Can we “close the loop” for scientific discovery for deep learning in computer vision?

Jimut Bahan Pal
22D1594

February 5, 2023

My research project deals with finding regions of interests within medical images that could help to aid doctors for faster treatment to a lot of patients using automation. The topic is essentially in the domain of medical image segmentation, but with the help of as less data and computational resources as possible without diminishing the precision of deep learning models.

We are not here to replace doctors, but to help doctors in their day to day tasks. In some cases deep learning model surpasses human accuracy [1], for example in classification, but in other cases, it fails disgracefully when supplied simple examples which are easily identified by humans. It is becoming important to check what deep learning models are learning and why they are failing to such examples when dealing with high-risk automations. Medical sectors involve high risks, so if a certain medical artifact is identified as tumor, then there might be high cost involved for the patient to remove it, where in reality the patient might not have tumor. Similarly, if a patient have tumor and the deep learning model fails to detect it, then the patient might go untreated. Deep learning models are very confident about their predictions [2], so when a model is messing up in classifying a sample, it confidently messes things up.

In case of computer vision, there is no such definition of interpretibility [3] unlike statistical machine learning, since deep learning models are humungous, and they deals with complex functions [4] that identifies textures and patterns in images. So, it is very hard to close the loop. The most one can do in this domain is to look at the regions for which the model is messing up the predictions and come up with explanations. Creating an explainable model is very challenging and people deals with expert systems when making interpretable models which in turn have a lot of prior and hand designed features. Deep learning is a very dormant field if we consider the way that the model will be able to come up with explanations, but way to go before we can find self explainable models.

Acknowledgements

The author is thankful to Amuthan A. Ramabathiran for the stimulating discussions he has shared for ML and Scientific Method.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[2] Eamon Duede. Deep learning opacity in scientific discovery. *arXiv preprint arXiv:2206.00520*, 2022.

[3] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[4] Jessica Zosa Forde and Michela Paganini. The scientific method in the science of machine learning. *arXiv e-prints*, pages arXiv–1904, 2019.