

Pepper the  
aussie pup

Text  
Encoder

$T_1$

$T_2$

$T_3$

...

$T_N$

Image  
Encoder

$I_1$

$I_2$

$I_3$

$\vdots$

$I_N$

$I_1 \cdot T_1$

$I_1 \cdot T_2$

$I_1 \cdot T_3$

...

$I_1 \cdot T_N$

$I_2 \cdot T_1$

$I_2 \cdot T_2$

$I_2 \cdot T_3$

...

$I_2 \cdot T_N$

$I_3 \cdot T_1$

$I_3 \cdot T_2$

$I_3 \cdot T_3$

...

$I_3 \cdot T_N$

$\vdots$

$\vdots$

$\vdots$

$\ddots$

$\vdots$

$I_N \cdot T_1$

$I_N \cdot T_2$

$I_N \cdot T_3$

...

$I_N \cdot T_N$