

For question-answer pair generated from visual content, please use **exactly one bounding box** to cover the entire area from which the QA pair is generated. If your QA pair is generated based on the entire painting, please label it with a bounding box that cover the entire image. Please look at the two examples below:



**Question from visual content:** Which fruit are depicted in this painting?

**Answer for the Question:** Grapes, Cherries, Peaches

**Source of the QA pair:** Please see the bounding box



**Question from visual content:** How many women are there in the painting?

**Answer for the Question:** Zero

**Source of the QA pair:** Since there is no woman in the painting, we treat the source of the QA pair as entire image in case like this. Thus, the bounding box should cover the entire image.