CNN-Based Chinese Character Recognition with Skeleton Feature

Wei Tang[1,2,3], Yijun Su[1,2,3], Xiang Li[1,2,3], Daren Zha[3], Weiyu Jiang[3], Neng Gao[3], and Ji Xiang[3]

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[2] State Key Laboratory of Information Security, Chinese Academy of Sciences, Beijing, China
[3] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{tangwei,suyijun,lixiang9015,zhadaren,jiangweiyu,gaoneng,xiangji}@iie.ac.cn

Abstract. Recently, the convolutional neural networks (CNNs) show the great power in dealing with various image classification tasks. However, in the task of Chinese character recognition, there is a significant problem in CNN-based classifiers: insufficient generalization ability to recognize the Chinese characters with unfamiliar font styles. We call this problem the Style Overfitting. In the process of a human recognizing Chinese characters with various font styles, the internal skeletons of these characters are important indicators. This paper proposes a novel tool named Skeleton Kernel to capture skeleton features of Chinese characters. And we use it to assist CNN-based classifiers to prevent the Style Overfitting problem. Experimental results prove that our method firmly enhances the generalization ability of CNN-based classifiers. And compared to previous works, our method requires a small training set to achieve relatively better performance.

Keywords: Chinese Character Recognition · Convolutional Neural Networks · Style Overfitting · Skeleton Feature.

1 Introduction

Chinese characters have been widely used (modified or extended) in many Asian countries such as China, Japan, Korea, and so on [22]. There are more than tens of thousands of different Chinese characters with variable font styles. Most of them can be well recognized by most people. However, in the field of artificial intelligence, Chinese character automatic recognition is considered as an extremly difficult task due to the very large number of categories, complicated structures, similarity between characters and the variability of font styles [3]. Because of its unique technical challenges and great social needs, during the last five decades there are intensive research in this field and a rapid increase of successful applications [15, 10, 17, 8]. However, higher recognition performance is continuously needed to improve the existing application and to exploit new applications.

Reducing Style Overfitting for Character Recognition via Parallel Neural Networks with Style to Content Connection

Wei Tang[1,2,3], Yiwen Jiang[1,2,3], Neng Gao[3], Ji Xiang[3], Jiahui Shen[3], Xiang Li[1,2,3], Yijun Su[1,2,3]
[1]*State Key Laboratory of Information Security, Chinese Academy of Sciences, Beijing, China*
[2]*School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China*
[3]*Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China*
{tangwei, jiangyiwen} @iie.ac.cn

***Abstract*—There is a significant style overfitting problem in neural-based character recognition: insufficient generalization ability to recognize characters with unseen styles. To address this problem, we propose a novel framework named Style-Melt Nets (SMN), which disentangles the style and content factors to extract pure content feature. In this framework, a pair of parallel style net and content net is designed to respectively infer the style labels and content labels of input character images, and the style feature produced by the style net is fed to the content net for eliminating the style influence on content feature. In addition, the marginal distribution of character pixels is considered as an important structure indicator for enhancing the content representations. Furthermore, to increase the style diversity of training data, an efficient data augmentation approach for changing the thickness of the strokes and generating outline characters is presented. Extensive experimental results demonstrate the benefit of our methods, and the proposed SMN is able to achieve the state-of-the-art performance on multiple real world character sets.**

***Index Terms*—character recognition, style overfitting, neural network**

I. Introduction

Characters have been widely used in our daily lives. There are probably tens of thousands of different characters with multifarious styles, and most of them can be well recognized by most people. But in the field of artificial intelligence, character recognition is considered as an extremely difficult task due to the very large number of categories, complicated structures, similarity between characters and the variability of styles. Because of its unique technical challenges and great social needs, there are intensive research in this field and a rapid increase of successful techniques, especially the end to end neural-based models such as Convolutional Neural Network (CNN), which is specifically designed to deal with the variability of 2D shapes of characters in documents [10]. Furthermore, this neural network has played an important role to improve recent image recognition studies [8], [14], [17].

However, the variability of styles still challenges character recognition due to the significant style overfitting problem: lack of generalization ability to recognize characters with unseen styles. For example, the trained models always perform poorly when test on the character sets with the styles that have never appeared in training data (see Sec. III-D for details). Therefore, in traditional applications, the training set needs to contain as many styles as possible to enhance the generalization ability of the recognizer, but this naive method is absolutely costly [8] and is contrary to the intention of building a human-level learning model to learn rich concepts from limited data [9].

To our best knowledge, few works are directly devoted to reducing style overfitting of character recognition. This is not surprising as the intractable challenge posed by style and content entanglement with such a large number of styles and content categories. The first study to disentangle two factors is using a bilinear model [19]. This bilinear model has been wildly used in zero-shot learning to associate visual representation and auxiliary class text description [1], [3], [21]. Recent study [26] successfully aplay this bilinear model to separate the content and style representations of characters. Furthermore, an intercross pair-wise optimization [16] have been proposed to extract relatively pure style representation by considering content to style feed. It is imperative take into account style to content feed for the purpose of extracting pure content representation to overcome style overfitting.

In this paper, we propose a novel framework named Style-Melt Nets (SMN), which disentangle the content and style factors of characters to extract pure content feature via style to content feed. More specifically, we take advantage of the superior capabilities of deep neural network for feature representation to build a style net and a content net, which learn the style and content representations respectively. To automatically disentangle the content feature from the style information, we first pre-train the style net supervised by the style label, and then feed the style feature to the content net besides the character images to extract relatively pure content feature.

Utilizing the domain knowledge of character structure is an effective method to reduce overfitting. It is based on an empirical observation: if two characters contain the same content but different styles, their structures are commonly invariant. Previous methods [11], [13], [16] usually cast the domain knowledge as the one-hot embedding. But such embedding methods require a lot of human resources for annotation.

SACR: Style-Aware Character Recognition via Feature Distillation and Correlation Learning

Wei Tang[1,2,3], **Yiwen Jiang**[1,2,3], **Neng Gao**[2], **Ji Xiang**[2], **Yijun Su**[2], **Xiang Li**[2]
[1]State Key Laboratory of Information Security, Chinese Academy of Sciences
[2]Institute of Information Engineering, Chinese Academy of Sciences
[3]School of Cyber Security, University of Chinese Academy of Sciences
No. 19 Yuquan Road, Shijingshan District
Beijing,China 100049
tangwei@iie.ac.cn

Abstract

In the task of character recognition, style and content are a pair of factors that interfere with each other but complement each other. This is disadvantageous to the traditional single-task model, because the entanglement of style features and content features will confuse the trained model. In this paper, we propose a novel multi-task model named Style-Aware Character Recognition (SACR) network that integrate both content and style supervision. More specifically, SACR use a feature distillation layer to disentangle the style and content features for interference eliminating, and it retain the beneficial correlation between them by a correlation learning layer to enhance the task-specific features. Experimental results prove that the proposed model is able to achieve the state-of-the-art performance on multiple character recognition tasks, and it obtains strong robustness in the case of insufficient training styles.

Introduction

Character has been widely used in our daily lives. There are probably tens of thousands of different characters with multifarious styles, and most of them can be well recognized by most people. But in the field of artificial intelligence, character recognition is considered as an extremely difficult task due to the very large number of categories, complicated structures, similarity between characters and the variability of styles (Long, He, and Yao 2018). Because of its unique technical challenges and great social needs, there are intensive research in this field and a rapid increase of successful techniques, especially the end-to-end trainable deep models such as Convolutional Neural Network (LeCun et al. 1998), which has played an important role to improve recent computer vision studies (Krizhevsky, Sutskever, and Hinton 2012; Simonyan and Zisserman 2015; Szegedy et al. 2015; He et al. 2016).

As deep models have achieved the great attention in computer vision community, it break the bottleneck of character recognition and achieve excellent performance even better than human on several popular benchmark datasets (Zhang, Bengio, and Liu 2017; Xiao et al. 2017; Wang and Liu 2018). But most deep learning models are data-thirsty, the generalization ability of the deep character recognition models will decline dramatically when the diversity of training styles is shrinking, which lead to a significant style overfitting problem: insufficient generalization ability to recognize characters with unseen styles (Long, He, and Yao 2018; Tang et al. 2018). In a broad sense, these styles could be font type, orientation and rotation angle, etc. The styles of characters are multifarious and arbitrary, so that in practice, to train a deep character recognition model, it is hard to provide a perfect training set with all styles. And it is absolutely expensive to use such massive amount of images to train a deep network. In view of this situation, a well trained model should obtain excellent generalization ability to recognize characters with unseen styles, which require the model became more aware of style.

Style is a very significant concept for character, but most of character recognition models ignore it. Note that people learning new concepts can often generalize successfully from just a single example, yet machine learning algorithms typically require tens or hundreds of examples to perform with similar accuracy, and people can also use learned concepts in richer ways than neural-based algorithms (Lake, Salakhutdinov, and Tenenbaum 2015). As a result of this fact, in traditional applications, to well train a deep model, the training set commonly contain as many sample with diverse styles as possible to avoid overfitting. Most single-task content supervised models ignore the concept of style, as the popular benchmark dataset commonly provide content label with out style annotation. Therefore, one of the most serious limitation of traditional content supervised models is incapable of disentangling content feature and style feature due to the absence of style supervision. Our motivation is to develop a style-aware model that could learn style concept for obtaining strong ability to resist overfitting.

Following this insight, in this paper, we propose a novel multi-task deep model named Style-Aware Character Recognition (SACR) network to separate the style feature and content features. Different from traditional approaches only supervised by content, SACR simultaneously leverage style and content supervision. More specifically, it adopt a shared convolution layer to extra shared feature of input

SCS: Style and Content Supervision Network for Character Recognition with Unseen Font Style

Wei Tang[1,2,3], Yiwen Jiang[1,2,3], Neng Gao[3], Ji Xiang[3], Yijun Su[1,2,3], and Xiang Li[1,2,3]

[1] State Key Laboratory of Information Security, Chinese Academy of Sciences, Beijing, China
[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[3] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
{tangwei,jiangyiwen,gaoneng,xiangji,suyijun,lixiang9015}@iie.ac.cn

Abstract. There is a significant style overfitting problem in traditional content supervision models of character recognition: insufficient generalization ability to recognize the characters with unseen font styles. To overcome this problem, in this paper we propose a novel framework named Style and Content Supervision (SCS) network, which integrates style and content supervision to resist style overfitting. Different from traditional models only supervised by content labels, SCS simultaneously leverages the style and content supervision to separate the task-specific features of style and content, and then mixes the style-specific and content-specific features using bilinear model to capture the hidden correlation between them. Experimental results prove that the proposed model is able to achieve the state-of-the-art performance on several widely used real world character sets, and it obtains relatively strong robustness when the size of training set is shrinking.

Keywords: Character Recognition · Convolutional Neural Networks · Style Overfitting · Style Supervision

1 Introduction

In our daily life, character is the most important information carrier. There are more than tens of thousands of different characters with variable font styles, and most of them can be well recognized by most people. But in the field of artificial intelligence, character recognition is considered as an extremely difficult task due to the very large number of categories, complicated structures, similarity between characters and the variability of font styles. Because of its unique technical challenges and great social needs, there are intensive research in this field and a rapid increase of successful techniques, especially the Convolutional Neural Network (CNN), which has played an important role to improve recent Computer Vision studies [9, 6, 14, 15, 4].

As CNNs have achieved the great success in general object recognition, face recognition and other image recognition tasks [7, 22, 14, 15, 3, 11], CNN-based