Current wisdom suggests more labeled image data is always better, and obtaining labels is the bottleneck. Yet curating a pool of sufficiently diverse and informa