Computer Vision Applications in Medical Diagnostics

Pascal C. Weinberger

1   Overview of Computer Vision

The field of computer vision has existed since the late 1940s, initially intending to help robots understand the environment they were operating in. The earliest versions of successful computer vision algorithms were largely utilizing pixel-based, statistical analysis methods, like, for example, averaging the gradients in a small area of the picture to detect boundaries of objects based on "sudden changes" of pixel-values. These methods were already good enough to help early robots navigate and were therefore used in experimental designs for video surveillance and self-driving cars, etc. However, these methods were lacking a few key features: They were relatively low-resolution and were not able to recognize smaller differences in objects, see e.g. (Minsky 1961).

These problems became solvable with the relatively recent advent of deep learning methods: Driven by innovations which enabled the scaling of the amounts of data we can create, store, and compute (like, for example, the invention of GPUs—graphics processing units) we are now able to process a much larger amount of correlations then before. This increased performance has enabled deep neural networks (DNNs), to work sufficiently fast to train large networks with millions (and even billions) of variable weights in reasonable times to allow for massive innovation in computer vision and the field of AI at large.

Deep neural networks in computer vision primarily rely on learning correlations of neighboring areas in an image in a hierarchical structure with increasing abstraction: These "convolutional neural networks", depicted in Fig. 1, have then been able to recognize for example handwritten digits with no prior human input in setting the parameters itself. These are learned through the method of "backpropagation"

P. C. Weinberger (✉)
Liederbach, Germany
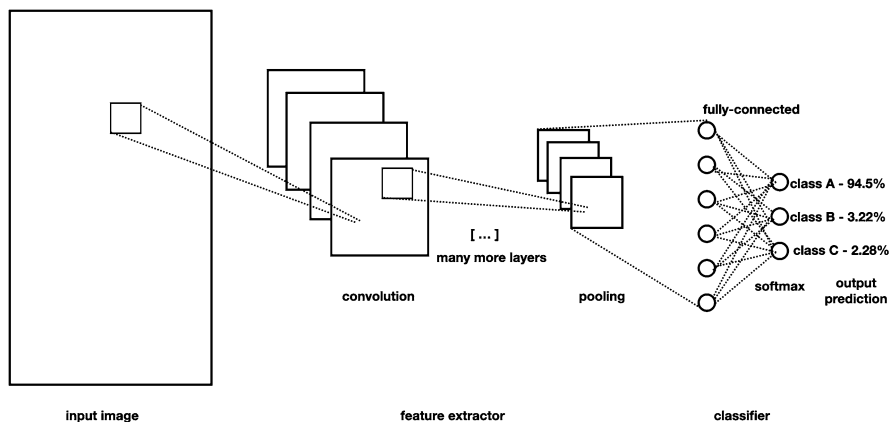e-mail: pascal@weinberger.pw

Fig. 1 Architecture of a convolutional network, which served as important part of the foundation for most computer vision work done today. This simplified sample network demonstrates the different stages and layers involved in a CNN: convolutional feature-extractors, pooling layers and finally a fully connected classifier with a softmax output layer. There are many other variations and architectures of CNN networks used today, but most of the underlying principles remain similar. Source: Author

(Rumelhart et al. 1986). This method was then applied to recognizing handwritten characters (LeCun 1986). That achievement unlocked a huge field of research and applications for networks to be trained end-to-end with only limited human input, which is mainly the case for the design of the architecture and learning functions of the DNNs.

Today, computer vision experts all over the world leverage deep learning, coupled with many more recent advances, like hyperparameter tuning, and neural architecture search, which both optimize the hyperparameters (general architecture and parameters like the model's learning rate, etc., which significantly can improve a model's performance if chosen correctly) and therefore allow for rapid scaling and experimentation on large (and small) datasets. Additionally, cloud providers today provide the computing infrastructure needed to train such DNNs very easily and with a limited upfront investment.

Notably, open and free access to data, deep learning frameworks, and pre-trained models is another trend which steadily promotes innovation. With the new platform of massive open online courses (MOOCs), which is largely driven through platforms built by computer vision experts,[1] made learning about computer vision, deep learning, and the skills needed to be active in the field, very easily accessible to anyone, anywhere. Since then, a lot of effort has gone into building frameworks, that simplify the usage of deep learning and other machine learning methods,

---

[1]Udacity, founded by Sebastian Thrun. Coursera, founded by Daphne Koller and Andrew Ng.

by providing a sufficient abstraction to the programmer. As these programming frameworks evolve, they become better and more efficient to use.

Today, the most used and most popular deep learning frameworks include Keras.io and TensorFlow (both maintained by Google engineers), PyTorch (maintained by Facebook), and MxNet (maintained by Amazon). These frameworks provide all the necessary tools to get started in computer vision with relatively low effort. One train a computer vision algorithm for Pneumonia Classification on X-ray images with less than 100 lines of code with Keras for example.[2]

Additionally, numerous "model hubs" are now freely available, which allow researchers or engineers to download pre-trained networks and then fine-tune them to specific use cases or automatically leveraging any of the "AutoML"—automated machine learning—frameworks to have their dataset modeled to the optimum performance.

Over the last 10 years, all of these innovations and platforms helped to unlock this field of research and development to hundreds of thousands of enthusiasts, engineers, and scholars. This has led to many interesting and impactful developments around the problems of medical diagnosis as well, some of which we will discuss in detail now.

2   Computer Vision in Healthcare Diagnostics: Applications

The problems of healthcare diagnostics are some of the most interesting, and impactful areas of research and application in computer vision.

### 2.1   Applying Computer Vision in Practice

When applying computer vision in practice, one often follows the workflow depicted in Fig. 2.

Generally, in computer vision, we categorize into different types of problems:

- Classifications with supervised learning: When you give the learning algorithm all the labels of your training examples to learn from.
- Classifications with unsupervised learning: When you do not have the labels of your training examples when learning, so the learning algorithm tries to identify what classes exist.
- Classifications with semi-supervised learning: When you have some, but not all labels; sort of a mix of supervised and unsupervised learning.
- Classifications with weakly supervised learning: When you try to use the signal from a small, often noisy, labeled dataset to infer the labels of a broader unlabeled dataset, to then use that for supervised learning.

---

[2]https://keras.io/examples/vision/xray_classification_with_tpus/
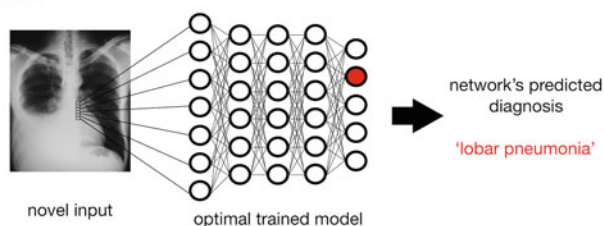
Training phase
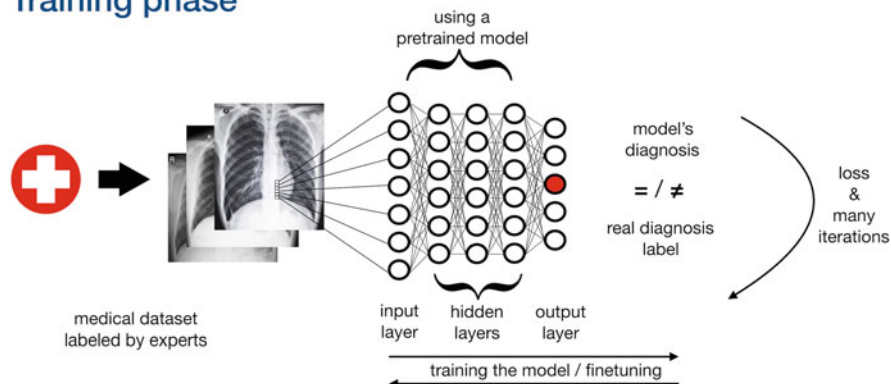
Diagnostic tool

Fig. 2 Common workflow in practice: (1) Get specific training data suitable for the problem at hand. (2) Load pre-trained model. (3) Use problem-specific training data to fine-tune the model. (4) Iterate over the data many times, minimizing the training error. Then use the model as a diagnostic tool by loading the model and running it on new data to get predictions. Source: Katharina Thoene and author

Most study problems one would encounter when learning computer vision methods will be using supervised learning: You have a lot of labeled training data (usually already labeled standard datasets) and you train a DNN to find the optimal mapping from the input vector (your image data) to a probability distribution over the possible labels (your models' prediction). This method works well if it is possible and effective (think about the time and cost it takes to label often hundreds of thousands of examples) to provide labeled training data.

However, in reality, this is very rarely the case: Often datasets are only partially labeled, and also these partial labels will be noisy and have some errors in them, like, for example, when you are getting clinical X-ray data from a hospital where the labels, or diagnosis, are only sometimes provided and then there might be also mistakes in them or errors due to the transfer from one format (often paper notes) to another (digital labels with the images). That is why a pure supervised learning algorithm will rarely be used in practice. The majority of times you will be using weakly supervised or semi-supervised learning algorithms.

2.2   Examples of Applications

Most applications today in medical diagnostics can be classified as one of these following categories:

- Detection of certain abnormal tissue, e.g. cancer detection, etc.
- Detection of (micro-)fractures, e.g. in bones, veins, etc.
- Counting of certain types of cells or tissues, e.g. count of white blood cells in probes, etc.
- Segmentation of certain tissues from the rest, e.g. seeing where and how big a cancer is, etc.
- Guided surgery through (3D) models of the tissue to help perform the surgery higher precision, often using computer vision-assisted robot arms to perform the surgery.
- Functional brain analysis, segmentation, and classification on fMRI data to identify areas of dysfunction.
- Image super-resolution, using generative adversarial networks (GANs) to produce higher resolution images, e.g. for MRI resolution (Sood et al. 2018).
- Classification-based image retrieval, using computer vision to "tag" images and data based on their content to allow for faster easier retrieval, especially with rare conditions and disorders (Müller et al. 2005).
- Substructure segmentation, to assist in identification and examination of different substructures, for example, of blood vessels in the retina which are hard to spot for humans (Fu et al. 2016).
- Motion tracking of human movements to identify abnormalities in movements like, for example, gait analysis to diagnose onset of Parkinson's disease (Kour and Arora 2019).

Early pioneers in this field include Daphne Koller,[3] Kunio Doi,[4] and many more. It is very interesting to go through their early publications to see the evolution and changes in the approaches in the field over time.

Because of the increasing media coverage and hype about AI taking over many jobs, many medical professionals fear that their jobs could be replaced soon. However, the overall goal of using computer vision in medical diagnosis is to help the medical professional to perform better and faster diagnosis, not to replace them.

Over the last few years of research, it became evident that deep learning methods are actually quite complementary to the human strengths. Therefore, there is a lot of emphasis on the idea of human-machine collaboration: Often the skills that are fairly easy for us are the hardest to learn for machines and vice versa. Humans are very strong at "seeing the bigger picture" and taking all different aspects of a patient's condition and symptoms into account when making the diagnosis, while

---

[3]http://dags.stanford.edu/publications.cgi

[4]https://profiles.uchicago.edu/profiles/display/13712530

even the best experts sometimes tend to overlook minute details in the vast amounts of available data and are limited in their capacity to compare a single patients case with many others in real time.

In contrast, trained machines detect picture anomalies within milliseconds, but fail to find the context beyond pixels. These are exactly the advantages of deep learning-based systems to help in medical diagnosis, as the machine learning algorithms are much better at using hundreds to millions of examples and other data points to base their conclusion on and can be much more sensitive to very small details in these large amounts of data. The human experts and machine learning systems can therefore form a very powerful "team" to optimize for the best possible diagnosis and outcome for the patient.

Many startups and solutions are already being integrated in today's clinical practice using this human-in-the-loop approach for medical diagnostics. Some notable ones include:

- Smart Reporting: building a reporting tool suite where radiologists are assisted by deep learning.
- Caption Health: using deep learning to help interpret and guide ultrasound examinations.
- Deep Pathology: helping pathologists evaluate their samples using deep learning to count, segment, and classify cell and tissue types.
- CellmatiQ: various products to help dentists and orthodontist identify and classify different types of problems using deep learning.
- Athelas: using computer vision to classify and count white blood cells and Lymphocytes to be able to detect infections earlier and much cheaper than usually done.
- There are many more great startups in the field of medical diagnostics using machine learning, most of which utilize the principles and methods discussed here.

With all of these applications, as well as the many more which are still at the research stage or pending approval from the regulating bodies in their markets, the most common challenge for them is very similar: Getting access to sufficient amounts of clean, well-annotated (at least partially) training data. Accessing, structuring, and assembling these training datasets is often the biggest challenge when developing a new solution using machine learning, due to the cost and time needed to get the data. Another difficulty lies in the approval to use the data for this project, and the domain experts time to label the training examples properly.

Even when that is all done properly and with many good training images, the DNNs still tend to "overfit" (e.g. model too closely to the specifics of the training datasets which were provided) which can lead to problems when scaling

the approaches to new sets of information.[5] This can be due to either the equipment used to generate the imaging data might have a small inherent bias (like different color/pixel gradings, etc. which might get picked up by the DNNs) or the underlying patient population has slightly different, which can lead to a significant reduction in the accuracy of the machine learning-based approach, and is often one of the biggest challenges when developing and bringing these applications to market.

3 Computer Vision in Healthcare Diagnostics: Opportunities and Challenges

Since the amount and quality of the available training data for each computer vision task is so critical, this therefore offers one of the biggest opportunities for improvement. There are several (sometimes opposing) forces currently at play to improve the handling of medical imaging data.

On the one hand, to make access to training data harder (by means of privacy enforcing regulations like GDPR, generally increased privacy awareness, etc.), and on the other hand to make it easier to get the right training data (open sharing platforms of anonymized data, conferences pushing to publish data along with the papers, increasing awareness of hospitals to gather and structure data, etc.). Additionally, innovations, like zero shot learning[6] or active learning techniques (Brust et al. 2018) make algorithms much more data-effective, allowing for better results with less data.

Furthermore, the interfaces of human-in-the-loop systems are becoming much better over time, allowing for more effective collaboration—both, in training models and also in the clinical production setting—and therefore higher data throughput.

The biggest challenges, besides ensuring access and quality of training data for each diagnostic problem, are currently:

- Bias: solving for unbalanced features in datasets (e.g. due to the ethnicity or gender of the majority of the patients in a training data set, etc.).
- Robustness: Correcting for overfitting due to characteristics of the individual or the imaging device.
- Scalability and access: Currently most of these advanced methods are only available in the best-resourced hospitals and areas.
- Interpretability of models: When basing the decision about medical treatments on the diagnosis provided, one has to have high confidence in the algorithm. For that, having interpretable (or explainable) algorithms is critical, but this remains one of the biggest challenges to overcome when using deep learning methods.

---

[5]https://www.sciencemag.org/news/2019/06/artificial-intelligence-could-revolutionize-medical-care-don-t-trust-it-read-your-x-ray

[6]https://www.learnopencv.com/zero-shot-learning-an-introduction/

- Regulation: ensure proper and appropriate regulation and standards for learning algorithms in diagnostics applications.
- Human–computer interaction (HCI): make the interfaces better and allow for the best possible collaboration (more below).
- Synthetic training data: Using computer graphics and simulations to generate new training data automatically, based on previously known data and principles, in order to make algorithms better and more robust without needing more expensive real-world data.
- New learning paradigm: Overall, the deep learning paradigm has offered a lot of opportunities and improvements over classical computer vision methods, however, it still lacks a few fundamental features which will be needed to enable a lot more applications and more reliability.

    It is currently unclear how exactly this new paradigm will look like, but it has become evident that one is needed beyond deep learning. Personally, I believe it will be developed out of the neuro-symbolic machine learning community.

Some of these challenges will be improved upon by building new model architectures, introducing new learning functions, and better hyperparameter tuning, etc. But some others are more fundamental and teach us a lot about the strengths and weaknesses of both human experts and learning algorithms on these tasks.

It becomes evident that the future most likely will be shaped by the collaboration of experts and machines, leveraging each other's strengths to deliver the best, fastest, and most accurate diagnostics possible. Therefore, the user experience of the interfaces providing such algorithms must be tuned to optimize the interaction between the human-skills (critical thinking, putting the data in context, having the holistic view of the patient) and the machine skills (extreme attention to detail, ability to compare with and learn from millions of other patients, steady or increasing performance with time, etc.). It might make sense to think of "artificial intelligence" as "augmented intelligence," augmenting the human capabilities, such that it becomes a tool for us to become better, just like we use watches as a tool to help us keep the time more accurately.

4   Conclusion

In this chapter, a short history and explanation of the most commonly used methods of computer vision has been given. Afterwards we laid out the most well-developed areas of application for computer vision methods and gave some examples of companies pushing the adoption of these novel techniques to the market. Then we went on to discuss the most important challenges and shortcomings of the current technology and gave some ideas for future directions of development. Overall it became evident, that there are still many improvements to be made and a lot of problems to solve, but computer vision is already a very helpful and impactful tool in the process of medical diagnostics and has contributed to saving and improving countless lives.

References

Brust, C. A., Käding, C., & Denzler, J. (2018). Active learning for deep object detection. arXiv preprint arXiv:1809.09875.

Fu, H., Xu, Y., Lin, S., Wong, D. W. K., & Liu, J. (2016). Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International conference on medical image computing and computer-assisted intervention* (pp. 132–139). Cham: Springer.

Kour, N., & Arora, S. (2019). Computer-vision based diagnosis of Parkinson's disease via gait: A survey. *IEEE Access, 7*, 156620–156645.

Le Cun, Y. (1986). Learning process in an asymmetric threshold network. In *Disordered systems and biological organization* (pp. 233–240). Berlin: Springer.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE, 49*(1), 8–30.

Müller, H., Rosset, A., Garcia, A., Vallée, J. P., & Geissbuhler, A. (2005). Benefits of content-based visual data access in radiology. *Radiographics, 25*(3), 849–858.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Sood, R., Topiwala, B., Choutagunta, K., Sood, R., & Rusu, M. (2018). An application of generative adversarial networks for super resolution medical imaging. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 326–331). IEEE.