Japanese Journal of Radiology (2019) 37:103–108
https://doi.org/10.1007/s11604-018-0804-6

INVITED REVIEW

Recent technical development of artificial intelligence for diagnostic medical imaging

Norio Nakata[1]

Received: 8 November 2018 / Accepted: 14 December 2018 / Published online: 31 January 2019

Abstract
Deep learning has caused a third boom of artificial intelligence and great changes of diagnostic medical imaging systems such as radiology, pathology, retinal imaging, dermatology inspection, and endoscopic diagnosis will be expected in the near future. However, various attempts and new methods of deep learning have been proposed in recent years, and their progress is extremely fast. Therefore, at the initial stage when medical artificial intelligence papers were published, the artificial intelligence technology itself may be old technology or well-known general-purpose common technology. Therefore, the author has reviewed state-of-the-art computer vision papers and presentations of 2018 using deep learning technologies, which will have future clinical potentials selected from the point of view of a radiologist such as generative adversarial network, knowledge distillation, and general image data sets for supervised learning.

Keywords Artificial intelligence · Deep learning · Computer vision

Introduction

Currently, AI applications in the field of image diagnosis are mainly using deep learning. If computer scientists would like to collect information about the latest technology of deep learning, surveying CV field is the quickest way. For that purpose, it is necessary to survey academic conferences at the top level in the world in the CV field such as Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), and European Conference on Computer Vision (ECCV) [1]. Therefore, the author wishes to survey trends in such CV field and to predict the latest technology used for AI applications in image diagnosis in the near future.

In CVPR2018, papers related to "deep learning" continue to increase every year, and 424 (92.4%) papers are related to deep learning. 459 papers are published in arxiv (https://arxiv.org/) among 979 papers accepted for this year [2]. On comparing the trends in the percentage of major methods for papers published at CVPR, ICCV, and ECCV, there were 25% of papers on general deep learning in 2017, and in 2018 it was 20%; deep leaning may become out of fashion as a whole. On the other hand, GAN has rapidly increased to 8% in 2018. Long short-term memory (LSTM), which is a kind of deep learning utilized in natural language processing (NLP), decreased slightly to less than 5% in 2018 [3].

✉ Norio Nakata
nakata@jikei.ac.jp

1 Department of Radiology, The Jikei University, School of Medicine, 3-25-8, Nishi-shimbashi, Minato-ku, Tokyo 1058461, Japan

AI automation levels of diagnostic imaging

National Highway Traffic Safety Administration (NHTSA) classified the level of automation into five levels to clarify the level of automation in automated vehicles [4]. If AI in diagnostic imaging also has similar criteria in the development and evaluation process, it will be useful as a measure of evaluation. Therefore, the author has proposed the AI automation levels in diagnostic radiology from level 0 to 4 (Fig. 1) in the report produced by the meeting of the Ministry of Health, Labor and Welfare, the author of which was one of the members [5]. Level 0 is image preprocessing without a computer-assisted diagnosis. Level 0 is further classified into two types: image preprocessing with AI (level 0+) and image preprocessing without AI (level 0−). Recently, new researches of synthetic imaging using GAN, which is image preprocessing with AI (level 0+), have progressed rapidly. Level 1 is the computer-assisted diagnosis

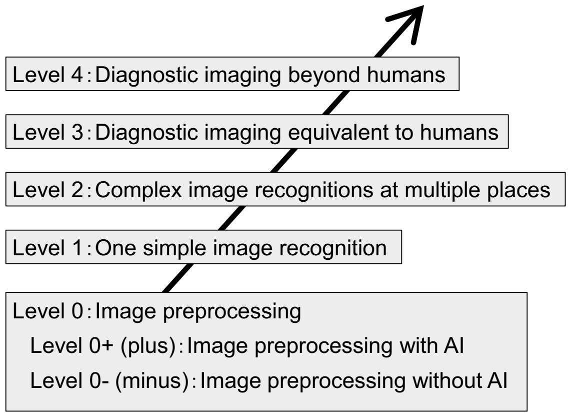

Fig. 1 AI automation levels of diagnostic imaging. Classification of automation levels is an important indicator for the evaluation of research and development processes. Recently, many AI studies of level 0 or 1 in radiology have been started

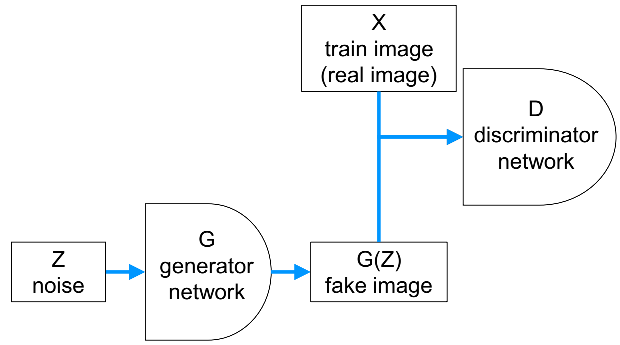

Fig. 2 Network overview of GAN. GAN has two networks, Generator and Discriminator. The Generator generates an image that causes the Discriminator to mistakenly be genuine, and the Discriminator has a role to distinguish whether it is genuine or fake. Two networks will hostile and learn. It is often compared to the creation of counterfeit bills and the police relationship to distinguish it

of only one kind of image recognition such as lung nodule detection of chest CT. Level 2 is complex image recognition at multiple places such as lung nodule, pneumonia lesions, and liver mass lesions. Level 3 is diagnostic imaging capabilities equivalent to humans. Level 4 is the diagnostic imaging capabilities beyond humans.

Changing biology and medicine using image preprocessing with AI

Level 0 technologies such as GAN are unpredictable in diagnostic radiology. However, highly advanced image preprocessing will be deeply involved and intervened in surgery, as three and four-dimensional images produced computer surgery. For example, a method of performing real-time 3D hand tracking using a GAN and convolutional neural network (CNN) from monocular RGB movies can be easily applied to surgical simulation and surgical computer-based monitoring [6]. In radiation therapy, if it is possible to perform image deformation corresponding to body movement in real time, it is considered to be useful for fine adjustment of the irradiation field. This technique is also considered to be useful for the analysis by the monitoring and structuring of the movement of the operator and the staff inside the interventional radiology suite. Additionally, highly intelligent three-dimensional images and four-dimensional image processing techniques using these AI methods are considered to be useful for research on geometrical structure analysis of proteins and others. Also, research on genome analysis, which was mere cryptanalysis, has the potential to transform it into a new discipline to analyze complex three-dimensional structure by AI.

GAN and derivation methods

GAN generates images learned from appropriate vectors by playing two networks of generator and discriminator. The generator is a network for generating images. Discriminator is a network for judging whether images are the fake or real. GAN generates images with the generator using noise as input and judges whether the discriminator is genuine or fake by the generator while mixing with real images [7] (Fig. 2). Various GAN methods derived from the original GAN have been devised as follows.

Deep convolutional generative adversarial networks (DCGAN)

The original paper of GAN does not refer to the composition of a specific network.

DCGAN has proposed a best practice to successfully learn by applying a convolution neural network to GAN [8].

Progressive growing of GAN (PGGAN)

In the learning of PGGAN, by devising the learning process so as to gradually achieve high-resolution learning, it becomes possible to generate a high-resolution image of 1024 × 1024 using PGGAN [9].

Conditional GAN (cGAN)

cGAN is a technique for generating the conditional image by giving input label to the image of the desired image to be generated [10].

Pix2Pix

Pix2pix is a type of cGAN. Given a training set which contains pairs of related images, a pix2pix model learns how to convert an image. GAN is judging a single image to input a vector. On the other hand, pix2pix gets an output image under the condition of image input and judges whether or not the pair of the input image and the output image is authentic [11].

CycleGAN

For example, when converting an image of an apple to an orange, cycleGAN does not need an image corresponding to learn; it only needs to have an apple image group and an orange image group. A similar correspondence orange image of the same posture as an apple is not necessary. The network is in a cycle shape and, in cycleGAN, an apple image is generated from the orange image and learns so that the accuracy becomes high when the apple image is returned to the orange image again [12]

StarGAN

Multiple version of cycleGAN is available for starGAN. Multiple images can be generated if cycleGAN of A and B and cycleGAN of B and C are performed so as to learn among plural domains. Therefore, starGAN is learning so that it can travel between multiple GANs in one GAN [13].

Auxiliary classifier generative adversarial network (ACGAN)

DCGAN was creating images from random noise sequences. In addition to this, ACGAN inputs image class label information to the generator and executes the auxiliary class identification task with the discriminator. This makes it possible to generate high-precision images. Learning uses 128×128 images and inputs images of ImageNet [14].

Self-attention generative adversarial networks (SAGAN)

In SAGAN, stability is improved by applying spectral normalization, which is a type of normalization, to each layer of the discriminator. Spectral normalization is also applied to the generator. By introducing a mechanism of self-attention, generation of the highest image quality up to now has been realized in the image generation of the specified class. The network using self-attention has not only local information but also a global view [15].

Stacked generative adversarial networks (StackGAN)

StackGAN generates 256×256 photo-realistic images conditioned on text descriptions.

StackGAN is composed of two stages of GAN; the first is a network which generates images from sentences, and the second is GAN which makes highly accurate images mostly written with the first one. Using StackGAN, it is possible to generate images from sentences considerably with high accuracy [16].

Progress inefficient learning method: knowledge distillation

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task [17]. Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem [18]. Recent researches on effective learning method of machine learning are not limited to conventional transfer learning and ensemble learning, but new learning methods have been proposed. These new learning methods are easy to apply to AI application development for diagnostic radiology. Knowledge distillation is model compression method in which a small model is trained to mimic a pre-trained, larger model or ensemble of models. This training set is sometimes referred to as “teacher-student”, where the large model is the teacher and the small model is the student (Fig. 3) [19]. Student modeling is one of the key factors that affect automated tutoring systems in making instructional decisions. A student model is a model to predict the probability of a student making errors on given problems. A good student model that matches with student behavior patterns often provides useful information on learning task difficulty and transfer of learning between related problems and, thus, often yields better instruction [20]. Higher precision can be expected compared with the case of normal learning without knowledge distillation. The precision comparable to the teacher model and, in some cases, the precision that exceeds the teacher are also reported [21, 22]. It has also been reported that there is a strong regularization effect when soft targets are added by distillation of knowledge. When learning was performed using only 3% of the training data, it is reported that the distortion of knowledge has correctly converged

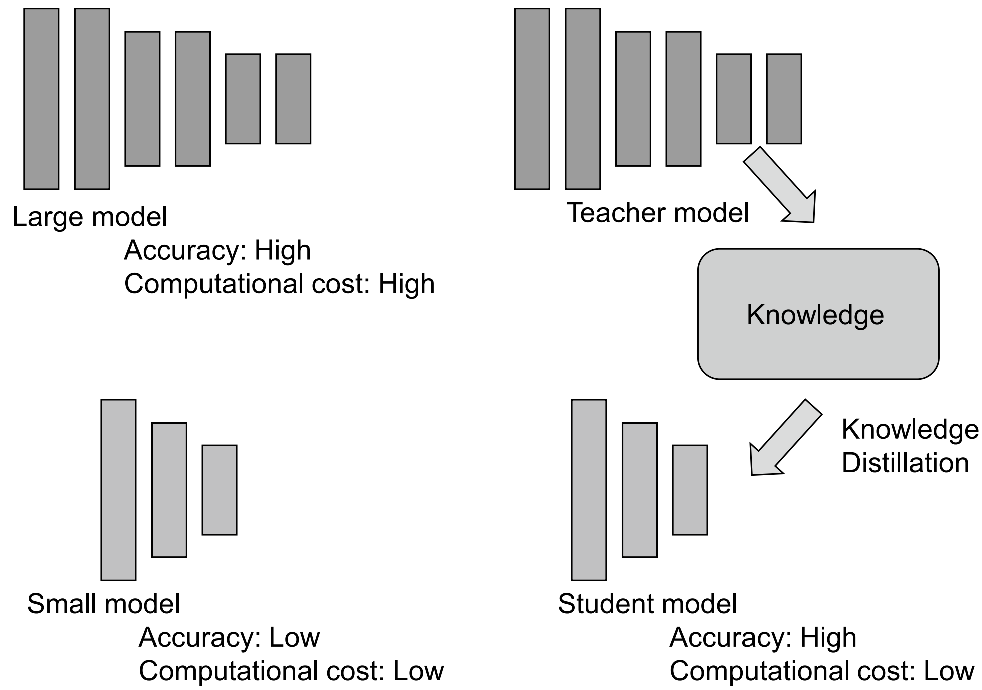

Fig. 3 The concept of knowledge distillation. Prepare a large model with high prediction accuracy as a teacher model and use that knowledge for learning a lightweight and easy-to-deploy student model. By this method, it is expected that a model with accuracy comparable to that of a teacher model can be obtained even though it is lightweight. This is the basic idea of "knowledge distillation"

where knowledge is distilled if it learns excessively without knowledge distillation [23]. Even in the case where there are huge amount of learning data with a very large number of classes, which take time to learn with a single model, the problem is divided and multiple teacher models are learned. Next, using this knowledge, it is possible to learn efficiently [23, 24].

General image data sets for supervised learning

As with progress in efficient learning method, transfer learning is also important methods for supervised learning of diagnostic medical imaging. Before supervised learning using medical images for diagnostic radiology AI, training with general image data sets is important. Therefore, some general image data sets frequently used in machine learning are listed below [25].

LabelMe

LabelMe is a project created by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) which provides a dataset of digital images with annotations. The dataset is dynamic, free to use, and open to public contribution. The most applicable use of LabelMe is in computer vision research. As of October 31, 2010, LabelMe has 187,240 images, 62,197 annotated images, and 658,992 labeled objects [26] [27].

Video annotation tool from Irvine, California (VATIC)

VATIC is a free, online, interactive video annotation tool for computer vision research that crowdsources work to Amazon's Mechanical Turk [28, 29].

ImageNet

ImageNet is an image database organized according to the WordNet hierarchy, in which each node of the hierarchy is depicted by hundreds and thousands of images [30, 31].

CoPhIR

The Content-based Photo Image Retrieval (CoPhIR) test collection has been developed to make significant tests on the scalability of the SAPIR project infrastructure Search In Audio Visual Content Using Peer-to-peer IR (SAPIR) for similarity search. The data collected so far represent the world largest multimedia metadata collection that is available for research on scalable similarity search techniques. CoPhIR consists of 106 million processed images [32, 33].

Tiny images dataset

This page has links for downloading the Tiny Images dataset, which consists of 79,302,017 images, each being a 32 × 32 color image [34].

SUN dataset

The goal of the SUN database project is to provide researchers in computer vision, human perception, cognition and neuroscience, machine learning and data mining, computer graphics, and robotics, with a comprehensive collection of annotated images covering a large variety of environmental scenes, places and the objects [35].

The MNIST database

The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples [36].

MegaFace and MF2

A face data set consists of about 5 million images. There are 700,000 people and an average of 7 images per person is prepared. Data of the bounding box surrounding the face are also provided [37].

Conclusion

Deep learning is progressing rapidly and its application range is going to extend not only to computer-assisted diagnosis in radiology but also to computer surgery and radiomics. Therefore, recent technical development of AI for diagnostic medical imaging such as various types of GAN, knowledge distillation, and general image data sets for supervised learning is reviewed.

Compliance with ethical standards

Conflict of interest The Conflict of Interest form and the Ethical Statement: no potential conflict of interest and of the commercial entities.

References

1. Top Conferences for Image Processing & Computer Vision. http://www.guide2research.com/topconf/computer-vision. Accessed 9 Nov 2018.
2. CVPR2018 http://cvpr2018.thecvf.com/. Accessed 9 Nov 2018.
3. Jordi Pont-Tuset's site—CVPR 2018: are GANs the new deep? http://jponttuset.cat/are-gans-the-new-deep/. Accessed 9 Nov 2018.
4. National Highway Traffic Safety Administration, Preliminary Statement of Policy, Concerning Automated Vehicles https://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf. Accessed 9 Nov 2018.
5. Japanese Ministry of Health, Labor and Welfare, Promotion of AI in health care field report https://www.mhlw.go.jp/file/05-Shingikai-10601000-Daijinkanboukouseikagakuka-Kouseikagakuka/0000169230.pdf. Accessed 9 Nov 2018.
6. Mueller F, Bernard F, Sotnychenko O, Mehta D, Sridhar S, Casas D, Theobalt C. Ganerated hands for real-time 3d hand tracking from monocular RGB. In: O'Conner L (ed) Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Computer Society Conference Publishing Services; 2018. pp. 49–59.
7. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Bengio Y. Generative adversarial nets. Adv Neural Inf Process Syst. 2014;2014:2672–80.
8. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
9. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of GANs for improved quality. Stability, and variation. arXiv preprint arXiv:710.10196.
10. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
11. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004.
12. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593.
13. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J (2017) Stargan: unified generative adversarial networks for multi-domain image-to-image translation. arXiv preprint, 1711.
14. Odena A, Olah C, Shlens J (2017) Conditional image synthesis with auxiliary classifier gans. arXiv preprint arXiv:1610.09585.
15. Zhang H, Goodfellow I, Metaxas D, Odena A (2018) Self-Attention Generative Adversarial Networks. arXiv preprint arXiv:1805.08318.
16. Huang X, Li Y, Poursaeed O, Hopcroft JE, Belongie SJ. Stacked generative adversarial networks. CVPR. 2017;2:3.
17. A Gentle Introduction to Transfer Learning for Deep Learning https://machinelearningmastery.com/transfer-learning-for-deep-learning/. Accessed 9 Nov 2018.
18. Ensemble learning—scholarpedia http://www.scholarpedia.org/article/Ensemble_learning. Accessed 9 Nov 2018.
19. Knowledge—Neural Network Distiller https://nervanasystems.github.io/distiller/knowledge_distillation/index.html. Accessed 9 Nov 2018.
20. Li N, Cohen WW, Koedinger KR, Matsuda N (2011) A machine learning approach for automatic student model discovery. In Edm 31–40. https://wwcohen.github.io/postscript/edm-2011.pdf. Accessed 9 Nov 2018.
21. Romero A, Ballas N, Kahou SE, Chassang A, Gatta C, Bengio Y (2014) Fitnets: hints for thin deep nets. arXiv preprint arXiv:1412.6550.
22. Furlanello T, Lipton ZC, Tschannen M, Itti L, Anandkumar A (2018) Born again neural networks. arXiv preprint arXiv:1805.04770.
23. Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

24. Gao J, Li Z, Nevatia R (2017) Knowledge concentration: learning 100 K object classifiers in a single CNN. arXiv preprint arXiv:1711.07607.
25. arXiv Times/datasets at master, arXivTimes/arXivTimes, GitHub https://github.com/arXivTimes/arXivTimes/tree/master/datasets. Accessed 9 Nov 2018.
26. LabelMe. The Open annotation tool http://labelme.csail.mit.edu/Release3.0/. Accessed 9 Nov 2018.
27. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. Int J Comput Vision. 2008;77(1–3):157–73.
28. VATIC—video annotation tool—UC Irvine http://www.cs.columbia.edu/~vondrick/vatic/. Accessed 9 Nov 2018.
29. Vondrick C, Patterson D, Ramanan D. Efficiently scaling up crowdsourced video annotation. Int J Comput Vision. 2013;101(1):184–204.
30. ImageNet http://www.image-net.org/. Accessed 9 Nov 2018.
31. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: Flynn P, Mortensen E (eds.) Proceedings / CVPR, IEEE computer society conference on computer vision and pattern recognition (CVPR 2009), 20–25 June 2009, Miami, Florida, USA, pp 248–55.
32. CoPhIR—COntent-based Photo Image Retrieval http://cophir.isti.cnr.it/. Accessed 9 Nov 2018.
33. Bolettieri P, Esuli A, Falchi F, Lucchese C, Perego R, Piccioli T, Rabitti F (2009) CoPhIR: a test collection for content-based image retrieval. arXiv preprint arXiv:0905.4627.
34. Tiny Images Dataset http://horatio.cs.nyu.edu/mit/tiny/data/. Accessed 9 Nov 2018.
35. SUN Database https://groups.csail.mit.edu/vision/SUN/. Accessed 9 Nov 2018.
36. MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges http://yann.lecun.com/exdb/mnist/. Accessed 9 Nov 2018.
37. MegaFace and MF2: Million-Scale Face Recognition http://megaface.cs.washington.edu/. Accessed 9 Nov 2018.