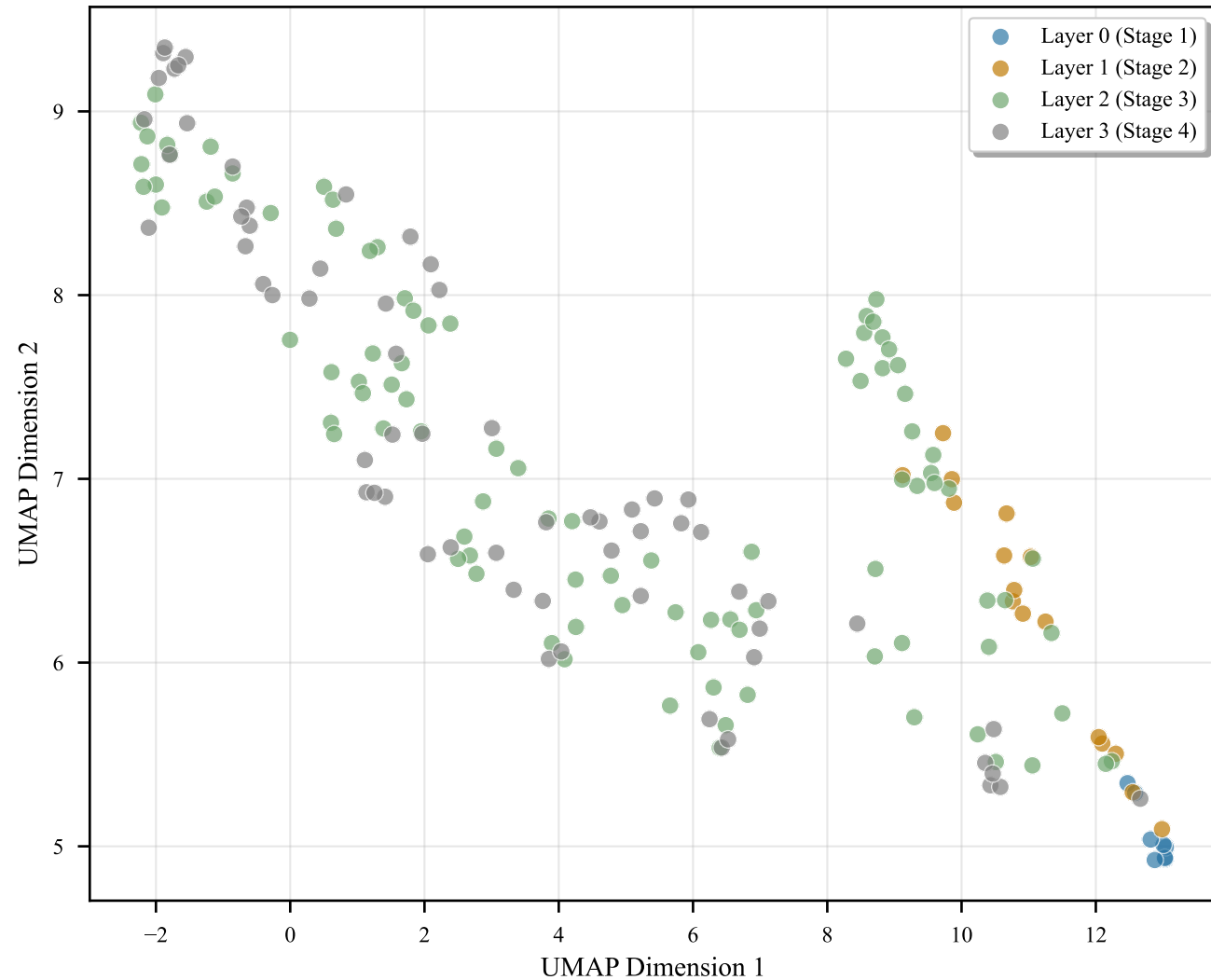(a) Head Clustering by Layer

(b) Head Clustering by Dimensionality