Model architecture

The atom position detection is done by a U-net type CNN with attention gates in skip connections. The CNN predicts a 3D image grid where atom positions are marked by Gaussian peaks. A peak-finding algorithm finds the positions of these peaks and returns a point cloud of atom positions. The molecule graph is then constructed by picking one new node position at a time and adding it to the graph. At each step the current graph is encoded by a GNN into a feature vector that is combined with an encoding vector for the AFM image. The combined information is then used to predict the new node class and bond connections to the existing nodes in the graph.