Component Awareness in Convolutional Neural Networks

Brian Kenji Iwana*, Letao Zhou*, Kumiko Tanaka-Ishii†, Seiichi Uchida*
*Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan
Email: {brian, shu, uchida}@human.ait.kyushu-u.ac.jp
†Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan
Email: kumiko@cl.rcast.u-tokyo.ac.jp

***Abstract*—In this work, we investigate the ability of Convolutional Neural Networks (CNN) to infer the presence of components that comprise an image. In recent years, CNNs have achieved powerful results in classification, detection, and segmentation. However, these models learn from instance-level supervision of the detected object. In this paper, we determine if CNNs can detect objects using image-level weakly supervised labels without localization. To demonstrate that a CNN can infer awareness of objects, we evaluate a CNN's classification ability with a database constructed of Chinese characters with only character-level labeled components. We show that the CNN is able to achieve a high accuracy in identifying the presence of these components without specific knowledge of the component. Furthermore, we verify that the CNN is deducing the knowledge of the target component by comparing the results to an experiment with the component removed. This research is important for applications with large amounts of data without robust annotation such as Chinese character recognition.**

I. Introduction

In recent times, artificial neural networks have been successful in many domains and in particular, Convolutional Neural Networks (CNN) [1] have achieved state-of-the-art results in image recognition [2]–[4]. However, these methods tackle entire image classification. As images become more complex and computer vision tasks address multiple possible objects, individual components need to be detected.

Object detection attempts to tackle this by identifying and annotating objects within an image. However, it is a difficult task because of object variation and interference from the background and other objects. Within an image frame, objects can vary in size, appearance, and location. They can also suffer from occlusion, complex backgrounds, have multiple instances, and be overshadowed by other objects. Traditionally, region-based CNN object detection methods use bounding boxes to isolate region proposals for instance-based classification [5]–[7]. Another approach is to use pixelwise likelyhood maps for semantic segmentation [8]. However, these methods require instance-level supervision during training.

Acquiring the large amounts of image data with detailed object annotations that is required can be difficult and costly to obtain. In this work, we focus on the use of *weakly* supervised labels consisting only entire image-level class annotations of images containing multiple objects. The primary contribution of this paper is to investigate the ability of CNNs to learn the awareness of the components with only the inference that it exists somewhere within the context of the image.

Fig. 1: Examples of Chinese characters with their components highlighted in color.

Secondly, we demonstrate the use of CNNs to detect the components, or radicals, that constitute Chinese characters. Chinese characters are logograms that can be deconstructed into one or more graphical components, shown in Fig. 1. The components are a combination of strokes that form the primitives of a character and the recognition of these components is important for the character's identification. The proposed research detects the presence of these components within Chinese characters with only character-level component supervision.

In addition, to further verify the CNN's capability of deducing the structure of the components, we evaluate the model on images of the Chinese characters with the component removed. By testing images with the target component removed, we ensure that the CNN was able learn using inferred knowledge of the object and not solely rely on surrounding semantic indicators. This confirms the usability of data lacking prior knowledge of the target's form for detection tasks.

The remaining of this paper is organized as follows. In Section II, we briefly summarize the previous work in object detection and Chinese character component recognition. Section III elaborates on the structure of Chinese characters and Section IV reviews CNNs and discusses their use. In Section V, we will expand on the details of our approach and report the results of the experiment. Section VI demonstrates the effect that removing the component has on the classification of characters. Finally, we draw a conclusion and describe future work in Section VII.

II. Related work

A. Object Detection

Beyond classification, object detection and segmentation with convolutional models is an active field. A conventional approach to tackle object detection use of a CNN with a sliding window view of the input images [9]. However, this approach is limited to a constrained class, e.g. faces, and is a brute force method to address scale invariant object detection and localization.

Multi-scale region-based methods isolate regions of interest containing objects within an image using bounding boxes [5], [6]. Regions with CNN features (R-CNN) [7] generates category-independent region proposals for the CNN. The extracted region proposals are warped and classified independently. Furthermore, the bounding box can be trained using Support Vector Machines (SVM) [7], multi-task loss with backpropagation [10], and Region Proposal Networks [11].

Another approach to component identification is the use of Fully Convolutional Networks (FCN) [8] for end-to-end for pixelwise prediction and segmentation. To achieve pixel-to-pixel training, the FCN is provided with full images with an equal sized segmentation ground truth. FCNs have efficiently achieved state-of-the-art results in semantic segmentation for natural images [8] and scene text detection [12], [13].

The difference between these methods and the proposed research is that they require detailed ground truth of the target objects or their locations. Object detection and localization using weakly supervised models also has a rich history [14]–[16]. Recently, CNNs have also been used for the localization of objects with weakly supervised training [17]–[20]. For instance, Bazzani et al. [21] localizes objects by monitoring a CNN recognition accuracy when masking regions.

In addition, some word spotting methods use a query-by-string (QbS) approach. This approach searches documents for a target word having only a textual representation. Some of these methods include bag-of-features Hidden Markov Models [22] and using Pyramidal Histogram of Characters (PHOC) as labels enabling QbS [23], [24].

B. Chinese Character Component Detection

In literature, there have been past attempts to detect or extract Chinese character components. Component-based Chinese character recognition is a promising field of online [25]–[27] and offline [28] handwritten Chinese characters due to the high variation of handwriting. These methods generally use predefined templates or features to recognize the Chinese character components. Ren et al. [29] uses statistical rules to determine the layout of Chinese characters in order to tackle scene text detection. To the best knowledge of the authors, this is the first trial in using a CNN to detect the presence of Chinese character components.

III. Composition of Chinese Characters

Chinese characters are logograms that represent words or part of words, written within the framework of a square. The characters can be separated into two categories, basic characters and compound characters. Compound characters are characters that combine two or more components within the bounds of a single character square and they account for more than 95% of Chinese characters. For example, the Chinese character “明” has two components, “日” and “月.” These components are often semantic indicators, suggesting the meaning of the character, or phonetic indicator, indicating the pronunciation of the character. However, in some cases, the components of modern characters have lost their original phono-semantic meanings due to obfuscation through simplification, evolution, and stylization [32].

TABLE I: Chinese Compound Character Structures

Description	Example Characters
Left to Right	頌 泙 剖
Above to Below	磊 冗 甾
Left to Middle to Right	椰 蜘 獄
Above to Middle to Below	菅 蠱 亨
Full Surround	団 図 瓦
Surround from Above	岡 閑 尽
Surround from Below	凶 鼎 函
Surround from Left	匣 玉 区
Surround from Upper Left	友 庄 屐
Surround from Upper Right	弍 句 司
Surround from Lower Left	咫 勉 遅
Overlaid	坐 丗 夾

While the structures of most compound characters are easily identified, there are many instances which can be difficult. For example, there are some cases where the components intersect, characters without evident regions, and characters comprising of three or more individual components.

In general, compound characters can be grouped into 12 different structures (see Table I) [33]. The previous example, “明” is *Left to Right* because it is structured in a way that “日” is on the left and “月” is on the right. In addition, Chinese characters can contain multiple of these structures. For example, “昆” is *Above to Below* with “日” on top and “比” on the bottom. Furthermore, the “比” component can be further decomposed *Left to Right* into “上” and “匕.”

There are 214 different standardized components, also known as radicals, defined by the Kangxi dictionary [34]. They range from primitive strokes (e.g. “一” and “丶”) to

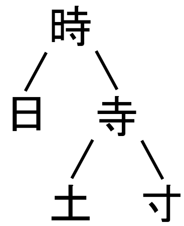

Fig. 2: Example of a component tree.

complex structures (e.g. “麻” and “龠”). One difficulty in component classification is the similarity of components. While distinct components, “口” (“mouth”) and “囗” (“enclosure”) are similar in appearance. Likewise “土” (“earth”) and “士” (“scholar”) as well as “自” (“self”) and “目” (“eye”) are different components. Another consideration required for identifying components is that they can consist of several separated strokes such as “氵,” “灬,” and “心.” In addition, there is an uneven representation of each component among the total number of characters. For instance, out of the 6,411 characters in the database described in Section V-A, only 3 contain the component “龠” and 1,785 contain the component “口.”

IV. Convolutional Neural Networks

CNNs [1] are a class of artificial neural networks which use a structured connection pattern with shared weights, namely convolutional kernels, which enforces a rigid sparse connectivity between the neurons maintaining the spatial features and structures of the data. Recently, they have been widely used in many recognition tasks due to their record breaking successes in image recognition [3], [4], [30], [31]. We implement a CNN to classify characters by the presence of a component.

V. Experiment and Results

A. Dataset

Chinese characters and components provide an ideal source of data for the experiment. Unlike natural images, components of Chinese characters are not always dependent on common surrounding components. Also, as previously stated, Chinese character components come in different morphologies, scales, and locations.

Tanaka-Ishii and Godon [35] created a database of 6,411 individual Chinese characters labeled by components.[1] In the database, every character is represented by a component tree structure with the root node as the entire character and each subsequent branch as a recursive decomposition of child components. An example of the component decomposition tree is shown in Fig. 2. Each character in the database was annotated with up to 10 different component labels.

To determine the component awareness of CNNs, we implemented a one-vs-others evaluation with the target component compared against examples of Chinese characters without the component. Each component was trained with a 90% - 10% split between training and test. However, in the database, there is a large discrepancy between the representation of a component and the representation of the other characters. In one-vs-others training, a heavily unbalanced training set with a low number of the target class would favor the “others.” To overcome this, characters in the training set that contain the target component were repeated until there was an equal distribution of the target component and the “others.” It should be noted that the test set is not augmented.

The dataset used for the CNN experiment contains images of the isolated Chinese characters rendered to 56px by 56px with 50pt black text on a white background.

[1] The database and proposed experiment are based in the Japanese version of Chinese characters.

B. CNN Architecture

For the experiments, a CNN similar to a LeNet [1] is used. The first three layers convolutional layers have 5×5 convolutional kernels at stride 1, each with proceeding 2×2 stride 2 maxpooling layers. They are made of 32, 64, and 128 nodes respectively. The following two fully-connected layers have 1024 nodes each and use dropout with a keep probability of 0.5. Rectified Linear Units (ReLU) is used for the activation for all hidden layers. The network is trained in batches of 25 using Adam optimizer [36] with an initial learning rate of 0.001 for 30,000 iterations.

C. Evaluation

In order to evaluate the effectiveness of the CNN, we calculated the accuracy, precision, and recall for the test set. Accuracy alone is not robust enough to determine the component awareness due to the previously mentioned unbalanced representation of the classes in the test set. Thus, precision is used to measure the relevancy of the characters identified to contain the component and recall is used to determine the rate that the target component was selected. For the experiment, precision is defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

and recall is

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{2}$$

Overall accuracy is defined as

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Patterns}}. \tag{3}$$

D. Results

The results of the evaluation of the model in Table II demonstrate that the proposed model is effective in detecting character components with near perfect accuracy. In general, for each component, the precision of the CNN was higher than the recall. This suggests that the CNN has a high accuracy in discriminating between components but suffers when detecting the presence of certain components.

Component Shape. Analysis of the results show that the CNN was able to accurately deduct the shape of the components despite only being trained on the knowledge that the

TABLE II: Chinese Character Component Detection Results

Component	Number of Characters	Accuracy (%)	Precision (%)	Recall (%)
亻	363	99.4	91.9	97.1
艹	437	99.4	100.0	89.5
⺍	316	100.0	100.0	100.0
扌	235	99.8	100.0	94.7
氵	369	100.0	100.0	100.0
二	251	96.7	83.3	51.7
亠	579	96.9	93.2	70.7
儿	328	99.4	93.5	93.5
冂	244	98.9	100.0	70.8
厂	374	98.0	91.9	77.3
口	1,785	95.8	99.3	85.4
小	556	98.0	96.6	83.8
山	198	98.9	100.0	53.3
月	404	98.8	97.1	82.9
木	813	99.8	98.9	100.0

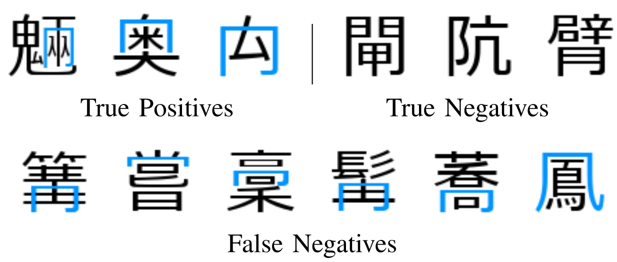

Fig. 3: Classification results from the test set for the component “冂.” The target component is highlighted in blue. There were no false positive results.

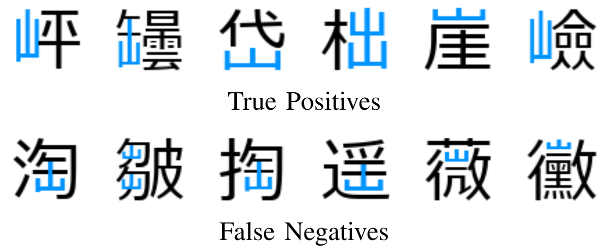

Fig. 4: Examples of characters with the component “山” classified by the CNN. The “山”s are highlighted in blue. When the component was large proportionally to the character, they were successfully classified. When the component was small or had large overlap with other components, they were falsely discriminated as negative.

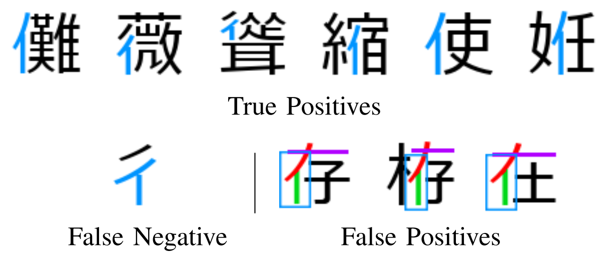

Fig. 5: The false negatives and false positives of “亻” detection as well as a sample of the true positives. The blue squares indicate regions which have similar structures, localization, and context as “亻”s, but are actually constructed of three individual components. “一,” “丨,” and “丿.”

component exists somewhere within the characters. For example, Fig. 3 shows the evaluated classification results for “冂.” The network was able to identify “冂”s at various locations and with appearances. Also, despite the true negatives in Fig. 3 having components with similar shapes to “冂,” the CNN was able to discriminate them from the target component. However, while the results achieved a perfect precision for “冂,” there were instances of failing to detect the component which resulted in only a 70.8% recall. The false negatives reveal that two characters use uncommon shape variations of “冂” and two characters have structures that are *Overlaid* with the component “土.”

Object size also has an effect on the accuracy of the results. Unlike many other object detection methods which use warped images from bounding boxes for classification, this paper focuses on component awareness in full-sized images. Thus, components that are too small in size relative to the training data can fail to be detected. Figure 4 shows that the low recall rate of 53.3% for the component “山” was due to misclassifying characters with small “山”s or instances of significant overlap with other components. Conversely, when the component was sufficiently large, the CNN was able to derive the shape of the component, even for the conjoined cases.

Component Context. Some Chinese character components have strong correlations to the surrounding components. Shown in Fig. 5, components, such as “亻,” are almost always accompanied by components on the right. The single false negative, “亻,” is the only instance in the database where the “亻” is isolated from left or right components. Figure 5 also reveals how similarly shaped components in similar contexts can lead to incorrectly detecting components. In general, components with context and localization relationships, similar to “亻” (e.g. “艹,” “⺍,” “氵,” etc.), had high precision and recall rates.

Another example where object detection depends on surrounding components is “二.” The difficulty with recalling “二” stems from the difficulty in discriminating between characters with horizontal lines from “二” and horizontal lines from other components (Fig. 6). In addition, “二” is a common sub-component to many other components such as “会,” “矢,” and “雨.” Consequently, the CNN learned to detect “二” contained those particular parent components with a high accuracy.

VI. Analysis with Component Removal

To further support the idea that a CNN can deduce the structure and location of weakly labeled components, the CNNs were tested on images of characters with the target component

Fig. 6: Examples of detection results of “二.”

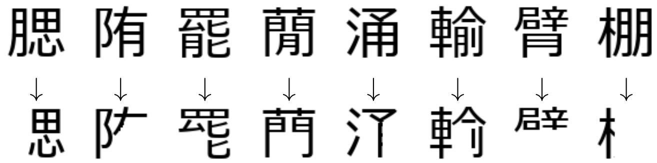

Fig. 7: Test patterns with the component "月" removed.

removed. If the positive classifications of characters with target components depend on the existence of those components, then by removing them, the expected outcome would be only negative classifications.

A. Data Preparation

To test the models described in Section V-B on Chinese characters with the target component removed, a new test set was created. The test set for the component removal experiment consists of only characters with the target component. The images were prepared manually by removing only the relevant component from the previously defined test set. Figure 7 demonstrates the component removal for "月."

B. Results

Table III confirms the assertion that removing the target component prevents most object detection by the CNN. This indicates that the CNN relies on the deduced structures of components for detection.

With exception to "二," the component removal process prevented all but a few characters from being identified as positive. As stated in Section V-D, "二" faces particular difficulties in that it is often integrated into other components as well as existing as a common simple structure in Chinese characters. Therefore, a higher ratio of positives reactions were detected when compared to the other components. Figure 8 reveals the positive detections despite the component being removed. The unexpected positive results in the figure are due to instances where the "二" was isolated or was a sub-component of uncommon components. Alternatively, with exception to the left two negative results, the results show that removing "二" changed the structure of "会," "矢," and "雨" parent components enough to prevent detection. These two factors affirm the conjecture from Section V-D that when detecting "二," the literal structure of "二" is not learned and rather the presence of parent components is learned. It is interesting to note that the characters enclosed in red boxes in Fig. 8 were previously identified as not containing a "二," but after its removal, the network now detects the presence of one.

VII. Conclusion

In this work, we demonstrated the ability that weakly supervised CNNs have on detecting component objects. Utilizing Chinese characters and their components, we showed that CNNs are able to develop awareness of objects without prior knowledge of their structures. Using weakly labeled

TABLE III: Chinese Character Component Detection Results with Component Removed

Component	Positives	Negatives	Positive → Negative Example
亻	0	35	喉 → [illegible]
艹	0	38	驩 → [illegible]
灬	0	28	鳳 → [illegible]
扌	0	19	控 → 空
氵	0	30	涛 → 寿
二	9	20	喉 → [illegible]
亠	5	53	陪 → [illegible]
儿	2	29	現 → [illegible]
冂	1	23	魎 → [illegible]
厂	0	44	茂 → [illegible]
口	2	176	綢 → [illegible]
小	4	64	原 → [illegible]
山	0	15	岱 → 代
月	1	40	罷 → [illegible]
木	1	87	襟 → [illegible]

components without specific annotation allows for the ease of training models for larger datasets.

Our proposed approach achieved a near perfect accuracy in one-vs-all Chinese character component detection. We showed that component detection is possible for a variety of components including simple components (e.g. "亻," "艹," "二," "亠," "冂," "厂," and "口"), complex components (e.g. "山," "月," and "木"), and disconnected components (e.g. "灬," "氵," "二," "儿," and "小"). The analysis of the results also reveal the inferred aspects of components that contribute to awareness, such as shape, size, and location.

To further verify that CNNs grounded detection on the components rather the surroundings, we conducted an experiment testing the Chinese characters with the components removed. The results showed that by removing the components, CNNs have difficulty detecting them.

Future research will be done in moving from single-label classification to multi-label classification. Chinese characters have one or more components, thus it is important to de-

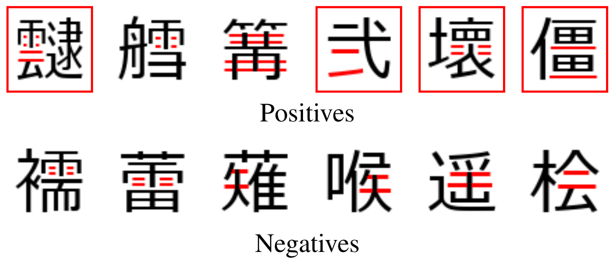

Fig. 8: Detection result examples with the component "二" removed. The red strokes indicate regions that were removed prior to testing. Due to the component removal, negative results are expected and positive results are erroneous. The enclosed red characters are characters that were identified to not contain "二" in the previous experiment.

velop an end-to-end solution to detect all of the components simultaneously. This research leads to future work in Chinese character recognition using component detection and awareness. By detecting the presence of components that make up characters, Chinese character recognition can be compartmentalized into searches for a smaller set of classes than full-character recognition which requires a very large number of classes.

Acknowledgments

This research was partially supported by the Institute of Decision Science for a Sustainable Society, Kyushu University.

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. on Comput. Vision*, 2015, pp. 1026–1034.

[3] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of neural networks using dropconnect," in *Proc. 30th Int. Conf. Mach. Learning*, 2013, pp. 1058–1066.

[4] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[5] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2015, pp. 5325–5334.

[6] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in Neural Inform. Process. Syst.*, 2013, pp. 2553–2561.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2014, pp. 580–587.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[9] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, 1998.

[10] R. Girshick, "Fast r-cnn," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1440–1448.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Inform. Process. Syst.*, 2015, pp. 91–99.

[12] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai, "Multi-oriented text detection with fully convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2016, pp. 4159–4167.

[13] T. He, W. Huang, Y. Qiao, and J. Yao, "Accurate text localization in natural image with cascaded convolutional text network," *arXiv preprint arXiv:1603.09423*, 2016.

[14] D. J. Crandall and D. P. Huttenlocher, "Weakly supervised learning of part-based spatial models for visual object recognition," in *European Conf. Comput. Vision*. Springer, 2006, pp. 16–29.

[15] O. Chum and A. Zisserman, "An exemplar model for learning object classes," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[16] R. Gokberk Cinbis, J. Verbeek, and C. Schmid, "Multi-fold mil training for weakly supervised object localization," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2014, pp. 2409–2416.

[17] H. Bilen, M. Pedersoli, and T. Tuytelaars, "Weakly supervised object detection with posterior regularization," in *British Mach. Vision Conf.*, vol. 3, 2014.

[18] H. O. Song, Y. J. Lee, S. Jegelka, and T. Darrell, "Weakly-supervised discovery of visual pattern configurations," in *Advances in Neural Inform. Process. Syst.*, 2014, pp. 1637–1645.

[19] C. Jiang and B. Zhang, "Weakly-supervised vehicle detection and classification by convolutional neural network," in *Int. Congr. Image and Sig. Proc., BioMedical Eng. and Informatics*. IEEE, 2016, pp. 570–575.

[20] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free?-weakly-supervised learning with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2015, pp. 685–694.

[21] L. Bazzani, A. Bergamo, D. Anguelov, and L. Torresani, "Self-taught object localization with deep networks," in *IEEE Winter Conf. Applicat. Comput. Vision*. IEEE, 2016, pp. 1–9.

[22] L. Rothacker and G. A. Fink, "Segmentation-free query-by-string word spotting with bag-of-features hmms," in *Int. Conf. Document Anal. and Recognition*. IEEE, 2015, pp. 661–665.

[23] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Word spotting and recognition with embedded attributes," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 36, no. 12, pp. 2552–2566, 2014.

[24] S. Sudholt and G. A. Fink, "Phocnet: A deep convolutional neural network for word spotting in handwritten documents," in *Int. Conf. Frontiers in Handwriting Recognition*. IEEE, 2016, pp. 277–282.

[25] L.-L. Ma and C.-L. Liu, "A new radical-based approach to online handwritten chinese character recognition," in *Int. Conf. Pattern Recognition*. IEEE, 2008, pp. 1–4.

[26] L.-L. Ma and C.-L. Liu, "On-line handwritten chinese character recognition based on nested segmentation of radicals," in *Chinese Conf. Pattern Recognition*. IEEE, 2009, pp. 1–5.

[27] E. Ni, M. Jiang, and C. Zhou, "Radical extraction for handwritten chinese character recognition by using radical cascade classifier," in *Elect., Inform. Eng. and Mechatronics 2011*. Springer, 2012, pp. 419–426.

[28] A.-B. Wang and K.-C. Fan, "Optical recognition of handwritten chinese characters by hierarchical radical matching method," *Pattern Recognition*, vol. 34, no. 1, pp. 15–35, 2001.

[29] X. Ren, Y. Zhou, J. He, K. Chen, X. Yang, and J. Sun, "A convolutional neural network-based chinese text detection algorithm via text structure modeling," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 506–518, 2017.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.

[31] B. Graham, "Fractional max-pooling," *arXiv preprint arXiv:1412.6071*, 2014.

[32] G. Sampson and Z. Chen, "The reality of compound ideographs," *J. Chinese Linguistics*, vol. 41, pp. 255–272, 2013.

[33] ArchChinese.com, "Arch chinese - chinese character structure," http://www.archchinese.com/arch_character_structure.html, accessed: 2016-12-12.

[34] E. P. Wilkinson, *Chinese History: A New Manual*, 4th ed., ser. Harvard-Yenching Institute Monograph Series. Cambridge, MA: Harvard University Asia Center, 2015.

[35] K. Tanaka-Ishii and J. Godon, "Kansuke: A logograph look-up interface based on a few modified stroke prototypes," *ACM Trans. Comput.-Human Interaction*, vol. 16, no. 2, p. 11, 2009.

[36] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learning Representations*, 2015.