View Reviews

Paper ID
321

Paper Title
Guiding Scene Text Question Answering with Region Priors

Track Name
Main Track

Reviewer #1

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This paper proposes a Region-aware OCR query model. The model uses query to conduct cross attention over regions first to obtain a region-aware query vector. This vector is further used in the region-ocr query module to compute attention over OCR tokens. Final decision is generated based on both region aware query vector and the output of region-ocr query. The model achieves new state-of-the-art performance on both TextVQA and STVQA datasets.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
Strengths: The model achieves new state-of-the-art performance in two widely used dataset.
Weakness: The novelty is quite limit. In general the entire model is not different from transformer, and the spatial self-attention block is also not a new method.

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
1. In the ablation study, why you take the M4C model as baseline ?
2. Although each module has a special name, I didn't see the differences between them and the transformer encoder/decoder, could you explain it a bit ?
3. Table 3 is bit of confusing. It is not clear how you use the proposed each module in the baseline model.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Fair: The paper contributes some new ideas or represents incremental advances.

5. {Evaluation: Quality} Is the paper technically sound?
Fair: The paper has minor technical flaws. For example, the proof of a theorem has some fixable errors or the experimental evaluation is weak.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Poor: The paper is likely to have minimal impact on AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Fair: The paper is somewhat clear, but some important details are missing or unclear.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Fair: key resources (e.g., proofs, code, data) are unavailable and/or some key details (e.g., proof sketches, experimental setup) are unavailable which make it difficult to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)

Fair: The shared resources are likely to be of some use to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Good: The paper adequately addresses most, but not all, of the applicable ethical considerations.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Borderline reject: Technically solid paper where reasons to reject, e.g., poor novelty, outweigh reasons to accept, e.g. good quality. Please use sparingly.

13. (CONFIDENCE) How confident are you in your evaluation?
Very confident. I have checked all points of the paper carefully. I am certain I did not miss any aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Mostly Knowledgeable: This paper has little overlap with my current work. My past work was focused on related topics and I am knowledgeable or somewhat knowledgeable about most of the topics covered by the paper.

Reviewer #7

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
The paper tackles the problem of scene text question answering (TextVQA). It involves the OCR information into the method and achieves better results.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
The paper has the following strengths:
1. It proposed a region query module (RQ) to find the salient region associated with answering the question.
2. The paper proposed an OCR query module (OQ) to identify salient OCR tokens from the salient region.

The paper has the following weaknesses:
1. The paper did not explain how to extract the OCR tokens, like using what OCR system? The OCR encoder seems to take as input the salient images, not the OCR tokens as shown in Figure 2.
2. The paper is hard to follow as it is misleading with the modules not well explained, especially the OCR-OCR query module.
3. Some typos. The first sentence of the abstract, TextVQA, "attracting" increasing interests.

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
1. For figure1, does the order of the OCR tokens matter in the generation process? It seems like the figure (c) does not show the right order.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Fair: The paper contributes some new ideas or represents incremental advances.

5. {Evaluation: Quality} Is the paper technically sound?
Fair: The paper has minor technical flaws. For example, the proof of a theorem has some fixable errors or the experimental evaluation is weak.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Fair: The paper is likely to have modest impact within a subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Fair: The paper is somewhat clear, but some important details are missing or unclear.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Fair: key resources (e.g., proofs, code, data) are unavailable and/or some key details (e.g., proof sketches, experimental setup) are unavailable which make it difficult to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Poor: The shared resources are unlikely to be useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Fair: The paper addresses some applicable ethical considerations but fails to address some important ones.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Borderline reject: Technically solid paper where reasons to reject, e.g., poor novelty, outweigh reasons to accept, e.g. good quality. Please use sparingly.

13. (CONFIDENCE) How confident are you in your evaluation?
Somewhat confident, but there's a chance I missed some aspects. I did not carefully check some of the details, e.g., novelty, proof of a theorem, experimental design, or statistical validity of conclusions.

14. (EXPERTISE) How well does this paper align with your expertise?
Knowledgeable: This paper has some overlap with my current work. My recent work was focused on closely related topics and I am knowledgeable about most of the topics covered by the paper.

Reviewer #8

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This work proposes a new modeling approach for TextVQA which explicitly accounts for salient regions of an image and the text contained in them by using self and cross attention modules.

There are two modeling branches.

[Branch-1] Salient regions (objects) in the image are used to build an improved query representation (via Question-Region Query) network. This network is trained via bounding box supervision (Region-aware loss).

[Branch-2] OCR tokens (words) detected are passed through a custom self-attention network which also accounts for their spatial proximity which groups them.

Both the branches are tied together using the Region-OCR query module that performs cross-attention on their outputs to generate the final answer.

Authors provide an ablation study of these components, and show results outperforming previous works.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
Strengths.

1. I believe modeling interactions between groups of words, and objects is a good idea to improve TextVQA systems.

2. The idea to use soft-labels (section 3.2) to enforce region-aware supervision seemed well thought. I also liked the formulation of spatial self-attention.

3. Ablation study clearly indicates a positive trend after adding each of the proposed components.

Weaknesses.

1. Beyond the two (somewhat novel) ideas of soft supervision (region aware loss), and OCR grouping (via spatial self-attention), other parts such as building region-aware queries with cross-attention seem to have existed in literature [1].

2. The OCR grouping method via spatial self-attention seemed like an overkill, is it not possible to just group the tokens using their proximity and simple color statistics? 2

3. I also feel the writing of this work can be improved. One thought is to use simpler and more intuitive terms (and avoid words like “region-aware OCR query”), and not introduce so many new abbreviations.

Typos.
1. Figure 2, question says “plart”.

[1] Hierarchical Question-Image Co-Attention for Visual Question Answering

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
1. Would it be possible to compare this approach’s model size and training time with previous works?

2. Could you try a simpler OCR grouping mechanism -- say only via proximity or color statistics b/w OCR patches.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Fair: The paper contributes some new ideas or represents incremental advances.

5. {Evaluation: Quality} Is the paper technically sound?
Good: The paper appears to be technically sound. The proofs, if applicable, appear to be correct, but I have not carefully checked the details. The experimental evaluation, if applicable, is adequate, and the results convincingly support the main claims.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Fair: The paper is likely to have modest impact within a subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Fair: The paper is somewhat clear, but some important details are missing or unclear.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper’s reproducibility checklist.)checklist.)
Fair: key resources (e.g., proofs, code, data) are unavailable and/or some key details (e.g., proof sketches, experimental setup) are unavailable which make it difficult to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper’s reproducibility checklist)

Poor: The shared resources are unlikely to be useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Not Applicable: The paper does not have any ethical considerations to address.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Weak Accept: Technically solid, modest-to-high impact paper, with no major concerns with respect to quality, reproducibility, and if applicable, resources, ethical considerations.

13. (CONFIDENCE) How confident are you in your evaluation?
Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Very Knowledgeable: This paper significantly overlaps with my current work and I am very knowledgeable about most of the topics covered by the paper.

Reviewer #9

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
The authors focus on the TextVQA task in this paper which aims to generate the answer to questions relevant to scene text and visual components in a scene. They propose a novel idea for this task by utilizing the objects as the bridge between the question and the OCR tokens. Corresponding modules/loss functions are designed to encode the relationship and guide the learning process.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
Strength: 1. novel ideas to tackle the TextVQA task by treating objects as the bridge between the question and OCR tokens.
2. two modules are designed to encode the relationships between the question, the objects, and the OCR tokens.
3. pseudo target regions are created to guide the learning of salient object regions to a question.

Weakness:
1. there are a few typos in the paper, e.g., attracting increasing interest (line 1 in the abstract) instead of attacking.
2. would be better to compare this work with more recent works in 2022 instead of papers from 2021.

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
any evaluation of the performance of the object extractor and OCR token extractor? These performances are quite important for the final question-answering task.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

5. {Evaluation: Quality} Is the paper technically sound?
Good: The paper appears to be technically sound. The proofs, if applicable, appear to be correct, but I have not carefully checked the details. The experimental evaluation, if applicable, is adequate, and the results

convincingly support the main claims.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Fair: The paper is likely to have modest impact within a subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Good: The paper is well organized but the presentation has minor details that could be improved.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Good: key resources (e.g., proofs, code, data) are available and sufficient details (e.g., proofs, experimental setup) are described such that an expert should be able to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Good: The shared resources are likely to be very useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Fair: The paper addresses some applicable ethical considerations but fails to address some important ones.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Accept: Technically solid paper, with high impact on at least one sub-area of AI or modest-to-high impact on more than one area of AI, with good to excellent quality, reproducibility, and if applicable, resources, and no unaddressed ethical considerations. Top 60% of accepted papers.

13. (CONFIDENCE) How confident are you in your evaluation?
Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Knowledgeable: This paper has some overlap with my current work. My recent work was focused on closely related topics and I am knowledgeable about most of the topics covered by the paper.