View Reviews

Paper ID
23

Paper Title
Multiscale Memory Comparator Transformer for Few-Shot Video Segmentation

Reviewer #13

Questions

1. Is the topic of this paper appropriate for the L3DIVU workshop?
Related

2. How do you rate the novelty of this paper?
Known method applied

3. How do you rate the experiments of this paper?
Lack some minor aspects

4. Comments about the paper (summary, strengths, and weakness; Visible to authors)
Summary:

The study introduces a Multiscale Memory Comparator (MMC) for few-shot video segmentation, which enhances performance by combining information across scales within a transformer decoder. By leveraging rich semantics and precise localization, MMC outperforms existing baselines and achieves state-of-the-art results in this task.

Strength:

+ Addressing limitations of conventional methods by utilizing both rich semantics and precise localization, enhancing the overall comparisons between query and support features.
+ Demonstrating that the proposed MMC outperforms existing baselines, validating its effectiveness in tackling the few-shot video segmentation problem.

Weakness:

- The motivation for the proposed multi-scale memory decoding is not clear. The limitations of the previous multiscale transformer decoders are not well explained.
- How the proposed method is able to preserve the spatiotemporal dimensions is not clear.
- The computational complexity aspect of the method should be discussed and compared with other methods.
- It would be good to show more visual results to demonstrate the effectiveness of the proposed method.

5. Overall score
Borderline

6. Confidence on the review
Confident

Reviewer #14

Questions

1. Is the topic of this paper appropriate for the L3DIVU workshop?
Related

2. How do you rate the novelty of this paper?
Known method applied

3. How do you rate the experiments of this paper?
Lack some minor aspects

4. Comments about the paper (summary, strengths, and weakness; Visible to authors)
Summary:

This paper proposes a novel multiscale memory comparator (MMC) module for few-shot video segmentation, which facilitates information exchange across scales. The MMC replaces the key and value inputs of transformer layers with memory entries across all decoding layers and scales, achieving significant performance improvements over state-of-the-art methods.

Strengths:

1.The MMC module design is well-explained and easy to follow.

2.The results achieved are promising and surpass state-of-the-art methods.

Weaknesses:

1.The manuscript lacks a related work section, which is necessary to provide context and position the contribution within the field.

2.The font used for tensors in Eq. (1) (e.g., Q, K, V) appears irregular. Suggest using \mathbf{Q, K, V} to standardize the notation.

3.While the study introduces a new transformer architecture, its relationship to few-shot video object segmentation task is not sufficiently explained. Clarify why this architecture is applied specifically in the few-shot VOS task and its potential benefits in this context.

4.The study should include experiments for the K-shot setting to better evaluate its performance in various scenarios.

5.The authors should conduct an ablation study to investigate how the performance changes as they increase the number of MMC layers applied in the decoder.

5. Overall score
Borderline

6. Confidence on the review
Weakly confident

Reviewer #15

Questions

1. Is the topic of this paper appropriate for the L3DIVU workshop?
Very related

2. How do you rate the novelty of this paper?
Adapted/Extended method

3. How do you rate the experiments of this paper?
Solid experiments

4. Comments about the paper (summary, strengths, and weakness; Visible to authors)
[Summary]
This is a four pages short paper. The authors proposed a novel meta-learned Multi-scale Memory Comparator (MMC) Transformer to tackle the temporal consistency of the extracted feature and confusion between novel class and background for few-shot video segmentation. The multi-scale memory decoding approach enables information exchange across different scales on the dense feature maps to capture detailed information. In addition, this method can preserve spatiotemporal information rather than a highly compressed representation. Their experiments demonstrate that their method outperforms the current SOTA method on YouTube-VIS for few-shot video object segmentation and three different automatic video object segmentation datasets.

[Pros]
1. Propose a novel multi-scale memory comparator transformer to tackle the problem in current multi-scale transformer decoders that learn from a highly compressed representation.
2. Propose solution is simple but effective.
3. The proposed method outperforms the current SOTA methods and demonstrates the effectiveness of multi-scale memory decoding design and meta-learned multi-scale comparator in the ablation studies part.

[Cons/Questions]
1. The author should clarify the operation of “+” in section 2. For example, equation 2 and equation 3.
2. The details of boundary refinement are missing in section 2, for example, the architecture design or the reference work that is related to this module.
3. The conclusion part should elaborate more. For example, you can mention what kind of problem you can solve in the FS-VOS by using an MMC transformer. In addition, you can briefly conclude your findings or results in the ablation studies.
4. The qualitative comparison of the newly proposed method and SOTA methods is missing in the experiment session.

5. Overall score
Borderline

6. Confidence on the review
Confident