CLIP
Image Encoder

I_1 I_2 ... I_N

Pepper the
aussie pup

CLIP
Text Encoder

T_1
T_2
...
T_N

I_1T_1	I_2T_2	...	I_NT_1
I_1T_2	I_2T_2	...	I_NT_2
...	...	...	...
I_1T_N	I_2T_N	...	I_NT_N