Artificial intelligence in orthopedics: three strategies for deep learning with orthopedic specific imaging

Sunho Ko^{1,2} · Ayoosh Pareek³ · Du Hyun Ro^{1,2} · Yining Lu³ · Christopher L. Camp³ · R. Kyle Martin⁴ · Aaron J. Krych³

Received: 24 November 2021 / Accepted: 2 December 2021 / Published online: 13 January 2022

© The Author(s) under exclusive licence to European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2021

Artificial intelligence (AI) has achieved remarkable results throughout society, including within the field of medicine. As the techniques advance, it is not uncommon for AI to outperform clinicians under certain conditions [2]. A branch of AI, known as machine learning, denotes the ability of a machine to identify relationships between data without explicit criteria. This relationship identification typically improves with increasing experience and data and allows algorithms to model relationships which may otherwise be too complex for standard statistical methods. Deep learning is a field of machine learning that refers to a model with an artificial neural network structure and mimics the human brain's neural connections (Fig. 1). The most important determinant of conventional machine learning algorithm performance (other than data quality and quantity) is appropriate selection of features. If the feature selection process is executed appropriately, it is possible to achieve sufficiently effective performance regardless of type of model used. On the other hand, if feature selection is unsuccessful, it is difficult to achieve adequate performance, irrespective of the popularity or purported capacity of the candidate algorithm. Currently, there is no gold standard for the process of feature selection. Thus, there is still a need for careful methodology that outlines the technical and medical knowledge when utilizing traditional machine learning algorithms. Conversely, deep learning has the advantage of end-to-end analysis using input data without the feature selection process. It provides the advantage of not having to rely strictly on feature

selection as it utilizes all available parameters. However, deep learning also has an entry barrier that requires data preparation for training. In addition, securing a high-performance graphics processing unit (GPU) for an efficient experiment is important as model training times and costs can often become increasingly burdensome [3].

Computer vision associated deep learning, especially convolutional neural networks (CNN), is more widely used than conventional machine learning algorithms due to improved performance. A CNN is a multi-layer network constituted of (1) convolutional layer, (2) pooling layer, and (3) fully connected layer. Convolutional and pooling layers are used for feature extraction and dimension reduction, respectively. The fully connected layer receives a reduced feature map from the aforementioned layers and provides the final outcome of interest (Fig. 2). A CNN does not provide intrinsic interpretability (i.e., black box), so additional algorithms have been developed to address this problem. Gradient-weighted Class Activation Mapping (Grad-CAM) method is one such popular computer vision algorithm with built-in interpretability provisions [11]. Grad-CAM produces images which highlight important regions that serve as input for prediction capabilities.

There are three main strategies for addressing computer vision problems with deep learning: (1) classification, (2) object detection, and (3) segmentation. First, classification is used to identify the appropriate class the input image corresponds to. Next, object detection identifies the presence or absence of a specific object and marks its location with a bounding box. Lastly, segmentation reports the exact pixel-wide margin of an object. For example, let us think of identifying a meniscus tear from knee MRI, the classification network is responsible for assigning the label of “meniscus tear” vs “no meniscus tear”, (Fig. 3a), the object detection network locates and classifies the meniscus tear via the use of a bounding box (Fig. 3b), and the segmentation network can demonstrate the exact location and extent of the meniscus tear (Fig. 3c). With this editorial, we will introduce

✉ Aaron J. Krych
krych.aaron@mayo.edu

¹ Department of Orthopedic Surgery, Seoul National University Hospital, Seoul, South Korea

² CONNECTEVE Co., Ltd, Seoul, South Korea

³ Department of Orthopedic Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

⁴ Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN, USA

Fig. 1 Venn diagram demonstrating the relationships and typical examples of machine learning and deep learning. Deep learning is a field of machine learning that refers to a model with an artificial neural network structure that mimics the human brain's neural connections

the deep learning technologies used in orthopedic medical imaging in relation to the three strategies above and provide examples of how they are used in addition to the associated algorithms/neural networks. This simplified workflow can be extended to other more complex imaging tasks. It is meant to be a framework to think about computer vision problems, though is not meant to be comprehensive.

Classification

ResNet is one of the most well-known and recognized classification neural networks [4]. A classification network is commonly used to assess the presence of abnormalities, classify the type of abnormality, or grade the severity. The labeling process of the classification network is the easiest among the three strategies, as only the class of each image need to be labeled. For example, Chung et al. used 1,891 plain anteroposterior shoulder radiographs to determine the presence of proximal humerus fracture (yes/no) and the type of fracture [1]. The deep learning model perfectly identified

the presence of fracture (accuracy = 1.00), and the network's performance on classifying the type of fracture was at a similar level to that of an orthopedic surgeon specializing in the shoulder (accuracy 65–86%).

Object detection

The most widely known object detection neural network is YOLO (You Only Look Once) [9]. YOLO annotates a target object by a bounding box and reports the category the target object belongs to. As inferred by its name, the YOLO network works rapidly, allowing real-time analysis of images and videos. Due to this advantage, it is mainly used to assess surgical videos in medical fields, such as in the real-time detection of surgical instruments, anatomical structures, and the stage of surgery. However, compared to the classification network, YOLO network takes a relatively longer time to prepare (label) the data as the class of the objects needs to be annotated. For example, Hossain et al. developed a deep learning model that detects surgical instruments in real-time from 16 total knee arthroplasty images recorded at 25 fps [6]. This deep learning model classified 31 surgical instruments at 87.6% mean average precision (MAP). Real-time analysis of surgical videos can be further developed into automated intraoperative assistance and have great potential such as being an automated feedback system for trainees.

Segmentation

U-net is one of the most powerful and established segmentation neural networks [10]. A segmentation network reports the exact pixel-wise probability of the presence of the target object. Due to this advantage, it is mainly used for precise tasks—such as identifying the contour of organs or the extent of tumors. However, to train the network, pixel-wise annotation of train images is required, which can consume

Fig. 2 Architecture of a Convolutional Neural Network (CNN). CNN is multi-layer network constituted of a (1) convolutional layer, (2) pooling layer, and (3) fully connected layer. Convolutional and pooling layers are used for feature extraction and dimension reduction, respectively, while fully connected layers receive a reduced feature map and provide the outcome of interest

Fig. 3 Three approaches to solving computer vision problems with deep learning: **a** classification, **b** object detection, and **c** segmentation. Classification is used to identify the appropriate class the input image corresponds to. Object detection identifies the presence or absence of a specific object and marks its location with a bounding box (red). Segmentation reports the exact pixel-wide margin of an object (green)

(a) Classification

Meniscus tear
Present

(b) Object Detection

(c) Segmentation

significant resources, and is both humanly and computationally expensive. For example, Norman et al. developed a model for segmentation of cartilage and meniscus from 638 knee MR imaging [8]. The volume and thickness of cartilage were automatically assessed by the network with a good performance (Dice coefficients 0.770–0.878 in cartilage and 0.809, 0.753 for lateral and medial meniscus, respectively). In another example, Hemke et al. utilized a segmentation network to analyze body composition. Pelvic muscles, fat, and bone was segmented from a pelvic CT image with an excellent performance (Dice score 0.91–0.97) [5].

Combination strategies

These three strategies are commonly used to address computer vision problems and achieve deep learning solutions in the orthopedic field. Although introduced separately, they can also be used in combination with one another and often are. For example, Liu et al. used a segmentation network from a knee MR image to identify the cartilage region [7]. The cartilage regions were cropped into a small square patch for the classification network to detect the cartilage lesion. Identification of cartilage area allowed the classification network to focus on cartilage area, yielding better performances.

Conclusion

In summary, (1) there are three common strategies to assess medical images using deep learning: classification, object detection, and segmentation, and (2) appropriate deep learning algorithms/networks should be chosen in response to the purpose of the study. Currently, despite the significant implications deep learning technology can have on orthopedics given our dependence on medical imaging for diagnosis and management, deep learning research in orthopedic surgery remains relatively sparse. Therefore, while caution should be exercised in adoption of new technology, we must prioritize deep learning as a tool to maintain the leadership role of orthopedic surgery in the musculoskeletal space.

Funding There is no funding sources.

Declarations

Conflict of interest A.J.K. has received research support from Aesculap/B.Braun, the Arthritis Foundation, Ceterix, Exactech, Gemini Medical, and Histogenics; consulting fees from Arthrex, DePuy, JRF, Musculoskeletal Transplant Foundation, Responsive Arthroscopy, and Vericel; speaking fees from Arthrex and Musculoskeletal Transplant Foundation; royalties from Arthrex; and honoraria from Vericel and JRF. C.L.C. has received education payments from Arthrex, nonconsulting fees from Arthrex, and hospitality payments from Stryker and Zimmer Biomet.

Ethical approval The study was exempt from institutional approval due to lack of data/biological material utilized and was performance in accordance with ethical standards.

References

1. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP et al (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop* 89:468–473
2. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318:2199–2210
3. García-Martín E, Rodrigues CF, Riley G, Grahm H (2019) Estimation of energy consumption in machine learning. *J Parallel Distrib Comput* 134:75–88
4. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016
5. Hemke R, Buckless CG, Tsao A, Wang B, Torriani M (2020) Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. *Skeletal Radiol* 49:387–395
6. Hossain M, Nishio S, Hiranaka T, Kobashi S (2018) Real-time surgical tools recognition in total knee arthroplasty using deep neural networks. Paper presented at: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR); 25–29 June 2018
7. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A et al (2018) Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* 289:160–169
8. Norman B, Pedoia V, Majumdar S (2018) Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology* 288:177–185
9. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016
10. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015, Cham
11. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vision* 128:336–359

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.