Logo-2K+: A Large-Scale Logo Dataset for Scalable Logo Classification

Jing Wang[1], Weiqing Min[2], Sujuan Hou[1], Shengnan Ma[1], Yuanjie Zheng[1], Haishuai Wang[3], Shuqiang Jiang[2]
[1] School of Information Science and Engineering, Shandong Normal University
[2] Institute of Computing Technology, Chinese Academy of Sciences
[3] Department of Computer Science and Engineering, Fairfield University

Abstract

Logo classification has gained increasing attention for its various applications, such as copyright infringement detection, product recommendation and contextual advertising. Compared with other types of object images, the real-world logo images have larger variety in logo appearance and more complexity in their background. Therefore, recognizing the logo from images is challenging. To support efforts towards scalable logo classification task, we have curated a dataset, Logo-2K+, a new large-scale publicly available real-world logo dataset with 2,341 categories and 167,140 images. Compared with existing popular logo datasets, such as FlickrLogos-32 and LOGO-Net, Logo-2K+ has more comprehensive coverage of logo categories and larger quantity of logo images. Moreover, we propose a Discriminative Region Navigation and Augmentation Network (DRNA-Net), which is capable of discovering more informative logo regions and augmenting these image regions for logo classification. DRNA-Net consists of four sub-networks: the navigator sub-network first selected informative logo-relevant regions guided by the teacher sub-network, which can evaluate its confidence belonging to the ground-truth logo class. The data augmentation sub-network then augments the selected regions via both region cropping and region dropping. Finally, the scrutinizer sub-network fuses features from augmented regions and the whole image for logo classification. Comprehensive experiments on Logo-2K+ and other three existing benchmark datasets demonstrate the effectiveness of proposed method. Logo-2K+ and the proposed strong baseline DRNA-Net are expected to further the development of scalable logo image recognition, and the Logo-2K+ dataset can be found at https://github.com/msn199959/Logo-2k-plus-Dataset.

Introduction

Logo classification (Chen, Leung, and Gao 2003; Hoi et al. 2015; Bianco et al. 2017; Fehrvri and Appalaraju 2019) aims to recognize the logo name of an input image, and can be viewed as a special case of image recognition from a computer vision perspective. It has been a well-studied subject for decades for its various real-world commercial applications, such as brand retrieval (Joly and Buisson 2009), brand monitoring (Liu, Dzyabura, and Mizik 2018; Liao et al. 2017), trademark infringement detection, product recommendation and intelligent transportation.

Traditional methods for logo classification rely on hand-crafted features and keypoint-based detectors (S. Romberg 2013). Recently deep learning has been used for logo classification for its superior performance with end-to-end pipeline automation (Hoi et al. 2015; Bianco et al. 2017; Fehrvri and Appalaraju 2019). One standard approach is that one image is fed into one deep neural object detector (e.g., fast R-CNN (Girshick 2015)) and one classifier outputs its prediction, where the ImageNet trained CNN is fine-tuned on FlickrLogos-32 (Romberg et al. 2011) dataset. However, such methods suffer from the lack of a big quality logo dataset, leading to lower generalization ability of these models. Although some datasets like WebLogo-2M (Su, Gong, and Zhu 2017) and PL2K (Fehrvri and Appalaraju 2019) are large, WebLogo-2M is very noisy for its unsupervised annotation while the recent proposed PL2K is not publicly available.

To address these problems, we introduce a new logo dataset, Logo-2K+ for logo classification. Compared with existing public available datasets, such as FlickrLogos-32, Logo-2K+ has three distinctive characteristics: (1) Large-scale. It consists of 167,140 images with a total number of 2,341 categories. (2) High-coverage. To our knowledge, Logo-2K+ consists of 2,341 categories and has the highest coverage of the logo categorical space. (3) High-diversity. The logo images are collected from different websites. There are different types of logo images with different logo appearance, scale and background. Figure. 1 shows some samples.

Furthermore, we propose a Discriminative Region Navigation and Augmentation Network (DRNA-Net) to first localize informative logo regions, and then explores the potential of data augmentation for region-oriented data augmentation. The proposed DRNA-Net consists of four sub-networks: the navigator sub-network navigates the model to focus on informative regions. For each region, the teacher sub-network evaluates its probability belonging to ground-truth class to select the most informative ones. Region-oriented data augmentation subnetwork is mainly used to augment these selected regions via both region cropping and region dropping. The scrutinizer sub-network extracts and

Figure 1: Some logo images from Logo-2K+ dataset.

fuses features from augmented regions and the whole image to make logo classification. The four sub-networks benefit each other and can be trained in an end-to-end way.

In summary, this paper has three main contributions. First, we introduce a large-scale publicly available logo dataset Logo-2K+ with 2,341 logo classes and 167,140 images. To our knowledge, this is the largest publicly available high-quantity dataset for logo classification. Second, we propose a novel Discriminative Region Navigation and Augmentation Network (DRNA-Net) for scalable logo classification. It can automatically locate logo-relevant informative regions and conduct region-oriented data augmentation to extract more discriminative features from these augmented regions. Third, we conduct extensive evaluation on four datasets, including newly proposed Logo-2K+, and other three datasets with different scales, namely BelgaLogos (Neumann, Samet, and Soffer 2002), FlickrLogos-32 (Romberg et al. 2011) and WebLogo-2M (Su, Gong, and Zhu 2017). The experimental results verified the effectiveness of the proposed method on all these datasets.

Related Work

In this section, we review related efforts towards (1) Logo-centric datasets, and (2) Logo classification.

Logo-centric Datasets

Several existing publicly available logo datasets have been used for classification, such as FlickrLogos-32 (Romberg et al. 2011), BelgaLogos (Neumann, Samet, and Soffer 2002) and Logos in the wild (Andras Tzk and Beyerer 2018). These datasets do not include a wide range of logo images and are lack of diversity and coverage in logo categories. Therefore, they are not sufficient to support complex statistical models, such as deep learning models for scalable logo classification. Although some larger datasets are proposed, such as LOGO-Net (Hoi et al. 2015), WebLogo-2M (Su, Gong, and Zhu 2017) and PL2K (Fehrvri and Appalaraju 2019). Unfortunately, most of these datasets are either very noisy or not publicly available. In contrast, our proposed Logo-2K+ is a large-scale high-coverage, high-quantity dataset with over two thousand logo categories and about 170 thousand logo images, where there are at least 50 images for one category. Table 1 summarizes the statistics of main logo datasets.

Logo Classification

Logo classification has a long history in computer vision. Traditional logo classification methods resort to hand-crafted features, such as SIFT features and keypoint-based methods (Joly and Buisson 2009; S. Romberg 2013; Li et al. 2014; Su, Gong, and Zhu 2017). To improve the robustness of features, some methods adopt a symbolic representation of features (Su, Zhu, and Gong 2018) using global features for classification. Recently, deep learning methods (Hoi et al. 2015; Iandola et al. 2015; Su, Zhu, and Gong 2017) have been proposed for logo classification. For example, Bianco *et al.* (Bianco et al. 2017) propose a logo classification pipeline, which is composed of a logo re-

Table 1: Comparison between Logo-2K+ and existing logo datasets.

Dataset	Logos	Images	Availability
FlickLogos-27 (Romberg et al. 2011)	27	1,080	√
FlickLogos-32 (Romberg et al. 2011)	32	8,240	√
BelgaLogos (Neumann, Samet, and Soffer 2002)	37	10,000	√
FlickLogos-47 (Romberg et al. 2011)	47	8,240	√
LOGO-Net (Hoi et al. 2015)	160	73,414	×
TopLogo-10 (Su, Zhu, and Gong 2017)	10	700	√
Logo-405 (Hou et al. 2017)	405	32,218	×
Logos in the wild (Andras Tzk and Beyerer 2018)	871	11,054	√
WebLogo-2M (Su, Gong, and Zhu 2017)	194	1,861,177	√
PL2K (Fehrvri and Appalaraju 2019)	2,000	295,814	×
Logo-2K+(Ours)	**2,341**	**167,140**	√

gion proposal followed by a Convolutional Neural Network (CNN) specifically trained for logo classification. In addition, there are also some works for logo detection (Su, Gong, and Zhu 2017; Oliveira et al. 2016; Li et al. 2018; Oliveira et al. 2016), where the state-of-the-art deep neural detector models, such as Faster R-CNN are utilized.

Our work also focuses on logo classification. Different from previous work, our work mainly focuses on learning discriminative logo-relevant regions in a self-supervised manner for logo classification. In addition, inspired by the spatial transformation for data augmentation (Hu et al. 2019), we introduce a region-oriented data augmentation strategy to further augment informative regions for more discriminative feature learning.

Dataset: Logo-2K+

Table 2: Data statistics on Logo-2K+.

Root Category	Logos	Images
Food	769	54,507
Clothes	286	20,413
Institution	238	17,103
Accessories	210	14,569
Transportation	203	14,719
Electronic	191	13,972
Necessities	182	13,205
Cosmetic	115	7,929
Leisure	99	7,338
Medical	48	3,385
Total	**2,341**	**167,140**

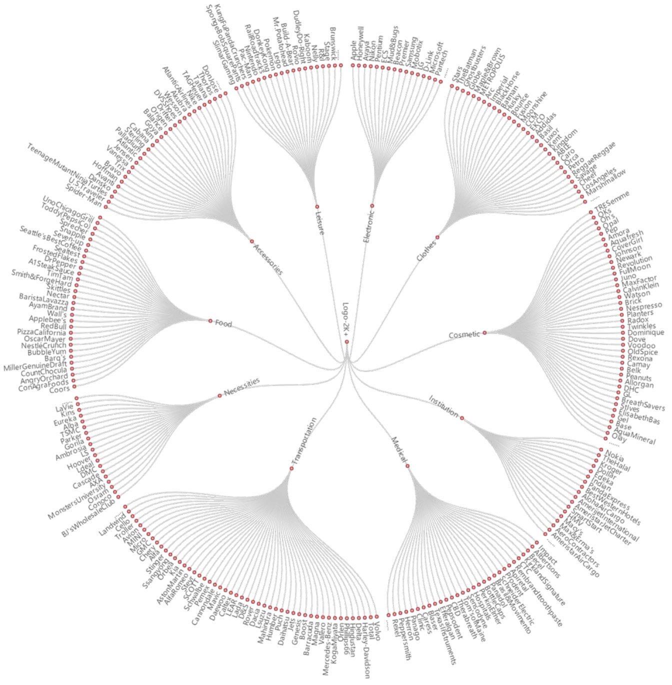

Figure 2: The boxable logo class hierarchy. Parent nodes represent more generic ten categories than their children (Best View under Magnification).

As noted above, to date, there are no publicly available large-scale high-quantity logo datasets. In order to perform scalable logo classification, we construct a large-scale logo dataset, Logo-2K+, which covers a diverse range of logo classes from real-world logo images. The construction of Logo-2K+ is composed of two steps, from constructing the logo category list to collecting and cleaning logo images.

Constructing the Logo Category List. The first asset of a high-quality dataset should be a high-coverage of the categorical space. We construct a logo list based on frequent appeared logo scenes and objects from the following 10 categories, namely **Food, Clothes, Institution, Accessories, Transportation, Electronic, Necessities, Cosmetic, Leisure, Medical.** For each root category, we further build a list of logo classes based on some logo websites, leading to the resulting 2,341 logo classes. Logo-2K+ has imbalanced class amounts in different categories. For example, there are 769 logo classes in the "Food" category while only 48 classes in the "Medical" category. Figure. 2 shows the results of logo hierarchy visualization.

Collecting and Cleaning Logo Images. We crawl candidate images from the Google image search website. In order to obtain more relevant logo images, we expand search terms by adding keywords, such as "logo" and "brand". For example, "Apple" means not only a famous brand of electronic products, but also a category of fruits. Therefore, we choose "Apple logo" and "Apple brand" as the search term to crawl logo images. We then manually check the quality of each class by removing duplicated images, images with terrible aspect ratio, too small logo ratio in the whole image. For logo class with less 50 images, we further enlarge the amount of logo images from the other websites, such as Baidu Images.

Data Statistics. Our resulting logo dataset contains 167,140 images with 10 root categories and 2,341 categories. There are at least 50 images for each logo class. The statistical comparison of 10 root categories from Logo-2K+ is shown in Table 2. Examples are shown in Figure 1. Figure 3 shows the distribution of images across logo categories in Logo-2K+. We can see that the distribution of images for different logo categories is not uniform. Examples of logo categories with more images are Apple, Asus, Corona, to name a few. Examples with few images are Brub, Nike, KHS, to name a few.

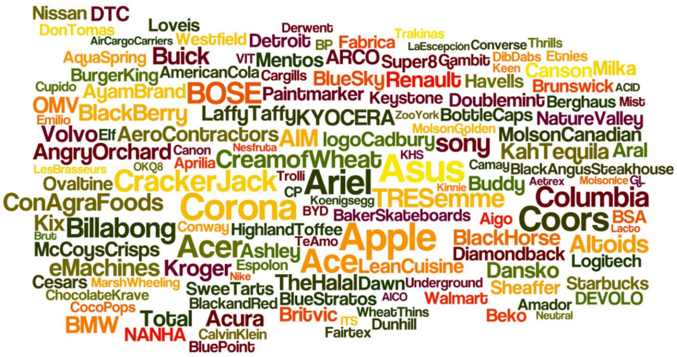

Figure 3: Visualization of logo classes based on the number of images in each class. Larger font size indicates more images in the corresponding class. Color is used randomly for visualization purpose only.

Approach

Overview

In this section, we will introduce the proposed Discriminative Region Navigation and Augmentation Network (DRNA-Net) for logo classification. Figure 4 illustrates the architecture of DRNA-Net, which is mainly divided into four parts, namely navigator sub-network, teacher sub-network, region-oriented data augmentation sub-network, and scrutinizer sub-network. Given an input image, the navigator sub-network first computes the informativeness of all regions generated from the pre-defined anchors. Then, the teacher sub-network evaluates each region's confidence that it belongs to the ground-truth class to enable the navigator sub-network to select the most informative logo-relevant regions. Next, the region-oriented data augmentation sub-network can augment the selected regions to localize more relevant logo regions. Finally, the scrutinizer sub-network fuses features from augmented regions and the whole image into the final feature representation for logo prediction.

Navigator Sub-network

Given an input image X, the navigator sub-network adopts convolutional layers, ReLU activation and max-pooling and a top-down feature pyramid network (Lin et al. 2017) with lateral connections to generate regions A with different scales and ratios. Each region $R \in A$. The non-maximum suppression on regions is used to reduce regional redundancy based on the informativeness, leading to top-M informative regions as $R = \{R_1, R_2, \ldots, R_M\}$ with corresponding informativeness $I = \{I_1, I_2, \ldots, I_M\}$. These regions are then fed into the teacher sub-network to generate the most informative regions.

Teacher Sub-network

The teacher sub-network can calculate the confidence $\{C(R_1),\ C(R_2), \ldots, C(R_M)\}$ based on detected regions $R = \{R_1, R_2, \ldots, R_M\}$ from the navigator sub-network. The regions with a higher probability belonging to the ground-truth class should have higher confidence, that is

$$\mathrm{if\ C(R_1) > C(R_2),\ then\ I(R_1) > I(R_2); R_1,\ R_2 \in A} \tag{1}$$

We optimize the navigator sub-network and teacher sub-network to make $\{C(R_1),\ C(R_2), \ldots, C(R_M)\}$ and $\{I(R_1),\ I(R_2), \ldots, I(R_M)\}$ keep the same order using Eq. 1 via the following pair-wise ranking loss between predicted confidence and the ground-truth class:

$$Loss_I(I, C) = \textstyle\sum_{(i,j):C_i<C_j} f(I_j - I_i) \tag{2}$$

where i, j are the region index and $f(x) = \max\{1 - \mathrm{x}, 0\}$. $Loss_I(I, C)$ encourages if $C_i < C_j$, then $I_i < I_j$. The loss function enforces consistency between informativeness of region and probability being ground-truth class.

Then, the teacher sub-network $Loss_c$ is defined to minimize the difference between losses of all regions $\log C(R_i)$ and the full image $\log C(X)$ as follows:

$$Loss_c = -\textstyle\sum_{i=1}^{M} \log C(R_i) - \log C(X) \tag{3}$$

Region-oriented Data Augmentation Sub-network

We obtain the most informative logo-relevant regions $\{C(R_1),\ C(R_2), \ldots, C(R_K)\}$, where K is the number of selected regions through both the navigator sub-network and teacher sub-network. Region-oriented data augmentation sub-network is then introduced to augment these selected regions to obtain more relevant regions via both region cropping and region dropping (Hu et al. 2019). We first normalize R_K to generate augmentation maps R_k^* as follows:

$$R_k^* = \frac{R_k - min\left(R_k\right)}{max\left(R_k\right) - min\left(R_k\right)} \tag{4}$$

(1) Region cropping: with augmentation map R_k^*, we can zoom into the region and extract more informative local feature. We set a threshold $\theta_c \in [0, 1]$ to determine whether the element $R_k^*\left(i, j\right)$ belongs to 0 or 1, the formula as follows,

$$C_k\left(i, j\right) = \begin{cases} 1, & if\ R_k^*\left(i, j\right) > \theta_c \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where C_k is the 0-1 crop mask, and a bounding box on this region can be found to cover the selected positive location of C_k.

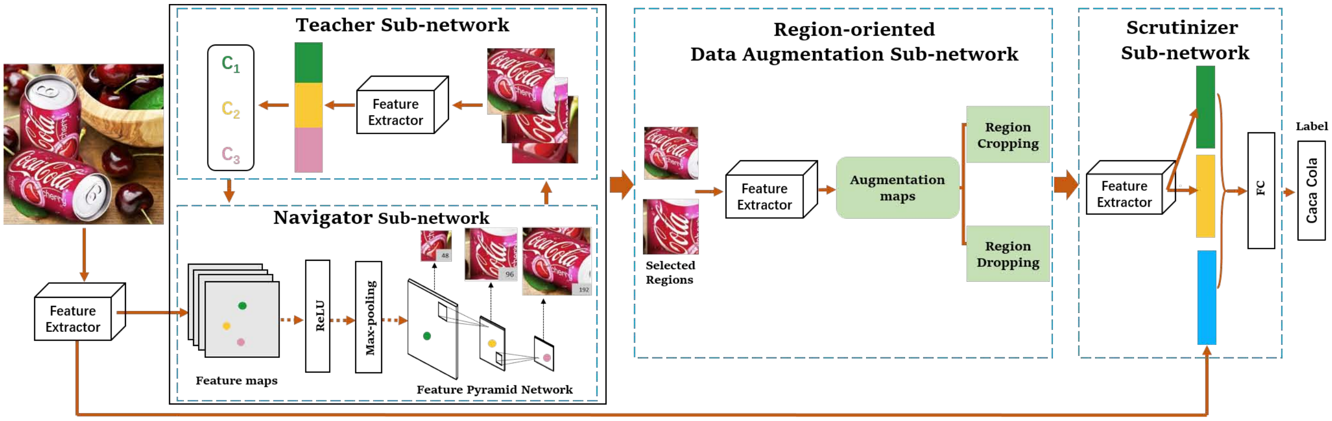

Figure 4: An overview of proposed Discriminative Region Navigation and Augmentation Network (DRNA-Net). For the input image, these feature maps from the feature extractor are fed into the navigator sub-network to compute the informativeness of all regions. Then, these regions from the full image are also fed into the teacher sub-network to get the confidence. We optimize the navigator sub-network to make informativeness and confidence consistent. Next, we put selected regions with high informativeness into region-oriented data augmentation sub-network to augment these regions via both region cropping and region dropping. Finally, the scrutinizer network fuses features of those augmented regions and the full image to predict the label. FC denotes the full-connected layer.

(2) Region dropping: region dropping is used to encourage region maps to represent multiple discriminative logo' parts, and is formulated as follows:

$$D_k\left(i,j\right) = \begin{cases} 1, & if\ R_k^*\left(i,j\right) > \theta_d \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

The k_{th} selected region will be dropped by masking region with D_k.

We denote E as the augmentation function and adopt cross entropy $Loss_a$ as follows,

$$Loss_a = -\sum_{i=1}^{K} \log E(R_k^*) - \log E(X) \qquad (7)$$

where E maps the augmentation selected region to its probability being ground-truth class label.

Scrutinizer Sub-network

The scrutinizer sub-network obtains K augmented informative regions, and uses a feature extractor to extract and fuse the features of the full image X and regions $\{R_1^*,\ R_2^*,\ ...,\ R_K^*\}$. We concatenate K augmentation features vector and input image X feature vector, and feed them into the full convolution layer. The scrutinizer sub-network finally gives the prediction results $P = F(X,\ R_1^*,\ R_2^*,\ ...,\ R_K^*)$, where F is a transformation, we define the classification cross entropy loss $Loss_s$ is used.

Joint Training Loss

Finally, we use stochastic gradient descent method to optimize the total loss as follows:

$$L_{total} = Loss_I + \alpha \cdot Loss_s + \beta \cdot Loss_c + \gamma \cdot Loss_a \quad (8)$$

where α, β and γ are hyper parameters.

DRNA-Net is inspired by work (Yang et al. 2018), but introduced region-oriented data augmentation sub-network for selected regions to further enhance final feature representation. Therefore, DRNA-Net achieves two-level coarser-to-finer localization to efficiently discover more discriminative regions better for logo classification.

Experiments

Experimental Setup

For our method, we adopt ResNet-50 and ResNet-152 pretrained on ILSVRC2012 as the feature extractor. The thresholds of region cropping and dropping θ_c and θ_d are both set to 0.5. We empirically set $M = 4$ in the navigator subnetwork and $K = 2$ in the teacher sub-network. In Eq. 8, hyper-parameters weights $\alpha = \beta = \gamma = 1$ without prior. It is optimized using the stochastic gradient descent with a momentum of 0.9, a batch size of 8 and weight decay of 0.0001. We adopt the Pytorch framework to train the network and implement our algorithm on a NVIDIA Tesla V100 GPU (32GB). Both Top-1 accuracy and Top-5 accuracy are adopted as evaluation metrics.

Experiments on Logo-2K+

Implementation Detials. 70%, 30% of images are randomly selected for training and testing in each logo category. All the models are trained for 100 epochs with an initial learning rate of 0.001 and decreased after 20 epochs to 0.0001.

For baselines, we evaluate different networks on Logo-2K+, such as AlexNet (Krizhevsky, Sutskever, and E. Hinton 2012), GoogLeNet (Szegedy et al. 2015), VGGNet (Simonyan and Zisserman 2014) and ResNet (He et al. 2016) for logo classification. We also list some efficient training and optimization method named Label Smoothing Regularization (LS) (He et al. 2018).

Table 3: Comparison of our model and baselines on Logo-2K+ (%).

Method	Top-1 Acc.	Top-5 Acc.
AlexNet	48.80	78.45
GoogLeNet	62.36	88.33
VGGNet-16	62.83	89.01
ResNet-50	66.34	91.01
ResNet-152	67.65	91.52
VGGNet-16+Efficient+LS (He et al. 2018)	65.45	90.12
ResNet-50+Efficient+LS (He et al. 2018)	66.94	91.30
ResNet-152+Efficient+LS (He et al. 2018)	67.99	91.68
NTS-Net(ResNet-50) (Yang et al. 2018)	69.41	91.95
DRNA-Net(ResNet-50)	71.12	92.33
DRNA-Net(ResNet-152)	**72.09**	**93.45**

The evaluation results on Logo-2K+ dataset are shown in Table 3. We can see that (1) The best single network is ResNet-152, with a Top-1 accuracy of 67.65% and a Top-5 accuracy of 91.52%, and achieves a 67.99% Top-1 accuracy using some tricks. (2) Our method produces the best 72.09% in Top-1 accuracy and 93.45% in Top-5 accuracy, surpassing the NTS-Net about 1%. This indicts that the effectiveness of region-oriented data augmentation strategy.

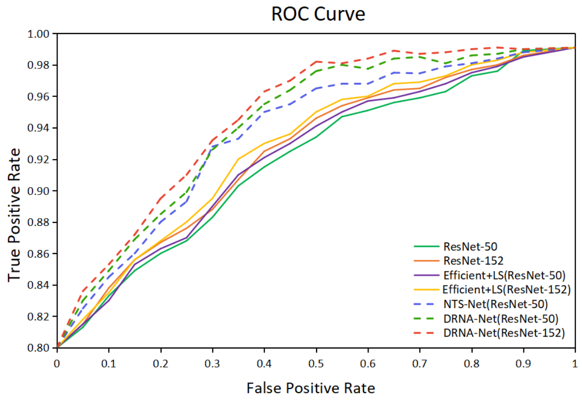

Figure 5: ROC curves of different logo classification models on Logo-2K+.

Considering image samples from different logo classes are unbalanced, to further comprehensively evaluate the performance of DRNA-Net, we further draw the ROC curves of all the methods in Figure 5, where the dotted green and red curve represent the performance of our method and the purple curve refers to the NTS-Net method. We can see that the true positive rate remains high on the DRNA-Net compared to NTS-Net for test logo classes, and DRNA-Net obtains the best performance in terms of overall quality for logo classification.

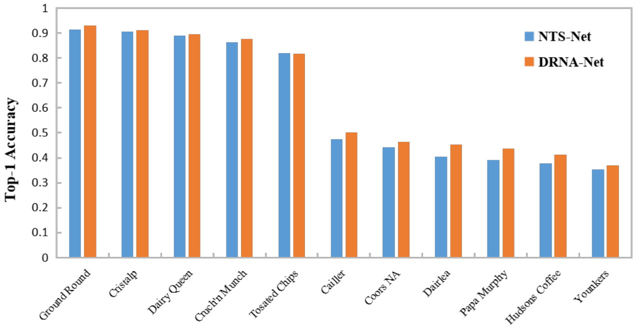

Figure 6: Selected classification Top-1 Accuracy: The 5 best and 5 worst performing classes are shown.

In addition, we select 10 classes in the test phase to further evaluate our method and NTS-Net on the ResNet-50 base CNN. Particularly, we listed the Top-1 Accuracy of both 5 best and 5 worst performing classes in Figure 6. As shown in Figure 6, we find that for 9 classes among 10 logo classes, there is the performance improvement for our method and the performance from 1 additional classes is degraded compared to NTS-Net. The improvements for single classes are visibly more pronounced than degradations, indicating the effectiveness of introduced region-oriented data augmentation sub-network. In addition, for the categories with bad classification results, we observe that most of these logo classes contain fewer samples, or the logo images contain the smaller logo region or complex background.

Visualization of Localized Regions. To analyze the localization effect of our method, we visualize the regions from some logo samples by navigator sub-network and region-oriented data augmentation sub-network in Figure 7. First, we adopt the navigator sub-network to coarser localization regions guided by teacher sub-network. Then, the regions are feed into the region oriented data augmentation sub-network to achieve the finer localization. We can see that after two-level coarser-to-finer localization, our method can find more relevant logo regions to support logo classification.

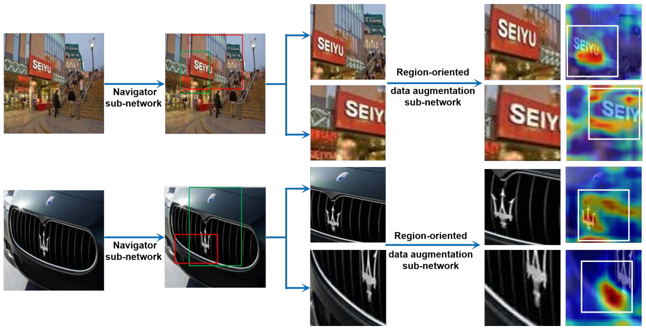
Navigator sub-network
Region-oriented data augmentation sub-network
Navigator sub-network
Region-oriented data augmentation sub-network

Figure 7: Examples of the visualization of discovered regions by DRNA-Net.

Table 4: Performance comparison of methods on BelgaLoges(%).

Method	Top-1 Acc.	Top-5 Acc.
RCNN(CaffeNet) (Girshick et al. 2014)	91.80	-
FRCN(VGGNet-16) (Girshick 2015)	87.30	-
SPPnet(ZF) (K. He and Sun. 2014)	87.70	-
NTS-Net(ResNet-50) (Yang et al. 2018)	93.33	96.15
DRNA-Net(VGGNet-16)	92.41	95.96
DRNA-Net(ResNet-50)	94.44	97.11
DRNA-Net(ResNet-152)	**95.82**	**98.40**

Experiments on Other Datasets

Besides Logo-2K+, we also conduct the evaluation on another publicly available benchmark datasets, BelgaLoges, FlickrLogo-32 and WebLogo-2M to further verify the effectiveness of our method. **BelgaLoges** contains 37 logo class and a total of 10,000 images. It is a small-scale logo image dataset for product brand recognition. **FlickrLogos-32** contains 32 different logo brands, 8,240 logo images. **WebLogo-2M** contains 1,867,177 logo images in 194 different logo classes. This dataset is annotated via an unsupervised process, and thus is very noisy. For all three datasets, 70%, 30% of images are randomly selected for training and testing for each logo category.

Experiments on BelgaLoges We list experimental results of baselines and our proposed method in Table 4. These models are trained for 80 epochs with an initial learning rate of 0.001 and divided by 10 after 20 epochs. As we can see from Table 4, our method achieves the best performance: the 94.44% Top-1 accuracy with ResNet-50 and the 95.82% Top-1 accuracy with ResNet-152. In addition, we also compare the FRCN (VGGNet-16) with our method with VGGNet-16, and find that there is about 5% improvement in Top-1 accuracy. These experimental results can further demonstrate the effectiveness of our proposed method.

Experiments on FlickrLogos-32 All models are trained for 80 epochs with an initial learning rate of 0.001 and divided by 10 after 20 epochs, and the batch size is 32. The classification accuracy of different methods is summarized in Table 5. We can see that our method achieves the best performance in both Top-1 accuracy and Top-5 accuracy. Compared with NTS-Net, the proposed strategy obtains 1.2% improvement, which demonstrates that introducing region-oriented data augmentation method can help improve the performance.

Experiments on WebLogo-2M In order to better prove the effectiveness of our method, we also carry out experiments on the WebLogo-2M dataset with large-scale but noisy images. The some experimental setup are the same as the settings of Logo-2K+. The experimental results of baselines and our method is summarized in Table 6. We can see that our method achieves the best performance with ResNet-50 compared with other baselines, such as NTS-Net. We can also see that although there are only 194 logo classes WebLogo-2M with many images for each logo class, the performance on WebLogo-2M is lower. The probable reason is that there are many noisy images in the WebLogo-2M.

Table 5: Comparison of our model and state-of-the-art methods on FlickrLogos-32 (%).

Method	Top-1 Acc.	Top-5 Acc.
FRCN + AlexNet (Iandola et al. 2015)	75.00	-
(C. Eggert 2015)	84.60	-
(Bianco et al. 2015)	88.40	-
(Bianco et al. 2017)	91.70	-
SIFT (S. Romberg 2013)	94.10	-
NTS-Net(ResNet-50) (Yang et al. 2018)	94.14	96.29
DRNA-Net(ResNet-50)	95.33	97.17
DRNA-Net(ResNet-152)	**96.63**	**98.80**

Table 6: Comparison of our model and baselines on WebLogo-2M (%).

Method	Top-1 Acc.	Top-5 Acc.
AlexNet	50.99	70.62
GoogLeNet	62.14	80.21
VGGNet-16	62.88	83.23
ResNet-50	62.93	83.32
NTS-Net(ResNet-50) (Yang et al. 2018)	63.67	84.31
DRNA-Net(ResNet-50)	**64.82**	**86.12**

Conclusion

In this paper, we have introduced a novel large-scale logo benchmark dataset Logo-2K+ with over 2000 logo classes and about 170,000 logo images. This dataset should further the state of the art in scalable logo image recognition. As one strong baseline, we have also presented a method to discover and augment discriminative regions to enhance final feature representation, and have shown its effectiveness on Logo-2K+ and other three existing logo benchmarks. Future work includes conducting the bounding-box annotation of logos in logo images or synthesizing logo-annotated images without expensive manual labeling (Su, Zhu, and Gong 2017) to support large-scale logo detection task based on Logo-2K+.

References

[Andras Tzk and Beyerer 2018] Andras Tzk, Christian Herrmann, D. M., and Beyerer, J. 2018. Open set logo detection and retrieval. In *IEEE Winter Conference on Applications of Computer Vision*, 284–292.

[Bianco et al. 2015] Bianco, S.; Buzzelli, M.; Mazzini, D.; and Schettini, R. 2015. Logo recognition using cnn features. In *In International Conference on Image Analysis and Processing*, 438–448.

[Bianco et al. 2017] Bianco, S.; Buzzelli, M.; Mazzini, D.; and Schettini, R. 2017. Deep learning for logo recognition. *Neurocomputing* 245:23–30.

[C. Eggert 2015] C. Eggert, A. Winschel, R. L. 2015. On the benefit of synthetic data for company logo detection. In *ACM Conference on Multimedia Conference*, 1283–1286.

[Chen, Leung, and Gao 2003] Chen, J.; Leung, M. K.; and Gao, Y. 2003. Noisy logo recognition using line segment hausdorff distance. *Pattern Recognition* 36(4):943 – 955.

[Fehrvri and Appalaraju 2019] Fehrvri, I., and Appalaraju, S. 2019. Scalable logo recognition using proxies. In *IEEE Winter Conference on Applications of Computer Vision*, 715–725.

[Girshick et al. 2014] Girshick, R.; Donahue, J.; Darrell, T.; and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

[Girshick 2015] Girshick, R. 2015. Fast R-CNN. In *IEEE International Conference on Computer Vision*, 1440–1448.

[He et al. 2016] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

[He et al. 2018] He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; and Li, M. 2018. Bag of tricks for image classification with convolutional neural networks. *CoRR* abs/1812.01187.

[Hoi et al. 2015] Hoi, S. C. H.; Wu, X.; Liu, H.; Yue, W.; Wang, H.; Hui, X.; and Qiang, W. 2015. LOGO-Net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks. *CoRR* abs/1511.02462.

[Hou et al. 2017] Hou, S.; Lin, J.; Zhou, S.; Qin, M.; Jia, W.; and Zheng, Y. 2017. Deep hierarchical representation from classifying logo-405. *Complexity* 2017(99):1–12.

[Hu et al. 2019] Hu, T.; Qi, H.; Huang, Q.; and Lu, Y. 2019. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification. *arXiv:1901.09891*.

[Iandola et al. 2015] Iandola, F.; Shen, A.; Gao, P.; and Keutzer, K. 2015. Deeplogo: Hitting logo recognition with the deep neural network hammer. *arXiv:1510.02131*.

[Joly and Buisson 2009] Joly, A., and Buisson, O. 2009. Logo retrieval with a contrario visual query expansion. In *ACM international conference on Multimedia*, 581–584.

[K. He and Sun. 2014] K. He, X. Zhang, S. R., and Sun., J. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision*, 346–361.

[Krizhevsky, Sutskever, and E. Hinton 2012] Krizhevsky, A.; Sutskever, I.; and E. Hinton, G. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 1097–1105.

[Li et al. 2014] Li, K. W.; Chen, S. Y.; Su, S.; and Duh, D. J. 2014. Logo detection with extendibility and discrimination. *Multimedia Tools and Applications* 72(2):1285–1310.

[Li et al. 2018] Li, Y.; Shi, Q.; Deng, J.; and Fei, S. 2018. Graphic logo detection with deep region-based convolutional networks. In *Visual Communications and Image Processing*, 10–13.

[Liao et al. 2017] Liao, L.; He, X.; Ren, Z.; Nie, L.; Xu, H.; and Chua, T.-S. 2017. Representativeness-aware aspect analysis for brand monitoring in social media. In *The Twenty-Sixth International Joint Conference on Artificial Intelligence*, 310–316.

[Lin et al. 2017] Lin, T.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Computer Vision and Pattern Recognition (CVPR)*, 936–944.

[Liu, Dzyabura, and Mizik 2018] Liu, L.; Dzyabura, D.; and Mizik, N. 2018. Visual listening in:extracting brand image portrayed on social media. In *The Thirty-Second AAAI Conference on Artificial Intelligence*, 71–77.

[Neumann, Samet, and Soffer 2002] Neumann, J.; Samet, H.; and Soffer, A. 2002. Integration of local and global shape analysis for logo classification. *Pattern Recognition Letters* 23(12):1449 – 1457.

[Oliveira et al. 2016] Oliveira, G.; Frazao, X.; Pimentel, A.; and Ribeiro, B. 2016. Automatic graphic logo detection via fast region-based convolutional networks. In *International Joint Conference on Neural Networks*, 985–991.

[Romberg et al. 2011] Romberg, S.; Pueyo, L. G.; Lienhart, R.; and Zwol, R. V. 2011. Scalable logo recognition in real-world images. In *ACM International Conference on Multimedia Retrieval*, 18–20.

[S. Romberg 2013] S. Romberg, R. L. 2013. Bundle minhashing for logo recognition. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, 113–120.

[Simonyan and Zisserman 2014] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[Su, Gong, and Zhu 2017] Su, H.; Gong, S.; and Zhu, X. 2017. Weblogo-2m: Scalable logo detection by deep learning from the web. In *IEEE International Conference on Computer Vision Workshop*, 270–279.

[Su, Zhu, and Gong 2017] Su, H.; Zhu, X.; and Gong, S. 2017. Deep learning logo detection with data expansion by synthesising context. In *IEEE Winter Conference on Applications of Computer Vision*, 530–539.

[Su, Zhu, and Gong 2018] Su, H.; Zhu, X.; and Gong, S. 2018. Open logo detection challenge. In *In Proceedings of the British Machine Vision Conference*, 111–119.

[Szegedy et al. 2015] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions. *CoRR, abs/1409.4842* 7–12.

[Yang et al. 2018] Yang, Z.; Luo, T.; Wang, D.; Hu, Z.; Gao, J.; and Wang, L. 2018. Learning to navigate for fine-grained classification. In *European Conference on Computer Vision*, 438–454.