Transformer on Inverse Handwritten Signature Verification

Yao Yichen, Fu Jie, Li Heyang, Wang Tiankai, Wang Yancong
2019533140 2019533142 2019533157 2019533139 2019533159
yaoych/fujie/lihy3/wangtk/wangyc@shanghaitech.edu.cn

Abstract

We present a new method that verify handwritten signature. Our approach effectively removing the need for hand-designed attentions. The main ingredients of the structure is a feature detector called Transformer FeatureNet. It uses transformer to extract attention information and share the attention between two images. Our work demonstrates accuracy performance on par with previous work based on Siamese network with hand-designed attention. Moreover, our network can be trained in a resonable time.

1. Introduction

Handwritten signature is often used in daily life to represent one's authorization. There is no doubt that it is important to identify whether the signature is fake or real. Teaching everyone to identify the genuine and forged ones is impossible, while machine learning is a feasible way. We choose to use deep learning method to achieve this task. The difficulty for this task is that every time you sign, your signature is different and at the same time fake signatures can be really similar. So we turned to Neural Network, which performance well at extracting features.

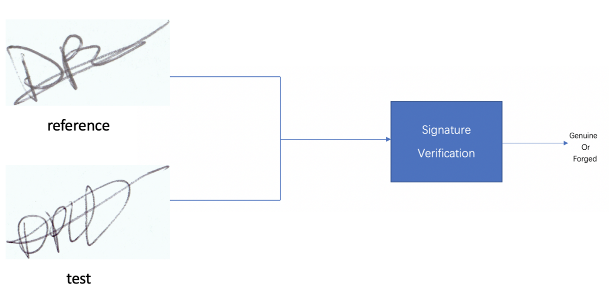

Figure 1. Illustration of handwritten signature verification.

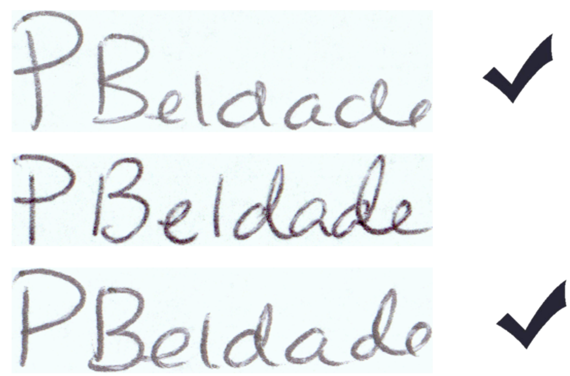

Figure 2. Instance of different signatures. Two images with mark are signed by one person.

2. Related Work

2.1. Other's method

Inverse Discriminative Networks for Handwritten Signature Verification[9] provides a method to verify signatures that uses attention information. That work implements its own attention layer and introduced inverse which inverse the image to enhance the dataset to provide more reliable output. As for datasets, there are many datasets that were publicly released, such as CEDAR[5] ,MCYT-75[4] , BH-Sig [6], and GPDS [3] [2] . However, there are no large-scale Chinese signature datasets existing in the current community, and because of this limit, we can't get access to any Chinese signature dataset, which is a pity. Also collecting dataset is almost impossible for us and we don't have enough human resources or financial utilities. So we turn to use datasets that have been publicly released.

2.2. Our method

Neural network based on attention have been developed since Inverse Discriminative Networks for Handwritten Signature Verification[9]. One of the examples is transformer[8] and some other works related to that. These works have achieved high performance in many areas, such as uncertainty estimation, transforming latent features to

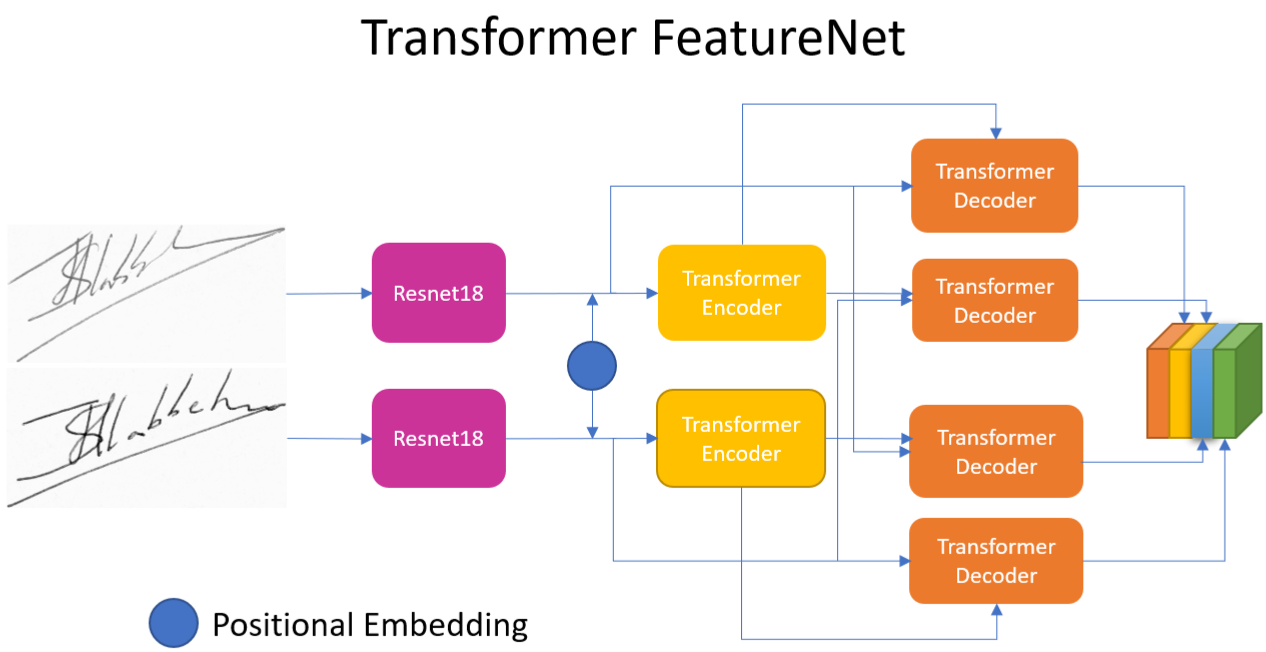

Figure 3. Pipeline of transformer featurenet

point clouds, NLP and so on[7][10][8]. So we choose to use transformer to replace its own attention layer. Transformer is a model that first used in NLP field. To make it work on the signature verification is a computer vision problem. We followed the method used in DE:TR[1], using a CNN network to extract features.

3. Model

We choose to a structure similar to DE:TR[1] to build our Transformer FeatureNet. We made two major difference. Firstly, we designed the input and output of transformer encoder, the target of the transformer decoder to have the same shape. For image1, we can get feature1 after CNN and positional embedding. For image2, we can get feature2. After passing feature1 and feature2 to transformer encoder we can get memory1 and memory2. Finally we pass feature1 and memory1, feature2 and memory2 and also feature1 and memory2 and feature2 and memory1 to transformer decoder. Since memory contain attention information, we can gain some features that focus on the parts that other image have from the image. Secondly we use Resnet18 instead of Resnet50 as our CNN network, to reduce parameter we need to learn and without sacrifice too much performance. For detail, we use 8 head, 6 layer transformer encoder and decoder with dmodel 256.

With the Transformer FeatureNet shown above, we can build our model. For any input image A and image B, we first extract features F_a and F_b from A and B by Transformer FeatureNet. F_a and F_b are 4 channel features, Through a 1x1 conv, we shrink the feature to 1 channel, After that we let the features go through AdaptiveAvgPool, fully connected layer and sigmoid function to get the outcome which use Inverse Discriminative Networks for Handwritten Signature Verification[9] as reference.

Inspire by Inverse Discriminative Networks for Handwritten Signature Verification[9], we also pass inverse the picture and pass the inversed image A and original B and original A and inversed B into net to get 2 more output.

When we use voting mechanism. We treat three output equally when voting.

3.1. Loss Function

In train,we apply L2loss on the all the outputs, and 0.4 weight on the output from two original image and 0.3 weight on the output from one inverse image and one origin image.

$$L2Loss_{=} \sum_{i=1}^{n} (y_i - f(x_i))^2 \quad (1)$$

4. Result

4.1. Implementation Detail

We use SGD as optimizer, $1e^{-3}$ as learning rate. We train the model on desktop computer RXT3080ti and AMD

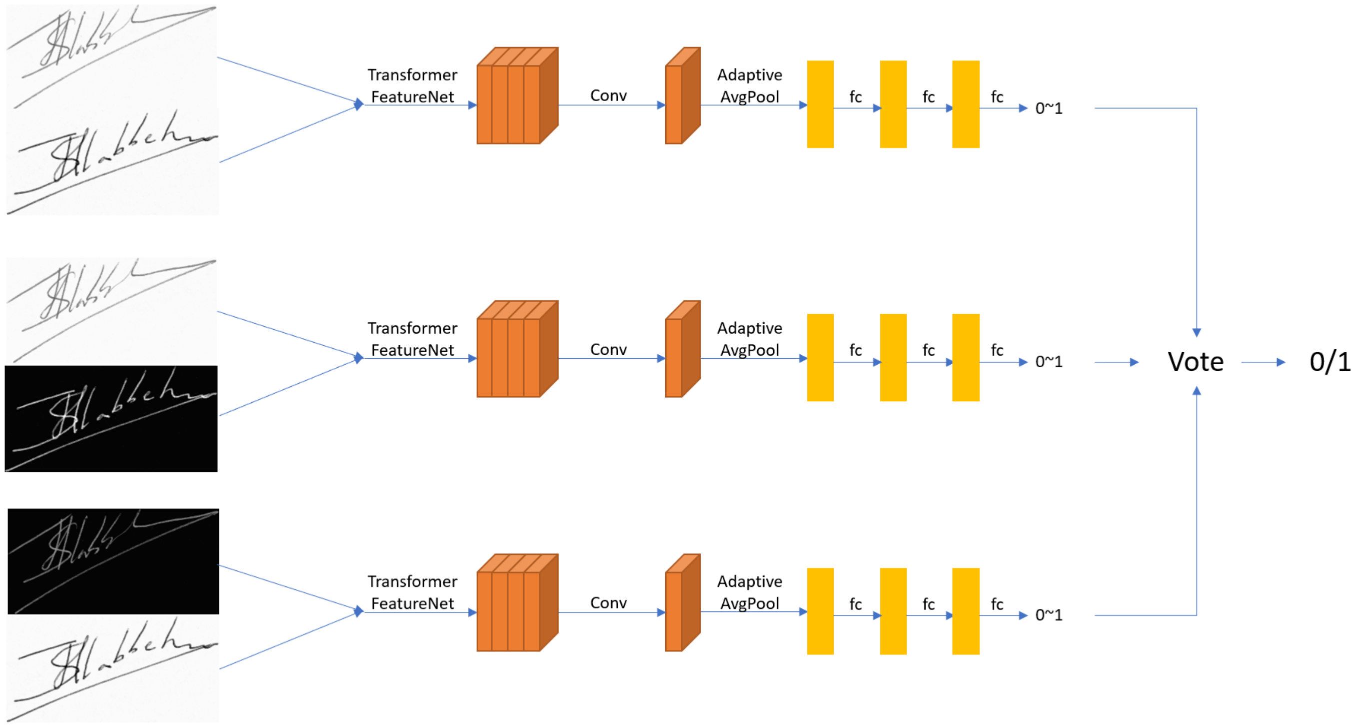

Figure 4. Total Pipeline

Ryzen 7 5800X and G of RAM. It takes about 5 to 10 hours to train.(Depend on the size of dataset)

4.2. Performance

Our model can run 15 images per second on a laptop with i7 8750H and GTX1060 gpu. GTX1060 is a commonly used economic level graphic card, which means our model can processed locally and almost real time on any machine.

4.3. Evaluation Metrics

We use False Rejection Rate (FRR), False Acceptance Rate (FAR),Accuracy (Acc) to evaluate our model and compare it with IDN[9]. False Rejection Rate (FRR) is the ratio of the number of wrong prediction on genuine samples divided by the number of genuine samples and False Acceptance Rate (FAR) is the ratio of the number of wrong prediction on forged samples divided by the number of forged samples. Accuracy(Acc) is the ratio of the number of correct predictions divided by the number of all test samples.

4.4. CEDAR Dataset

The CEDAR signature dataset contains signature samples of English names. It consists of 55 individuals' samples with each individual having 24 genuine and 24 forged signatures. Following other works, we choose 50 individuals' samples for training and 5 individuals' samples for test. For each individual, we get 276 reference-genuine pairs by matching all of the different ones. 276 reference-forged pairs are created by generating all possible pairs and randomly choosing 276 ones. In this case the number of genuine samples and number of forged samples are balanced.

We compare our model with IDN[9], which is the best model we find among all the other solutions. The data shows that our model outperforms its approach.

CEDAR(ENGLISH)			
	FAR	FRR	ACC
Our	0.20%	0	99.90%
IDN	5.87%	2.18%	95.98%

Table 1. Comparison on CEDAR dataset

4.5. ICDAR2011 Dataset

The ICDAR2011 signature dataset has Dutch signature samples. It contains 69 individuals' signatures smaples. We choose 59 individuals' samples for training and 10 individuals' samples for test. While the number of genuine samples and forged samples are not the same, making the reference-genuine set and reference-gorged set unbalanced.

So we created pairs as much as possible by generating all pairs of the smaller set, and randomly choose the same number of pairs in the bigger one. In this case, the number of reference-genuine pairs and that of reference-forged pairs of an individual will be the same, thus the total number of samples will be evenly balanced. We compare our model with IDN[9] on this dataset, and it shows that our model performs better.

ICDAR2011(DUTCH)			
	FAR	FRR	ACC
Our	0.79%	0	99.60%
IDN	5.91%	0	97.05%

Table 2. Comparison on ICDAR2011 dataset

4.6. Merged Dataset

To test the ability of our model to hold the condition of different language, we combine the CEDAR dataset(English) and ICDAR2011 dataset(Dutch). The result shows that the performance doesn't fluctuate a lot when more languages are added.

ICDAR2011(DUTCH) and CEDAR(ENGLISH)			
	FAR	FRR	ACC
Our(train on both language)	1.82%	0.65%	98.76%

Table 3. Signature verification accuracy on Merged dataset

4.7. Cross-language Test

Two datasets of different language are used in this work. We would like to know if it can correctly identify signatures of another language without knowing it in training. So we train our model on the English dataset and test it on the Dutch one, and also the reverse one. The result shows that the accuracy drops considerably, even lower than 50 percent.

test\train	CEDAR (ENGLISH)	ICDAR2011 (DUTCH)	CEDAR & ICDAR2011
CEDAR (ENGLISH)	99.90%	41.35%	100%
ICDAR2011 (DUTCH)	57.79%	97.29%	95.08%

Table 4. Signature verification accuracy on Cross-language Test

4.8. Quantitative Results

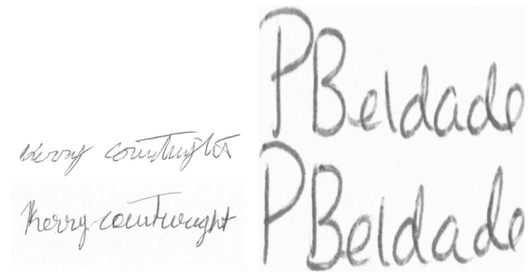

Figure 5. Successful detect genuine signature

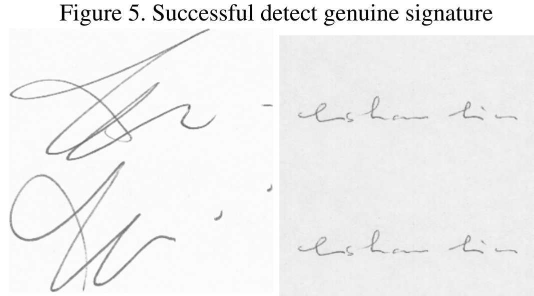

Figure 6. Successful detect forged signature

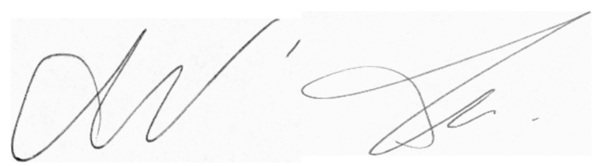

Figure 7. Failed to detect forged

5. Conclusion and Future work

Our network has promising performance on the language it is trained on. When training on the merged dataset, it can also reach high performance on each dataset. But when it encounters a language it has never seen, its outcome will be unreliable. The network runs fairly fast on desktop and laptops, but there is no way for it to run on mobile device, such as phone. Improving the performance and finding whether shrinking the scale of the network will be our future work.

References

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers, 2020.

[2] M.A. Ferrer, J.B. Alonso, and C.M. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):993–997, 2005.

[3] Miguel A. Ferrer, Moises Diaz-Cabrera, and Aythami Morales. Synthetic off-line signature image generation. In *2013 International Conference on Biometrics (ICB)*, pages 1–7, 2013.

[4] Julian Fierrez-Aguilar, Loris Nanni, Jaime Lopez-Peñalba, Javier Ortega-Garcia, and Davide Maltoni. An on-line signature verification system based on fusion of local and global information. In Takeo Kanade, Anil Jain, and Nalini K. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, pages 523–532, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[5] MEENAKSHI K. KALERA, SARGUR SRIHARI, and AIHUA XU. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(07):1339–1360, 2004.

[6] Srikanta Pal, Alireza Alaei, Umapada Pal, and Michael Blumenstein. Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, pages 72–77, 2016.

[7] Jiahuan Pei, Cheng Wang, and György Szarvas. Transformer uncertainty estimation with hierarchical stochastic attention, 2021.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[9] Ping Wei, Huan Li, and Ping Hu. Inverse discriminative networks for handwritten signature verification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5757–5765, 2019.

[10] Kaiyi Zhang, Ximing Yang, Yuan Wu, and Cheng Jin. Attention-based transformation from latent features to point clouds, 2021.