Proposal of DRIA Group

Group Member: Zhiqing Zhang(Kevin), Yifei Yang, Morui Zhu

We will focus on XAI, which means explainable artificial intelligence, on the network of DRIU. XAI is a relatively new issue in recent years. Researchers proposed this field to balance the transparency and the performance of DNNs. The DNNs with great performance won't be put into widespread use until their mechanisms are understood by people, especially in terms of medical use. Therefore, we are interested in how the network works and making that mechanism visualized. We will read two review of XAI published in this year first, *A Survey on Explainable Artificial Intelligence (XAI): towards Medical XAI* and *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. We will try to find some related works via these two papers. Afterwards we want to use approaches to visualize what each layer does in the network. We may try drawing the feature map, or building an activation network by adding symmetrical layers such as deconvolution layers, transposed convolution layers, unpooling layers, etc. According to our investigation, there is no work for explaining specially how the network recognize objects such as vessels. Furthermore, considering the network for DRIU is mainly based on VGGNet, if we can successfully explain the mechanism for recognizing vessels, we will have a chance to interpret the mechanism in VGGNet. Besides, we may enhance the performance more theoretically if we know how the network works.