An explainable deep learning approach to kidney cortex cell classification

Thomas Y. Chen

Academy for Mathematics, Science, and Engineering, Rockaway, USA

Deep learning algorithms, which consist primarily of neural networks, have emerged as key assets in analyzing visual imagery and video. One specific area in which deep learning models are useful is phenotype classification of cells, tissue, and organisms. However, a recurring challenge that researchers encounter is interpreting these models, i.e. understanding the inner decision-making processes within the various layers and nodes. Therefore, in this contribution, we take an existing dataset of human kidney cortex cell imagery, train a convolutional neural network (CNN) classifier on the data, and further investigate its interpretability by visualizing input weights, hiding important image parts, creating gradient-weighted class activation maps, and visualizing similar outputs in two dimensions. We used image dataset BBBC051, available from the Broad Bioimage Benchmark Collection. Our results include figures that demonstrate the severity of which parts of the image were most crucial to a particular prediction by the network (saliency maps).