Input Sticker $\frac{H}{4} \times \frac{W}{4} \times C_1$ Stage 2 Stage 4 $\frac{H}{32} \times \frac{W}{32} \times C_3$ Stage 1 Stage 3 $\frac{H}{16} \times \frac{W}{16} \times C_2$ $\frac{H}{8} \times \frac{W}{8} \times C_2$ Yes! (\mathcal{O}) FLA-Module FLA-Module FLA-Module FLA-Module Encoder Encoder Encoder Emb CLS $H \times W \times 3$ **TGCA-Module TGCA-Module TGCA-Module TGCA-Module Sticker Topic LERA-Module LERA-Module LERA-Module LERA-Module** Panda bear **Topic ID** k-layer attention weights Visual Feature **Textual Feature** Feature Input **Feature Concatenate** \oplus Element-wise Add **Topic ID** att¹ att^{k} **Dense Layer Dense Layer Input Text** 0 **Position Embedding** Conv 1D 1x1 **G Grouped by Topic ID Pre-trained** Yes! **BERT** Select Tokens **Topic Context Aware Group Mean TG-Attention** \otimes **Matrix Multiply** Global Max **MLP Class Token** CLS TGCA-LERA-**Prediction Head Enhanced Local Feature** Feature Map **Module** Module