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

Yizhou Tan , Lan Yang , Honggang Zhang
Beijing University of Posts and Telecommunications, Beijing, China

Introduction

Sketch recognition is essential in sketch-related researches. Different from the natural image, the sparse pixel distribution of sketch discards the visual texture which brings a huge obstacle to prior methods on ImageNet.

Compared with traditional comprehension of sketch, we further split the temporal information of sketch into two types of feature, invariable structural feature (ISF) and drawing habits feature (DHF) with the aim of finer feature extraction in temporal information.

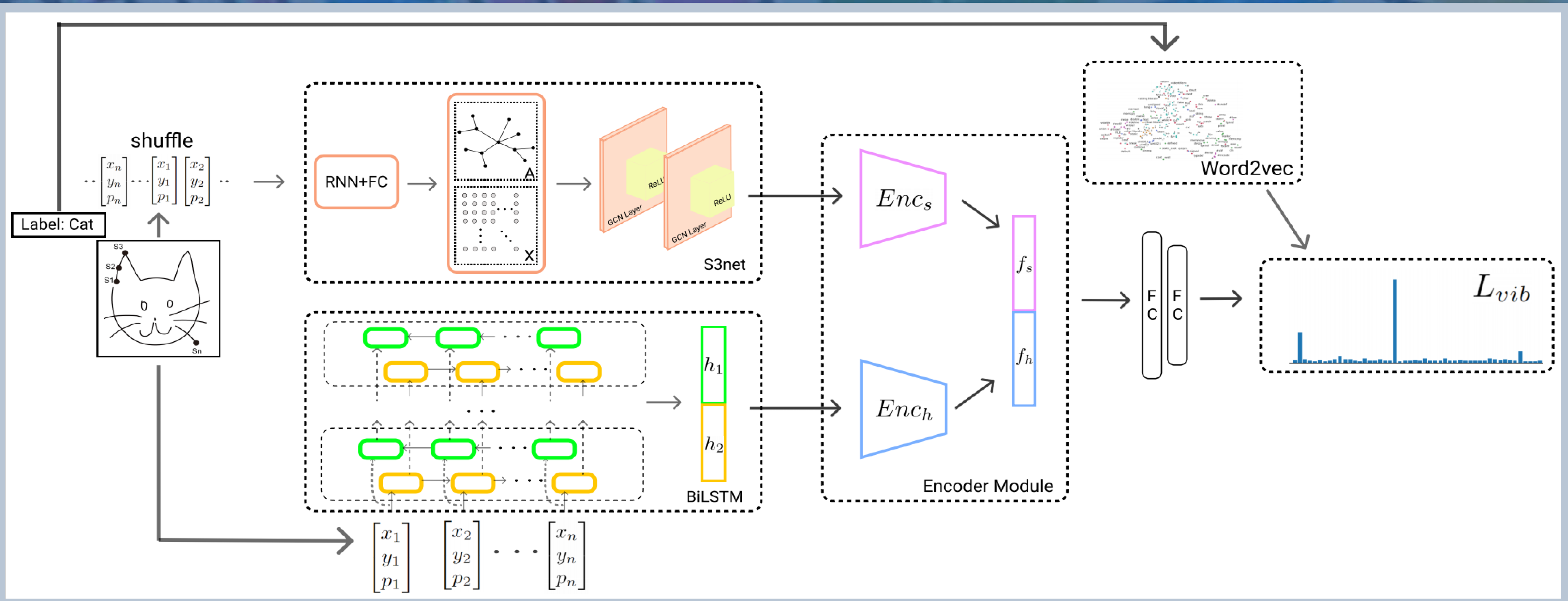
We propose a two-branch GCN-RNN network, Sketch-SNet, to extract two types of feature respectively, which surpasses state-of-the-art by a large margin.

Architecture

We propose a two-branch GCN-RNN network utilizing GCN to extract the structural characteristic and RNN to learn the common drawing habits.

As shown right hand, Sketch-SNet consists of three modules:

(i) parallel inputs of forward and strokes shuffled sequence in two-branch GCN-RNN network
(ii) two encoders which distinguish ISF and DHF
(iii) soft-label loss based on word2vec which generates soft semantic label.



Structural feature (ISF) & Drawing habits feature (DHF)

The ISF reflects the specific structural relationships between strokes – a car is made up of body and wheels.
And the DHF contains the specific temporal relationships between strokes – people prefer to draw body of a car first instead of wheels.

Most previous models crudely take raw temporal information as input to learn a feature for classification. We think that this operation will entangle DHF and ISF which increases the probability for the model to be stuck at local optimum and weaken the robustness of the model.

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%**

Extensive experiments on Quick-Draw dataset show that separating ISF and DHF in a two-branch network greatly improves the robustness of Sketch-Snet.

And Sketch-SNet reaches the highest accuracy of 90.42% which is over 5% above the S3Net (84.22%) and S3Net (Stroke-5) (85.10%).

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%

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

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.