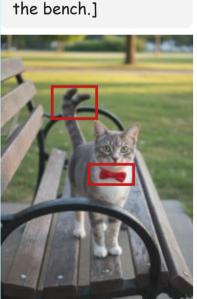**Prompt:**
Given a caption and an image generated based on this caption, please analyze the provided image in detail.
Caption: [An image of a cat wearing a green bow tie standing on the bench.]
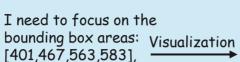
😰 **PickScore:** 19.5   **ImageReward:** 1.87   **CLIPScore:** 0.97
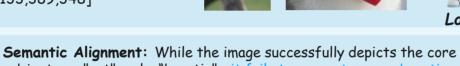
ImageDoctor (Ours)

I need to focus on the bounding box areas:
[401,467,563,583],
[238,133,389,348]

Visualization →

Look

Semantic Alignment: While the image successfully depicts the core subjects—a "cat" and a "bow tie"—it fails to generate green bow tie.
Aesthetic: The composition is well-balanced with the cat as a clear focal point, and the lighting is natural, contributing to a pleasant visual experience.
Plausibility: While the cat and the bench generally appear realistic, the cat has an extra tail, that reduce the image's overall realism.
Overall Impression: Image presents a clearly identifiable cat with good detail and a clean presentation. However, but the failure to accurately represent the specified 'green' bow tie color detract from its overall quality.

Think

4D Scores + Heatmaps:
Semantic Alignment Score: 0.67
Aesthetic Score: 1.0
Plausibility Score: 0.75
Overall Score: 0.7

😀

Misalignment   Artifact

Predict

Flow-GRPO w/ PickScore   **DenseFlow-GRPO w/ ImageDoctor** 🏅

Sign that says "**AI Image go brrrrr**".

selfie photo of a monkey wearing a fedora hat.

A man with a dog mask with a **chainsaw**.