With the rapid increase of large-scale, real-world datasets, it becomes critical to address the problem of long-tailed data distribution (i.e., a few classes account f