Learning image transformations via convolutional neural networks: a review

Nicholas J. Tustison[1], Brian B. Avants[1], and James C. Gee[2]

[1]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

[2]Department of Radiology, University of Pennsylvania, Philadelphia, PA

Corresponding author:

Nicholas J. Tustison

ntustison@virginia.edu

Abstract

Recent methodological innovations in deep learning and associated advancements in computational hardware have significantly impacted the various core subfields of quantitative medical image analysis. The generalizability, computational efficiency, and open-source availability of deep learning algorithms and related software, particularly those utlizing convolutional neural networks, have produced paradigm shifts within the field. This impact is evident from topical prevalance in the literature, conference and workshop themes, and winning methodologies in relevant competitions. In this work, we review the various state-of-the-art approaches to learning and prediction and/or optimizing image transformations using convolutional neural networks. Although of primary importance within the quantitative imaging domain, image registration algorithmic development, in the context of these deep learning strategies, has received comparatively less attention than its counterparts (e.g., image segmentation). Nevertheless, significant progress has been made in this particular subfield which has been presented in various research venues. We contextualize these contributions within the broader scope of deep learning advancements and, in so doing, attempt to facilitate the leveraging and further development of such techniques within the medical imaging research community.

Key words: deep learning, diffeomorphisms, image registration, spatial normalization

Introduction

Determining the spatial correspondence between imaging domains is frequently a critical component in quantitative image analysis workflows. The trajectory of image registration theoretical and technological development has led to increasingly high quality transformational mappings that have significantly improved performance in related processing tasks (e.g., image segmentation via joint label fusion [1]) and imaging-based statistical analysis involving template-based normalization (e.g., voxel-based morphometry [2] and sparse canonical correlation analysis [3]). Several reviews [4–9] have charted this chronology and provided insight into related issues such as algorithmic classification, available implementations, evaluation strategies, and speculation concerning possible future directions of the field. While prescient in many respects, such speculation vis-à-vis the resurgence of deep learning is understandably limited due to its recent explosion in popularity and research focus.

The foundational concepts that form the basis for contemporary deep learning research dates back decades (e.g., [10]). Since this early seminal work, major developmental milestones include the *Neocognitron*, an early neural network for character recognition [11], and convolutional neural networks ("CNNs" or "ConvNets") utilized in speech [12] and visual signal processing [13], largely inspired by the visual cell types of the feline visual cortex [14]. Historical neural networks are differentiated from their modern progeny by the deep, or "hidden," layering that characterizes current architectures and is the reason for the extreme performance gains seen in the contemporary literature. The training of such architectures is made computationally tractable with gradient-based optimization using backpropagation (first performed in [13]) and the advent of GPU-based hardware [15]. Uptake by both industry and academia alike is further facilitated through the various neural network open-source software platforms (e.g., Tensorflow [16] and Keras [17]).

A key event in the widespread adoption of CNNs was the 2012 ImageNet Large Scale Visual Recognition Challenge for object classification [18]. The winning entry, a CNN-based architecture colloquially known as *AlexNet* [19], reduced the error rate by almost half over other entries . The following years' competitions were dominated by CNN variants such as VGG [20], GoogLeNet [21], and ResNet [22] with performance ultimately exceeding human performance in 2015 [23]. Additional competition outlets including conference-based venues (e.g., NeurIPS) and community-based platforms, such as Kaggle[1], continue to highlight the salience of CNNs as paradigmatic solutions to computational problems. This is in addition to the sheer number of formal research reports discussed in the same conferences and published in dedicated journals. Notable reviews by key figures in the field include those of Yann LeCun, Yoshua Bengio, Geoffrey Hinton [15], and Jürgen Schmidhuber [25].

Early CNN-based research tailored to medical imaging dates back to the 1990s with classification tasks providing the majority of use cases (e.g., lung nodule classification [26, 27] and breast tissue differentiation [28, 29]). Despite the early adoption by certain research groups, widespread uptake did not occur until much later. Several deep learning overviews specific to medical imaging have been presented in the recent research literature

- in editorial form [30];
- specific to generative adverserial networks (GANs) [31];
- focusing on MRI [32] and specific to neuro applications [33];
- for issues related to radiation therapy [34];
- concentrating on applications [35]; and
- as general reviews [36–40].

Despite the thorough treatment contained in these reviews, discussion of chronological adoption within the community is limited. Regardless, one can informally gauge this evolution from utilization of alternative machine learning techniques to predominately CNN-based approaches from the various competitions held simultaneously with medical imaging conferences. For example, the annual Multimodal Brain Tumor Segmentation (BraTS) Challenge has taken place under the auspices of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) since 2012 wherein large sets of training data are provided to the competitors who attempt to perform a voxelwise labeling of the constituent components of tumors from multimodal MR image data. The winning entries from the first two years employed random forest classifiers for segmentation [41]. Although variations of the traditional random forest scheme continued to be well represented in the 2014 Challenge, convolutional neural networks made an appearance [42]. By 2018, CNN-based pipelines were, by far, the most common [43] with specific preference being that of the

[1]Following the 2017 ImageNet challenge, in which the vast majority of teams surpassed the 5% classification error rate threshold, the ImageNet organizers ceded management to the Kaggle community which maintains a running performance assessment in ostensible perpetuity [24].

U-net architecture [44, 45] which, as we describe below, features prominently in image registration.

Conspicuously, coverage of the topic of deep learning-based image registration, relative to the related algorithmic categories of image classification and segmentation, has not been as extensive in the reviews mentioned above, despite its prominence in the broader research literature. This disparity seems to be similarly reflected in the quantity of published research for those respective categories [31, 38]. This review is meant to address this disparity and thus provide an overview of the current state-of-the-art of this burgeoning subfield. We first provide a description of key network components that are crucial to certain image registration architectures, or perhaps which might find utility in future architectures. Second, we discuss current approaches to the deep learning image registration categorized by ...

Preliminaries

Convolutional Neural Networks

The major elements of CNNs are localized connections, convolutions, and subsampling (or "pooling") [15].

An illustration of a bare-bones CNN configuration is provided in Figure 1 which depicts the core components of convolution and max pooling. Architectural novelty derives from innovative arrangements of these core (and other) network components and the connections between them.

Spatial transformer networks

In 2015 Jaderberg and his fellow co-authors described a powerful new module, known as the spatial transformer network (STN) [46], which figures prominently in many of the image registration approaches that we review below. Although concepotually relatively straightforward, the obvious influence reflected in recent work will most likely continue in future research which is why we review this important network component.

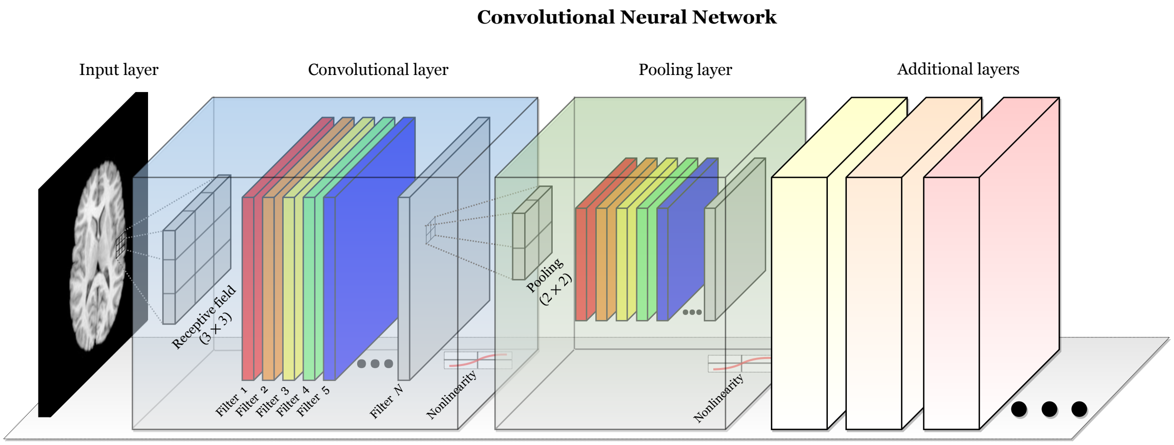

Figure 1: The basic elements of the prop convolutional neural network. The convolutional layer comprises several filters which are optimized in terms of their responses to various features found in the input layer. Pooling is used to extract salient features and reduce computational complexity and passed on to subsequent layers.

Generally, STNs enhance CNNs by permiting a flexibility which allows for an explicit spatial invariance that goes beyond the implicitly limited translational invariance associated with the architecture's pooling layers. In many image-based tasks (e.g., localization or segmentation), designing an algorithm that can account for possible pose or geometric variation of the object(s) of interest within the image is crucial for maximizing performance. The STN is a fully differentiable layer which can be inserted anywhere in the CNN to learn the parameters of the transformation of the input feature map (not necessarily an image) which renders the output in such a way to optimize the network based on the specified loss function. The added flexibility and the fact that there is no manual supervision or special handling required makes this module an essential addition for any CNN-based toolkit.

An STN comprises three principal components: 1) a localization network, 2) a grid generator, and 3) a sampler (see Figure 3). The localization network uses the input feature map to learn/regress the transformation parameters which optimize a specified loss function. In many examples provided, this amounts to transforming the input feature map to a quasi-canonical configuration to facilitate, for example, classification. The actual architecture of the localization network is fairly flexible and any conventional architecture, such as a fully connected network (FCN), is suitable as long as the output maps to the continuous estimate of the transformation parameters. These transformation parameters are then applied to the output of the grid generator which are simply the regular coordi-

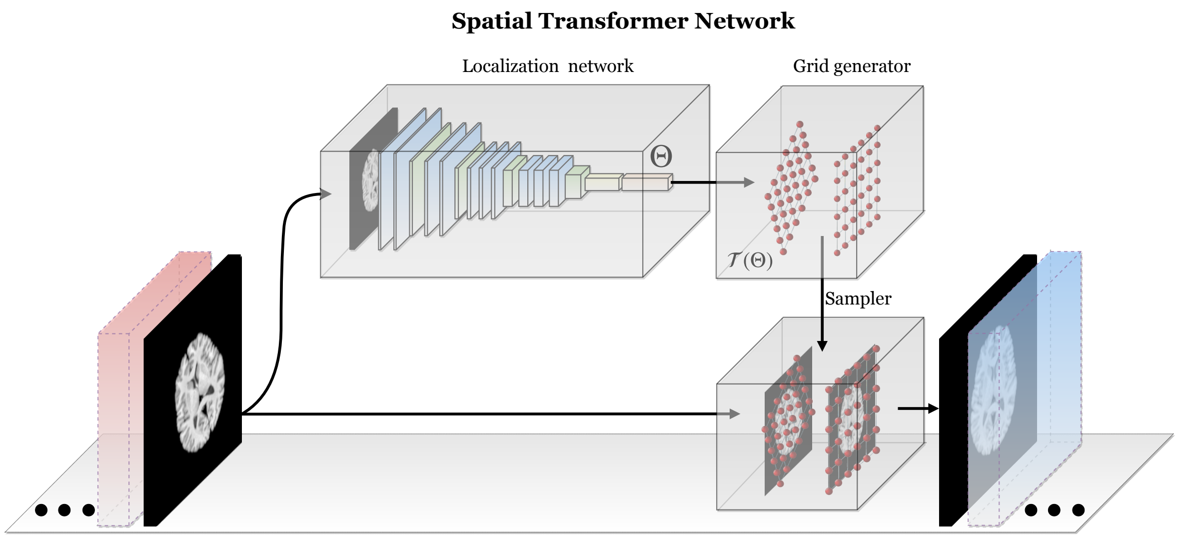

Figure 2: Diagrammatic illustration of the spatial transformer network. The STN can be placed anywhere within a CNN to provide spatial invariance for the input feature map. Core components include the localization network used to learn/predict the parameters which transform the input feature map. The transformed output feature map is generated with the grid generator and sampler.

nates of the input image (or some normalized version thereof). The sampler, or interpolator, is used to map the transformed input feature map to the coordinates of the output feature map.

Inverse compositional transformer networks

[47] Inspired by the IC-LK algorithm, we advocate an improved extension to the STN framework that (a) propagates warp parameters, rather than image intensities

Diffeomorphic transformer networks

Although discussion of transform generalizability was included in the original STN paper [46], discussion was limited to affine, attention (scaling + translation), and thin-plate spline transforms which all fill the requirements of differentiability. This was later extended to encompass a diffeomorphic transformer network (DTN) [48] based on continuous piecewise affine-based (CPAB) transformations [49]. **This section needs to be expanded**.

Siamese Networks

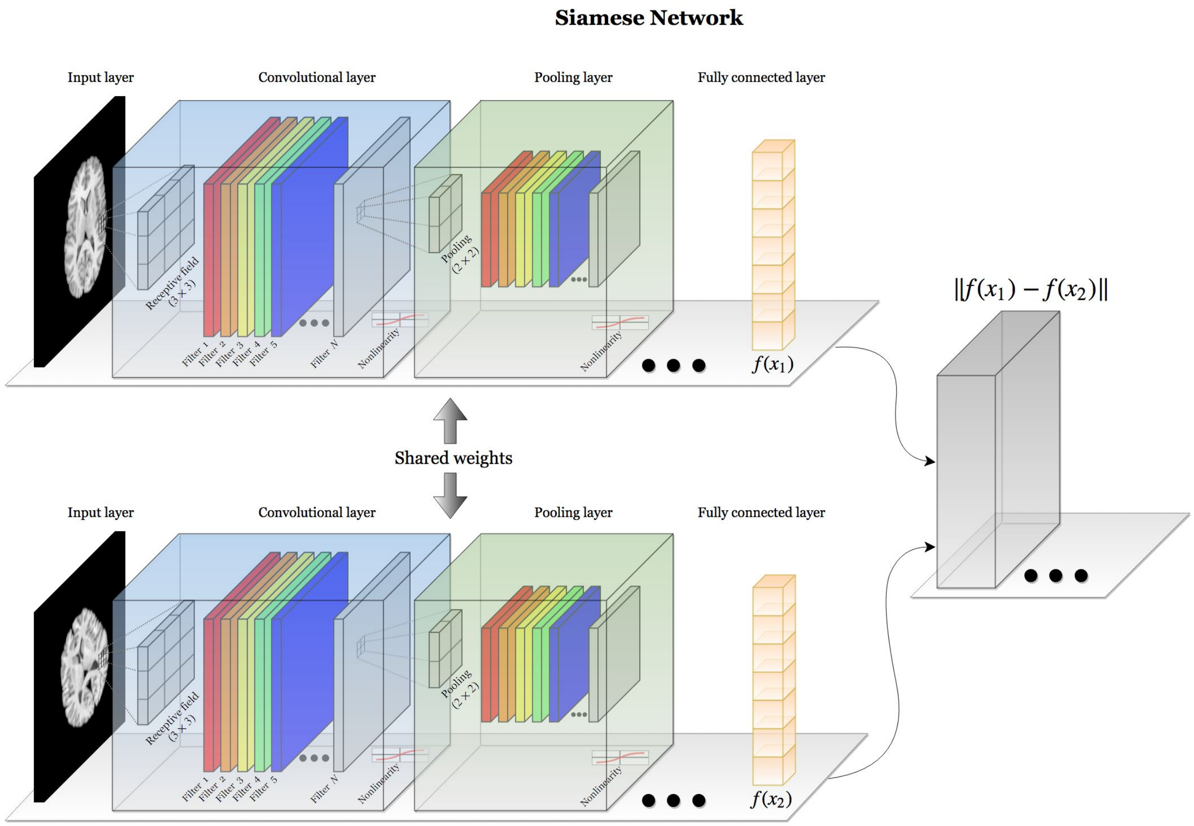

Figure 3: Diagrammatic illustration of the spatial transformer network.

Image Registration with Deep Learning

Geodesic shooting with Quicksilver

The large deformation diffeomorphic metric mappings (LDDMM) framework for image matching derives from the theoretical foundations underlying diffeomorphic *flows* [50–52]. Such diffeomorphisms are sufficiently differentiable bijective mappings, or transformations, which have sufficiently differentiable inverses. Specifically, the set of possible diffeomorphic mappings, $\phi(\mathbf{x}, t)$ ($\mathbf{x} \in \Omega$, $t \in [0, 1]$), between two images, I and J can be described as the collection of *paths* connecting the two images on a manifold determined by the equation

$$\int_0^1 \|v(t)\|_L^2 dt + \int_\Omega |I \circ \phi^{-1}(x,1) - J|^2 d\Omega. \tag{1}$$

v is a time-dependent smooth field dictated by the functional norm L and determines the mapping via the ordinary differential equation

$$\frac{d\phi(\mathbf{x},t)}{dt} = v(\phi(\mathbf{x},t),t), \phi(\mathbf{x},0) = \mathbf{Id}. \tag{2}$$

The optimal diffeomorphic transformation between I and J can be described as a geodesic [53] connecting the two images. Traditionally, computational approaches to determining this geodesic path involve discretization of the velocity field followed by numerical integration. This is performed for a given number of iterations where, presumably, convergence implies arrival at this geodesic (i.e., optimal) path. Alternatively, based on the work of [54], the Euler-Lagrange equations for Equation (1) can be written as a system incorporating a "momentum" term. It was further demonstrated that the initial momentum determined the entire geodesic path. This alternative perspective engendered a new approach to determining the diffeomorphic solution between two images, known as *geodesic shooting* (e.g., [53, 55]). Although initially formulated in terms of scalar momenta [55], a vector formulation was proposed in [56] which tends towards superior numerical behavior.

The supervised deep learning technique of Yang et al. [57], known as *Quicksilver*, leverages this geodesic shooting/vector momentum optimization approach for determining optimal diffeomorphic transformations. The network architecture consists of two parallel encoders for separate fixed/moving image patches ($15 \times 15 \times 15$ voxels) feature learning. The output is then concatenated and sent through three identical decoder branches (one for each dimension) which comprises the inverse operations as the single encoder branch. Thus, the output consists of the predicted vector momentum map which, as described above, determines the total transformation. In order to improve accuracy of the predicted momentum maps, a follow-on correction network is also proposed. This correction network, trained by inverting the mapping produced by the predicted momentum and computing the residual error, is meant to account for large deformations across patch boundaries. Of note, Quicksilver, written in PyTorch [58], is one of the handful of algorithms surveyed which has been made publicly available[2].

[2]https://github.com/rkwitt/quicksilver

Acknowledgments

We gratefully acknowledge the support of the NVIDIA Corporation with the donation of two Titan Xp GPUs used in support of this research.

References

1. Iglesias, J. E. and Sabuncu, M. R. “**Multi-Atlas Segmentation of Biomedical Images: A Survey**” *Med Image Anal* 24, no. 1 (2015): 205–219. doi:10.1016/j.media.2015.06.012

2. Ashburner, J. and Friston, K. J. “**Voxel-Based Morphometry–the Methods**” *Neuroimage* 11, no. 6 Pt 1 (2000): 805–21. doi:10.1006/nimg.2000.0582

3. Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C., and Grossman, M. “**Dementia Induces Correlated Reductions in White Matter Integrity and Cortical Thickness: A Multivariate Neuroimaging Study with Sparse Canonical Correlation Analysis**” *Neuroimage* 50, no. 3 (2010): 1004–16. doi:10.1016/j.neuroimage.2010.01.041

4. Brown, L. G. “**A Survey of Image Registration Techniques**” *ACM Comput. Surv.* 24, no. 4 (1992): 325–376. doi:10.1145/146370.146374, Available at http://doi.acm.org/10.1145/146370.146374

5. Maintz, J. B. and Viergever, M. A. “**A Survey of Medical Image Registration**” *Med Image Anal* 2, no. 1 (1998): 1–36.

6. Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. “**Mutual-Information-Based Registration of Medical Images: A Survey**” *IEEE Trans Med Imaging* 22, no. 8 (2003): 986–1004. doi:10.1109/TMI.2003.815867

7. Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., and Gopinath, K. “**Brain Functional Localization: A Survey of Image Registration Techniques**” *IEEE Trans Med Imaging* 26, no. 4 (2007): 427–51. doi:10.1109/TMI.2007.892508

8. Viergever, M. A., Maintz, J. B. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. W. “**A Survey of Medical Image Registration - Under Review**” *Med Image Anal* 33, (2016): 140–144. doi:10.1016/j.media.2016.06.030

9. Keszei, A. P., Berkels, B., and Deserno, T. M. “**Survey of Non-Rigid Registration Tools in Medicine**” *J Digit Imaging* 30, no. 1 (2017): 102–116. doi:10.1007/s10278-016-9915-8

10. Ivakhnenko, A. G. “**Polynomial Theory of Complex Systems**” *IEEE Transactions on Sys-*

tems, Man, and Cybernetics SMC-1, no. 4 (1971): 364–378.

11. Fukushima, K. "**Neocognitron: A Self Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position**" *Biol Cybern* 36, no. 4 (1980): 193–202.

12. Waibel, A. "**Phoneme Recognition Using Time-Delay Neural Networks**" *Meeting of the Institute of Electrical, Information and Communication Engineers (IEICE).* (1987):

13. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jacke, L. D. "**Backpropagation Applied to Handwritten Zip Code Recognition**" *Neural Computation* 1, no. 4 (1989): 541–551.

14. Hubel, D. H. and Wiesel, T. N. "**Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex**" *J Physiol* 160, (1962): 106–54.

15. LeCun, Y., Bengio, Y., and Hinton, G. "**Deep Learning**" *Nature* 521, no. 7553 (2015): 436–44. doi:10.1038/nature14539

16. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. "**TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**" (2015):

17. Chollet, F. and others. "**Keras**"

18. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. "**ImageNet Large Scale Visual Recognition Challenge**" *International Journal of Computer Vision* 115, no. 3 (2015): 211–252.

19. Krizhevsky, A., Sutskever, I., and Hinton, G. E. "**ImageNet Classification with Deep Convolutional Neural Networks**" *Proceedings of the 25th International Conference on Neural Information Processing Systems* (2012): 1097–1105. Available at http://dl.acm.org/citation.cfm?id=

2999134.2999257

20. Simonyan, K. and Zisserman, A. "**Very Deep Convolutional Networks for Large-Scale Image Recognition**" *CoRR* abs/1409.1556, (2014): Available at http://arxiv.org/abs/1409.1556

21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. "**Rethinking the Inception Architecture for Computer Vision**" *CoRR* abs/1512.00567, (2015): Available at http://arxiv.org/abs/1512.00567

22. He, K., Zhang, X., Ren, S., and Sun, J. "**Deep Residual Learning for Image Recognition**" *CoRR* abs/1512.03385, (2015): Available at http://arxiv.org/abs/1512.03385

23. He, K., Zhang, X., Ren, S., and Sun, J. "**Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015):

24. Available at https://www.kaggle.com/c/imagenet-object-localization-challenge

25. Schmidhuber, J. "**Deep Learning in Neural Networks: An Overview**" *Neural Networks* 61, (2015): 85–117.

26. Lo, S. C., Freedman, M. T., Lin, J. S., and Mun, S. K. "**Computer-Aided Detection of Mammographic Calcifications: Pattern Recognition with an Artificial Neural Network**" *Proc. spie: Medical imaging: Image processing* 1898, (1992): 859–869.

27. Lo, S. C., Freedman, M. T., Lin, J. S., and Mun, S. K. "**Automatic Lung Nodule Detection Using Profile Matching and Back-Propagation Neural Network Techniques**" *J Digit Imaging* 6, no. 1 (1993): 48–54.

28. Chan, H. P., Lo, S. C., Sahiner, B., Lam, K. L., and Helvie, M. A. "**Computer-Aided Detection of Mammographic Microcalcifications: Pattern Recognition with an Artificial Neural Network**" *Med Phys* 22, no. 10 (1995): 1555–67. doi:10.1118/1.597428

29. Sahiner, B., Chan, H. P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D., and Goodsitt, M. M. "**Classification of Mass and Normal Breast Tissue: A Convolution Neural Network Classifier with Spatial Domain and Texture Images**" *IEEE Trans Med Imaging* 15, no. 5 (1996):

598–610. doi:10.1109/42.538937

30. Greenspan, H., Ginneken, B. V., and Summers, R. M. "**Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique**" *IEEE Trans Med Imaging* 35, no. 5 (2016): 1153–1159.

31. Yi, X., Walia, E., and Babyn, P. "**Generative Adversarial Network in Medical Imaging: A Review**" *arXiv preprint* (2018):

32. Mazurowski, M. A., Buda, M., Saha, A., and Mustafa R. Bashir. "**Deep Learning in Radiology: An Overview of the Concepts and a Survey of the State of the Art with Focus on Mri**" *Journal of Magnetic Resonance Imaging* (2018):

33. Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., and Lladó, X. "**Deep Convolutional Neural Networks for Brain Image Analysis on Magnetic Resonance Imaging: A Review**" *Artif Intell Med* (2018): doi:10.1016/j.artmed.2018.08.008

34. Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., Summers, R. M., and Giger, M. L. "**Deep Learning in Medical Imaging and Radiation Therapy**" *Med Phys* (2018): doi:10.1002/mp.13264

35. Ker, J., Wang, L., Rao, J., and Lim, T. "**Deep Learning Applications in Medical Image Analysis**" *IEEE Access* 6, (2018): 9375–9389.

36. Suzuki, K. "**Overview of Deep Learning in Medical Imaging**" *Radiol Phys Technol* 10, no. 3 (2017): 257–273. doi:10.1007/s12194-017-0406-5

37. Shen, D., Wu, G., and Suk, H.-I. "**Deep Learning in Medical Image Analysis**" *Annu Rev Biomed Eng* 19, (2017): 221–248. doi:10.1146/annurev-bioeng-071516-044442

38. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M. van der, Ginneken, B. van, and Sánchez, C. I. "**A Survey on Deep Learning in Medical Image Analysis**" *Med Image Anal* 42, (2017): 60–88. doi:10.1016/j.media.2017.07.005

39. Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M. K. "**Medical Image Analysis Using Convolutional Neural Networks: A Review**" *J Med Syst* 42, no. 11

(2018): 226. doi:10.1007/s10916-018-1088-1

40. Biswas, M., Kuppili, V., Saba, L., Edla, D. R., Suri, H. S., Cuadrado-Godia, E., Laird, J. R., Marinhoe, R. T., Sanches, J. M., Nicolaides, A., and Suri, J. S. "**State-of-the-Art Review on Deep Learning in Medical Imaging**" *Front Biosci (Landmark Ed)* 24, (2019): 392–426.

41. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C. R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K. M., Jena, R., John, N. M., Konukoglu, E., Lashkari, D., Mariz, J. A., Meier, R., Pereira, S., Precup, D., Price, S. J., Raviv, T. R., Reza, S. M. S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C. A., Sousa, N., Subbanna, N. K., Szekely, G., Taylor, T. J., Thomas, O. M., Tustison, N. J., Unal, G., Vasseur, F., Wintermark, M., Ye, D. H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., and Van Leemput, K. "**The Multimodal Brain Tumor Image Segmentation Benchmark (Brats)**" *IEEE Trans Med Imaging* 34, no. 10 (2015): 1993–2024. doi:10.1109/TMI.2014.2377694

42. "**Conference Proceedings of the 3rd MICCAI BraTS Challenge**" (2014):

43. "**Pre-Conference Proceedings of the 7th MICCAI BraTS Challenge**" (2018):

44. Ronneberger, O., Fischer, P., and Brox, T. "**U-Net: Convolutional Networks for Biomedical Image Segmentation**" *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* 9351, (2015): 234–241.

45. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., Diester, I., Brox, T., and Ronneberger, O. "**U-Net: Deep Learning for Cell Counting, Detection, and Morphometry**" *Nat Methods* 16, no. 1 (2019): 67–70. doi:10.1038/s41592-018-0261-2

46. Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. "**Spatial Transformer Net-**

works" (2015):

47. Lin, C.-H. and Lucey, S. "**Inverse Compositional Spatial Transformer Networks**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017):

48. Detlefsen, N. S., Freifeld, O., and Hauberg, S. "**Deep Diffeomorphic Transformer Networks**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018):

49. Freifeld, O., Hauberg, S., Batmanghelich, K., and Fisher, J. W. "**Transformations Based on Continuous Piecewise-Affine Velocity Fields**" *IEEE Trans Pattern Anal Mach Intell* 39, no. 12 (2017): 2496–2509. doi:10.1109/TPAMI.2016.2646685

50. Trouvé, A. "**Diffeomorphic Groups and Pattern Matching in Image Analysis**" *Int. J. Computer Vision* 28, (1995): 213–221.

51. Christensen, G. E., Rabbitt, R. D., and Miller, M. I. "**Deformable Templates Using Large Deformation Kinematics**" *IEEE Trans Image Process* 5, no. 10 (1996): 1435–47. doi:10.1109/83.536892

52. Dupuis, P., Grenander, U., and Miller, M. I. "**Variational Problems on Flows of Diffeomorphisms for Image Matching**" *Quarterly of Applied Mathematics* LVI, (1998): 587–600.

53. Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. "**Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms**" *International Journal of Computer Vision* 61, no. 2 (2004): 139–157.

54. Miller, M. I., Trouvé, A., and Younes, L. "**Geodesic Shooting for Computational Anatomy**" *J Math Imaging Vis* 24, no. 2 (2006): 209–228. doi:10.1007/s10851-005-3624-0

55. Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. "**Diffeomorphic 3d Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation**" *Int J Comput Vis* 97, (2012): 229–241.

56. Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. "**A Vector Momenta Formulation of Diffeomorphisms for Improved Geodesic Regression and Atlas Construction**" *Proc IEEE*

Int Symp Biomed Imaging 2013, (2013): 1219–1222. doi:10.1109/ISBI.2013.6556700

57. Yang, X., Kwitt, R., Styner, M., and Niethammer, M. "**Quicksilver: Fast Predictive Image Registration - a Deep Learning Approach**" *Neuroimage* 158, (2017): 378–396. doi:10.1016/j.neuroimage.2017.07.008

58. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. "**Automatic Differentiation in Pytorch**" *NIPS-w* (2017):