Action Detection via an Image Diffusion Process

What is the problem?

The problem tackled by the authors is **action detection in untrimmed videos**. Action detection involves identifying the start and end points of actions and predicting the type of actions in videos. The challenge arises because of complex motions, background clutter, and variations in lighting and viewpoints.

What has been done earlier?

Earlier methods for action detection used techniques like:

Action proposals: Extracting possible action segments and refining them.

Anchor-based methods: Using predefined anchors to detect actions.

Anchor-free methods: Predicting action boundaries without predefined segments.

Diffusion models: Applied in image generation and object detection

Aabhas Agarwal, B421002

What are the remaining challenges?

Handling complex motions: High variability within the same action class.

Noisy predictions: Existing models struggle to handle uncertainty in classification.

Inadequate methods for sequential actions: Difficulty in accurately detecting multiple actions in sequence, especially in cluttered environments.

What novel solutions are proposed by the authors?

The authors propose several novel solutions:

- **ADI-Diff framework**: Treats action detection as an image generation problem, using a diffusion model to generate three images representing action class, start point, and end point.
- **Discrete Action-Detection Diffusion Process**: Adapts diffusion models to handle classification tasks by ensuring that the noise added during training respects the discrete nature of action classification.
- **Row-Column Transformer**: Encodes temporal relationships between video frames and class relationships, improving accuracy and efficiency in action detection.

Aabhas Agarwal, B421002