2020 ICPR

Sketch-SNet: Deeper Subdivision of Temporal Cues for Sketch Recognition

Yizhou Tan* , Lan Yang* , Honggang Zhang

Beijing University of Posts and Telecommunications, Beijing, China

Sketch&Image

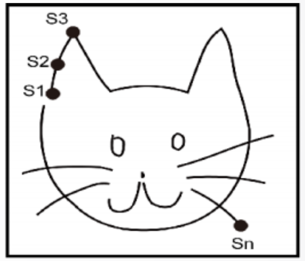
Sketch is a unique tool for human beings to exchange abstract information.

With the development of touch screen, researches on sketch understanding have proliferated. Sketch recognition [1] acts as an essential task, driving a series of sketch understanding developments such as sketch synthesis [2] and sketch based image retrieval (SBIR) [3].

Image

Texture
Dense data
Pixel matrix

Sketch



Non-Texture
Sparse data
Pixel matrix & Sequence

Subdivision of Temporal Cues

Compared with traditional comprehension of sketch, we further split the temporal information (sequence) of sketch into two types of feature:

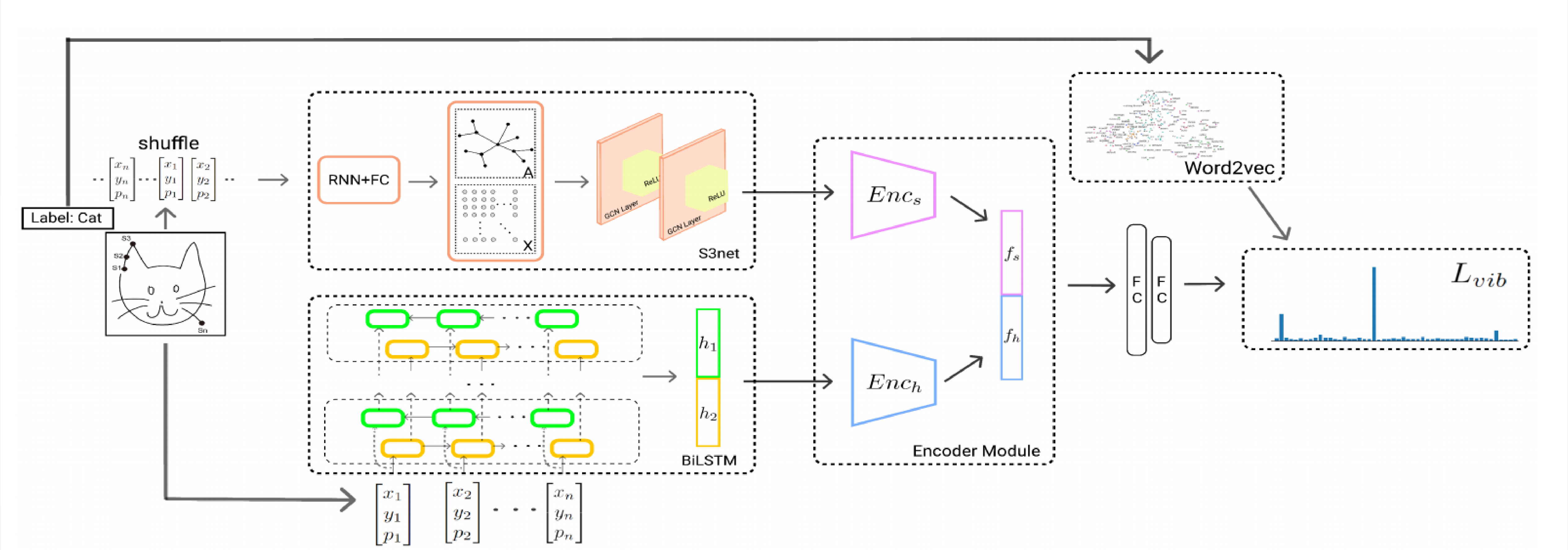
Invariable structural feature (ISF) ------------- GCN

drawing habits feature (DHF) ------------- RNN

ISF reflects the specific structural relationships between strokes --- a car is made up of body and wheels.

DHF contains the specific temporal relationships between strokes --- people prefer to draw body of a car first instead of wheels.

Network Architecture



We propose a two-branch GCN-RNN network, Sketch-SNet, to extract two types of feature respectively. The GCN branch is used to extract the ISF through receiving various shuffled strokes of an input sketch. The RNN branch takes the original order to extract DHF by learning the pattern of strokes' order.

Result

TABLE I

COMPARISON WITH STATE-OF-ART SKETCH CLASSIFICATION COMPETITORS ON QUICK-DRAW (PART OF DATA ARE CITED FROM [8])

Type	Method	Acc.
CNN-based	ResNet-50 [29]	78.56%
	AlexNet [30]	73.76%
	DenseNet-121 [31]	78.96%
	GoogleNet [32]	78.01%
RNN-based	LSTM [33]	78.35%
	BiLSTM [26]	79.96%
Extension of CNN/RNN	Sketch-a-Net [28]	68.71%
	SketchMate [7]	80.51%
	Doodle-Variant [34]	78.13%
	SketchFormer [36]	77.68%
GCN-based	SketchGCN [35]	70.04%
	S^3Net [8]	84.22%
	S^3Net (Stroke-5) [8]	85.10%
GCN-RNN branches	**Sketch-SNet**	**90.42%**

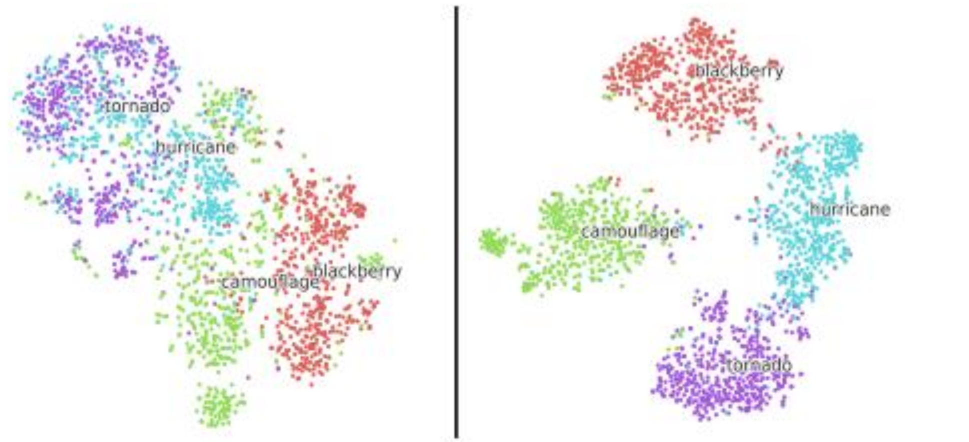


Fig. 2. Figure shows the feature distribution in feature space of S^3Net [8] method (Left) and Sketch-SNet (Right). Through T-SNE algorithm to reduce the high dimension, we visualize the space position relationship between different categories.

Shuffling operation analysis

Comparing traditional model under shuffling operation, we want to find that if the improvement of Sketch-SNet is owing to the shuffling data augmentation

TABLE II
RESULTS OF SHUFFLING OPERATION IN EXPERIMENTS.

Method	Shuffled	No Shuffled
BiLSTM	78.74%	79.96%
S^3Net	84.80%	84.22%
Sketch-SNet	90.42%	89.57%

And the results show that, the shuffling operation doesn't improve the performance of BiLSTM. Because while the data are augmented in the shuffling operation, it also completely destroyed the DHF that we want to extract. In contrast, the Sketch-SNet and S3Net not only achieve a better result with shuffling operation, but also Sketch-SNet is far superior to other models through adding another RNN branch to extract DHF.

Reference

[1] M. Eitz, J. Hays, and M. Alexa, “How do humans sketch objects?” ACM Transactions on Graphics, vol. 31, no. 4, pp. 1–10, 2012.
[2] Y. Li, Y.-Z. Song, T. M. Hospedales, and S. Gong, “Free-hand sketch synthesis with deformable stroke models,” International Journal of Computer Vision, vol. 122, no. 1, pp. 169–190, 2017.
[3] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, “Sketch-based image retrieval: Benchmark and bag-of-features descriptors,” IEEE Trans. Vis. Comput. Graph., pp. 1624–1636, 2010.

Thanks