Hope We Are On The Same Page

Saturday, October 9, 2021 9:25 AM

The intention of this document is to eliminate misunderstanding as possible as I can, hoping **we are on the same page**. Especially when I am still struggling with my oral English, I hope to clarify something I think I don't make it very clear in our meetings. This document will be actively maintained.

Demystify Deep Learning

To the best of my knowledge, all supervisors have your own expertise in different scientific areas. I presume you might not have closely followed the advances made in the field of deep learning and its application areas, such as computer vision, automatic speech recognition, and natural language processing. Therefore, I feel like it is important to let you know my personal understanding of this field.

An accurate summary of the success of deep learning quoted from Aaron van den Oord, a Research Scientist at DeepMind, "Learning high-level representations from labelled data with layered differentiable models in an end-to-end fashion is one of the biggest successes in artificial intelligence so far."

There has been a **paradigm shift** in many research areas of computer science with the development of deep learning, i.e., from theoretical analysis to empirical study. Nowadays, many research ideas have been proposed solely based on observations from experiments or verified through rigorous designs of controlled experiments. ResNet, with citations more than 90000, has been proposed solely based on a very simple observation that the training error increases with the increase of the depth of neural networks, indicating that the problem is the difficulty of optimization rather than the commonly speculated overfitting. Nowadays, ResNet has become one of the most widely used neural network backbones in various application areas. By the way, the author of ResNet is Kaiming He, a Research Scientist at Facebook AI Research (FAIR), who has made significant contributions to the advance of deep learning. The recently published paper on Generative Adversarial Networks (GANs) GANSpace is another representative work following this paradigm. The work proposed to use principal component analysis (PCA) to automatically identify interpretable control directions existing in the latent space of GANs, which has never been explored previously. Despite the algorithmic simplicity, the findings are very insightful, especially for understanding the great success of StyleGAN. Besides, the experiment section has gradually become the most crucial section in papers on deep learning as the ideas are practically impossible to verify through theoretical analysis. Even though this can be somehow achieved, researchers are also expected to provide evidence that their proposed algorithms will work in real-world applications rather than some toy datasets because of the severely restricted conditions imposed on their mathematical models. Also, it can be frequently seen that there is a section called "**Ablation Study**" in papers about deep learning, which is the terminology from biology. And increasingly more practitioners in the field of deep learning prefer to call it "Art" rather than "Science".

I am not saying that the theoretical analysis or mathematics is not important in the field of deep learning. As far as I know, mathematicians have attempted to use stochastic differential equations (SDE) to interpret deep neural networks, which is far beyond my mathematical

capability. In my opinion, one cannot undertake such kind of research without a PhD degree in mathematics. Furthermore, if one wants to fully understand the theoretical foundation of GANs, at least they are expected to be familiar with measure theory, measure theory-based modern probability, and optimal transport theory. For example, the derivation leading to the optimal discriminator is based on Radon–Nikodym theorem. I have to admit that I have **zero knowledge** in the above-mentioned subjects. I hope I can gain a basic understanding of these subjects in 10 or 20 years, provided there is an instructor who is willing to help me. But it has little influence on me to apply GANs to real-world applications based on some **intuitive understanding**.

The reason why deep learning has transformed many research areas and achieved unprecedented success can be partially attributed to its algorithmic simplicity and the nature of being practically oriented. If you are a user of GitHub, you will notice the thriving development. Many coders, practitioners, and developers celebrate this festival by generously sharing their code repositories to the community (This is the code repository I shared on GitHub). As you can see from the below screenshot, researchers and coders are now considered to be equally important at least in industry, which is rare in the past.

Demystify application scenarios when deploying deep learning-based models

Another characteristic feature of deep learning is proving the superiority of end-to-end learned features over hand-crafted ones. One of the significant implications is substantially reducing barriers of various research areas because deep learning models can consume raw data. For example, before the advent of deep learning, Mel-Frequency Cepstral Coefficients (MFCCs) and filter banks were indispensable when it comes to speech recognition or speaker verification, which could easily deter enthusiastic novices from diving deep into this field. The phenomenon can be similarly observed in other research areas. But, nowadays, these ad-hoc techniques are rarely used because of the end-to-end learning scheme enabled by deep learning. In the field of multispectral analysis for agriculture crops, researchers have developed various indices, e.g., the well-known Normalized Difference Vegetation Index (NDVI) involving near-infrared and red spectral channels. When using deep learning models, however, the common practice is to let models directly consume spatio-spectro-temporal tensors of size TxCxHxW, with T the number of temporal observations, C the number of spectral channels, and H and W the dimensions of height and width, respectively.

As for the use of synthetic aperture radar (SAR), it is similar to that of multispectral images, just stacking four polarization modes (VV, VH, HH, HV), if available, without considering phase information (LAKE ICE DETECTION FROM SENTINEL-1 SAR WITH DEEP LEARNING). This does not represent domain expertise is no longer important. Preparing large-scale datasets of satellite images for training and evaluating deep learning models would heavily depend on experts in the field of remote sensing or geoscience. This is also why publicly available datasets are scarce and why some world-leading institutions or organizations dedicate to fostering collaboration between the computer vision and machine learning community and the remote sensing community.

In a nutshell, the boundaries between different research/application areas have become less distinct due to the less required domain expertise for entering into a field. As a result, it can be seen that many AI research groups span their research interests across a wide array of areas, such as computer vision, automatic speech recognition, natural language processing, medical image analysis and diagnostics, agriculture, chemistry, biology, and drug discovery. Clearly, we don't have the capacity and resources necessary to attain such a level. Also, it is far beyond my

capability. But what I want to say is that it is not necessary to limit our attention to only one application area under the broader framework of AI4EO. More importantly, it is expected that the proposed methods have good generalization ability, meaning that the effectiveness needs to be verified on different tasks with minor modifications.

The importance of becoming a member of the community

As far as I know, AI4EO is a rather new concept. Efforts have been made to attract researchers and practitioners outside the field of remote sensing to address challenges related to EO, especially facilitating the automatic interpretation of EO data. Some top-notch research institutions, groups, or individual researchers have already established their leading positions in recent years. Therefore, why do we refuse to be active participants by following rules they have set. I believe this is the best way to let you and your work be noticed and further recognized by the community. We need to actively participate in competitions and workshops they would organize. Without a doubt, the standards will become increasingly higher in this field just as what has happened to the computer vision community and other well-developed research areas.

Addressing concerns about publications

Firstly, I also want to get my work published and recognized. Actually, it is important for me to have a good publication track record even though I want to find a research position in industry after earning my PhD degree. Therefore, our objectives do not conflict.

The second problem is how to publish. I want to first enumerate some difficulties of preparing a deep learning paper. Provided you have an "excellent" idea, the first problem is to get it put into practice. Without reference to prior work and the associated, publicly released code repositories, it would be an arduous task. Unsurprisingly, you will see papers with similar ideas being published or having their preprint versions made public while you are stuck on code. Therefore, it has become the common practice nowadays that researchers upload their preprint versions on Arxiv and perhaps share them on social media such as Twitter so that there is a clear timeline to distinguish similar ideas when someone trying to dishonestly claim originality. Or, you are likely to find the models implemented according to this "excellent" idea achieve rather low scores on datasets. Therefore, this is the reason why I would like to stress the importance of getting hands-on experience. During my PhD interviews, one question that I was asked frequently is what you think is of significant importance for publishing your first paper? My answer is to **establish a standard pipeline that can allow you to swiftly deploy your models and test your ideas.** This pipeline can be considered as a toolkit, with which you can create datasets, preprocess data, make it ready for parallelizable training, create training logs, evaluation tools, and many other functions. This can partially explain why research-intensive private companies, such as Google, Facebook, DeepMind, OpenAI, and SenseTime, are playing the leading role in advances made in AI. I believe establishing this pipeline is my top priority. Clearly, taking part in competitions can force me to have it in position as you can imagine the process would be time-consuming. Besides, it is unavoidable to dive deep into code to fully understand a paper in the era of deep learning, which is also why making code publicly available has become an important criterion for publishing on top conferences and convincing other researchers that you do not cheat in your code.

My work style preference and vision of my PhD career

As I mentioned in our first meeting, I would like to use my strengths to achieve something in my PhD career. I have a reasonably good background in computer vision and deep learning, and more importantly I consider myself as a practitioner. Thanks to the reasons I mentioned above, I feel I am very lucky to be able to carry out some research in the field of AI4EO. And I hope this can become my future career, making my minor contributions to something really meaningful. Given the limited time and energy, I think it is better for me to approach this area from the perspective of computer vision and deep learning rather than remote sensing. Starting from scratch to familiarize myself with the knowledge in remote sensing is impractical. But I am aware that gaining some very basic knowledge in this field is beneficial for me to conduct research in this field. Furthermore, deep learning is a rapidly evolving area such that one can find it very hard to follow the latest advances. Therefore, I would rather devote myself to learning advanced topics in deep/machine learning, such as Probabilistic Graphical Models, Geometric Deep Learning/Graph Signal Processing, State-Space Models, etc., and sharpening my programming skills, such as advanced features in deep learning frameworks (PyTorch, TensorFlow, Keras, PyTorch Lightning), parallel programming (CUDA), and differentiable programming.

I think it would be ideal for me to spend the majority of my time on participating in competitions and submitting papers for conferences (I will share with you a list of AI conferences.). Preparing a journal paper, especially for those high-ranking journals, is time-consuming as normally the review process would last at least three months and the work needs to be very solid and even comprehensive. Besides, it is better to keep me motivated when there is a deadline that is 3 months away rather than 3 years away. And another point that I want to clarify is participating competitions dose not conflict with publishing papers as coming up with good solutions to these challenges would require one to read a substantial number of papers and the relevant code, considering that you would compete with world-wide top research teams. The accumulated practical experience would make you realize some tricky problems that you would not have identified in papers and easily judge the effectiveness and quality of published papers. After completing the competitions, we can extend the solutions to formal papers for submission. And if the solution is highly-ranked, one will be invited to submit an accompanying paper to relevant conferences or workshops and give a talk.

At last, I would like to refer you to a paper "Learning Disentangled Representations of Satellite Image Time Series", which I believe is a typical paradigm that I want to achieve in my PhD dissertation (at least outlining the general structure of it), that is focusing on addressing some common challenges, e.g., unsupervised representation learning, then applying the methods to different application scenarios under the umbrella of AI4EO. Expanding the potential application areas not only increases the credibility but also the influence of the work.