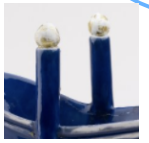
Bottom-up Attention

Artwork image: I

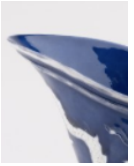
Sentence T:
Porcelain libation vessel in dark blue glaze with three legs, a loop handle and two short columns rising from the rim and with two five-clawed dragons

v_1

v_2



...
...
...



Stage 2: Attend to the important image fragment given a_i^t

Attended sentence vector a_1^t

Porcelain libation vessel dark blue glaze

five-clawed dragons

Stage 1: Attend to words

dark blue glaze
two short column

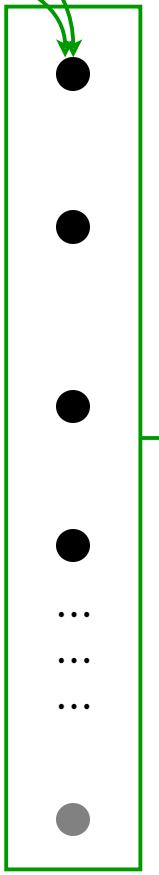
dark blue glaze three legs

dark blue glaze

five-clawed dragons

Porcelain libation vessel in dark blue glaze with three legs, a loop handle and two short columns rising from the rim and with two five-clawed dragons

Similarity:
$R(v_i, a_i^t)$



Pooling

Similarity
$S(I, T)$