1. Why use cross attention for feature fusion? What other method for feature fusion:

The effective fusion of features from multiple sources is crucial for land use classification tasks and a focal point of our work. Existing feature-level fusion methods mainly focus on concatenation, stacking, and some weighted fusion strategies. These methods often fail to conduct a thorough analysis of the disparities and congruences among different modalities, resulting in the adoption of a uniform fusion approach for multiple features. This overarching fusion strategy overlooks the nuanced similarities and differences among multimodal features, resulting in substantial loss of intermodal information. Consequently, this approach leads to mutual feature interference and an inability to extract diverse contextual information, thereby imposing a significant computational burden