network (DCNN)
  - Student model - Transformers
  - Contrastive learning loss

- Feature embeddings
- Implementation details

ACKNOWLEDGMENTS

This work was supported by NIH grant R01NS121154. L.G. is also supported in part by NIH grants R21EB029575, U01AG070112 and The Robert and Janice McNair Foundation.

AUTHOR CONTRIBUTIONS

Y.D. and L.G. wrote the manuscript. Y.D. ran all of the experiments, and S.P. wrote part of the code. X.L. proposed some instructions on medical imaging. S.P., L.G., and S.A.S. collected and preprocessed the data. Y.D. preprocessed the data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 9, 2023
Revised: December 20, 2023
Accepted: January 19, 2024
Published: February 1, 2024

REFERENCES

1. Amukotuwa, S.A., Straka, M., Dehkharghani, S., and Bammer, R. (2019). Fast automatic detection of large vessel occlusions on CT angiography. Stroke *50*, 3431–3438.
2. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., and Brox, T. (2014). Discriminative unsupervised feature learning with convolutional neural networks. Adv. Neural Inf. Process. Syst. *27*, 766–774.
3. Doersch, C., and Zisserman, A. (2017). In Multi-task self-supervised visual learning (Venice, Italy: IEEE), pp. 2070–2079. https://doi.org/10.1109/ICCV.2017.226.
4. Tschannen, M., Djolonga, J., Rubenstein, P.K., Gelly, S., Lucic, M. On mutual information maximization for representation learning Preprint at. arXiv. https://doi.org/10.48550/arxiv.1907.13625.
5. Olive-Gadea, M., Crespo, C., Granes, C., Hernandez-Perez, M., Pérez de la Ossa, N., Laredo, C., Urra, X., Carlos Soler, J., Soler, A., Puyalto, P., et al. (2020). Deep learning based software to identify large vessel occlusion on noncontrast computed tomography. Stroke *51*, 3133–3137.
6. Barman, A., Inam, M.E., Lee, S., Savitz, S., Sheth, S., Giancardo, L., et al. (2019). In Determining ischemic stroke from CT-Angiography imaging using symmetry-sensitive Convolutional networks (IEEE), pp. 1873–1877. https://doi.org/10.1109/ISBI.2019.8759475.
7. Czap, A.L., Bahr-Hosseini, M., Singh, N., Yamal, J.M., Nour, M., Parker, S., Kim, Y., Restrepo, L., Abdelkhaleq, R., Salazar-Marioni, S., et al. (2022). Machine learning automated detection of large vessel occlusion from mobile stroke unit computed tomography angiography. Stroke *53*, 1651–1656.
8. Stib, M.T., Vasquez, J., Dong, M.P., Kim, Y.H., Subzwari, S.S., Triedman, H.J., Wang, A., Wang, H.L.C., Yao, A.D., Jayaraman, M., et al. (2020). Detecting large vessel occlusion at multiphase CT angiography by using a deep convolutional neural network. Radiology *297*, 640–649.
9. Luijten, S.P.R., Wolff, L., Duvekot, M.H.C., van Doormaal, P.J., Moudrous, W., Kerkhoff, H., Lycklama A Nijeholt, G.J., Bokkers, R.P.H., Yo, L.S.F., Hofmeijer, J., et al. (2022). Diagnostic performance of an algorithm for automated large vessel occlusion detection on CT angiography. J. Neurointerv. Surg. *14*, 794–798.
10. Czap, A.L., Bahr-Hosseini, M., Singh, N., Yamal, J.-M., Nour, M., Parker, S., Kim, Y., Restrepo, L., Abdelkhaleq, R., Salazar-Marioni, S., et al. (2022). Machine Learning Automated Detection of Large Vessel Occlusion From Mobile Stroke Unit Computed Tomography Angiography. Stroke *53*, 1651–1656. https://doi.org/10.1161/STROKEAHA.121.036091.
11. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., and Loh, A. (2021). Big self-supervised models Advance medical image classification. In IEEE/CVF International Conference on Computer Vision (ICCV) (IEEE), pp. 3458–3468. https://doi.org/10.1109/ICCV48922.2021.00346.
12. Taleb, A., Kirchler, M., Monti, R., and Lippert, C. (2022). ContIG: self-supervised multimodal contrastive learning for medical imaging with genetics. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA (IEEE), pp. 20876–20889. https://doi.org/10.1109/CVPR52688.2022.02024.
13. Wang, Z., Wu, Z., Agarwal, D., and Sun, J. (2022). MedCLIP: contrastive learning from unpaired medical images and text. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 3876–3887.
14. Tiu, E., Talius, E., Patel, P., Langlotz, C.P., Ng, A.Y., and Rajpurkar, P. (2022). Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat. Biomed. Eng. *6*, 1399–1406.
15. Shlobin, N.A., Baig, A.A., Waqas, M., Patel, T.R., Dossani, R.H., Wilson, M., Cappuzzo, J.M., Siddiqui, A.H., Tutino, V.M., and Levy, E.I. (2022). Artificial intelligence for large-vessel occlusion stroke: a systematic review. World Neurosurg. *159*, 207–220.e1.
16. Giancardo, L., Niktabe, A., Ocasio, L., Abdelkhaleq, R., Salazar-Marioni, S., and Sheth, S.A. (2023). Segmentation of acute stroke infarct core using image-level labels on CT-angiography. Neuroimage. Clin. *37*, 103362.
17. Chen, X., and He, K. (2021). Exploring Simple Siamese Representation Learning. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15745–15753. https://doi.org/10.1109/CVPR46437.2021.01549.
18. Chen, T., Kornblith, S., Norouzi, M., Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. International conference on machine learning. 2020. pp.1597-1607.
19. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., and Uszkoreit, J. (2010). An image is worth 16x16 words: Transformers for image recognition at scale. Preprint at arXiv arXiv, 11929. https://doi.org/10.48550/arXiv.2010.11929.
20. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F.A. (2020). Shortcut learning in deep neural networks. Nat. Mach. Intell. *2*, 665–673.
21. Beare, R., Lowekamp, B., and Yaniv, Z. (2018). Image segmentation, registration and characterization in R with SimpleITK. J. Stat. Softw. *86*, 8.
22. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., and Beare, R. (2018). SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. J. Digit. Imaging *31*, 290–303.
23. Brett, M., Markiewicz, C.J., Hanke, M., Côté, M.-A., Cipollini, B., McCarthy, P., Jarecka, D., Cheng, C.P., Halchenko, Y.O., and Cottaar, M. (2022). nipy/nibabel: Version 4.0.0. Zenodo *10*. https://doi.org/10.5281/zenodo.6658382.

24. Islam, K.T., Wijewickrema, S., and O'Leary, S. (2021). A deep learning based framework for the registration of three dimensional multi-modal medical images of the head. Sci. Rep. *11*, 1860.

25. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning (PMLR), pp. 8748–8763.

26. Pachade, S., Datta, S., Dong, Y., Salazar-Marioni, S., Abdelkhaleq, R., Niktabe, A., Roberts, K., Sheth, S., and Giancardo, L. (2023). In Self-Supervised Learning with Radiology Reports, A Comparative Analysis of Strategies for Large Vessel Occlusion and Brain (Cartagena, Colombia), pp. 1–5. https://doi.org/10.1109/ISBI53787.2023.10230623.

27. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. Preprint at arXiv. https://doi.org/10.48550/arxiv.1309.0238.

28. Yang, T. (2022). Algorithmic foundation of deep X-risk optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.2206.00439.

29. Yuan, Z., Zhu, D., Qiu, Z.-H., Li, G., Wang, X., and Yang, T. (2023). A Deep Learning Library for X-risk Optimization. In 29th SIGKDD Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery), pp. 5487–5499. https://doi.org/10.1145/3580305.3599861.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
DeepSymNetv3	Giancardo et al.[16]	https://glabapps.uth.edu/
ViT	Dosovitskiy et al.[19]	https://github.com/google-research/vision_transformer
Self-Supervised Learning Approach for Registration Agnostic Imaging Models	This study	https://github.com/lgiancaUTH/registration_agnostic_ml_cta/.
SimSiam	Chen et al.[17]	https://arxiv.org/abs/2011.10566
SimCLR	Chen et al.[18]	https://arxiv.org/abs/2002.05709

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Luca Giancardo (Luca.Giancardo@uth.tmc.edu).

Materials availability statement

No reagents were generated in the study.

Data and code availability

- The code for this work can be found at: https://github.com/lgiancaUTH/registration_agnostic_ml_cta/.
- Requests for imaging data used in this work should be directed to lead contact. The availability of imaging data will be contingent upon the specific request, institutional policies, and the project requirements of NIH R01NS121154.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In this proposed work, we used a dataset of 402 CTA images of different subjects which were split into a train, validation, and test with the ratio of 60%, 20%, and 20%, respectively. The demographic information for subjects was listed in Table 1.

To obtain the unregistered images, we applied the imaging padding function on the three axes, then, we randomly generated rotations with an angle in the range of −15°–15° and translations with distance in the range of −30 mm–30 mm in 3 axes (x, y, z). Finally, we anisotropically resampled the images to 182 × 182 × 182 which reduces the memory requirement for training the deep learning approaches. To make a fair comparison and avoid our proposed model learning "shortcuts"[20] instead of LVO, we did transformations of the test dataset in two ways. Firstly (Dataset A), we conducted random transformation on a single subject, which means every subject has different transformations. Secondly (Dataset B), we applied the same random transformation on all of the subjects and generated 50 different transformations in total. The process was implemented using SimpleITK[21,22] and NiBabel.[23]

METHOD DETAILS

Image encoders

There are multiple individual image encoders in our proposed work. One is for registered 3D CTA which is named DeepSymNetv3, and another is DCNN for unregistered 3D CTA.

The teacher model is an adaptation of the original DeepSymNetv3[16] which adds non-symmetric paths and it is trained on LVO rather than stroke core. In summary, the model splits the registered brain CTA I_{reg} into left I^L_{reg} and right I^R_{reg} parts and then constructed several layers of VGG network for each part of the brain to obtain VI^L_{reg} and VI^R_{reg}. After that, the L1-norm was calculated based on the outputs of separated VGG nets for the merged layer as $Merged_{reg} = L1(VI^L_{reg}, VI^R_{reg})$. Then, we obtained concatenated layers with two separated layers from respective VGG and one L1-norm merged layer as $Cat(Merged_{reg}, VI^L_{reg}, VI^R_{reg})$. After the average pooling layer and fully connected layer, E_{reg} was extracted as registered image embedding.

Student model - Convolutional neural network (DCNN)

For unregistered images, we applied DCNN[24] as the image encoder. Before we put the unregistered images into the encoder, we conducted preprocessing to crop the x-, y-, and z axis to reduce the unregistered image's I_{raw} size. Then we put I_{raw} into DCNN to obtain unregistered

image embedding E_{raw}. There are 19 convolutional layers in the DCNN, except the last convolutional layer, each of them followed by a batch normalization layer and activated layer leaky ReLU. DCNN contains 5 max pooling layers. After the last convolutional layer, there is a global average pooling layer followed by a fully connected layer.

Student model - Transformers

In order to evaluate the student/teacher strategy proposed with an alternative image encoder, we also utilized ViT[19] as the unregistered image encoder. The preprocessing details are the same as DCNN experiments. I_{raw} is the input for ViT, the patch size is 26, and the input channel is 1. After obtaining the unregistered image embedding E_{raw}, we used the projection layers to get the unregistered images logit.

Contrastive learning loss

CLIP loss

We applied CLIP-based loss on projected image embeddings of registered images and unregistered images. The same project module with output dimensionality in 256 was used to obtain projected embeddings. CLIP model[25] firstly was applied in pairs of text and image to conduct contrastive self-supervised learning. Here, we applied the CLIP-based loss on pairs of images for contrastive self-supervised learning to obtain CLIP-based loss $L = CLIPLoss(PE_{reg}, PE_{raw})$, where PE_* represents projected embeddings for registered and unregistered images. In the CLIP-based loss, logits and targets were used as inputs of cross entropy loss, which were calculated as

$$x = \left(PE_{raw} \cdot PE_{reg}^{\top}\right) \Big/ \tau,$$

and

$$y = \sigma\left(\left(PE_{reg} \cdot PE_{reg}^{\top}\right) + \left(PE_{raw} \cdot PE_{raw}^{\top}\right)\right) \Big/ 2\tau\Big),$$

where x represents the logit obtained from projected embeddings of unregistered images and registered images, in our experiments, y represents the target, τ is the temperature parameter and $\tau = 1.0$ in our experiments. T represent the transpose. To calculate cross entropy loss,

$$L_{ij} = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{N} x_{ij} \log y_{ij};$$

$$L_{ji} = -\frac{1}{N}\sum_{j}^{N}\sum_{i}^{N} x_{ji} \log y_{ji};$$

where x_{ji} and y_{ji} are transposed of x_{ij} and y_{ij} respectively, $i, j \in (0, N)$, N is the batch size. And the final loss output is

$$L = \frac{\left(L_{ij} + L_{ji}\right)}{2}.$$

SimCLR loss

We also applied SimCLR[18] loss for comparison with CLIP loss. The normalized temperature-scaled cross entropy (NT-Xent) loss is used in the SimCLR study. However, our study used different sources as input for the SimCLR loss calculation, as opposed to the original SimCLR paper where the authors used different augmentations from the same source. Also, SimCLR ignores negative pairs. For the positive pairs of samples, the loss function is defined as:

$$L_{o,p} = -\log \frac{\exp\left(\left(s\left(x_o, x_p\right)/\tau\right)\right)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq 1} \exp\left(\frac{s(x_0, x_k)}{\tau}\right)},$$

where x_o represents the image logit obtained from projected embeddings of images, τ is the temperature parameter and $\tau = 0.5$. Here, $o \in (0, 2N)$, $p \in (0, 2N)$ and $k \in (0, N)$, N is the batch size. $s(x_o, x_k)$ represents the pairwise similarity.

SimSiam loss

SimSiam[17] loss was tested using registered images and unregistered images as inputs and then minimizing the negative cosine similarity between two image embeddings. The unmatched pairs are not considered as in SimCLR. The loss function follows:

$$D\left(P_{reg}, E_{raw}\right) = -\frac{P_{reg}}{\left\|P_{reg}\right\|_2} \cdot \frac{E_{raw}}{\left\|E_{raw}\right\|_2},$$

where $\| \bullet \|_2$ is L2-norm, and P_{reg} is output from prediction MLP, E_{raw} is unregistered image embedding. And symmetrized loss is defined as:

$$L = 0.5 \cdot D(P_{reg}, E_{raw}) + 0.5 \cdot D(P_{raw}, E_{reg}).$$

Feature embeddings

The teacher model (DeepSymNetv3) was initially trained using pairs of registered images and the radiologist's reports and then fine-tuned on LVO.[26] Then, we loaded the model as the teacher model and the weights of the model were frozen in the teacher-student experiments. DCNN was utilized as an unregistered image encoder, and it works as the student model. DCNN is one of the popular neural networks that showed outperformed performance among published models in computer vision. In our model, the deep CNN composed of 19 convolutional layers, 18 batch normalization layers, and maxpooling layers. Besides, we also conducted experiments using ViT, which is a popular Transformer-based architecture.[19] The size of image embedding for both registered images and unregistered images using DCNN is 72, and for unregistered images using ViT is 768. The same image projection structure was applied for CLIP-based loss, which has an output in dimensionality 256.

Implementation details

LVO detection is a binary classification task. In the implementation of teacher model training and fine-tuning experiments, the batch size is 14, with 100 epochs. Adam optimizer with a learning rate $1e-4$ was utilized in pretraining and a learning rate $1e-5$ was applied in fine-tuning experiments. In the validation, the loss function Binary Cross Entropy with Logits was used, and AUC scores were calculated by Scikit-Learn[27] function.

We set the batch size as 4, and the number of epochs is 200 with an early stopping strategy which has a 25 tolerance setting in the implementation of teacher-student model experiments. Besides, the loss function in training is CLIP-based loss and the optimizer is AdamW with a learning rate of $1e-4$. The teacher model's weights were frozen in the training phase. What's more, we applied AUCM loss[28] from LibAUC[29] in the validation which is a margin-based surrogate loss and has shown better performance in medical imaging tasks.

All of the experiments were implemented on a single NVIDIA T100 40GB GPU with AMD EPYC 7402 24-Core Processor. The whole Teacher-Student model structure is described as a pseudo-code in Algorithm 1.

Algorithm 1. Teacher-Student Model Structure

Data: Registered Images I_{reg}, Unregistered Images I_{raw}

for each mini-batch **do**

E_{reg} = TeacherModel(I_{reg}); /* Frozen weights

E_{raw} = StudentModel(I_{raw}); /* Eq1

P_{reg} = Projection(E_{reg}); /* Eq2

P_{raw} = Projection(E_{raw}); /* Eq3

$logits = \frac{P_{raw} \bullet P_{reg}^{\top}}{\tau}$; /* Eq4

$target = \sigma\left(\frac{sim(P_{reg}, P_{reg}^{\top}) + sim(P_{raw}, P_{raw}^{\top})}{2\tau}\right)$ /* Eq5

L_{raw} = CrossEntropy($logits, target$); /* Set gradients to zero

L_{reg} = $CrossEntropy(logits^{\top}, target^{\top})$; /* Compute the gradients

$L = \frac{(L_{raw} + L_{reg})}{2}$; optimizer.zerograd(); /* Parameters update

loss.backward();

optimizer.step();

end