Current fully-supervised video datasets consist of only a few hundred thousand videos and fewer than a thousand domain-specific labels. This hinders the progr