Data Drift Detection with KS Test using Attention Map

Tsunemi Nitta, Yuzhi Shi, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi

Chubu University, 1200 Matsumoto, Kasugai, Aichi, Japan
{ntsunemi0122, shi, hirakawa}@mprg.cs.chubu.ac.jp,
{takayoshi, fujiyoshi}@isc.chubu.ac.jp

Abstract. Data drift is a change in the feature distribution of input data during machine learning model training and during system operation. The data drift occurs regardless of the type of data and adversely affects model performance. The existing methods detects the data drift by using two-sample test for network output. However, these methods merely apply two-sample test with the distribution of class probabilities. Even though drifted input images are transformed by noise and/or geometric transformations, these methods does not consider such transformations. In addition to class probability, we believe that detecting drift for changes in the local region that the model is actually gazing at will improve accuracy. In this study, we propose a drift detection method based on attention branch network (ABN), which enables visualization of the basis of judgment in image classification. In our method, drift is detected using the class probabilities output by the attention branch and perception branch, which constitute the ABN, and the attention map. The results show that we can improve the detection ratio by introducing an attention map to drift detection in addition to class probability. We also observed that the attention map tended to shrink with drift.

Keywords: Drift detection · KS test · Attention branch network · Attention map · Maximum mean discrepancy

1 Introduction

Data drift is a change over time in the data distribution during model operation from the data distribution during model training. Drift occurs under various conditions, such as noise and out-of-focus due to the aging degradation of the camera lens, and misalignment and reversal caused when replacing the camera. Once such drift occurs, the data distribution becomes different one from that of the training data, which causes the performance degradation of the machine learning model [5]. These factors make it particularly important to monitor changes in data distribution and detect drift during visual inspection and surveillance camera operations.

To detect the data drift, we compare the data distribution used for model training with the data distribution obtained during operation. These distributions are compared by a two-sample test [8, 9]. Rabanser et al. [9] proposed a method for detecting data drift. They focused on the feature distribution of input data. Both methods detect drift by performing a two-sample test on the class probability distribution output by the trained model. However, these methods merely apply two-sample test with the distribution of class probabilities. Even though drifted input images are transformed by noise and/or geometric transformations, these methods does not consider such transformations.

In this study, we propose a novel drift detection method. Our method uses the attention branch network (ABN) [2] in addition to two-sample test. ABN outputs a couple of classification scores from attention and perception branches. Also, we can obtain an attention map during the inference process, that highlights on which the network model focus to make a decision. We leverage this property, that is, the proposed method detects drift by performing a two-sample test on the class probability distribution output of attention and perception branches and on the attention map that highlights a region contributed to model inference. Considering the several two-sample test results, we detect data drift, which enables us to accurately detect drift. Moreover, our ABN-based method can analyze the detection results by using the attention map. The experimental results with image classification datasets show that our method outperforms the existing drift detection method and provides detailed analysis for detection results by the attention map.

2 Related Work

2.1 Drift detection

Zhao et al. [12] have categorized *drift* in machine learning/pattern recognition communities into the following three types: *concept drift*, *label drift*, and *data drift*. The concept drift means that the interpretation or concept of input data change between model training and operation. The label drift is that the label distribution during model training changes from the label distribution during operation. It is known that seemingly trivial distribution changes can affect the model performance [14]. And also, in case that a model outputs a decision under uncertainty, even a change in the distribution of labels can affect the performance of the model [11]. The data drift is a change in the feature distribution of the input. For instance, noise due to age-related deterioration of camera lenses and out-of-focus images are ones of data drift, which would affects classification performance.

For detecting such drifts, we generally compare the data distributions between training data and test data that appears during system operation, and we use two-sample test to compare the distributions. Two-sample test determines if there is a significant difference between two distributions. Lipton et al. [8] proposed a label drift detection method. They used Kolmogorov-Smirnov (KS)

test and maximum mean discrepancy (MMD) on the class probability distribution output by the trained model. Rabanser et al. [9] conducted comprehensive studies aimed at detecting both label drift and data drift. In their method, drift is detected by performing dimensionality reduction on each data set and comparing the distributions with a two-sample test. The results show that the detection method using the KS test is the most accurate for the class probability distribution output by ResNet-18 [4].

The above drift detection methods merely consider the distribution of classification outputs. In contrast, our method also uses attention maps obtained from attention branch network (ABN) [2]. This enables us to detect drift considering the local data changes and to analyze the drifted data in details.

2.2 Visual explanation

Towards the realization of explainable artificial intelligence (XAI), *visual explanation* have been widely studied in the computer vision community. Visual explanation outputs an attention map that highlights on which a network model focuses to decide classification result. By visually inspect the obtained attention map, we can understand the reason for the model's decision making and can analyze the trained models.

One of major visual explanation method is class activation mapping (CAM) [13]. CAM can use the response values of the convolutional layer to create an attention map where the network contributed highly to recognition. This attention map is called the class activation map. CAM uses the average of the feature maps in each channel output by global average pooling (GAP) as weights and generates a class activation map from the weighted sum of each feature map. Gradient-weighted CAM [10] provides an attention map by computing gradient information from weights of each channels. Therefore, there are no restrictions on the model structure and the model can be generalized. However, these methods require processing such as replacing the fully connected layer with a convolution layer, which can cause performance degradation in image classification.

The above mentioned CAM has a problem of lower classification performance due to the restriction of network architecture. Attention branch network (ABN) [2] resolve this problem. ABN consists of feature extractor that extract a feature map from an input image, attention branch to generate attention map, and perception branch that outputs final classification results from weighted feature map by using attentin mechanism. ABN applies the attention map generated from the visual explanation model to the attention mechanism, and simultaneously improves the performance of the visual explanation model and visualizes the gazing area using the attention map. The attention mechanism is a method to improve the generalization performance of a network by emphasizing features in specific regions.

3 Preliminaries on Data Drift Detection

Given two samples $\boldsymbol{x} \in \{x_1, x_2, \ldots, x_n\} \sim P$ and $\boldsymbol{x}' \in \{x'_1, x'_2, \ldots, x'_m\} \sim Q$, where n and m are the number of samples in $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. In data drift detection, $\boldsymbol{x}$ is a dataset used for training a machine learning model, and $\boldsymbol{x}'$ is an unseen test dataset. The data drift detection task aims to decide whether $P(\boldsymbol{x})$ is equal to $Q(\boldsymbol{x}')$, that is, whether two samples $\boldsymbol{x}$ and $\boldsymbol{x}'$ arise from the same population probability density function. For example, in case of $P(\boldsymbol{x}) = Q(\boldsymbol{x}')$, we can assume that the data drift does not happen. Meanwhile, if $P(\boldsymbol{x}) \neq Q(\boldsymbol{x}')$, we can assume that the data drift would happen because those samples arise from the different population probability density functions. To detect data drift by using the above approach, we apply two-sample tests.

Rabanser et al. [9] have conducted comprehensive study on the effectiveness of two-sample tests for data drift detection. They used Kolmogorov-Smirnov (KS) test and maximum mean discrepancy (MMD) in their studies. Hence, we adopt those tests in our approach. Especially, we use KS test for the proposed method. Herein, we introduce KS test and MMD.

3.1 Kolmogorov-Smirnov test

The KS test determines whether there is a significant difference by using the maximum difference of the cumulative distributions obtained by accumulating the two distributions as the statistical test quantity. The test statistic D is defined as follows:

$$D = \sup_z |F_{\boldsymbol{x}}(z) - F_{\boldsymbol{x}'}(z)| \,, \tag{1}$$

where $F_{\boldsymbol{x}}(z)$ and $F_{\boldsymbol{x}'}(z)$ mean the cumulative density function obtained from the sample $\boldsymbol{x}$ and $\boldsymbol{x}'$ whose values are less than or equal to a certain value z, respectively. Therefore, $F_{\boldsymbol{x}}(z)$ used for KS test can be defined as

$$F_{\boldsymbol{x}}(z) = \frac{1}{n} \sum_{i=1}^{n} f(x_i; z), \tag{2}$$

where

$$f(x_i; z) = \begin{cases} 1 \ (x_i \leq z) \\ 0 \ (x_i > z) \end{cases} . \tag{3}$$

3.2 Maximum mean discrepancy

MMD estimates the distance between distributions using a kernel function to determine if there is a significant difference between two distributions. Assuming that p and q are based on the mean embeddings $\boldsymbol{\mu}_p$ and $\boldsymbol{\mu}_q$ of the distributions in a reproducing kernel Hilbert space $\mathcal{H}$, the MMD is formulated as

$$\mathrm{MMD}(\mathcal{H}, p, q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\mathcal{H}} \,. \tag{4}$$

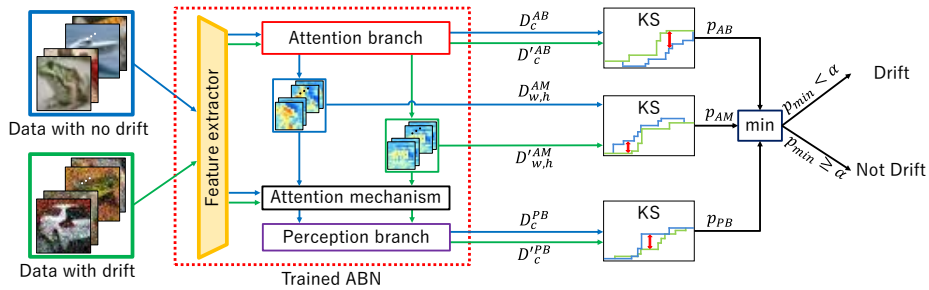


Fig. 1: Overview of the proposed drift detection method

Then, we can calculate an unbiased estimate of the squared MMD statistic as follows:

$$\begin{aligned}\mathrm{MMD}^2 =& \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(x_i, x_j) \\ &+ \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(x'_i, x'_j) \\ &- \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(x_i, x'_j),\end{aligned} \tag{5}$$

where $k(x, x')$ is a kernel function. In this study, we use the Gaussian kernel by following Rabanser et al. [9], which is defined as

$$k(x, x') = \exp\left(-\frac{1}{\sigma}\|x - x'\|^2\right), \tag{6}$$

where σ is a positive constant and we set σ as the median distance between the aggregate samples of p and q. Finally, we can obtain p-value by applying a permutation test for the kernel matrix.

4 Proposed Method

In this section, we introduce the details of the proposed data drift detection method. Figure 1 shows the overview of the proposed method. We first train the ABN with training data. Here, note that the ABN does not simulate drift on the training data. Then, we detect data drift. The specified number of samples is obtained from the data without drift and from the operational data and input into the trained ABN. For each data set, the class probability distribution of the attention branch and perception branch is calculated. The feature distribution of the attention map is calculated. The calculated distribution is temporarily stored for drift detection. Finally, we calculate p-values by KS test and decide if

the input is drifted or not. Hereafter, we introduce the details of the proposed method. Especially, we explain the way to conduct KS test in our ABN-based model.

The proposed method utilizes the attention branch network (ABN) [2]. In ABN, the attention map obtained from the attention branch is input to the attention mechanism, and inference is performed by highlighting features in specific regions. Therefore, the change in the attention map due to drift should enhance the features in response to changes in the input image and improves the accuracy of drift detection.

4.1 KS test for class probability distribution

The proposed method detects drift from three outputs, focusing on the different feature spaces of the layers that make up the ABN. The first is the class probability distribution output from softmax for the GAP of the attention branch, and the second is the class probability distribution output from the perception branch.

The cumulative distribution is expressed as the sum of these frequencies divided by the number of samples. Here, we assume that we handle C class classification problem. The output from a classification model for an input x_i is C-dimensional class probability (classification score) $\boldsymbol{s}_i = (s_{i,1}, \ldots, s_{i,C})^\top$. Because ABN has two output layers, attention and perception branches, we define the class probability outputs for each branch $\boldsymbol{s}_i^{AB}$ and $\boldsymbol{s}_i^{PB}$ as

$$\boldsymbol{s}_i^{AB} = \left(s_{i,1}^{AB}, \ldots, s_{i,C}^{AB}\right)^\top, \tag{7}$$

$$\boldsymbol{s}_i^{PB} = \left(s_{i,1}^{PB}, \ldots, s_{i,C}^{PB}\right)^\top. \tag{8}$$

The proposed method detects a data drift by using the class probability distribution obtained from both attention and perception branches. Specifically, we apply KS test for each category's class probability and for each branch $s_{i,c}^{AB}$ and $s_{i,c}^{PB}$. In case of class c of attention branch, we denote $\boldsymbol{s}_c^{AB} = \{s_{1,c}^{AB}, \ldots, s_{n,c}^{AB}\}$ and $\boldsymbol{s}_c'^{AB} = \{s_{1,c}'^{AB}, \ldots, s_{m,c}'^{AB}\}$ as sets of class probability of sample $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. The test statistic D_c^{AB} can be formulated as

$$D_c^{AB} = \sup_z \left| F_{\boldsymbol{s}_c^{AB}}(z) - F_{\boldsymbol{s}_c'^{AB}}(z) \right|, \tag{9}$$

where

$$F_{\boldsymbol{s}_c^{AB}}(z) = \frac{1}{n} \sum_{i=1}^{n} f(s_{i,c}^{AB}; z). \tag{10}$$

Likewise, with respect to class c of perception branch, we define as follows:

$$D_c^{PB} = \sup_z \left| F_{\boldsymbol{s}_c^{PB}}(z) - F_{\boldsymbol{s}_c'^{PB}}(z) \right|, \tag{11}$$

where

$$F_{\boldsymbol{s}_c^{PB}}(z) = \frac{1}{n} \sum_{i=1}^{n} f(s_{i,c}^{PB}; z). \tag{12}$$

Finally, we apply the KS test for the class probability of each class on $(D_c^{AB}, D_c'^{AB})$, and we decide the minimum value of them as p_{AB}, which is formulated as

$$p_{AB} = \min_{c \in C} \left\{ \mathrm{KS}(D_c^{AB}, D_c'^{AB}) \right\}, \tag{13}$$

The p value is the probability that the test statistic obtained from the KS is greater than or equal to that obtained from the KS under the assumption that the populations of the two distributions are the same. To perform multiple testing where the KS test is repeated for the number of classes and the p value is calculated, a Bonferroni correction [1] is performed for the significance level α used to determine the drift. For the class probability distribution output by the perception branch, p_{PB} is also calculated as

$$p_{PB} = \min_{c \in C} \left\{ \mathrm{KS}(D_c^{PB}, D_c'^{PB}) \right\}. \tag{14}$$

4.2 KS test for attention maps

In the proposed method, we also use an attention map highlighting the gazing region of the model for drift detection. By exploiting the attention map to drift detection, it is possible to capture the characteristics of changes in the actual gazing region of the model. Here, let $\boldsymbol{a}_i \in \mathbb{R}^{W \times H}$ be an attention map obtained from x_i, which is expressed as

$$\boldsymbol{a}_i = \begin{pmatrix} a_{i,(1,1)} & \cdots & a_{i,(W,1)} \\ \vdots & \ddots & \vdots \\ a_{i,(1,H)} & \cdots & a_{i,(W,H)} \end{pmatrix}. \tag{15}$$

Because drift detection using an attention map requires a KS test for each pixel (w, h), the test statistics $D_{w,h}^{AM}$ of the attention map is calculated by

$$D_{w,h}^{AM} = \sup_z \left| F_{\boldsymbol{a}_{w,h}^{AM}}(z) - F_{\boldsymbol{a}_{w,h}'^{AM}}(z) \right|, \tag{16}$$

where $\boldsymbol{a}_{w,h}^{AM} = \{a_{1,(w,h)}^{AM}, \ldots, a_{n,(w,h)}^{AM}\}$ and $\boldsymbol{a}_{w,h}'^{AM} = \{a_{1,(w,h)}'^{AM}, \ldots, a_{m,(w,h)}'^{AM}\}$. The feature distribution $F_{\boldsymbol{a}_{w,h}^{AM}}(z)$ is defined as follows:

$$F_{\boldsymbol{a}_{w,h}^{AM}}(z) = \frac{1}{n} \sum_{i=1}^{n} f\left(a_{i,(w,h)}^{AM}; z\right). \tag{17}$$

The $D_{w,h}'^{AM}$ is also calculated in the same manner as $D_{w,h}^{AM}$.

Next, we apply KS test for $(D_{w,h}^{AM}, D_{w,h}'^{AM})$ in the attention map and we find the minimum value p_{AM}, which is defined as

$$p_{AM} = \min_{w,h} \left\{ \mathrm{KS}(D_{w,h}^{AM}, D_{w,h}'^{AM}) \right\}. \tag{18}$$

As with drift detection using the class probability distribution, the significance level α is adjusted by the Bonferroni correction depending on the number of times the KS test is performed.

4.3 Drift detection from multiple p-values

As mentioned in above sections, we can obtain multiple p-values from the classification output of ABN' branches p_{AB} and p_{PB} and from attention map p_{AM} by applying KS test for them. For detecting data drift, we integrate the results. Given a set of p-values (p_{AB}, p_{AM}, p_{PB}), we take the minimum of them p_{min} which is defined as

$$p_{min} = \min\left(p_{AB}, p_{AM}, p_{PB}\right). \tag{19}$$

The p_{min} is the one determined to have the largest difference between the distribution of the training data and the operational data when the KS test is performed on the three outputs. This enables the three outputs to be used jointly to detect drift.

Finally, if $p_{min} < \alpha$, we conclude that drift has occurred, where α is the threshold value equivalent to Rabanser et al.'s method [9], and the Bonferroni correction is made on the basis of to the number of times the KS test is performed. This value is the significance level of $5^{\%}$ commonly used in two-sample tests.

5 Experiment

To investigate the effectiveness of the proposed method, we evaluate the proposed method. Especially, we evaluate the following aspects: (i) detection accuracy of data drift and (ii) qualitative analysis for obtained attention maps for each drifted data.

5.1 Datasets

We use MNIST [7] and CIFAR-10 [6] datasets in our experiments. The MNIST dataset consists of gray-scale handwritten images. MNIST dataset is originally divided into 60,000 training images and 10,000 test images. Among them, we further split the original training set into 50,000 training set and 10,000 validation set. Also, CIFAR-10 dataset consists of 10 object color images, which contains 50,000 training images and 10,000 test images. In our experiments, we split the original training set into 40,000 training set and 10,000 validation set and use them.

We use the training dataset of each dataset to train ABN in our method. The validation data is defined as the data during training, since it has the same data distribution as the training data. Test data shall be used as operational data after drift simulation.

5.2 Drift simulation

Because MNIST and CIFAR-10 dataset does not contain drifted samples. To make drifted samples and to use our evaluation, we simulate the following four data drifts.

Gaussian blur Gaussian blurring is the process of blurring an image using a Gaussian function. In this experiment, three levels of blur intensity are used, and the percentage of simulation applied to the test data for each intensity is set to {10%, 50%, 100%}.

Gaussian noise Gaussian noise is a process of adding noise by changing the luminance of each pixel in an image on the basis of a normal distribution. In this experiment, three levels of noise intensity are used, and the percentage of simulation applied to the test data for each intensity is set to {10%, 50%, 100%}.

Geometric transformation Geometric transformations combine rotation, horizontal and vertical translation, shear, scaling, and horizontal and vertical flipping on an image. In this experiment, there are three levels of strength for each element of the geometric transformation, and the percentage of simulation applied to the test data for each intensity is set to {10%, 50%, 100%}.

Class imbalance Class imbalance involves reducing the number of samples in one particular class. In our experiments, we have three levels of reduction of the number of samples for a particular class: {10%, 50%, 100%}.

5.3 Comparative method & evaluation metrics

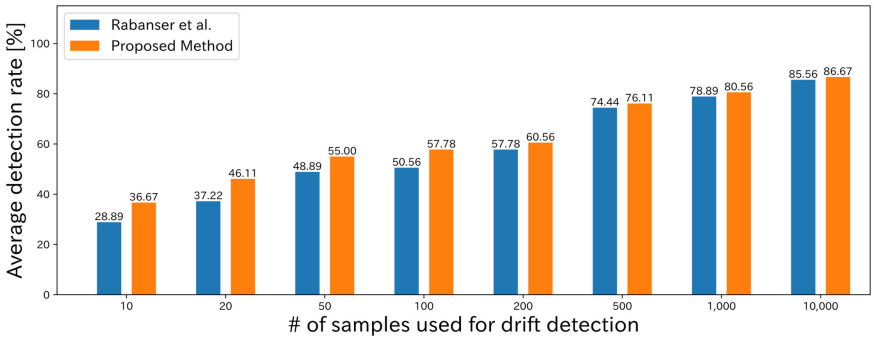
We adopt a drift detection method proposed by Rabanser et al. [9] as a comparative method. Meanwhile, ABN of our method is based on the ResNet-14 architecture for CIFAR dataset. To train the both networks, we used SGD optimizer and train networks in 200 epochs whose mini-batch size is 128.

To evaluate the number of samples required for drift detection in steps, we change the number of samples used for drift detection as $\{10, 20, 50, 100, 200, 500, 1,000, 10,000\}$. These samples includes the above mentioned four types of simulated drift data. We take average the detection rate for all drift simulations. In the experiment, the process involves obtaining a specified number of samples from the training and operational data and inputting them into ABN.

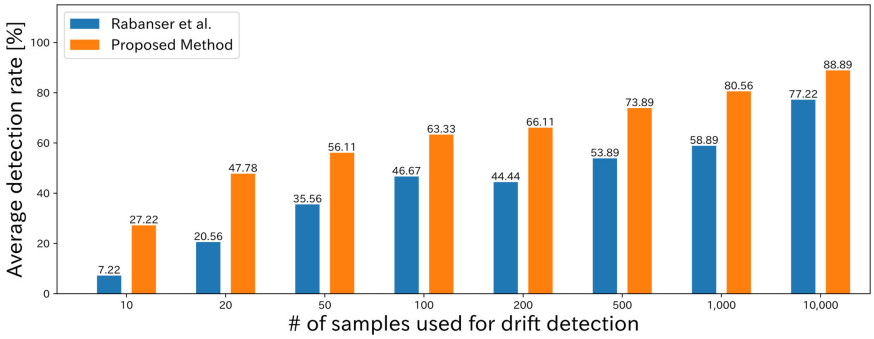
5.4 Comparison of drift detection rates

First, the drift mean detection rates of Rabanser et al.'s method and the proposed method on the MNIST dataset are compared as shown Fig. 2(a). As we can see, the proposed method generally improves the average detection rate compared to the conventional method. Next, a comparison of the drift mean detection rate on the CIFAR-10 dataset is shown in Fig. 2(b), which indicate an improvement in the average detection rate. Moreover, the improvement is greater for detection on the CIFAR-10 dataset compared to detection on the MNIST dataset. These results, demonstrate the effectiveness of the proposed method for in detecting drift.

Also, we compare the detection rate for each drift simulation method for the CIFAR-10 data set for further analysis on the average detection rate, which had a large improvement rate. To facilitate the evaluation of the detection rate, the



(a) MNIST



(b) CIFAR10

Fig. 2: Comparison of average detection rate of drift on (a) MNIST and (b) CIFAR-10 datasets.

number of samples used for drift detection is 100 and 1,000. The results of the experiment shown in Tab. 1 indicate that Gaussian noise is the drift simulation method that contributes the most to improving the average detection rate. The detection rate for 1,000 Gaussian blur images was improved by 26.6 pt., which is the next highest improvement rate after Gaussian noise.

5.5 Comparison of p-value adoption rates

Next, to verify the effectiveness of the proposed method against the detection rate in each drift simulation method, we investigate the contribution of each p value, (p_{AB}, p_{AM}, p_{PB}), calculated in the proposed method. The contribution is the ratio of the number of times each p value is selected for successful drift detection, divided by the number of times all p values are selected. The results of the experiment shown in Table 2 indicate that the overall contribution of p_{PB} is high. However, in Gaussian blurring, p_{AB} is utilized about 48 %, the highest among the three, while p_{AM} is utilized 12.6 %. It can be seen that P_{AM} is

Table 1: Comparison of each drift detection rate (CIFAR-10)

		# of samples to detect drift	
Drift type	Method	100	1000
Gaussian blur	Rabanser et al.	53.3%	57.8%
	Proposed	60.0%	84.4%
Gaussian noise	Rabanser et al.	24.4%	26.7%
	Proposed	53.3%	68.9%
Geometric transformation	Rabanser et al.	46.7%	73.3%
	Proposed	66.7%	86.7%
Class imbalance	Rabanser et al.	62.2%	77.8%
	Proposed	73.3%	82.2%

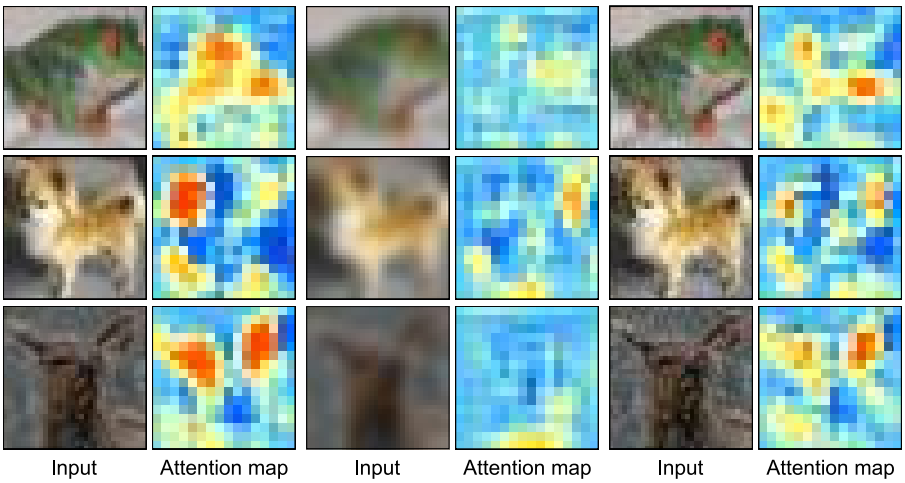
Table 2: Contribution of each p value (CIFAR-10)

Drift type	p_{AB}	p_{AM}	p_{PB}
Gaussian blur	47.9%	12.6%	39.5%
Gaussian noise	34.9%	8.9%	56.2%
Geometric transformation	41.6%	2.6%	55.8%
Class imbalance	46.7%	2.2%	51.1%

employed about 9 % in the Gaussian noise. This indicates that the method of integrating the three outputs is effective for drift detection. The attention map obtained from the ABN can be used for drift detection, and under certain conditions it is more effective than the method using the model class probabilities, contributing to an improvement in the average detection rate of drift.

5.6 Qualitative evaluation on attention map

Next, we determine the change in the attention map for the image that simulates the drift. Figure 3 shows the attention map for an image with no simulated drift and the attention map obtained for an input image with Gaussian blurring and Gaussian noise added. As seen in the figure, the gazing area tends to shrink when simulating the drift caused by Gaussian blurring and Gaussian noise. Even in cases where it is difficult to distinguish the change in the input image before and after the addition of Gaussian noise, as shown in Figure 3(c), it is possible to understand the change from the gazing region of the model by acquiring an attention map. Because the proposed method uses the ABN, which uses an attention map for performing inference, obtaining these changes should improve the accuracy of drift detection.



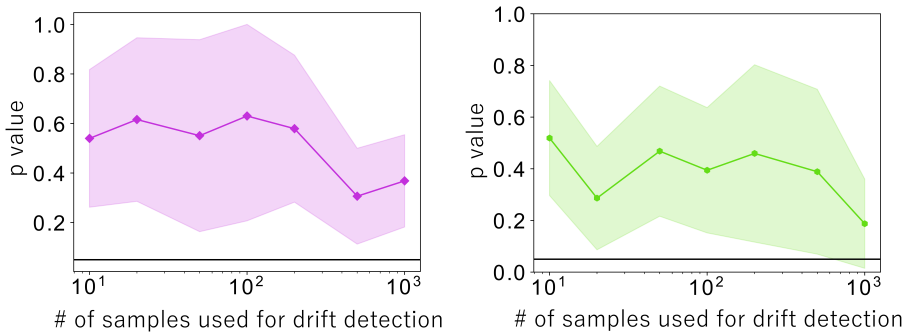
(a) No simulated drift (b) Gaussian blur (c) Gaussian noise

Fig. 3: Attention map changes due to drift

5.7 Difference of two-sample tests for attention maps

We then compare the two-sample test methods for the attention map. As investigated by Rabanser et al. [9], the two-sample test used for drift detection includes the KS test as well as MMD [3]. Rabanser et al. showed that the KS test is the optimal test method when performing a two-sample test on the class probabilities output of the model. However, no experiments have been conducted on the attention map. Therefore, the two-sample test for $(D^{AM}_{w,h}, D'^{AM}_{w,h})$ in Fig. 1 is changed from the KS test to MMD for our experiment. The drift simulation method and significance level are not changed, but the number of samples used to detect drift is limited to 1000, when using MMD, as in Rabanser et al.'s experiment. As an evaluation index, we use the change in p-values relative to the training data that showed a change in the experiment.

Fig. 4(a) shows the transition of p values when the KS test is used for the attention map, and Fig. 4(b) shows the transition of p values when the MMD is used for the attention map. The black line on the graph indicates the significance level. Because the experiment visualizes the evolution of p values for data with no simulated drift, a line graph and its surrounding range below the significance level indicates a false positive. As Fig. 4(a) and Fig. 4(a) show, there are no false positives when the KS test is used, whereas in Fig. 4(b), there are false positives when the MMD is used. It can also be seen that the results are less stable when using MMD. These results indicate that the KS test is a viable option when conducting a two-sample test on the attention map.



(a) KS test used for attention map (b) MMD used for attention map

Fig. 4: Comparison of two-sample test methods used for attention map

6 Conclusion

In this paper, we proposed a data drift detection method based on the ABN and Kolmogorov-Smirnov test. The proposed method detects drift by integrating the class probability distributions output by the attention and perception branches, which construct the ABN, and the results of the KS test using the attention map. Experiments verified that the proposed method improves the average detection rate of drift compared to the conventional method. The drift detection results were analyzed by obtaining an attention map for the drift data. We also investigated the extent to which each output contributes to drift detection and demonstrated the effectiveness of using both the class probability distribution and the attention map to detect drift. Our future work includes further analysis of drift detection results, additional drift simulation methods, and the introduction of two-sample tests appropriate for each distribution to increase accuracy.

References

1. Dunn, O.J.: Multiple comparisons among means. Journal of the American Statistical Association **56**(293), 52–64 (1961)
2. Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H.: Attention branch network: Learning of attention mechanism for visual explanation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
3. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. Journal of Machine Learning Research **13**(25), 723–773 (2012)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
5. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. In: International Conference on Learning Representations (2019)

6. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. Tech. rep., University of Toronto (2009)
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
8. Lipton, Z., Wang, Y.X., Smola, A.: Detecting and correcting for label shift with black box predictors. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 3122–3130 (2018)
9. Rabanser, S., Günnemann, S., Lipton, Z.: Failing loudly: An empirical study of methods for detecting dataset shift. In: Advances in Neural Information Processing Systems. vol. 32 (2019)
10. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
11. Zhang, K., Schölkopf, B., Muandet, K., Wang, Z.: Domain adaptation under target and conditional shift. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 819–827 (2013)
12. Zhao, S., Yue, X., Zhang, S., Li, B., Zhao, H., Wu, B., Krishna, R., Gonzalez, J.E., Sangiovanni-Vincentelli, A.L., Seshia, S.A., Keutzer, K.: A review of single-source deep unsupervised visual domain adaptation. IEEE Transactions on Neural Networks and Learning Systems **33**(2), 473–493 (2022)
13. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
14. Zügner, D., Akbarnejad, A., Günnemann, S.: Adversarial attacks on neural networks for graph data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2018)