An investigation into deep learning for the analysis of medical images

Chaohao Hu

Stony Brook Institute, Anhui University, Hefei, 230000, China

r32014063@stu.ahu.edu.cn

Abstract. In recent years, deep learning has emerged as a pivotal paradigm in the analysis of medical images, with convolutional networks serving as a cornerstone of this advancement. This paper delves into a comprehensive exploration of the fundamental principles underpinning deep learning and its applications within the domain of medical image analysis. Through a meticulous review of many contemporary contributions, this study synthesizes the latest developments in the field, emphasizing tasks like image classification, object detection, segmentation, and registration. The inquiry spans diverse medical disciplines, encompassing neurology, retinal imaging, pulmonary studies, digital pathology, breast and cardiac evaluations, and musculoskeletal analyses. As a culmination, the paper not only assesses the present state-of-the-art achievements but also critically discusses persistent challenges and illuminates promising avenues for future research endeavors.

Keywords: Convolutional Neural Networks, Deep Learning, Medical Imaging, Survey.

1. Introduction

From the moment medical images could be digitized and processed by computers, researchers embarked on the development of automated analysis systems. The journey commenced in the 1970s and continued through the 1990s, characterized by rule-based systems constructed through pixel-level processing and mathematical modeling. These systems, akin to expert systems of the era, marked the initial phase of medical image analysis. The late 1990s ushered in a shift toward supervised techniques, where training data fueled system development. Active shape models, atlas methods, and feature extraction with statistical classifiers exemplified this pattern recognition approach, forming the foundation for successful commercial medical image analysis systems.

In this evolution, the manual crafting of features gave way to systems that learn from example data and optimize decision boundaries. Convolutional neural networks (CNNs), particularly the breakthrough AlexNet, revolutionized image analysis. Their cascading layers enabled learning progressively higher-level features, making them the model of choice in computer vision.

This transition resonated within the medical image analysis domain, gradually replacing handcrafted features with learned ones. Deep learning techniques witnessed an exponential rise, evidenced by the surge in papers from 2015 onwards. While existing reviews captured substantial work, gaps persisted. The motivation for this comprehensive survey emerged, aiming to provide a panoramic overview encompassing applications and methodologies.

DOI: 10.54254/2755-2721/54/20241121

The survey, delves into diverse applications of deep learning in medical image analysis. The search encompassed prominent databases, conferences, and consultations, focusing on works involving convolutional or deep learning techniques. With this comprehensive review, the author aspire to demonstrate deep learning's pervasiveness in medical image analysis, pinpoint challenges, and spotlight contributions that address these hurdles. The survey unfolds by introducing key deep learning techniques (Section 2), exploring contributions to canonical medical imaging tasks (Section 3), analyzing results and challenges across application domains (Section 4), and concluding with a summary, critical discussion, and future outlook.

2. Survey of Deep Learning Techniques

2.1. Learning algorithms

Learning algorithms are the bedrock of machine learning and artificial intelligence, facilitating the extraction of patterns and insights from data to make informed decisions and predictions. These algorithms operate as mathematical models, adapting and improving over time as they process more information.

Supervised learning algorithms learn from labeled data, wherein input features are paired with corresponding output labels. By discerning correlations between inputs and labels, these algorithms create predictive models capable of assigning labels to new, unseen data points. Regression algorithms, used for predicting continuous values, and classification algorithms, employed for discrete categorization, are the two main branches of supervised learning.

Unsupervised learning algorithms, on the other hand, analyze data without predefined labels. They focus on uncovering underlying structures, patterns, or relationships within the data. Clustering algorithms group similar data points, while dimensionality reduction algorithms simplify complex datasets, making subsequent analysis more manageable.

Reinforcement learning algorithms follow a different approach, learning from interactions with an environment to maximize a reward signal. By iteratively exploring actions and receiving feedback on their outcomes, these algorithms develop strategies that lead to the most favorable outcomes.

Deep learning algorithms are a subset of machine learning, specifically utilizing neural networks with multiple layers to extract intricate features from raw data. Convolutional Neural Networks (CNNs) excel in image analysis, Recurrent Neural Networks (RNNs) specialize in sequence data, and Long Short-Term Memory (LSTM) networks overcome the limitations of traditional RNNs.

The selection of a learning algorithm depends on the problem's nature, the available data, and the desired outcome. Each algorithm comes with its strengths and limitations, paving the way for a dynamic and diverse landscape of machine learning applications across various industries, from healthcare and finance to manufacturing and entertainment.

2.2. Crucial techniques

Deep learning methods have significantly transformed medical image analysis by harnessing intricate neural networks to automatically extract intricate features from raw data. Here's a closer look at a few crucial techniques:

2.2.1. Artificial Neural Networks (ANNs). Artificial Neural Networks were inspired by the interconnected neurons of the human brain. They have evolved from simple models to complex architectures with deep layers. ANNs are widely used for various applications, including image recognition and medical image analysis. These networks can learn intricate patterns from data and adapt to variations, offering the advantage of generalization. However, they require large datasets and computational resources for training, and their complex structure can lead to overfitting.

2.2.2. Convolutional Neural Networks (CNNs). Convolutional Neural Networks were initially inspired by the visual cortex's receptive field mechanism. They have gained immense popularity due to their

DOI: 10.54254/2755-2721/54/20241121

remarkable performance in image analysis tasks. CNNs automatically learn hierarchical features from images, capturing intricate details. This characteristic makes them a perfect fit for medical image analysis tasks like segmentation and detection. While their automatic feature learning is advantageous, they demand substantial data for training and can be computationally intensive.

2.2.3. Recurrent Neural Networks (RNNs). Recurrent Neural Networks were designed to tackle sequence data, introducing memory cells to capture temporal dependencies. They have applications in time series analysis, speech recognition, and more. RNNs retain memory of past inputs, making them suitable for sequential data. However, traditional RNNs suffer from vanishing and exploding gradient issues, which limits their effectiveness in learning long sequences.

2.2.4. Long Short-Term Memory (LSTM) Networks. Long Short-Term Memory Networks (LSTMs) were developed to address the limitations of traditional RNNs. LSTMs employ gating mechanisms, such as update, forget, and output gates, to control information flow and mitigate vanishing gradient problems. They are well-suited for tasks involving sequence prediction, language modeling, and video analysis. LSTMs excel in capturing long-range dependencies in data, making them a powerful tool for dynamic data analysis.

These deep learning techniques, with their distinctive characteristics and capabilities, have reshaped the landscape of medical image analysis, propelling the field toward unprecedented accuracy and insights.

3. Deep Learning Uses in Medical Imaging

3.1. Deep Learning's Transformative Role in Medical Image Analysis

One of the pioneering areas where deep learning has made profound strides is medical image analysis, particularly in the realm of exam classification. In such scenarios, diagnostic assessments involve one or multiple images, each linked to a singular diagnostic outcome, such as the presence or absence of a specific medical condition. Despite working with relatively smaller datasets compared to broader computer vision tasks, the integration of deep learning, particularly through transfer learning, has significantly advanced the accuracy and efficiency of analysis.

Transfer learning, a pivotal strategy, capitalizes on the power of pre-trained networks, initially designed for natural images, to overcome the traditional data-intensive nature of deep network training. This approach manifests in two main strategies: (1) utilizing pre-trained networks as potent feature extractors and (2) refining their performance on medical data through a process known as fine-tuning. The former strategy bears the additional advantage of negating the need to extensively train a deep network, as the extracted features seamlessly fit into existing image analysis pipelines. While both strategies are widespread, comprehensive head-to-head assessments of their comparative efficacy remain limited.

Intriguingly, investigations led by and have yielded conflicting yet insightful outcomes[1-2]. Antony's work demonstrates the superiority of fine-tuning in assessing grades of knee osteoarthritis, yielding a 57.6% accuracy. Conversely, Kim's study showcases the prowess of convolutional neural networks (CNNs) as effective feature extractors in cytopathology image classification.

CNNs have emerged as the cornerstone of modern medical image analysis, progressively outshining alternative architectures. Their ascendancy is evident across diverse applications, spanning brain MRI, retinal imaging, digital pathology, and lung CT scans. A compelling trend is the development of custom CNN architectures from scratch, demonstrating a shift from sole reliance on pre-trained networks. Notably, some researchers have ingeniously tailored networks to exploit inherent structural characteristics in medical images, resulting in remarkable breakthroughs.

In summation, CNNs have solidified their status as the gold standard in exam classification within medical image analysis. Their adaptive nature, coupled with the ability to harness intrinsic data structures, underscores their pivotal role in redefining the boundaries of accuracy and insights in medical diagnostics and research.

3.2. Deep Learning's Precision in Medical Image Localization

Within the realm of medical image analysis, deep learning has brought significant advancements to the accurate localization of organs, regions, and landmarks – a crucial task for segmentation and clinical workflows such as therapy planning.

Tackling the intricacies of 3D data inherent in medical imaging, innovative methods have emerged. employed convolutional neural networks (CNNs) on distinct 2D MRI slices to discern landmarks within a 3D context [3]. Building on this, expanded the approach to localize regions of interest (ROIs) in 3D CT volumes through 2D parsing [4], overcoming data limitations through the application of pretrained CNNs and restricted Boltzmann machines (RBMs) [5-7].

The pursuit of direct localization led to intriguing strategies. Payer et al.deployed CNNs for landmark regression by predicting Gaussian representations[8],while explored reinforcement learning for landmark identification across varied tasks[9].

Navigating the complexities of 3D data, Zheng et al. streamlined convolutional approaches for detecting carotid artery bifurcations in CT scans [10]. Ghesu et al. introduced a sparse adaptive deep neural network, driven by marginal space learning [11], to pinpoint the aortic valve's location in 3D transesophageal echocardiograms.

Temporal localization also witnessed advancements. CNNs showcased their capability by identifying scan planes and focal points in videos [12]. Recurrent neural networks (RNNs), specifically long short-term memory (LSTM) models, capitalized on the temporal context of medical videos [13-14].

In conclusion, the dominance of CNNs in 2D classification for localization remains evident. Recent progress in refining learning processes for pinpoint precision holds promise. These innovative adaptations underscore the adaptability of deep learning across various localization tasks. Furthermore, RNNs, including multidimensional RNNs, display potential not only in temporal but also spatial localization.

3.3. Deep Learning's Prospects in Content-Based Image Retrieval for Medical Images

With the potential to unearth related case histories, clarify unusual illnesses, and ultimately improve patient treatment, content-based image retrieval (CBIR) has emerged as a promising method for sifting through large medical datasets. However, the main obstacle to the development of CBIR is the conversion of unprocessed pixel-level data into meaningful feature representations that may be connected to pertinent concepts.

The deep convolutional neural network (CNN) models have attracted a lot of interest in this endeavor because of their aptitude for collecting complicated characteristics at many levels of abstraction. Currently, CBIR techniques use CNNs that have already been trained to extract feature descriptors from medical pictures. In order to extract features from fully-connected layers, Anavi et al. and Liu et al. [15–16] used a five-layer CNN and focused their approaches on X-ray image databases. Anavi et al. used a one-vs-all support vector machine (SVM) classifier to construct the distance measure, using the last layer of a pre-trained network to get the best results. They specifically emphasized how adding gender information to CNN features improved performance.

Liu et al. (2016b) used a bespoke CNN and the penultimate fully-connected layer to categorize X-rays into 193 categories. Despite using descriptor binarization and Hamming separation values to produce descriptive feature vectors, their performance fell short of the state-of-the-art due to their modest patch sizes (96 pixels).

CNN feature descriptors and hashing-forests were combined in a novel CBIR technique by Shah et al. [17]. This technique produced a thorough feature matrix across all prostate MRI volumes by extracting 1000 features from overlapping areas within the volumes. This matrix was then compressed using hashing trees into unique descriptors for each volume.

DOI: 10.54254/2755-2721/54/20241121

While deep learning approaches have not yet been completely utilized in content-based image retrieval for medical pictures, triumphs in other fields indicate that game-changing innovations are just around the corner. Directly training deep networks that are specifically designed for the retrieval job itself would be an interesting direction for further investigation.

4. Versatility and Extensibility of Deep Learning in Varied Applications

The concluding section sheds light on papers that span a multitude of applications and diverse domains. A striking revelation is that a singular deep learning architecture or methodology can seamlessly transcend different tasks, underscoring the adaptability and extensive applicability of deep learning. Intriguingly, pre-trained architectures prove useful, even when initially tailored for images from entirely dissimilar realms. Authors often probe the ramifications of fine-tuning networks through training with a confined dataset from the specific application domain. A prevalent trend involves amalgamating CNN-derived features with conventional ones, ushering in a fusion of modern and traditional approaches.

The surge of obstetric applications captures attention, focusing primarily on fundamental aspects such as optimal frame selection from ultrasound (US) streams. This groundwork serves as a precursor to the anticipated upsurge of automated measurements within these US sequences through the prowess of deep learning. Another noteworthy domain undergoing rapid transformation is dermoscopic image analysis. Once regarded as a challenging feat for machines, the accurate diagnosis of skin cancer from photographs has made substantial strides with the aid of deep networks. This advancement extends to encompassing both standard photographic images and dermoscopic ones, signifying a remarkable evolution. Particularly noteworthy is the accomplishment of Esteva et al., who achieved parity with a panel of 30 board-certified dermatologists. This achievement was realized by training Google's Inception v3 on an extensive dataset of unprecedented scale, representing an exponential leap compared to prior research endeavors.

5. Conclusion

This survey has traversed the wide spectrum of deep learning applications in medical image analysis, unveiling a landscape ripe with transformative potential. Deep learning, and particularly Convolutional Neural Networks (CNNs), have emerged as powerful tools across diverse tasks, from image segmentation to disease classification.

The synthesis of medical images through Generative Adversarial Networks (GANs) has opened doors to augmenting datasets and enhancing model training. Integrating multi-modal data and leveraging temporal and spatial information through RNNs and 3D convolutional architectures has further broadened the horizons of medical image analysis.

Yet, challenges persist, especially in data-scarce domains. While transfer learning and fine-tuning have helped mitigate this limitation, there's room for improvement. Combining traditional feature extraction methods with deep learning showcases the synergy between these approaches.

Exploring organ and landmark localization, content-based image retrieval, and various other applications has illuminated deep learning's adaptability to diverse medical imaging needs. The prospect of directly training deep networks for specific tasks looms, promising tailored solutions.

However, the author acknowledges certain shortcomings in this survey. The breadth of deep learning applications demands in-depth individual exploration, and we've provided a broad overview. Future surveys could delve deeper into specific domains. Moreover, the rapid evolution of deep learning techniques requires ongoing updates to stay current.

Looking ahead, the future of medical image analysis with deep learning appears promising. Emphasizing robustness, interpretability, and addressing ethical concerns should be at the forefront of research. Collaboration between medical and computer science communities will be crucial in shaping the future of healthcare through innovative deep learning applications.

In summary, deep learning has already reshaped medical image analysis and holds immense potential for the future. Continual research, adaptation, and collaboration will be the driving forces propelling this field toward new horizons, ultimately benefiting patient care and healthcare as a whole.

References

[1] Antony, J., McGuinness, K., Connor, N. E. O., Moran, K., 2016. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. arXiv:1609.02469.
[2] Kim, E., Cortre-Real, M., Baloch, Z., 2016a. A deep semantic mobile application for thyroid cytopathology. In: Medical Imaging. Vol. 9789 of Proceedings of the SPIE. p. 97890A.
[3] Yang, D., Zhang, S., Yan, Z., Tan, C., Li, K., Metaxas, D., 2015. Automated anatomical landmark detection on distal femur surface using convolutional neural network. In: IEEE Int Symp Biomedical Imaging. pp. 17–21.
[4] de Vos, B. D., Wolterink, J. M., de Jong, P. A., Viergever, M. A., Isgum, I., 2016b. 2D image classification for 3D anatomy localization: employing deep convolutional neural networks. In: Medical Imaging. Vol. 9784 of Proceedings of the SPIE. p. 97841Y.
[5] Cai, Y., Landis, M., Laidley, D. T., Kornecki, A., Lum, A., Li, S., 2016b. Multi-modal vertebrae recognition using transformed deep convolution network. Comput Med Imaging Graph 51.
[6] Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P. A., 2015b. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J Biomed Health Inform 19 (5), 1627–1636.
[7] Kumar, A., Sridar, P., Quinton, A., Kumar, R. K., Feng, D., Nanan, R., Kim, J., 2016. Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks. In: IEEE Int Symp Biomedical Imaging. pp. 791–794.
[8] Payer, C., Stern, D., Bischof, H., Urschler, M., 2016. Regressing heatmaps for multiple landmark localization using CNNs. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci. pp. 230–238.
[9] Ghesu, F. C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., Comaniciu, D., 2016a. An artificial agent for anatomical landmark detection in medical images. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci.
[10] Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D., 2015. 3D deep learning for efficient and robust landmark detection in volumetric data. In: Med Image Comput Comput Assist Interv. Vol. 9349 of Lect Notes Comput Sci. pp. 565–572.
[11] Ghesu, F. C., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J., Comaniciu, D., 2016b. Marginal space deep learning: Efficient architecture for volumetric image parsing. IEEE Trans Med Imaging 35, 1217–1228.
[12] Baumgartner, C. F., Kamnitsas, K., Matthew, J., Smith, S., Kainz, B., Rueckert, D., 2016. Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci. pp. 203–211.
[13] Chen, H., Dou, Q., Ni, D., Cheng, J.-Z., Qin, J., Li, S., Heng, P.-A.,2015a. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: Med Image Comput Comput Assist Interv. Vol. 9349 of Lect Notes Comput Sci. Cham, pp. 507–514.
[14] Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S., 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci. pp. 264–272.
[15] Anavi, Y., Kogan, I., Gelbart, E., Geva, O., Greenspan, H., 2016. Visualizing and enhancing a deep learning framework using patients age and gender for chest X-ray image retrieval. In: Medical Imaging. Vol. 9785 of Proceedings of the SPIE. p. 978510.

DOI: 10.54254/2755-2721/54/20241121

[16] Liu, X., Tizhoosh, H. R., Kofman, J., 2016b. Generating binary tags for fast medical image retrieval based on convolutional nets and Radon transform. In: International Joint Conference on Neural Networks. ArXiv:1604.04676.

[17] Shah, A., Conjeti, S., Navab, N., Katouzian, A., 2016. Deeply learnt hashing forests for content based image retrieval in prostate MR images. In: Medical Imaging. Vol. 9784 of Proceedings of the SPIE. p. 978414.