

Data Programming for Imaging Problems: Leveraging Domain Expertise, Auxiliary Modalities, and Multiple Tasks

Jared Dunnmon, Stanford University

One of the most significant roadblocks to using modern machine learning models is collecting hand-labelled training data at the massive scale they require. In real-world settings such as imagery analysis, where domain expertise is needed and modeling goals change rapidly, the process of hand-labelling training sets is prohibitively slow, expensive, and static. For these reasons, practitioners are increasingly turning to weak supervision techniques wherein noisier, often programmatically-generated labels are used instead. In this talk, we will discuss recent applications of the data programming technique for weak supervision to imaging problems, and introduce a newly developed approach for extending these ideas to the multi-task setting.