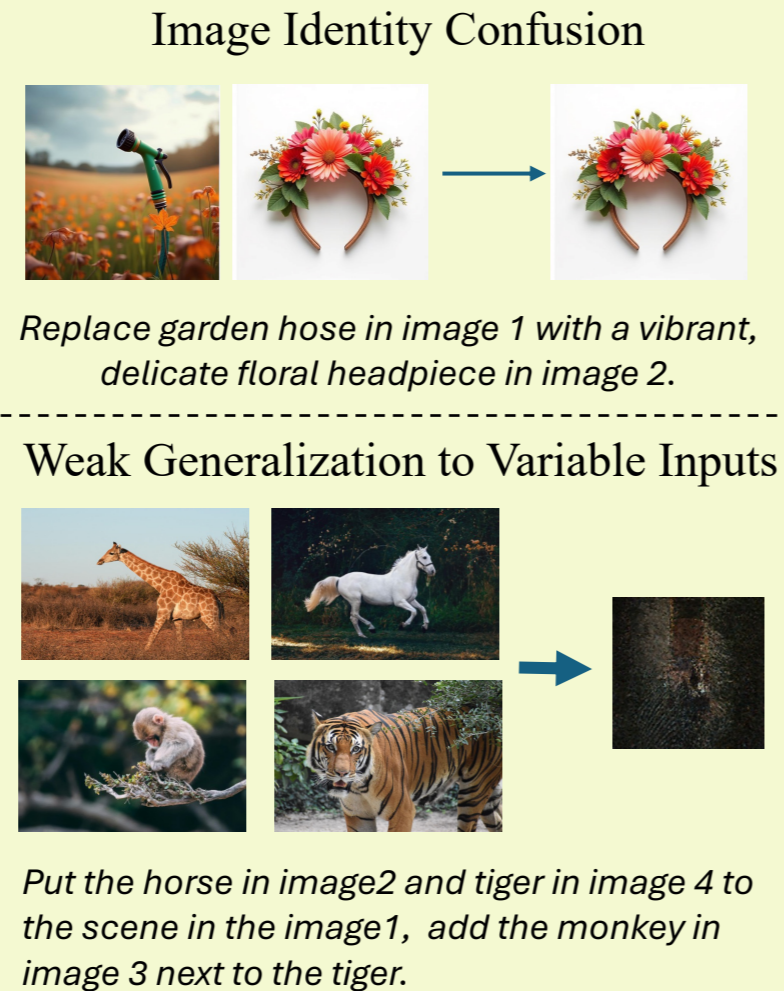
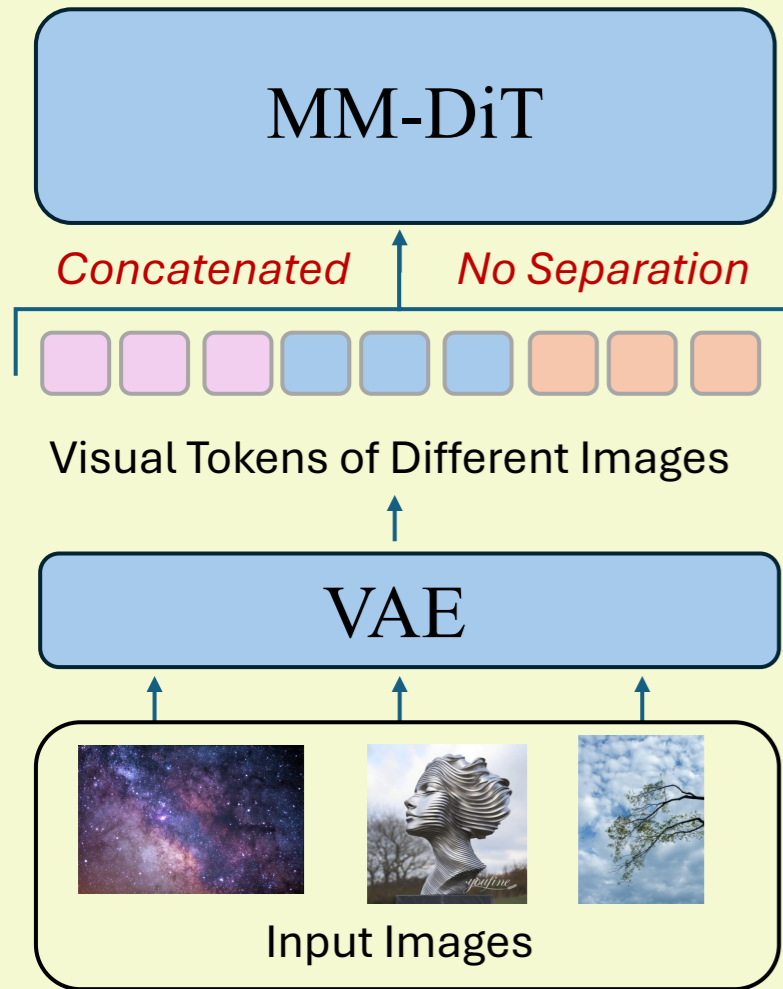


## Origin Visual Token Encoding in UMMs



## Generalized Absolute Image Indexing

### Sinusoidal Image Index Embedding

$$E_j^{(2k)} = \sin(\tilde{j}/\tau^{2k/d})$$

$$E_j^{(2k+1)} = \cos(\tilde{j}/\tau^{2k/d})$$

$E_j$

MM-DiT

