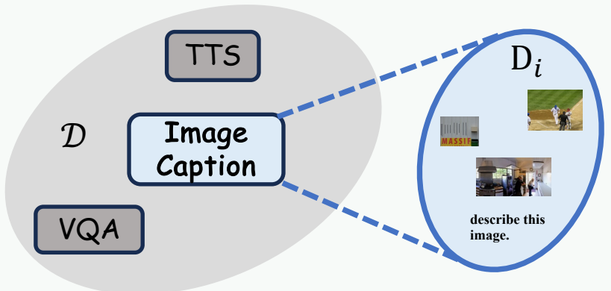
Separate Modalities

Input: Multi-modal sample $x_{(i,j)} \sim \mathrm{D}_i$

Output: Modality-specific tokens



$$\left(\text{<image>}, \text{describe this image.}\right) = x_{(i,j)} \sim \mathrm{D}_i$$

Data Preprocessor

$$\mathrm{T}^{(i,j)}_{\text{prompt}} = \text{<|im\_start|>system ....<|vision\_start|><|image\_pad|>...describe this image...}$$

$$T^{(i,j)}_{\text{text}} = [t^t_1 \dots] \qquad T^{(i,j)}_{\text{special}} = [t^s_1, t^s_2 \dots]$$

$$[\mathrm{T}^{(i,j)}_{\text{video}}, \mathrm{T}^{(i,j)}_{\text{audio}}, \dots] \qquad T^{(i,j)}_{\text{image}} = [t^i_1, t^i_2 \dots]$$

Calculate Importance Scores

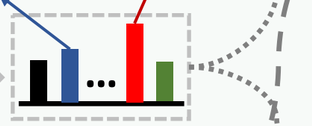
Input: Token set and neuron in FFN

Output: Importance score $\mathrm{IS}(u_{l,i})$ of i-th neuron in layer l

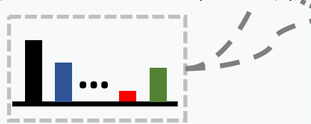
Importance Function

image component $\mathrm{IS}_t\left(T^{(i,j)}_{\text{image}}, u_{l,1}\right)$

text component $\mathrm{IS}_t\left(T^{(i,j)}_{\text{text}}, u_{l,1}\right)$



Importance Score $\mathrm{IS}_t(T^{(i,j)}, u_{l,1})$



Importance Score $\mathrm{IS}_t(T^{(i,j)}, u_{l,2})$

Importance Score $\mathrm{IS}_t(T^{(i,j)}, u_{l,N})$

Aggregate Importance Scores

Input: Importance scores of each neuron across modality token sets

Output: For each modality, sum ISM over datasets and samples

$$\mathrm{ISM}^{(i,j)}_{\text{text}} = \left[\mathrm{IS}_t\left(T^{(i,j)}_{\text{text}}, u_{l,n}\right)\right]_{l=1\sim L, n=1\sim N}$$

Importance Score Matrix (ISM)

$$\mathrm{ISM}^{(i,j)}_{\text{image}} = \left[\mathrm{IS}_t\left(T^{(i,j)}_{\text{image}}, u_{l,n}\right)\right]_{l=1\sim L, n=1\sim N}$$

Aggregation Operation

$$\mathrm{ISM}_* = \mathbb{E}_{\mathrm{D}_i\sim\mathcal{D}}[\mathbb{E}_{\mathrm{x}_j\sim\mathrm{D}_i}\mathrm{ISM}^{(i,j)}_*]$$

where $*\in\{\text{image}, \text{text}, \dots\}$

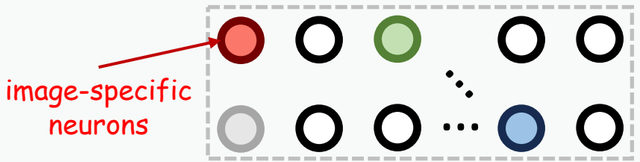
$[\mathrm{ISM}_{\text{image}}, \mathrm{ISM}_{\text{text}}, \dots]$

Select Modality-Specific Neurons

Input: ISMs and total $L\times N$ neurons in FFN

Output: Select top-K important neurons

- Uniform Selection
- Adaptive Selection
- Layer-Uniform Modality-Adaptive Selection
- Layer-Adaptive Modality-Uniform Selection



Selection Strategy

neuron mine with $L\times N$ total neurons