2019 International Collegiate Competition for Brain-inspired Computing

Technical Report

Team ID:	0048
Title:	Augmented Reality App for Interactively Learning and Exploring Chinese Characters
Topic:	Education
Embodiment:	Software
Affiliation:	TU Darmstadt / Tsinghua University
Team Members:	Jan-Christoph Klie (杨洋)
	Yelan Tao (陶晔澜)

Abstract

Chinese characters are an essential part of the Chinese language and culture. They are an integral part of becoming proficient in Mandarin. Learning even only the most common 2000-3000 characters is an arduous task that requires many hours of study. To support students, we implemented an augmented reality smart phone app for interacting with Chinese characters. When pointing the smart phone camera to a single character, we identify the character with machine learning and show its 3D representation above it. Its etymology, decomposition and mnemonic can be looked up. We use the knowledge about how the brain works and learns so that our users can better and more efficiently remember all the Hanzi. This is done by e.g. teaching character decompositions 六書, showing the decomposition in the 3D view or helping users coming up with mnemonics to connect Hanzi parts to a story.

The learning process implemented in Yangtao relies on the Dual Coding Theory of Reading and Writing which states that the brain best remembers if the item to remember, in this case Hanzi, is associated with as many different cues as possible. Therefore, we present the users with visual cues (the 3D Hanzi), audio cues (the pronunciation) and the logic behind the Hanzi (its decomposition, etymology and kind of formation).

Our prototype supports 250 selected characters. In the next iteration, we want to add online learning/lifelong learning capabilities so that the character recognition is improved by user feedback if it was wrong, adapting the app to the users' environment. This is similar to how the brain continually learns and adjusts to new stimuli.

The code is available under https://github.com/jcklie/iccbc-2019.

Keywords: Augmented Reality, Smartphone App, Chinese learning software, Gamification, Character Recognition, Hanzi

Contents

1 Background

Chinese is one of the fastest growing language in the world. More and more people decide to learn it as a foreign language for reasons like travel, business or better getting to know the rich and vast culture of China. Chinese characters (Hanzi) are an essential part of Chinese everyday life, history and culture. Therefore, when learning Chinese, learning Hanzi is of utmost importance to become a proficient Chinese speaker. Learning even only the most common 2000-3000 Hanzi is an arduous task that requires many hours of study. Native speakers are permanently exposed to Hanzi and spend many hours over several years in school to learn them. For second language learners, they are a major hurdle to overcome and perceived as one of the more difficult aspects of learning Chinese (Yang 2018). This gives an indication on what a challenging undertaking it is to learn Chinese. Students are in need of novel approaches to help learning the Hanzi. Thus, we aim at making this process more interesting and engaging to help students of Chinese to master the Hanzi by using their idiosyncrasies.

1.1 Hanzi

In order to more effectively and efficiently learn Hanzi and understand them more deeply, one has to comprehend their inner workings. Seeing Hanzi for the first time, they might appear intimidating and confusing. But they follow certain rules and their properties can be related to 6 principles of character formation. Understanding these is tremendously helpful in the quest of learning the most common Hanzi. These principles have been defined in the *Shuowen Jiezi* (说文解字) by ancient scholar Xu Shen 许慎 in the 2nd century.

A common misconception about Chinese characters is that they resemble images. While this is true for some, only a small subset of all Hanzi are categorized as pictograms (象形). Examples are 日 (sun), 月 (moon) or 大 (big). Related to them are ideographs (指事) which resemble more abstract concepts, e.g. (up), 下 (down) or 三 (three). Compound ideographs (会意) combine several pictographic or ideographic characters to suggest a related meaning. Examples include 森 (forest - three trees) or 明 (bright - sun and moon). Rebus (phonetic loan) characters (假借) are characters that borrow parts with a similar pronunciation but different meaning. 少 was originally used to describe sand, now it is used to express *few*, 要 was waist, now is used for *to want*. Over 90% of modern Chinese characters are phono-semantic compound characters (形声). They are constructed by two parts, a phonetic component (a character with approximately the correct pronunciation) and a semantic component (a character which supplies an element of meaning, often the radical). Radicals are a graphical components that is often not a character by itself, but frequently gives semantics, for instance 扌 (hand) or 忄 (heart). Some examples are 梅 (plum, pronounced mei3) with consists of the wood radical 木 for meaning and the phonetic component 每 (every, pronounced mei3). The last and least understood group are transfer characters (转注). They often had the same etymological root and a similar pronunciation, but developed into different characters, e.g. 父 and 爸 (father).

Using these principles allows learners to more efficiently learn Hanzi, as radicals and ideographs can give clues about meaning and knowing their pronunciation helps with phono-semantic compound characters. As Hanzi developed over time and were simplified, understanding the original shape of a character often helps to better relate the meaning to its shape. The source of origin of a character is called etymology. For instance, the original meaning of the character 西 now *west*, was a bird in a nest. Xu Shen said that the bird rested in the nest when the sun was in the west and therefore changed the meaning to *west*.

1.2 Yangtao

Learning Hanzi by rote requires large amounts of time and perseverance. Studies have shown that using several aspects of Chinese characters in combination, e.g. using the radicals, its etymology, mnemonics, phonetics, semantics and overall shape in general improves long-term recollection (Lam 2016). Our entry for the ICCBC 2019 therefore is an serious game app called `Yangtao` that uses augmented reality (AR) to stimulate the visual senses and make it interesting to interact with Hanzi. The user uses his or her phone to scan Hanzi in a wild,the camera image then is used to project a 3D model of the character. It then can be investigated as if it was a real object, including lighting and perspective (Fig. 2). Its decomposition is color coded; same parts in different Hanzi use the same color to further strengthen the memorability. Besides displaying Hanzi in 3D, we also provide users with the pronunciation of a character, its etymology, decomposition and allow them to define a custom mnemonic (Fig. 3). A mnemonic in our case is a story that connects properties of the character, for instance for 植 (plant), which is composed of 木 (tree) and 直 (straight), one could come up with a story like: *When you **plant** plants, you use a piece of **wood** to make the plant grow **straight***. While it is for some Hanzi easier to come up with a story than others, it is almost always possible to find one after enough iterations. This learning style is famously used in the *Remember the Hanzi series* by Heisig and Richardson 2008.

1.3 Inspired by the brain

Regarding the call for entries of the ICCBC 2019, the contest description states that

> The contest uses an open proposition. The theme of your entry should be a general brain-inspired intelligent system. It can be innovative, a market-oriented application,

Figure 1: The collection shows characters that can be scanned or already scanned (marked by a star)

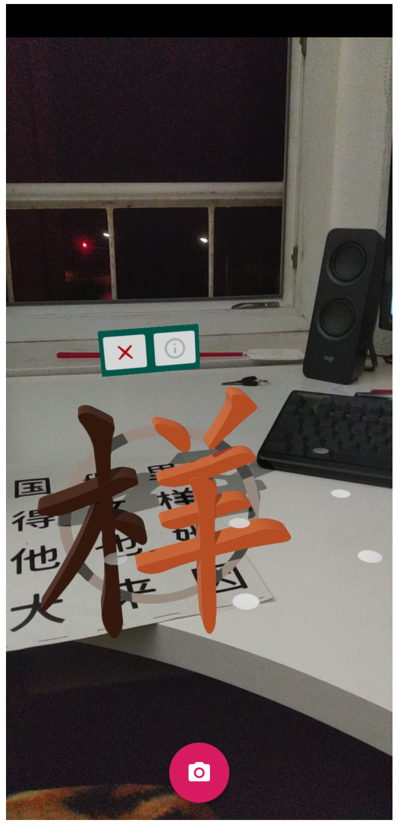

Figure 2: The augmented reality view allows users to scan and then explore characters in 3D. Clicking on the info sign leads the user to the details view.

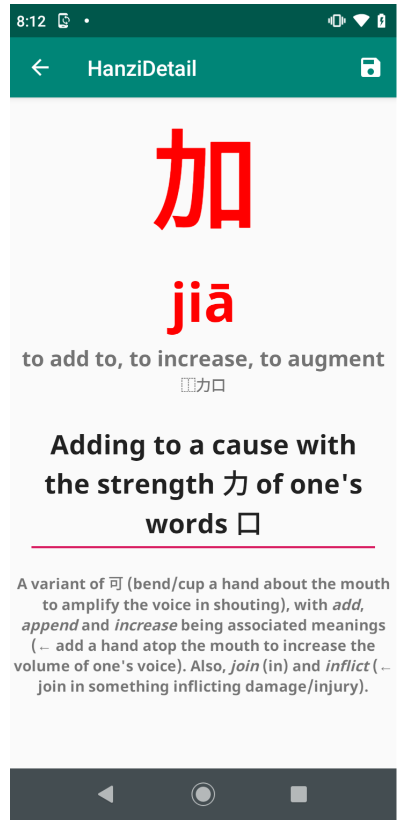

Figure 3: The detail view shows information regaring pronounciation, meaning, decomposition, etymology and user-defined mnemonic.

> or advanced research breakthrough, including but not limited to: robots, unmanned aerial vehicles (UAVs), intelligent wearable devices, natural language processing systems, human–computer interaction systems, smartphone apps, algorithms, and software. New ideas, innovative algorithms, and novel applications are highly encouraged. Your entry should reflect versatility, practical value, and application potential.

Our entry is a **novel application** in form of a smartphone app in the **education category**. It is brain-inspired by levering theories about how the brain and memory work for learning, e.g. the *A Dual Coding Theory of Reading and Writing* (Paivio 1969). In a nutshell, it describes that the brain e.g. stores the word *panda* and the image *panda* in a different form, both can be used to recall this information, increasing the chance to remember when two instead of only one clue is used. This understanding of the brain is what we use when providing users with many different aspects of Hanzi.

In addition to that, we use gamification and in a way reinforcement learning. The user is incentivized to scan more and more characters, thereby improving his knowledge and getting rewarding by a better Hanzi understanding. This kind of serious game has been reported to work in educational settings (Sailer and Homner 2019). A good overview about gamification can be found in (Hamari, Koivisto, and Sarsa 2014).

The scanning of characters for the augmented reality part uses machine learning. In the future, it is planned to use online learning/life long learning approaches to retrain the model after deployment, similar to how the brain permanently adjusts to new stimuli and experiences. This can improve the character classification in the field, as it is trained on actual data from real users.

2 Innovation

Our innovation lies in the use of augmented reality, machine learning for character recognition and the combinatorical aspect of Hanzi. We combine these aspects to develop an engaging smartphone app to learn and engage with Chinese characters both for second-language learners and native speakers. While there are systems that address one of these parts, we are to the best of our knowledge the first to combine these.

Our main contributions are

- We combine the six principles of character formation and augmented reality to engage students of Chinese to more and better learn Hanzi.
- We trained a machine learning model to recognize Chinese characters on the smartphone in real time.
- We developed a completely viable proof-of-concept Android app that supports the 250 most common characters in modern Mandarin, together with colored 3D models and dictionary entries consisting of etymologies, decomposition and mnemonics.

3 Method

3.1 Challenge Description

The following sections describe the key challenges this competition entry faced during development.

3.1.1 Character recognition

In order to display the character in 3D that the user is interested in, it has to be first localized and detected. While this is not the focus of this project, it is in itself a hard problem in which decades of research has already been conducted on. For latin scripts, the approaches are already very good and off-the-shelf tools are widely available. But for Chinese, it is much harder due to the vast number of characters and fonts that are used in the wild compared to the Western alphabet[1]. Therefore, systems often need more training data, larger networks and more computational resources to be successful.

It has also to be taken into account that Yangtao will run on a mobile device with limited computational resources. A model needs to be small and fast for the app to appear responsive and interactive to the user.

3.1.2 Augmented Reality

Augmented Reality (AR) describes enhancing real-world environments with computer-processed content, e.g. when navigating using a navigation smartphone app to add a directions arrow over the camera live feed. This is difficult, as it requires 3D scene understanding and tracking of objects in the real world, either because they are moving and/or the camera moves. Finally, the camera has to be calibrated so that dimensions in the real world correctly correspond to dimensions in the virtual world.

In order to display Hanzi via AR, they have to be converted to 3D models first. Its parts need to be highlighted to show its decomposition for learners. Creating these models by hand is not feasible for our resources available, therefore, a better way has to be found.

3.2 Solution

The following sections describe how the aforementioned challenges are solved in the final implementation of the Yangtao app.

3.2.1 Android App

As augmented reality is a difficult problem that is out of scope of this project, we use an off-the-shelf toolkit for it. After researching different frameworks, we settled for the ARCore framework from Google[2] in combination with their Sceneform AR utility library. These can be easily integrated

[1] A benchmark for OCR software can be .e.g found in https://ocr.space/blog/2015/03/best-ocr-software-for-chinese.html

[2] https://developers.google.com/ar/

into Android to discover planes on which 3D models can be placed[3]. We add our own user interface and controls to the standard camera AR view. The app itself is written in Kotlin. ARCore is also available for iOS, making it possible to port Yangtao later.

Designing 3D models for every Hanzi by hand and coloring it was infeasible for our available time and resources, therefore, we fully automated this step. As the basis, we leverage vector graphics that were generated from freely available fonts by the *Makemeahanzi* project[4]. For each stroke in every character, they provide information to which part in its decomposition the stroke belongs. This is used to color the decompositions of Hanzi. Fig. 4 shows the different stages. Then, we use the Python API of the 3D modeling software Blender[5] to extrude the 2D SVG into a 3D model.

Figure 4: Different stages when creating 3D models from SVG. First, we color the different parts in the decomposition, then we use Blender to generate a 3D model. As a side note, the depicted Hanzi 想 (xiang3, to want, to miss) here is composed of 心 (heart) which gives the meaning and 相 which gives the pronunciation (xiang1). It is important to note that 木 and 目 belong together to form 相, viewing them seperately would be incorrect. With the coloring, we make this important information available to learners.

For the character details, we use the *Makemeahanzi* dictionary that ships with the character stroke information. For the etymology, we use the freely available *Etymological Dictionary of Han/Chinese Characters* by Lawrence J. Howell[6].

[3] It requires a modern and supported smartphone though.
[4] https://github.com/skishore/makemeahanzi
[5] https://www.blender.org/
[6] http://nihongo.monash.edu/Etymological_Dictionary_of_Han_Chinese_Characters.pdf

3.2.2 Character recognition

For the character recognition itself, we decided to use neural networks, the state-of-the-art-approach for this task. When starting, we planned to perform character detection, that means detecting and localizing characters at the same time. But we realized that it is much more difficult than character recognition and also unnecessary, as the user can point the camera to the character he or she wants to scan. Therefore, the task we try to solve is, given an image that contains a character, predict the correct character.

For the detection part, we experimented with different datasets. As a first attempt we used the dataset by Yuan et al. 2019 which contains street view images from Chinese cityscapes where characters from e.g. street signs and shop fronts are annotated with a bounding box and the respective character. Based on this information, we extracted these single character patches and trained a MobileNet V2 (Sandler et al. 2018). We use the MobileNet here, as it is a convolutional neural network (CNN) optimized to be running on mobile devices by replacing computationally expensive operations and using quantized weights. We quickly came to a roadblock as extracted patches are very small (often below 20 pixel), classes are heavily imbalanced and only common characters are contained. Therefore we did not continue to use it in production.

In the end, we used the dataset by Zhong, Jin, and Feng 2015 which contains images of single Hanzi generated from font files. Each dataset contains characters from 120 respectively 280 fonts. We use the larger dataset to maximize training data. For training, we use a simple multilayer perceptron (MLP) with *Histogram of oriented gradients* (HOG) (Dalal and Triggs 2005) as input features. It has been shown that these are on-par with CNN while using much fewer resources during training and inference (Mishra, Alahari, and Jawahar 2012). We also observed this. The model itself is trained using Tensorflow 2.0. To run it on mobile, we convert it to the tensorflow lite format after training.

When opening the AR view, the user first has to wait for planes to be detected, for instance tables or the floor, as these are used to place 3D models on. Once tracked, the user has to point the camera at a Hanzi of interest. If he pressed the *scan* button, we capture the camera image and preprocess it first to remove noise and crop the character to fill the whole frame. The preprocessing steps are illustrated in Fig. 6. Then, we compute HOG features of the image and feed it to our previously offline trained classifier. The result is a list of the most likely characters the image could contain. The user can now either choose that the top predicted Hanzi is indeed the Hanzi he scanned, select from the top 10 list if it was recognized wrongly or abandon the current scan and try again, We decided to only let the user choose from the top 10 in order to not overwhelm him and to prevent cheating. This process can best be seen in the accompanying

Figure 5: Examples from the SPCCI dataset for the characters 他 and 东. For training, we use the 250 most common Hanzi with images created from 280 different fonts. It can be seen that fonts are every different even for simple characters. Also, some render the traditional version of a character. This variability makes it challenging for a machine learning model to properly classify.

video.

3.3 Result

This section shows the quality of the character recognition. As we do not have testing dataset available for our task of character recognition in the wild, we evaluate intrinsically on the test split of the used datasets as well as extrinsically by using the app.

3.3.1 Intrinsic evaluation

We use a multi-layer perceptron with hidden layers of size 512 and 256. Between these, we apply during training 50% dropout to prevent overfitting. As activation, we use Rectified Linear Units (ReLU). We train and test the model on different number of characters in order to observe how the model behaves with a larger classifier. Throughout the experiments, we do not change the model, all runs use the same parameters. We use early stopping once the validation loss does not improve anymore. We employ the same train/test split as (Zhong, Jin, and Feng 2015) in which each Hanzi has 240 images in the training set and 40 in the test set. For every model, the training time does not exceed 10 minutes when training on an Nvidia RTX 2060 Super GPU. The results can be seen in Table 1.

It can be seen that while the performance overall decreases with the number of classes, it still is reasonably accurate, especially for the Accuracy@10. This result means that even if the model wrongly predicts the top candidate, the user still can recover by selecting from the list of top 10 predictions, with often contains the correct one.

Figure 6: The different preprocessing steps performed after scanning: ① The original image ② Cropping to the region of interest ③ Binarization to black and white only ④ Using erosion and Dilation to remove specks and close contours ⑤ Computing a bounding box around detected contours and cropping to it

3.3.2 Extrinsic evaluation

For the extrinsic evaluation, we prepared an A4 sheet and printed 12 characters on it, each around 5×5 cm large. We scanned all characters on these sheets several times and were always able to scan it correctly after at most two attempts. In addition to that, we also tried a book that teaches Hanzi to children which contains large printed characters. While these are more difficult due to shadows, different fonts and the pages not being totally flat, we still were able to scan these in most cases.

We see that even if we not center the character in the focus, the preprocessing steps help us to align it properly for the classifier. Problems we observed are if users do not fully capture the character or if they scan it with too much of an angle, thereby warping it. Also, noisy backgrounds that cannot be filtered out fully confuse the classifier.

	Train		Test			
n	**Acc@1**	**Acc@10**	**Acc@1**	**Acc@10**	**#Train**	**#Test**
100	97.05	99.96	93.02	98.87	24,000	4,000
250	89.04	99.24	88.11	97.08	60,000	10,000
500	75.91	96.33	79.88	94.21	120,000	20,000
1000	57.31	89.08	67.62	88.99	240,000	40,000
2000	26.38	65.18	37.70	72.24	480,000	80,000

Table 1: Evaluation of the character detection component of Yangtao for different numbers of characters to predict **n**. We report Accuracy@1 (is the top prediction the gold label?) and Accuracy@10 (is the gold label in the top 10 predictions of the model?). One can see that while Accuracy@10 drops quickly when increasing the number of characters, the Accuracy@10 drops much slower. Also listed are the number of data points in the training and test set for each run.

4 Conclusion

With the rapid economic growth of China, more and more people are learning Mandarin. Hanzi are a key component of this. Learning the Hanzi is difficult, time-consuming and it is easy to forget them when not using them. That is why we use the understanding of Chinese characters and how the brain better learns to build a smartphone app that is novel. interesting and engaging for users. The user can use his smartphone to scan characters he sees in the wild, we then use AR to project the Hanzi right in the smartphone view. It can then inspected and its properties can be looked up in a dictionary we also ship. Predefined mnemonics can be used or custom ones can be defined to never forget the scanned Hanzi ever again.

The popularity of AR games, e.g. *Pokémon Go*[7] or *Harry Potter: Wizards Unite*[8] has shown that using this technology creates a very immersive experience. Therefore, we hope that our app can be helpful to students of Chinese and individuals interested in learning more about Hanzi.

5 Outlook

For this competition entry, we built a fully working prototype of an AR Android app. There are several extensions and projects we want to perform in the future to make it even better. These are described in the following section.

[7]https://www.pokemongo.com/en-us/

[8]https://www.harrypotterwizardsunite.com/

5.1 Technical Outlook

We have shown that our classifier based on HOG features works well and is quickly to train. As the performance drops when extending to support more characters, we plan to incorporate more features. It has been shown that e.g. Gabor filters work well for character recognition (Qiang Huo, Yong Ge, and Zhi-Dan Feng 2001). These can easily be added to our existing MLP, thereby keeping its fast training and inference times which would be lost when using e.g. a CNN.

It would be also interesting to add a second model that can detect Chinese handwriting. In addition to learning Hanzi, this would also aid by learning to read Chinese handwriting, which is quite difficult for foreigners.

Other new functionalities we want to add are animating the 3D models to show the stroke order and a mode in which the Hanzi can be drawn by the user in 3D.

Finally, the detection of characters right now relies on good offline training. It requires data from the target domain, that is the characters that users are most likely scan in the environment of the user. If the training data does not contain a certain setting the user will try, e.g. street signs, then the model performance will likely not be good. That is why we intend to incorporate online training/lifelong training, in which the feedback of the user is used to train the model further on his smartphone. To prevent catastrophic forgetting, that is overfitting to the new domain and forgetting about the original training, special care and algorithms have to be employed.

5.2 Application Outlook

After the contest takes part, we intend to deploy it to the Android app store and make it also available as apk from our personal homepage. We want to add a social component in which mnemonics can be shared. This enables users that might not have a good mnemonic to use one from another user. This social aspect hopefully further incentivises students of Hanzi to engage more with the app and Chinese characters. The code is available under https://github.com/jcklie/iccbc-2019.

6 Team Introduction

Jan-Christoph Klie (杨洋) is a second year PhD student at TU Darmstadt, Germany in Natural Language Processing. Currently, he researches how text annotation can be interactively supported by machine learning. He was responsible for the app development and the technical document. In his free time, he likes to swim, read, program and to learn Chinese.

Yelan Tao (陶晔澜) is a senior undergraduate from School of Vehicle and Mobility at Tsinghua, China. She converted characters into 3D models and designed different material properties for them with Blender, prepared the presentation and the video as well as managed the team. Her hobbies include playing soccer, reading, drawing and learning German.

References

Cai, Qing and Marc Brysbaert (June 2010). "SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles". In: *PLOS ONE* 5.6, pp. 1–8.

Dalal, N. and B. Triggs (2005). "Histograms of oriented gradients for human detection". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1, 886–893 vol. 1.

Hamari, Juho, Jonna Koivisto, and Harri Sarsa (2014). "Does gamification work? - A literature review of empirical studies on gamification". In: *Proceedings of the 47th Annual Hawaii International Conference on System Sciences, HICSS 2014*, pp. 3025–3034.

Heisig, James W. and Timothy W. Richardson (2008). *Remembering Simplified Hanzi 1: How Not to Forget the Meaning and Writing of Chinese Characters*. University of Hawaii Press.

Lam, Carol Xiang (2016). "Investigating Semantic Alignment in Character Learning of Chinese as a Foreign Language: The Use and Effect of the Imagery Based Encoding Strategy". PhD thesis. University of Portland.

Mishra, A., K. Alahari, and C. V. Jawahar (2012). "Top-down and bottom-up cues for scene text recognition". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2687–2694.

Paivio, Allan (1969). "Mental Imagery in Associative Learning and Memory". In: *Psychological Review* 76.3, pp. 241–263.

Qiang Huo, Yong Ge, and Zhi-Dan Feng (2001). "High performance Chinese OCR based on Gabor features, discriminative feature extraction and model training". In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*. Vol. 3, 1517–1520 vol.3.

Sailer, Michael and Lisa Homner (2019). "The Gamification of Learning: a Meta-analysis". In: *Educational Psychology Review*.

Sandler, Mark et al. (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yang, Juan (2018). "What makes learning Chinese characters difficult? The voice of students from English secondary schools". In: *Journal of Chinese Writing Systems* 2.1, pp. 35–41.

Yuan, Tai-Ling et al. (2019). "A Large Chinese Text Dataset in the Wild". In: *Journal of Computer Science and Technology* 34.3, pp. 509–521.

Zhong, Z., L. Jin, and Z. Feng (2015). "Multi-font printed Chinese character recognition using multi-pooling convolutional neural network". In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 96–100.

Appendices

A Included characters

We use the 250 most common characters mostly based on Cai and Brysbaert 2010. They are

一三上下不与世业东两个中为主么义之也了事二于些产人什从他代以们件任会但位体何作你使信做儿先入全公关其内再军几出分利别到制前力加务动化十原去又及反发受变口只可各合同名后向员和四回因国在地场声处外多大天太头女她好如子学它安定实家对将小少尔就工己已常平年并应度建开当很得心必性总情想意感成我或战所手才打把报接提政教数文斯新方无日时明是更最月有本机条来果样次正此比民气水没法活海点然物特现理生用由电的目相看真着知神种立第等系经结给美老者而能自行表被西要见解认论话说走起身过还这进通道那部都里重量长门问间面题高