TransSketchNet: Attention-based Sketch Recognition using Transformers

Gaurav Jain[1], Shivang Chopra[1], Suransh Chopra[1], Anil Singh Parihar

Machine Learning Research Laboratory, Department of Computer Science & Engineering
Delhi Technological University, New Delhi 110042, India

Abstract

Sketches have been employed since the ancient era of cave paintings for simple illustrations to represent real-world entities. The abstract nature and varied artistic styling makes automatic recognition of drawings more challenging than other areas of image classification. However, dealing with images as a sequence of small information makes it challenging. In this paper, we propose a Transformer-based network, dubbed as TransSketchNet, for sketch recognition. This architecture incorporates ordinal information to perform the classification task in real-time through vector images.

Contributions

- **Sketch Recognition using Transformers:** This is the first approach to the best of our knowledge that employs transformers for sketch recognition. We leverage the attention mechanism of Transformers to identify objects as a sequence of strokes in real-time.
- **Attention-based analysis of sketches:** We isolate parts of the sketches necessary for object classification and analyze these characteristic fragments.

Data and Preprocessing

Sketches are represented in the vector-image format. $\mathcal{S}$ is a sketch with sequence of strokes s_i, where $s_i = \{\Delta x_i, \Delta y_i, p_i\}$, $\forall i \in \{1, 2, ..., n\}$, such that $(\Delta x_i, \Delta y_i)$ is the offset distance in the x and y direction. Pen-state, p_i, is a binary variable indicating if the pen is in contact with surface or lifted.

Proposed Method

Fig. 1 illustrates the proposed architecture, which consists of two modules:

1. Auto-encoder module extracts features from the input, while simultaneously facilitating a larger set of features to be fed into the transformer module.

2. Transformer module processes these features to attentively capture the characteristic information from the strokes at each time step.

Overall Architecture

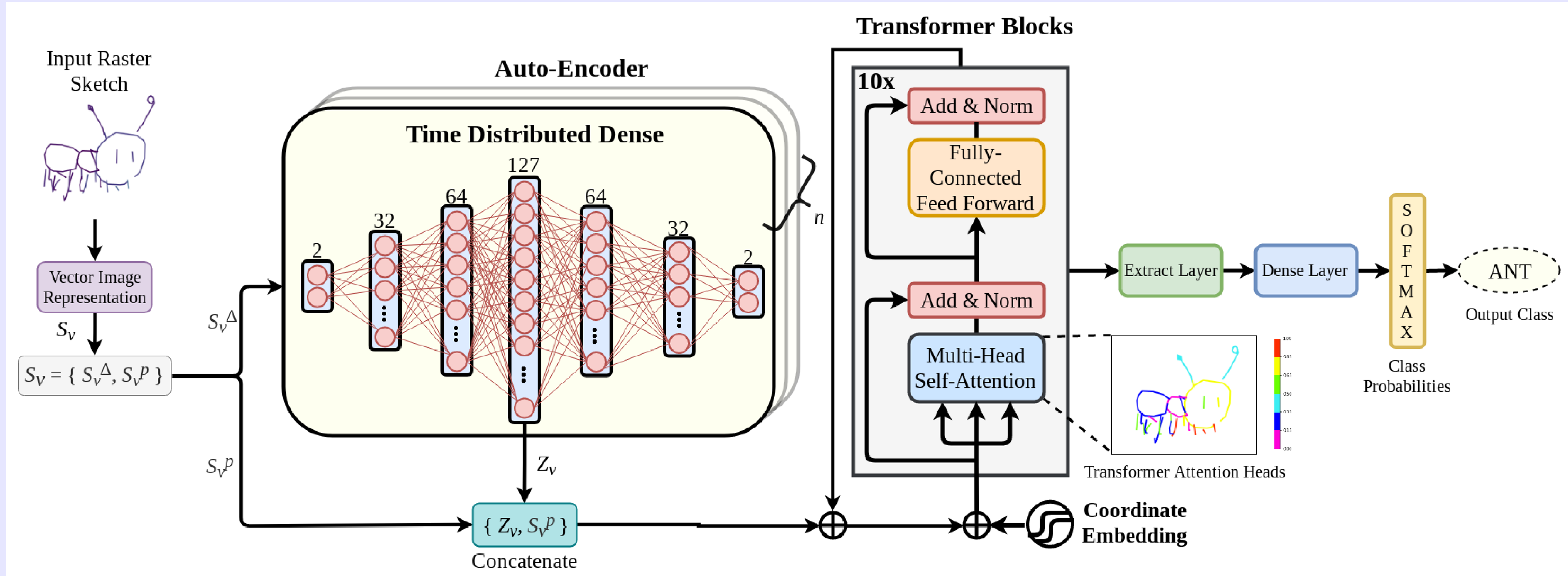

Figure 1:Overview of the proposed Network depicting the pre-processing of raster sketch to form the vector image representation. Projecting the input to a higher latent dimension using the Auto-Encoder and the final classification using the attention maps from the Transformer blocks.

Attention Heatmaps

Fig. 2 shows attention heatmaps. For the *bat* class, attention on both the wings are equally focused, which is intuitively the most characteristic feature in a bat. In the *star* class, equally high attention is given to the complete structure, which indicates that symmetric structures are identified based on the overall view of the sketches.

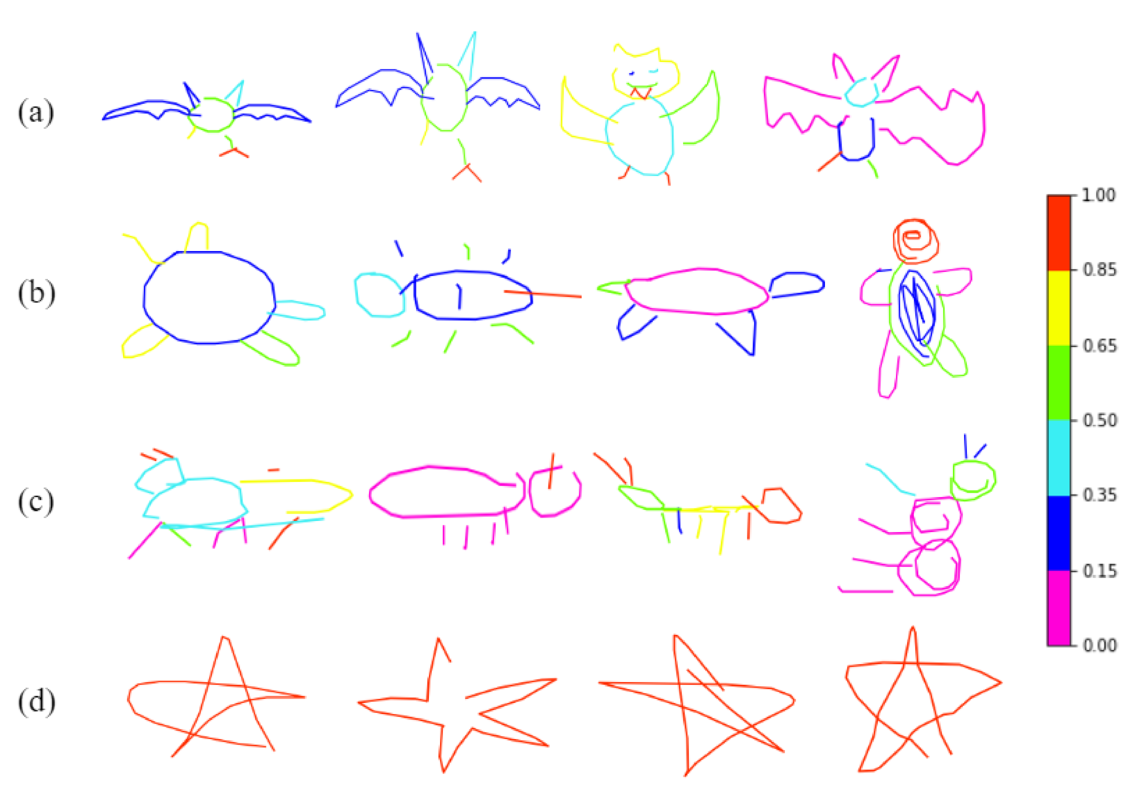

Figure 2:Attention heatmaps depicting the relative importance of strokes while inference of few classes in the QuickDraw [1] dataset, (a)Bat, (b) Sea Turtle, (c) Ant, (d) Star.

Order of Strokes Analysis

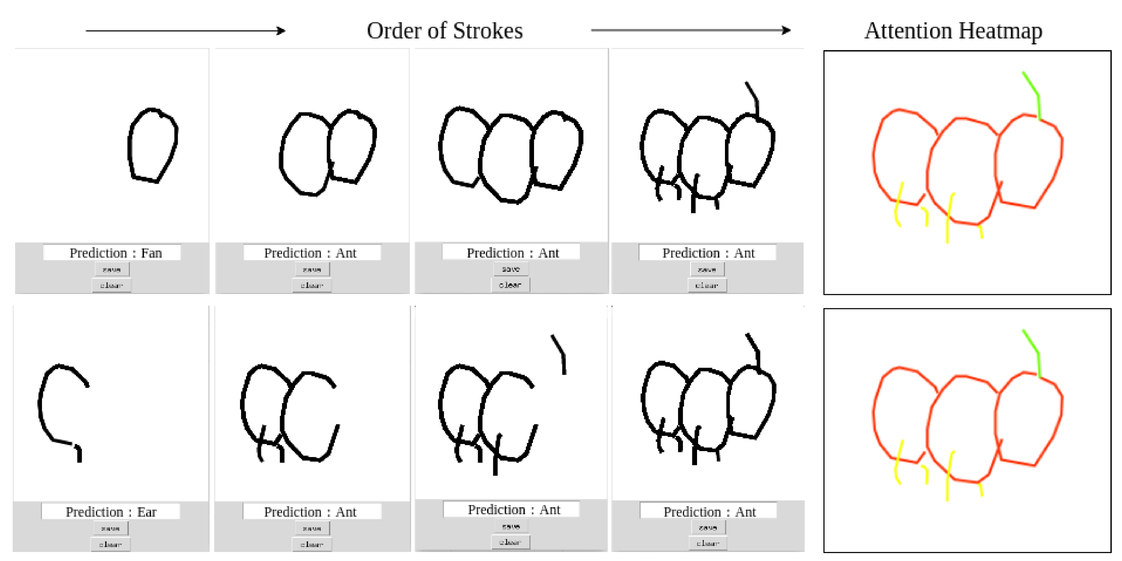

Figure 3:Experiments to evaluate the effect of order of strokes on attention heat-maps and classification (drawn through a User Interface in real-time).

The usage of vector image raises another important question about the importance of *order of strokes*, which was previously immaterial in the case of raster images. Since the input is a real-time set of strokes, unlike images that are fed into the network after completion, it is important to assess the effect of the same for classification results. Fig. 3 displays two different ways to render the sketch of a `ant`. From the attention heat-map, it can be inferred that there is no effect of sequence of strokes for the final classification or attention.

Results

Method	**5**	**20**	**50**
HOG-SVM	75.21%	66.79%	63.22%
Fisher-Vectors	79.53%	75.80%	72.90%
AlexNet	77.18%	75.22%	73.06%
Sketch-a-Net v2	94.78%	88.64%	85.19%
Resnet50-CNN	**96.47%**	90.06%	86.20%
TransSketchNet	96.21%	**90.31%**	**88.72%**

Table 1:Comparative evaluation of recognition accuracy on the Quick Draw dataset, with (a) 5 classes, (b) 20 classes, (c) 50 classes. Values in **bold** depict the best accuracy.

We compared the proposed TransSketchNet with three types of approaches, (1) traditional classifiers, (2) CNN-based approaches, and (3) RNN-based approaches. Table 1 reports the recognition accuracy.

Conclusion

- Sketch recognition using vector images performs favourably against state-of-the-art approaches.
- Transformers effectively extract characteristic features from sketches for recognition.
- Order of strokes play a minor role in determining the important parts of a sketch.

References

[1] David Ha and Douglas Eck.
A neural representation of sketch drawings.
CoRR, abs/1704.03477, 2017.

Contact Information

- Gaurav Jain
Email: gauravjain13298@gmail.com
- Shivang Chopra
Email: shivangchopra11@gmail.com
- Suransh Chopra
Email: suransh2008@gmail.com
- Anil Singh Parihar
Email: parihar.anil@gmail.com
Delhi Technological University, New Delhi, India