ALGORITHM 1: Text-Matched Patch Selection Patch set P **Input**: Image patch set **P** and source text T^{src} **Output**: Patches corresponding to the source text Psel **Parameter**: K: # of patches in **P**, M: # of patches similar to source text **Image** Isrc 1: $\mathbf{S}, \hat{\mathbf{S}} \leftarrow \emptyset, \emptyset$ **Encoder** 2: **for** i = 1 to K **do** 3: $f_i \leftarrow E_I(P_i)$ 4: $s_i \leftarrow \frac{f_i \cdot E_T(T^{\text{src}})}{\|f_i\| \cdot \|E_T(T^{\text{src}})\|}$ Tsrc. $\begin{array}{|c|c|c|c|c|c|}\hline P_1 \cdot T_{src} & P_2 \cdot T_{src} & P_3 \cdot T_{src} & P_4 \cdot T_{src} & P_5 \cdot T_{src} \\\hline \end{array}$ **Text** 5: $\mathbf{S} \leftarrow \mathbf{S} \cup \{s_i\}$ Encoder Building 6: end for 7: $\mathbf{I} \leftarrow \{i \mid s_i \geq \text{the } M^{\text{th}} \text{ largest value in } \mathbf{S}\}$ 8: $f_{\text{avg}} \leftarrow \frac{1}{|\mathbf{I}|} \sum_{i \in \mathbf{I}} f_i$ P_1 P_2 P_3 P_4 P_5 9: **for** j = 1 to K **do** 10: $\hat{s}_j \leftarrow \frac{f_j \cdot f_{\text{avg}}}{\|f_j\| \cdot \|f_{\text{avg}}\|}$ $\hat{\mathbf{S}} \leftarrow \hat{\mathbf{S}} \cup \{\hat{s}_i\}$ 12: end for 13: $\hat{s}_k \leftarrow \text{the } (\text{round}(\frac{K}{2}))^{\text{th}} \text{ largest value from } \hat{\mathbf{S}}$ 14: $\mathbf{J} \leftarrow \{j \mid \hat{s}_i \geq \hat{s}_k \text{ and } \hat{s}_i > 0.8\}$ 15: **return** $\mathbf{P}^{\text{sel}} \leftarrow \{P_i \mid j \in \mathbf{J}\}$

Top-k select

Psel