Reviews

Review 1

Strengths of the Paper

-This paper deals with a rather old and complex problem that is to find the online signal trajectory corresponding to an offline handwriting (image).

-The approach is an interesting extension and combination of existing approaches by targeting a decomposition at the sub-stroke level.

-The paper is well written and the results show contribution compared to the state of the art

Weaknesses of the Paper

-The illustrations of the qualitative results are difficult to understand: figures 6 and 7 need to be improved (presentation, size, color, graphic highlighting of differences)…
-What is missing is a discussion of what kind of information the model uses (learns) to better reconstruct the trajectory. If the model implicitly incorporates recognition capabilities to better reconstruct trajectories it is likely that reconstruction errors are correlated with recognition errors. This can then limit the generalization capabilities of the model to other types of handwriting...

Originality **4**: (good)

Related Work **4**: (good)

Technical Merit **4**: (good)

Evaluation **3**: (fair)

Writing and Presentation **4**: (good)

Overall evaluation

2: (accept)
This paper deals with a rather old and complex problem that is to find the online signal trajectory corresponding to an offline handwriting (image). The approach is an interesting extension and combination of existing approaches by targeting a decomposition at the sub-stroke level.

The paper is well written and the results show contribution compared to the state of the art.
What is missing is a discussion of what kind of information the model uses (learns) to better reconstruct the trajectory. If the model implicitly incorporates recognition capabilities to better reconstruct trajectories it is likely that reconstruction errors are correlated with recognition errors.
This can then limit the generalization capabilities of the model to other types of handwriting (to be verified and discussed at the end of the paper).

The implicit embedding of recognition capabilities in the reconstruction of the trajectory can also limit the potential applications mentioned in the introduction of the paper.

In addition, one could even be tempted to recognize before reconstructing the trajectory... it all depends on what is the final goal of the reconstruction task : this discussion is missing in the introduction and conclusion of the paper.

The illustrations of the qualitative results are difficult to understand: figures 6 and 7 need to be improved (presentation, size, color, graphic highlighting of differences)…

Reviewer's confidence **5**: (expert)

Review 2

Strengths of the Paper

1. Innovative use of transformer for offline to online conversion. While image embedding does not strictly require transformer (e.g. CNNs), the use of auto-regressive SORT transformer to output order given past sub-strokes is a great fit for transformer.

2. Detailed description of the dataset used and the use of MyScript ink recognition engine (better reproducibility).

3. Clear logic for the most part. It is easy to visualize the pipeline and data flow overall.

Weaknesses of the Paper

1. SORT network: this is arguably the most important piece of the pipeline, yet also the most opaque. Some questions/suggestions are:
Given that we have a detailed figure for sub-stroke cutting algorithm, it would be helpful for SORT to have its own diagram as well.
Further, please elaborate on the sub-strokes set. Is its size always fixed? Sub-stroke cutting algorithm implies the size should be dynamic, hence the confusion, which also makes it difficult to visualize the probability distribution output.
In equations (1), please explicitly define the dimensions (e.g. T, L, etc.)

2. Oracle not clearly defined: The authors mention "this oracle answer is a satisfying approximation of the original online" -- what is the difference? Please elaborate on this oracle concept.

3. SDTW: the authors introduced this metric as an alternative to DTW as it "detect under/over-segmentation issues". Given this metric is later used to overrule DTW to justify this approach's performance, please elaborate on the differences with DTW and intuitively, why it is the better choice.

Originality **4**: (good)

Related Work **3**: (fair)

Technical Merit **3**: (fair)

Evaluation **3**: (fair)

Writing and Presentation **3**: (fair)

Overall evaluation

2: (accept)
Overall, the authors propose a good idea using transformers operating on sub-strokes and auto-regressive order prediction. The logic and is clear for the most part, though some sections (e.g. SORT, SDTW) could really benefit from more elaboration and clarity. The experiments demonstrate this approach is promising, though the authors may want to look into additional datasets such as [1] to better support the results.

[1] Y. Seki, "Online and Offline Data Collection of Japanese Handwriting," 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), Sydney, NSW, Australia, 2019, pp. 13-18, doi: 10.1109/ICDARW.2019.70135.

Page 2, the second paragraph, "To overcome this issue, rasterization or online data to offline ..." should be "rasterization of ...".

5.2 Results section, can you explain what "Oracle" means?

Reviewer's confidence **4**: (high)

Review 3

Strengths of the Paper

1) The transformer framework is a natural solution to this longstanding problem
2) Interesting results are shown, esp. in the case of the equations
3) The separation between the SET & SORT stages is a good way to realize 'separation of concerns' and keep an already complicated approach under control.

Weaknesses of the Paper

1) Online is not online: there are many variants of sampling: irregular mouse-event based with an unclear dt, equidistant in space (e.g. 1pix) up to a real approximation to what happened along the time axis (equidistant in time, with a good approximation of human writing times).
2) There is more literature on this (cf.: Stefan Jaeger)
3) The one-pixel thickness heuristic entails an intrinsic vulnerability which the authors admit between the lines, but which may play a detrimental role in low-quality historical handwriting.

Originality **5**: (excellent)

Related Work **3**: (fair)

Technical Merit **4**: (good)

Evaluation **3**: (fair)

Writing and Presentation **4**: (good)

Overall evaluation

3: (strong accept)
'Strong accept', but ... good papers deserve attention to details, in their revised version.

A reference to the work of Stefan Jaeger is missing (the German counterpart of Lallican's dissertation).
See below.

The reader does not get an idea concerning the robustness of results. How many samples in test sets? k-fold evaluation? Standard deviations on performances?

S. Jaeger. Recognizing Handwritten Words by Recovering Pen Movements. Contribution to the Second Conference on Perception, published by H. Buelthoff, M. Fahle, R. Gegenfurtner, and A. Mallot, Tuebingen, 1999.
S. Jaeger. Motor Control, Curvature Minimization, and the Traveling Salesman. 8th Biennial Conference of the International Graphonomics Society, edited by A. Colla, F. Masulli, and P. Morasso, Genova, 1997.
S. Jaeger. A Psychomotor Method for Tracking Handwriting. 4th International Conference on Document Analysis and Recognition, pages 528-531, Ulm, 1997.
S. Jaeger. Recovering Writing Traces in Off-Line Handwriting Recognition: Using a Global Optimization Technique. 13th International Conference on Pattern Recognition, volume C, pages 150-154, Vienna, 1996.

Reviewer's confidence **4**: (high)

Metareview

Metareview for paper 2503
Title: SET, SORT! A Novel Sub-Stroke Level Transformer for Offline Handwriting to Online Conversion
Authors: Elmokhtar Mohamed Moussa, Thibault Lelore and Harold Mouchère
Text: The reviewers consistently gave good scores.