Metadata-guided Consistency Learning for High Content Images

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

High content imaging assays can capture rich phenotypic response data for large sets of compound treatments, aiding in the characterization and discovery of novel drugs. However, extracting representative features from high content images that can capture subtle nuances in phenotypes remains challenging. The lack of high-quality labels makes it difficult to achieve satisfactory results with supervised deep learning. Self-Supervised learning methods have shown great success on natural images, and offer an attractive alternative also to microscopy images. However, we find that self-supervised learning techniques underperform on high content imaging assays. One challenge is the undesirable domain shifts present in the data known as *batch effects*, which may be caused by biological noise or uncontrolled experimental conditions. To this end, we introduce Cross-Domain Consistency Learning (CDCL), a novel approach that is able to learn in the presence of batch effects. CDCL enforces the learning of biological similarities while disregarding undesirable batch-specific signals, which leads to more useful and versatile representations. These features are organised according to their morphological changes and are more useful for downstream tasks – such as distinguishing treatments and mode of action.

Keywords: Representational Learning, Domain Shifts, High Content Screening

© 2023 Anonymous.