

**ALGORITHM 1: Text-Matched Patch Selection****Input:** Image patch set  $\mathbf{P}$  and source text  $T^{\text{src}}$ **Output:** Patches corresponding to the source text  
 $\mathbf{P}^{\text{sel}}$ **Parameter:**  $K$ : # of patches in  $\mathbf{P}$ , $M$ : # of patches similar to source text

- 1:  $\mathbf{S}, \hat{\mathbf{S}} \leftarrow \emptyset, \emptyset$
- 2: **for**  $i = 1$  to  $K$  **do**
- 3:    $f_i \leftarrow E_I(P_i)$
- 4:    $s_i \leftarrow \frac{f_i \cdot E_T(T^{\text{src}})}{\|f_i\| \cdot \|E_T(T^{\text{src}})\|}$
- 5:    $\mathbf{S} \leftarrow \mathbf{S} \cup \{s_i\}$
- 6: **end for**
- 7:  $\mathbf{I} \leftarrow \{i \mid s_i \geq \text{the } M^{\text{th}} \text{ largest value in } \mathbf{S}\}$
- 8:  $f_{\text{avg}} \leftarrow \frac{1}{|\mathbf{I}|} \sum_{i \in \mathbf{I}} f_i$
- 9: **for**  $j = 1$  to  $K$  **do**
- 10:    $\hat{s}_j \leftarrow \frac{f_j \cdot f_{\text{avg}}}{\|f_j\| \cdot \|f_{\text{avg}}\|}$
- 11:    $\hat{\mathbf{S}} \leftarrow \hat{\mathbf{S}} \cup \{\hat{s}_j\}$
- 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:  $\mathbf{J} \leftarrow \{j \mid \hat{s}_j \geq \hat{s}_k \text{ and } \hat{s}_j > 0.8\}$
- 15: **return**  $\mathbf{P}^{\text{sel}} \leftarrow \{P_j \mid j \in \mathbf{J}\}$

