Large-scale face datasets usually exhibit a massive number of classes, a long-tailed distribution, and severe label noise, which undoubtedly aggravate the diffic