Textual Input X

The color of the banana is ...

Original Visual Input V

$logits(y|x,v)$

Black	**16.72**
Dark	14.45
Yellow	11.27
Green	12.30

Distorted Visual Input V'

$logits(y|x,v')$

Black	12.57
Dark	11.84
Yellow	**14.74** ↑
Green	13.57 ↑