Title page

Title : A Machine Learning Approach Towards Standardizing Microscopic Agglutination Test for Diagnosis of Leptospirosis

Authors : Yuji Oyamada[1†], Ryo Ozuru[2†*], Toshiyuki Masuzawa[3], Satoshi Miyahara[4], Yasuhiko Nikaido[4], Fumiko Obata[2], Mitsumasa Saito[4], Sharon Yvette Angelina M. Villanueva[5] and Jun Fujii[2]

Affiliations :
[1] Department of Electrical Engineering and Computer Science, Faculty of Engineering, Tottori University, Tottori, Japan
[2] Division of Bacteriology, Department of Microbiology and Immunology, Faculty of Medicine, Tottori University, Yonago, Japan
[3] Laboratory of Microbiology and Immunology, Faculty of Pharmaceutical Sciences, Chiba Institute of Science, Choshi, Japan
[4] Department of Microbiology, School of Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan
[5] Department of Medical Microbiology, College of Public Health, University of the Philippines Manila, Manilla, Philippines

[†] These authors contributed equally to this work.
* Corresponding author: Ryo Ozuru; ozuru@tottori-u.ac.jp

Abstract

Leptospirosis is a zoonosis caused by the pathogenic bacterium *Leptospira*. The Microscopic Agglutination Test (MAT) is widely used as the gold standard for diagnosis of leptospirosis. In this method, diluted patient serum is mixed with serotype-determined Leptospiras, and the presence or absence of aggregation is determined under a dark-field microscope to calculate the antibody titer. Problems of the current MAT method are 1) a requirement of examining many specimens per sample, and 2) a need of distinguishing contaminants from true aggregates to accurately identify positivity. Therefore, increasing efficiency and accuracy are the key to refine MAT. It is possible to achieve efficiency and standardize accuracy at the same time by automating the decision making process. In this study, we built an automatic identification algorithm of MAT using a machine learning method to determine aggregation within microscopic images. The machine learned the features from 316 positive and 230 negative MAT images created with sera of Leptospira-infected (positive) and non-infected (negative) hamsters, respectively. In addition to the acquired original images, wavelet-transformed images were also considered as features. We utilized a support vector machine (SVM) as a proposed decision method. We validated the trained SVMs with 210 positive and 154 negative images. When the features were obtained from original or wavelet-transformed images, all negative images were misjudged as positive, and the classification performance was very low with sensitivity of 1 and specificity of 0. In contrast, when the histograms of wavelet coefficients were used as features, the performance was greatly improved with sensitivity of 0.99 and specificity of 0.99. We confirmed that the current algorithm judges the positive or negative of agglutinations in MAT images, that it gave the further possibility of making a much suitable algorithm by adding more indices (*e.g.*, antibody titers and bacterial counts) in the future.

Introduction

Leptospirosis, an infectious disease caused by the pathogenic species of *Leptospira*, is one of the most widespread zoonoses in the world. The World Health Organization (WHO) estimates one million leptospirosis cases and 58,900 deaths worldwide each year, of which more than 70% occurring in the tropical regions of the world [1]. Nonspecific and diverse clinical manifestations make clinical diagnosis difficult, and it is easily misdiagnosed with many other diseases in the tropics, such as dengue fever, malaria, and scrub typhus [2].

Microscopic agglutination test (MAT) is considered as the standard test for serological diagnosis of leptospirosis [3]. Leptospires have over 250 serovars [4] and MAT is usually used to diagnose patients based on the *Leptospira* serotypes that infect humans or animals. The principle of MAT is simple, it consists of mixing the serially diluted test serum with a culture of leptospires and then evaluating the degree of agglutination due to immunoreaction using a dark-field microscope [2]. The highest serum dilution agglutinating 50% or more of the leptospires is considered to be the antibody titer. However, the procedures involved in MAT, especially judging the results (i.e., whether positive or negative) requires highly-trained personnel that making it difficult to adopt as a general test [5]. Furthermore, the liquid handling such as transferring all the samples from each well of multi plates onto slide grasses is complicated and takes time. Although the International Leptospirosis Society has been implementing the International Proficiency Testing Scheme for MAT for several years now [6], worldwide standardization of MAT is yet to be achieved. This is partially because not only devices used for MAT such as dark-field microscopes, objective lenses, illuminations and cameras, but the testing condition (the dilution range of serum, incubation time and magnification of an objective lens used) of MAT is diverse at the various laboratories.

This kind of diagnosis is solved as binary classification whether each data is categorized either as positive or negative. Binary classification consists of training and test steps. Figure 1 shows an example of 2-dimensional non-linearly separable data (Fig. 1 (a)). The training step makes several potential boundaries (curves in Fig. 1 (b)) and selects the boundary that separates the training data with the highest classification score (the green curve in Fig. 1 (b)). As shown in the figure, the selected boundary perfectly separates the negative and positive data. The test step evaluates the selected boundary with test data (Fig. 1 (c)). As shown in the figure, the selected boundary both negative and positive data are perfectly separated, which means the trained classifier is good. The success of binary classification heavily depends on the distribution of data. Figure 2 shows how badly distributed data worsen the performance of binary classification. Contrast to Fig. 1, positive and negative data are inseparably distributed (Fig. 2 (a)). None of the potential boundaries cannot perfectly separate even the training data (Fig. 2 (b) and (c)).

For binary classification on images, we use image features instead of raw images because raw images are too complex and redundant, which is potentially badly distributed data. An image feature, extracted from an image, encodes interesting information of the image into a sort of numerical values such as vector and matrix [7]. Image features appropriate for binary classification are sensitive to class differences but not to other factors, which means each negative and positive data are distributed separately.. Therefore, it takes an important role to use or design appropriate image features for the target problem.

Our ultimate goal is to fully automate the MAT procedure. This study proposes a machine learning technique in classifying MAT images towards the fully automated MAT. This is the first report using machine learning techniques for MAT diagnosis that uses a binary classification where each MAT image is taken either as agglutination positive or negative. This study aims to: (1) establish a machine learning based method for reading MAT results to aid in serodiagnosis; (2) design a MAT image feature that is appropriate for binary classification; and, (3) conduct a validation test with MAT images obtained from animal experiments.

Materials and Methods

This section describes the proposed method that uses machine learning techniques for MAT image classification that takes each MAT image either as agglutination positive or negative. Our idea is to follow a typical machine learning based image classification pipeline so that we can efficiently and effectively classify MAT images. The key contribution of this paper is to design an appropriate MAT image feature making binary classification feasible. The design of the image feature is based on how expert examiners evaluate MAT images. They consider how agglutination occupies a MAT image and therefore the proposed method extracts occupancy like information from a MAT image. The remainder of this section first explains how datasets were collected for this research and then the major steps, pre-processing, feature extraction, and binary classification issues were described.

Animal Ethics Statement

Animal experiments were reviewed and approved by the Ethics Committee on Animal Experiments at the University of Occupational and Environmental Health, Japan (Permit Number: AE15-019). The experiments were carried out under the conditions indicated in the Regulations for Animal Experiments of the said University and Law 105 and Notification 6 of the Government of Japan.

Preparation of image datasets

Animal infection experiment and acquisition of MAT images

In this study, a total of three golden syrian hamsters, male, 3-week-old were used (purchased from Japan SLC, Inc., Shizuoka, Japan) . Serum from a hamster subcutaneously infected with 10^4 Leptospira interrogans serovar Manilae at day 7 post infection was used as a MAT positive sample, and sera from other two hamsters at day 0 post infection was used as a MAT negative. MAT was performed according to the standard manual (Terpstra et al., 2003). In brief, each serum was primarily diluted 50-fold and then serially diluted (2-fold) until 25,600-fold dilution. Leptospires (1×10^8) was added to each diluted serum and incubated at 30°C for 2-4 hours. Samples from each well were transferred to slide glasses and covered with coverslips. Each sample was observed under a dark-field microscope (OPTIPHOT, Nikon). Ten images from 20× or 40× objective lens fields per slide were obtained with a CCD camera (DP21, Olympus).

Pre-processing

In this step, training and test datasets were prepared from the original image datasets as shown in Fig. 2. Asit was mentioned earlier, the proposed method utilizes the occupancy of agglutination area in a MAT image as an image feature. This assumes that all the images are taken under the same condition such as lighting condition and meter-resolution scale. Furthermore, images are supposed to contain enough resolution to measure the occupancy of areas with agglutination but not too large in order not to overwhelm the computer. To limit sizes of images suitable to analyze, all images were resized into smaller patches prior to the feature extraction.

Scale normalization

The scale normalization is done in order to make all the images in the dataset have the same micrometer-resolution scale. The original dataset is made of a series of negative images captured under 20X objective lens (50 micrometer scale) and of positive images with 40X objective lens (20 micrometer scale) (Fig. 4(a)). They contain the same resolution; however, the pixels applied to 50 micrometer in the negative images are 290, whereas to 20 micrometer in the positive images are 232. Based on the micrometer-resolution scale indicator shown in each image, this preprocessing resizes them to have the same micrometer-pixel scale (Fig. 4(b)).

Image patching

In the next step, each image was cut into smaller patches. Each patch must be large enough to count the amount of agglutination and also must avoid too large resolution to fulfill the minimum requirement of computers. In the succeeding experiments, different patch sizes were compared and the effect of the patch size difference was validated.

Patch-level standardization

The last step of the pre-processing was to apply data standardization, a frequently used data normalization method, on each patch. This process was applied to satisfy an assumption by the later classification step that all feature vectors are centered around zero and have similar variance. For instance, if the data acquisition process affects the brightness of observed pictures, then the standardization is applied to cancel those unwilling effects. Suppose we have a feature vector $x \in R^d$ representing an image patch. Data standardization normalizes the feature vector such that the normalized vector has zero mean and unit variance as $\hat{x} = \frac{\{x-\mu_x\}}{\sigma_x}$, where $\mu_x$and σ_x denote the mean and standard deviation of a set of x.

Image features

Here, the design of the image feature that takes the key role of image classification was described. There was still difficulty in the pre-processed MAT image patches such as images that might contain non-agglutination objects such as dust that has strong reflection. Consequently, non-agglutination objects from the pre-processed MAT image patches are excluded and the occupancy of the agglutination area is measured. Specifically, the designed image feature is the histogram of multi-level wavelet coefficients of a MAT image patch. The multi-level wavelet transform (Mallat, 1989) was used to extract objects on MAT image patches that have similar size to agglutination. By using the histogram of its coefficients, the behavior of the coefficients can be efficiently represented. In the remainder of this part, raw images (Image), multi-level wavelet coefficients (Wavelet), and the histogram of multi-level wavelet coefficients (HoW) will be explained step by step.

Raw images (*Image*)

One of the simplest image features is raw image itself.
A gray-scale image is represented as a 2D matrix $I \in R^{X \times Y}$, where X and Ydenote the image width and height. A pixel intensity at pixel(x, y) is represented as $I_{x,y}$ and the value is

usually expressed by an eight-bit integer, 0, 1, 2, …, 255. When raw images are used as an image feature, comparison between two images was done by comparing their corresponding pixel intensities. For instance, mean square error between two images $I$and $I'$is defined as

$$E_{MSE} = \frac{1}{XY}\sum_{y=1}^{Y}\sum_{x=1}^{X}\left(I_{x,y} - I'_{x,y}\right)^2$$

.

Raw images are easy to use and require less computer power. However, the pixel-wise comparison is inappropriate to consider the occupancy of the agglutination areas.

Multi-level wavelet coefficients (*Wavelet*)

To extract agglutination of a certain size, multi-level wavelet transformation that is usually used to decompose images into base signals in different scales was utilized.

The idea of multi-level wavelet transformation (Mallat, 1989), which decomposes an image to a combination of a base signal with different directions and resolutions, is similar to the Fourier transformation. Applying multi-level wavelet transformation to a MAT image patch, any objects on the patch are separately extracted and stored into coefficients at different scales as larger coefficients. The lowest scale coefficients $W_0 \in R^{X_l \times Y_l}$ is the mean image of the patch while the other coefficients $\{W_l \in R^{X_l \times Y_l} \mid l = 1, \ldots, L\}$ contain large values where image objects of specific size exist at the location, where X_l and Y_l denote the width and height of l-th level as

$$X_l = \frac{X}{2^{(L-l+1)}}$$

and

$$Y_l = \frac{Y}{2^{(L-l+1)}}$$

.

Namely, lower level coefficients have the information of smaller image objects and larger level coefficients have one of larger image objects. In other words, lower level coefficients correspond to coarser image components such as rough shape objects and larger level coefficients to finer components.

Figure 4 shows multi-level wavelet coefficients of a MAT image patch with Haar wavelet as base signal and 4 levels. As shown in the figure, image objects of different sizes are extracted in different levels, coarser features are in lower level and finer in higher. In these figures, zero coefficients are depicted as gray pixels and strong features are depicted as white and black pixels. Trained MAT examiners confirmed that *Leptospira*-like objects showed brighter pixels in levels 2 and 3 but not in the other levels.

Multi-level wavelet transformation to MAT image patches was applied and decided their coefficients of a level might be a candidate of the image feature. Benefit of this feature is that we can exclude non-agglutination objects based on their size. The drawback is the same as the *Image* feature that *Wavelet* feature also applies pixel-wise comparison.

Histogram of Multi-level wavelet coefficients (*HoW*)

To represent the occupancy of the agglutination areas in a MAT image patch, a normalized histogram of multi-level wavelet coefficients per level with a sum of 1 was constructed. The shape of such histograms represents the occupancy of the agglutination areas.

Constructing the histogram of wavelet coefficients is equivalent to count agglutination-like objects and therefore it is appropriate to measure the occupancy of the agglutination areas. Multi-level wavelet transformation was first applied and the histogram of the coefficients per level was constructed. Next, the histogram with a sum of 1 was normalized. By normalizing the histogram, histograms of different image resolutions are comparable. In total, we have a set of $L+1$ histograms as $\{H_l \in R^B \mid l = 0, \ldots, L\}$ where H_l denotes the normalized histogram of l-th level wavelet coefficients and$B$ denotes the number of histogram bins. Benefits of this feature are its robustness against the existence of non-agglutination objects.

Figure 5 shows HoWs at different levels of a MAT image patch. As shown in the figure, histograms at different levels have different sharpness.Since each level of wavelet coefficients has zero at most of the pixels, their histogram has their peak at the center. When the patch contains image objects of a specific size, their corresponding histograms have gentle peaks.

Image feature classification

Here, we explain the details of binary classification. The proposed method utilizes the Support Vector Machine (SVM) (Boser et al. 1992), which is one of the supervised learning models for classification problems and has been well-used in practical situations because of its generalization performance against unknown data. Moreover, grid search hyper-parameter optimization and K-fold cross validation were combined in order to obtain better training effects.

SVMs

The proposed method utilizes Support Vector Machines (SVMs) [8], which is one of well-used supervised learning models for solving classification problems.
Suppose we have a set of Ntraining data $D = \{(x_i, y_i) \mid x_i \in R^M, y_i \in \{-1, 1\} \mid i = 1, \ldots, N\}$, where x_i denotes feature vector and y_i denotes its label. The label $y_i = -1$ indicates$i$-th feature is negative data and $y_i = 1$ indicates it is positive data. In our case, x_i is a MAT image feature, and negative and positive data means not-agglutinated and agglutinated data, respectively. Given the set of training data D, an SVM finds a boundary that maximizes its margin, which is the largest distance to the nearest data of both classes.

Using kernel trick, SVMs accomplish even non-linear classification. Using a kernel function, An SVM alters original data x_i to higher dimensional space, and linearly separates the calculated data in the projected domain. The kernel function decides the complexity of boundary shape. Typical kernel functions for non-linear classifications are polynomial functions

$$k_{poly}(x_i, x_j) = \gamma(x_i \cdot x_j)^d,$$

where γdenotes the scale factor and d the dimensionality of the polynomial and radial basis functions

$$k_{rbf}(x_i, x_j) = exp\,(-\gamma||x_i - x_j||^2),$$

where γdenotes the non-negative scale factor.

Grid Search Hyper-parameter tuning

The proposed method utilizes grid search for tuning hyperparameters of SVMs.

The hyperparameters of SVMs are parameters of kernel function, e.g., γ of polynomial and radial basis functions, and general SVM parameter C. Grid search is one of hyperparameter tuning methods that exhaustively considers all potential combinations of hyperparameters.

K-fold cross validation

The proposed method applies K-fold cross validation to maximize training effects.

One of key concerns about supervised learning models is its generalization performance against unseen data. SVM training is similar to how we tune all parameters for the given training data but not for unseen data. So, there is no guarantee that the trained SVM performs well against unseen data. Situations that a trained supervised model performs well to training data but not to unseen data is called overfitting.

K-fold cross validation is a well-used practical solution to avoid such overfitting. Figure 6 shows the concept of K-fold cross validation. All data is splitted into either training or test datasets and the training dataset is further divided into$K$ group, each of which is called a fold. For each fold, an SVM is trained with the remaining $K - 1$ folds and is validated with the fold. The validation errors of all folds are averaged to measure generalization performance of the model while cancelling overfitting for specific data. After the training, the trained SVM is further validated with the test dataset.

Results

We conducted three evaluations to validate the effectiveness and efficiency of the proposed method: (1) the elapsed time of the feature extraction process, (2) qualitative evaluation of the MAT image feature classification, and (3) quantitative evaluation of the MAT image feature classification.

Experimental conditions

Here, we describe the experimental conditions. We conducted all the experiments on a computer with an 8 core central processing unit (CPU) (Intel(R) Xeon(R) CPU E5-2620 v4) and 128 GB random access memory (RAM). All the data is stored on a hard disk drive (HDD). We implemented the proposed method using Python with standard packages such as numpy and scikit-learn and PyWavelet for Multi-level Wavelet Transformation [9].

Table 1 and 2 show the data properties. We have 295 images in total, which consist of 32 negative and 263 positive images. All images are in a resolution of 2448 x 1920. The micrometer-pixel ratio is 5.8 and 11.6 in negative and positive images, respectively. We resized all positive data so that the micrometer-pixel ratio is consistent across the classes.

In these experiments, we test two patch sizes 256 x 256 and 512 x 512. The number of patches for each patch size is shown in Table 3.

For image features, L, the level of multi-level wavelet transformation, is set to 4, whereas B, the number of histogram bins, is set to 64. We tested image features: *Image, Wavelet* and *HoW* features at each level, denoted as *Wavelet-l* and *HoW-l.* Moreover, we combined all levels of *Wavelet* and *HoW*, denoted as combined *Wavelet* and combined *HoW,* and tested as well.

Table 4 shows the dimensionality of each feature with different patch sizes. *Wavelet-l* dimensionality becomes larger as the level increases, while *HoW-l* has constant dimensionality.

Elapsed time of feature extraction

The first experiment measures the elapsed time of the feature extraction process. Figure 7 (a) and (b) visualize the elapsed time of the pre-processing and feature extraction respectively. Note that the pre-processing is executed per raw image while the feature extraction is executed per patch. Thus, patch size affects the elapsed time only on the feature extraction. The pre-processing is roughly 4 Hz that is far from real-time applications that require at least 30 Hz. The feature extraction is dominated by Wavelet transformation; however total elapsed time for *HoW* computation is roughly 60 Hz even for the larger patch size.

To improve the computational speed, we can use both hardware and software level techniques. The hardware technique is to use solid state drive (SSD) instead of HDD so that the process *Image Load* is fastened. One of potential software techniques is parallel

programming that simultaneously processes more than a single image such as multi-core programming and graphics processing unit (GPU) programming. To improve and globalize the computing environment, a potential application can be a cloud-based system that users send raw MAT images via the internet and all the processes are performed on the server PC with multi-core CPU and GPU.

Qualitative evaluation of MAT image feature classification

The second experiment is a qualitative evaluation of the MAT image feature classification. In this experiment, we visualize the distribution of all the data in each image feature domain to see which image feature is suitable for image classification.
For this evaluation, we use T-distributed Stochastic Neighbor Embedding (t-SNE) [10]. In t-SNE, a non-linear dimensionality reduction method, it embeds high dimensional data into lower dimensions of two or three dimensions in such a way that similar data are distributed closer and dissimilar data are distributed further in the lower dimension with high probability. When t-SNE embeds image features of each class into distinctly isolated clusters, the features have potential ability to be good features for classification.

We set the hyper-parameter of t-SNE as follows: perplexity is 50 and the number of iterations is 3000. We applied t-SNE visualization for all combinations of image features and resolutions of 256×256 and 512×512.

Figure 9 compares t-SNE visualization of each image feature. In each plot, red x and purple + symbols represent negative and positive data respectively and values of both axes are omitted because the scale does not matter in the embedded lower dimensional spaces. As shown in the figure, both negative and positive data of *Image* and *Wavelet* are embedded into inseparable clusters. On the other hand, higher level *HoW*, levels 2 to 4, and all levels are embedded into distinctly isolated clusters as expected. Lower level *Wavelet*, levels 0 and 1, show intermediate embedding of aforementioned ones that the embedded features look inseparable with lower resolution 256×256 but more separable with higher resolution 512×512.

For further investigation, Figure 10 compares *HoW* features of the training data with patch size 256×256. The shape of *HoW*-0 varies even in the same category while *HoW*-4 seems to have a unique shape for each category. This indicates that SVM well-classifies higher level *HoW*s but mis-diagnoses lower level *HoW*s.

Quantitative evaluation of MAT image feature classification

The third experiment is a quantitative evaluation of the MAT image feature classification. In this experiment, we conducted an image classification experiment with K-fold cross validation to quantitatively evaluate the image features.

For SVM, we use radial basis function as kernel and potential hyperparameters which are described as below:

$C \in \{10^i \mid i = 0, 1, 2, 3\}$ and $\gamma \in \{10^i \mid i = -3, -2, -1, 0\}$.

For K-fold cross validation, we use 60% of data for training and the remaining 40 % for test data. The number of folds K is set to 5. As evaluation criteria, we use Matthews Correlation Coefficient (MCC) [11]) that ranges from -1 to 1, -1, where -1 means all predictions are wrong, 0 means the predictions are equivalent to random prediction, and 1 means all predictions are correct. Contrary to F-measure, more frequently used criteria, MCC considers balance ratios of two classes that is crucial for our case.

Figure 11 (a) and (b) compare confusion matrices and MCC values of each image feature. Definitions of graph keys are as follows: True Negative, negative data is predicted as negative; False Positive, negative data is predicted as positive; False Negative, positive data is predicted as negative; True Positive, positive data is predicted as positive. In the figure, horizontal bars with \ symbol represent true prediction while ones with / symbol do false prediction. In both resolutions, higher-level *HoWs* result in higher MCC while lower-level *HoW*, Image, and *Wavelet* result in lower MCC.
Figure 12 shows failed test cases of *HoW* features. Comparing those failed features to the training data shown in Figure 9, it is convincing that the SVM mis-classified those features. Table 5 and 6 show the elapsed time for the training and test respectively. Note that the elapsed time of the test is measured per each patch while the one of the training is measured per all the training dataset. Comparing the data between patch sizes, the elapsed time of 256×256cases is 10 times and 5 timers larger than 512×512 cases, on training and test cases respectively. Considering the elapsed time and classification performance, we can say that HoW efficiently codes the characteristics of MAT images. This result indicates that SVM with higher-level *HoWs* work as a good MAT image classifier.

Discussion

MAT is a standard test for diagnosis of leptospirosis and a determination training is required for clinical technologists, though there have still been variations among technicians or facilities. An alternative to MAT is the macroscopic slide agglutination test (MSAT), developed by Galton et al.[12]. The MSAT is considered a rapid, practical, easy and accessible test, and is as sensitive as the MAT [13]. The MSAT was originally developed for the serological diagnosis of leptospirosis in humans, mainly for the screening of acute and recent cases of infection [14]. The MSAT suggests the possible infective serovars using an antigen in suspension, which may include a pool of up to three inactivated serovars. The MSAT can be used for the diagnosis of leptospirosis in both humans [15,16] and animals [17–19]. Other serological assays, such as the enzyme-linked immunosorbent assay (ELISA) [20,21] can also be used to detect infection. However, these alternatives are just simple rapid screening tests, and MAT is needed as ever in high-density areas where precisely testing with more serovars. We aim to automate and standardize MAT, the conventional and standard diagnostic method, rather than developing a new method to replace MAT. We propose to introduce a machine learning approach for determination of antibody titer.

In this study, we primarily attempted to standardize MAT by automating the determination process. Our experimental results showed that the proposed machine learning-based pipeline with the derived image feature well-recognized the agglutination in MAT images, especially combined *HoW* resulted in the MCC score of 1.0, which means perfect prediction. This indicates that the proposed method could substitute the skilled examiners' knowledge and experience by using machine learning techniques. As we showed in the current paper, adding parameters progress capability of analysis, further tuning of the image feature may improve the classification performance.

There still remains some space to improve the proposed method. The standard MAT examination procedure by comparing MAT images to a reference image, while the proposed method classifies each MAT image based on the amount of agglutination areas within. To refine the proposed method, we need to establish an appropriate way to compare reference and MAT images in image feature space. We plan to apply several histogram comparison methods and select one with the best classification performance. Or we will design another image feature that shows a clear difference between positive and negative data when we compare reference and MAT images.

Although a lot of image data was acquired as a result, the limitation of this study is that we tested only one serotype condition, serovar Manilae, with a positive and a negative serum, and an image acquisition device. We need to study and validate with data from other serovars and devices but conclude at least that the direction of the image analysis has been adequately investigated.

A potential future direction is to directly tackle the situation even with smaller numbers of data. One of the recent trends in the machine learning community focuses on algorithms working with less data. To name a few, semi-supervised learning and transfer learning are good candidates for our problem. Semi-supervised learning algorithms are designed for situations that small amounts of samples are known for their positivity or negativity. Transfer

learning algorithms alter knowledge that was built with a large amount of labeled data for a specific problem to another problem where we have access to a small amount of labeled data.

The current study suggests that MAT will be fully automated in the future. Usually, MAT requires to transfer samples from 96-well plates to glass slides, and this is the most complicated handling in all processes. However, if we automate all the manual procedures, it is possible to develop a quick MAT method.

We picture the final product of the proposed method as a cloud-based system. By launching this kind of system on the cloud, anyone can utilize the system via the internet and therefore we can support people in poor resource situations. Furthermore, we can collect a huge amount of data from all over the world that is an image of the future research style.

Conclusion

This paper aimed to automate the MAT procedures for the diagnosis of Leptospirosis. Our idea was to introduce a typical machine learning based image classification pipeline that represents images by an appropriate feature and uses an SVM to classify each MAT image based on the difference in the feature space. The conducted experiments validated that *HoW* feature is an efficient and effective feature for MAT image classification. From these evidences, we concluded that the machine learning-based image classification pipeline has potential power for fully automated classification.

Acknowledgements

We are grateful to Mr. Shoji Tokunaga for his advice.

Author Contributions

Conceptualization; Y.O., R.O.
Data curation; Y.O., R.O.
Formal analysis; Y.O., R.O., T.M., S.M., Y.N., M.S., S.Y.A.M.V.
Funding acquisition; Y.O., R.O., J.F.
Investigation; Y.O., R.O.
Methodology; Y.O., R.O.
Project administration; R.O.
Resources; T.M., S.M., Y.N., M.S., S.Y.A.M.V.
Software; Y.O.
Supervision; J.F.
Validation; Y.O., R.O., T.M., S.M., Y.N., F.O., M.S., S.Y.A.M.V., J.F.
Visualization; Y.O.
Writing - original draft; Y.O., R.O.
Writing - review & editing; Y.O., R.O., T.M., S.M., Y.N., F.O., M.S., S.Y.A.M.V., J.F.

Funding

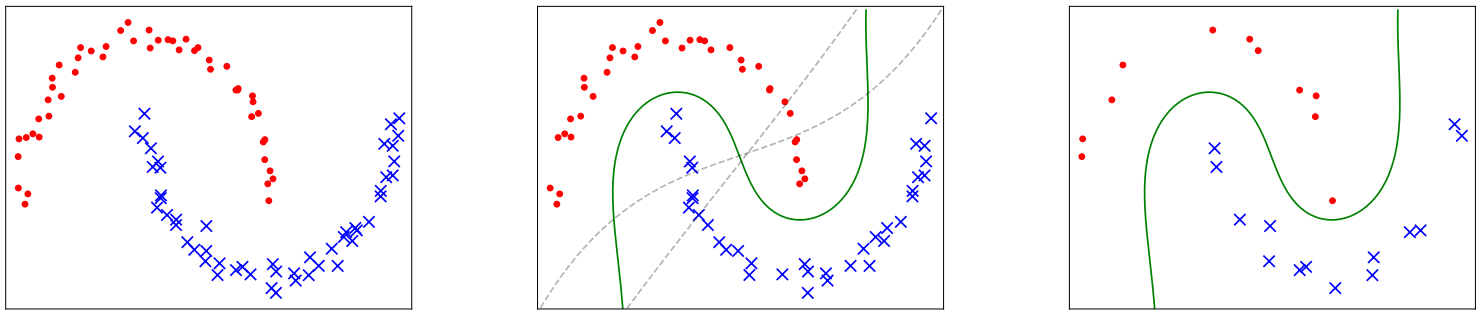
This work was partially supported by JSPS KAKENHI Grant Number 18K16174 to R.O., the discretionary fund of Tottori University President to Y.O. and R.O, and the Research Program of the International Platform for Dryland Research and Education, Tottori University to J.F.

References

1. Costa, F. *et al.* Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Negl. Trop. Dis.* **9**, e0003898 (2015).
2. Terpstra, W. J., WHO, World Health Organization & International Leptospirosis Society. *Human Leptospirosis: Guidance for Diagnosis, Surveillance and Control*. (World Health Organization, 2003).
3. Levett, P. N. Leptospirosis. *Clinical Microbiology Reviews* vol. 14 296–326 (2001).
4. Levett, P. N. Systematics of leptospiraceae. *Curr. Top. Microbiol. Immunol.* **387**, 11–20 (2015).
5. Goris, M. G. A. & Hartskeerl, R. A. Leptospirosis serodiagnosis by the microscopic agglutination test. *Curr. Protoc. Microbiol.* **32**, Unit 12E.5. (2014).
6. Chappel, R. J., Goris, M., Palmer, M. F. & Hartskeerl, R. A. Impact of proficiency testing on results of the microscopic agglutination test for diagnosis of leptospirosis. *J. Clin. Microbiol.* **42**, 5484–5488 (2004).
7. Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **60**, 91–110 (2004).
8. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92* (1992) doi:10.1145/130385.130401.
9. Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K. & O'Leary, A. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software* **4**, 1237 (2019).
10. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
11. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).
12. Galton, M. M., Powers, D. K., Hall, A. D. & Cornell, R. G. A rapid macroscopicslide

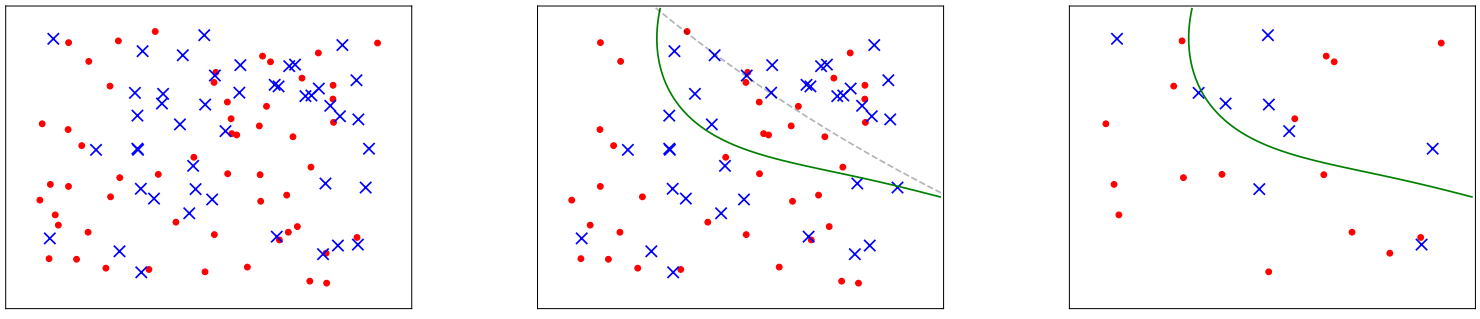
screening test for the serodiagnosis of leptospirosis. *Am. J. Vet. Res.* **19**, 505–512 (1958).

13. Chirathaworn, C., Kaewopas, Y., Poovorawan, Y. & Suwancharoen, D. Comparison of a slide agglutination test, LeptoTek Dri-Dot, and IgM-ELISA with microscopic agglutination test for Leptospira antibody detection. *Southeast Asian J. Trop. Med. Public Health* **38**, 1111–1114 (2007).
14. FAINE & S. Guideline for control of leptospirosis. *World Health Organization Geneva* **67**, 129 (1982).
15. Sumathi, G., Pradeep, K. S. C., Shivakumar, S. & Chinari K, S. MSAT - a screening test for leptospirosis [correspondence]. *Indian J. Med. Microbiol.* **15**, 84 (1997).
16. Brandão, A. P., Camargo, E. D., da Silva, E. D., Silva, M. V. & Abrão, R. V. Macroscopic agglutination test for rapid diagnosis of human leptospirosis. *J. Clin. Microbiol.* **36**, 3138–3142 (1998).
17. Solorzano, R. F. A comparison of the rapid macroscopic slide agglutination test with the microscopic agglutination test for leptospirosis. *Cornell Vet.* **57**, 239–249 (1967).
18. Lilenbaum, W., Ristow, P., Fráguas, S. A. & da Silva, E. D. Evaluation of a rapid slide agglutination test for the diagnosis of acute canine leptospirosis. *Rev. Latinoam. Microbiol.* **44**, 124–128 (2002).
19. Guedes, I. B. *et al.* Development of a pooled antigen for use in the macroscopic slide agglutination test (MSAT) to detect Sejroe serogroup exposure in cattle. *J. Microbiol. Methods* **166**, 105737 (2019).
20. Ngbede, E. O., Raji, M. A., Kwanashie, C. N. & Okolocha, E. C. Serosurvey of leptospira spp serovar Hardjo in cattle from Zaria, Nigeria. *Rev. Med. Vet.* **164**, 85–89 (2013).
21. Derdour, S.-Y. *et al.* Prevalence of The Main Infectious Causes of Abortion in Dairy Cattle in Algeria. *J Vet Res* **61**, 337–343 (2017).



(a) All data (b) Training data and boundaries (c) Test data and boundaries

Figure 1: A sketch of binary classification problem. (a) Red circle and blue cross symbols represent negative and positive data in an image feature space. (b) A classifier obtains boundaries, represented by curves, from the training data and the green curve is selected as the best boundary. (c) The best boundary and the training data.



(a) All data (b) Training data and boundaries (c) Test data and boundaries

Figure 2: A sketch of binary classification problem with a bad image feature. Symbols and curves represent same as Fig. 1. (a) The negative and positive data are inseparably distributed in a bad image feature space. (b) The best boundary and the training data. (b) The best boundary and the test data.

Table 1: The specification of raw MAT images.

	#Data	Resolution [pixels]	Scale [μm/pixel]
Negative	32	2448×1920	5.8
Positive	263	2448×1920	11.6

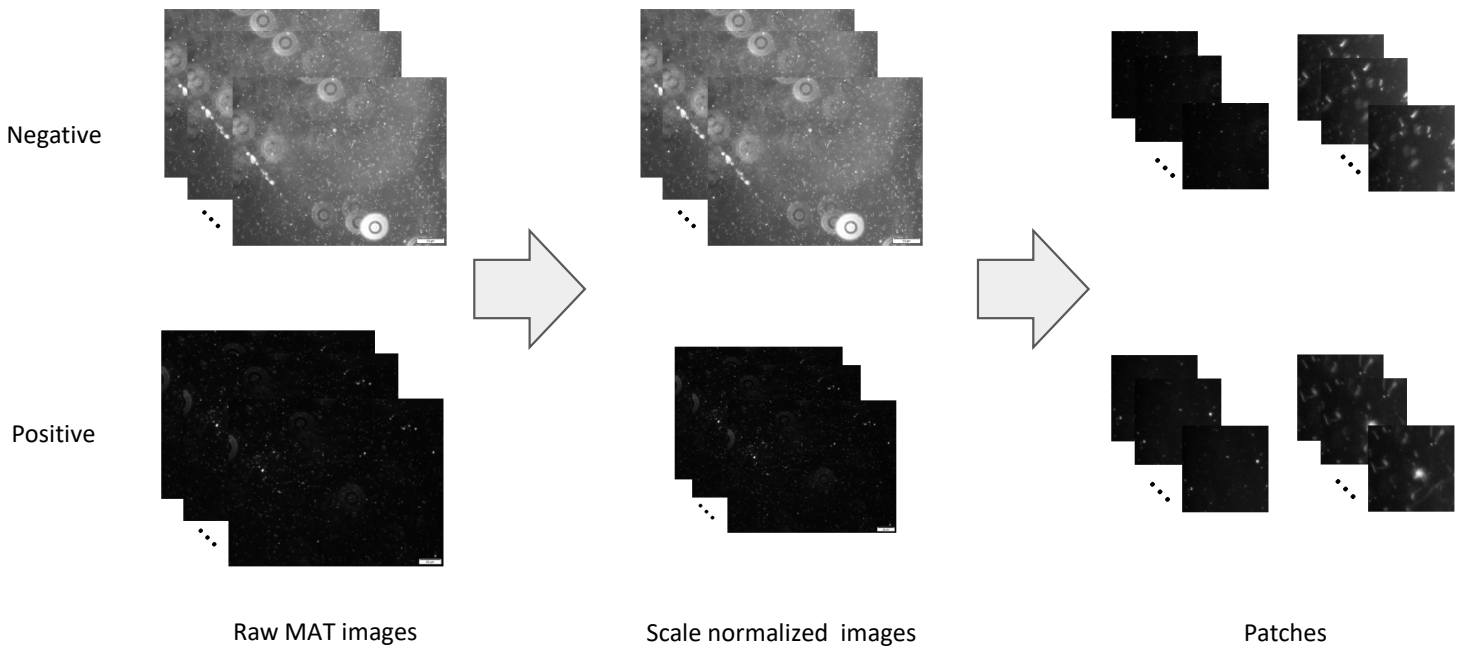


Figure 3: A flow of image patching.

Table 2: The specification of scale normalized MAT images.

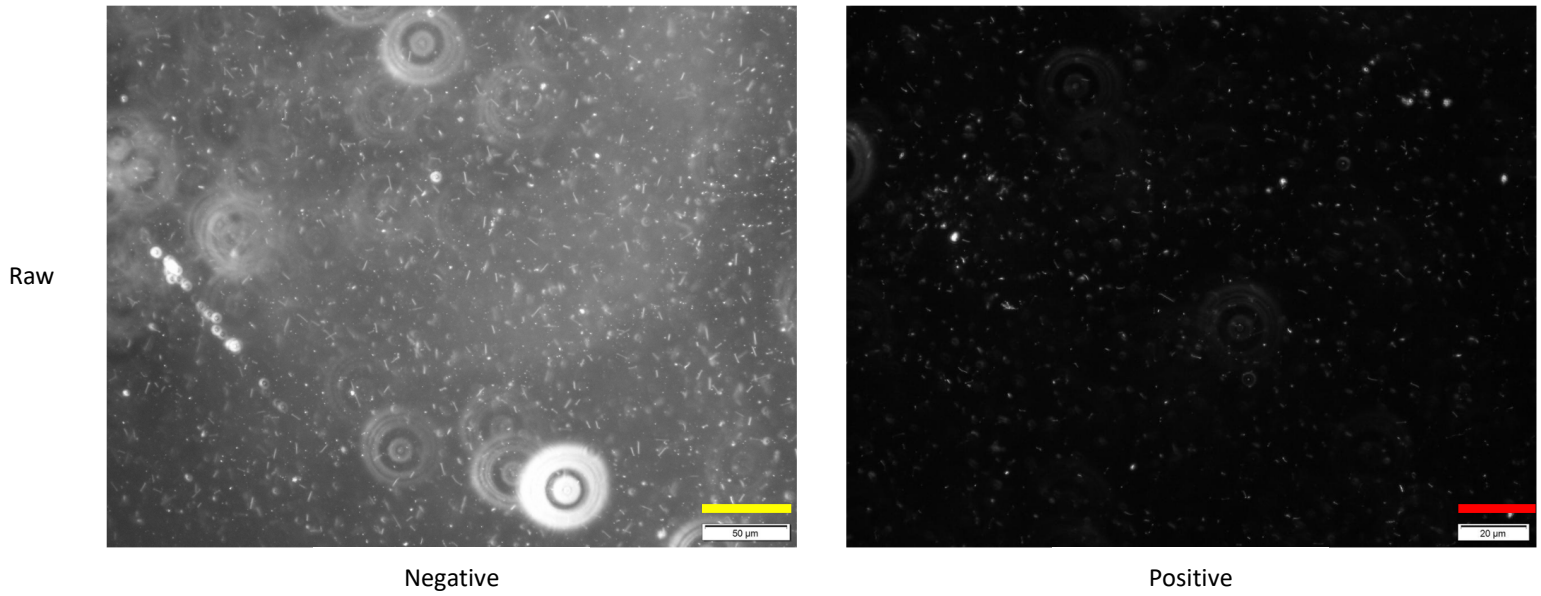
	#Data	Resolution [pixels]	Scale [μm/pixel]
Negative	32	2448×1920	5.8
Positive	263	1224×960	5.8

Table 3: The number of patches used for all the experiments.

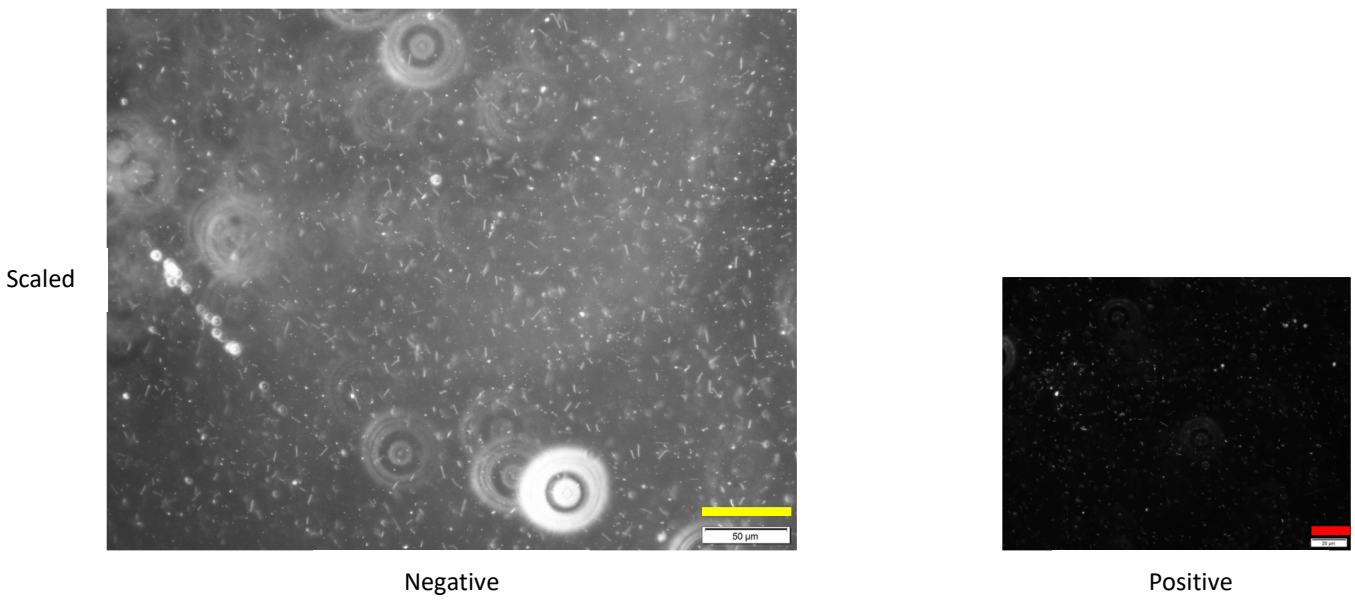
Data label \ Patch size	256×256	512×512
Negative	1680	360
Positive	2688	448
Total	4368	808

Table 4: The dimensionality of the image features.

	256×256	512×512
Image	65536	262144
Combined *Wavelet*	65536	262144
Wavelet-0	256	1024
Wavelet-1	768	3072
Wavelet-2	3072	12288
Wavelet-3	12288	49152
Wavelet-4	49152	196608
Combined *HoW*	320	320
HoW-0	64	64
HoW-1	64	64
HoW-2	64	64
HoW-3	64	64
HoW-4	64	64



(a) Raw images.



(b) Scale normalized images.

Figure 4: The scale normalization. Yellow lines on the negative images represent 50μm while red lines on the positive images represent 20μm.

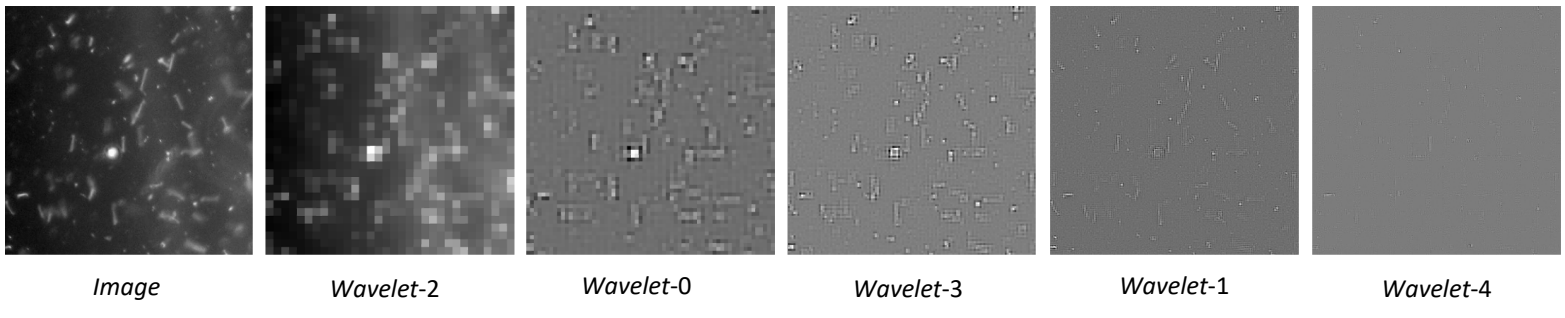


Figure 5: An image and its multi-level wavelet coefficients. *Wavelet-l* means l-th level wavelet coefficients. Lower level wavelet coefficients extract coarser image characteristics while higher ones finer characteristics.

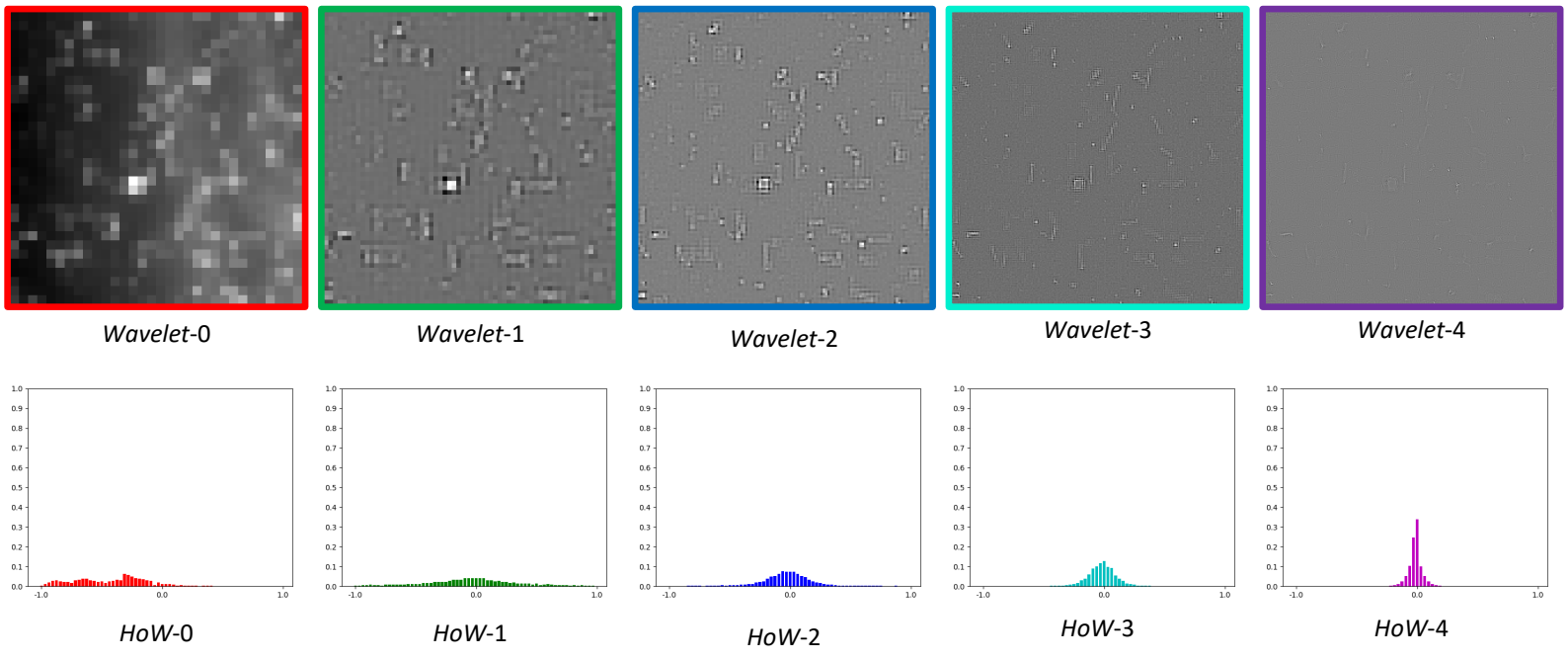


Figure 6: Wavelet coefficients and their histograms. Images on top and bottom show *Wavelet* and *HoW* features. From left to right, depicted each feature of level 0 to 4 are shown.

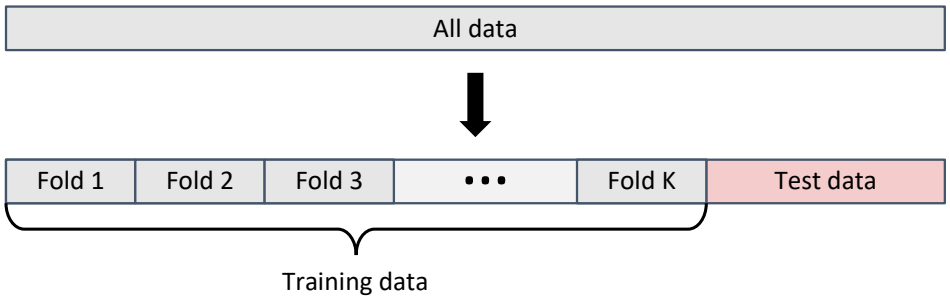
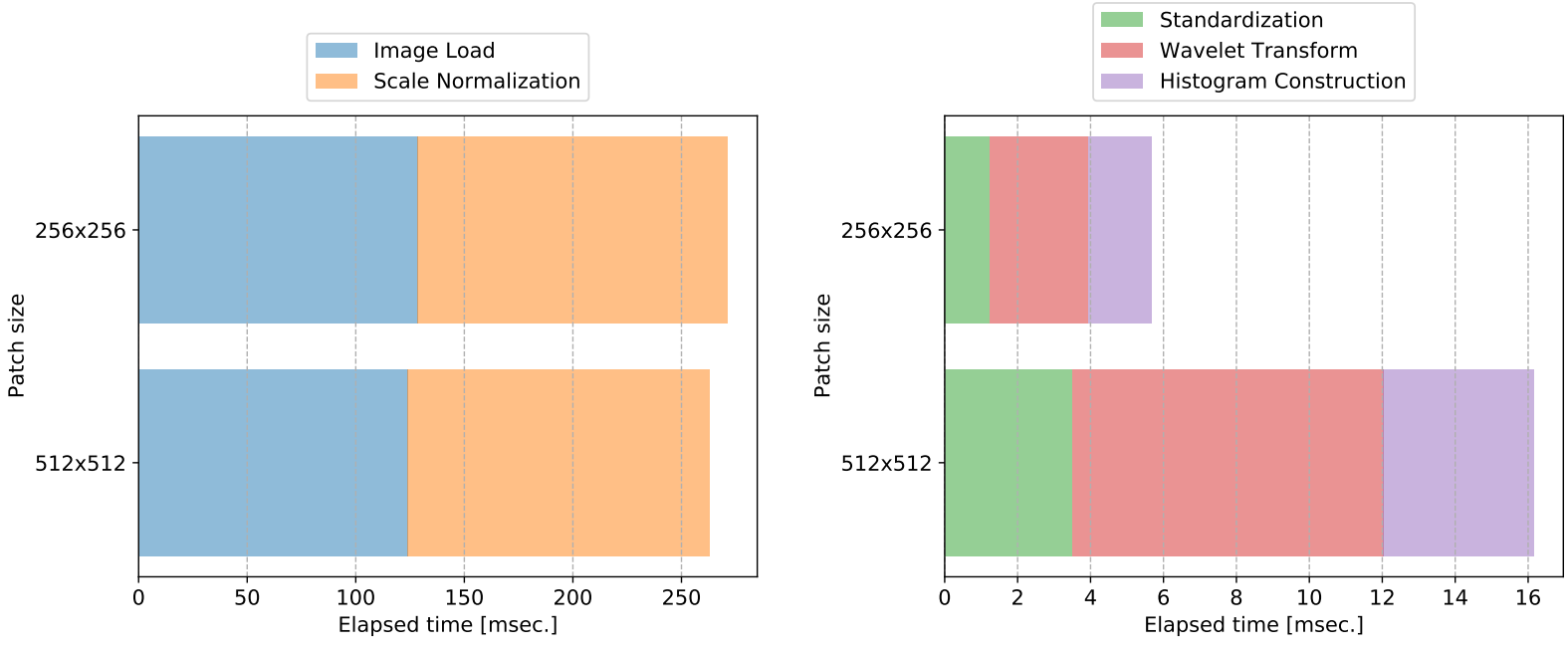


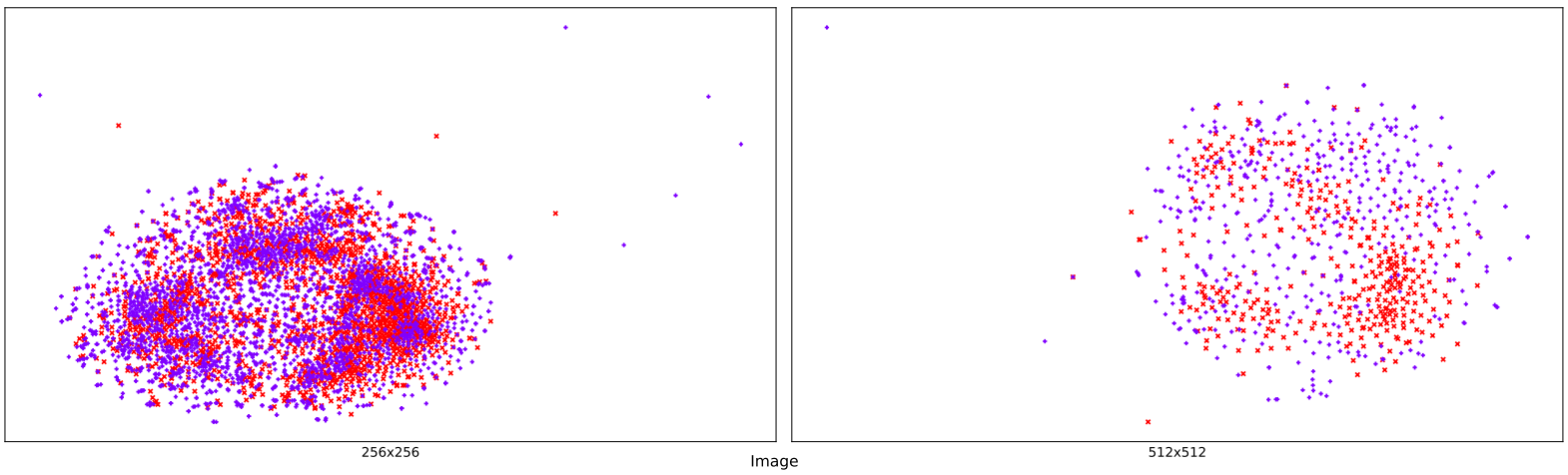
Figure 7: A sketch of K-fold cross validation. We first split all the data into training and test data. The training data is used only for maximizing the classification performance of the SVM while the test data is used only for evaluating the classification performance of the trained SVM. The training data is further splitted into K folds. For each fold, the fold is used as validation data while the remaining folds are used as training data. After all the folds are used, the average of all the evaluation metrics among the folds is used as the evaluation metric.



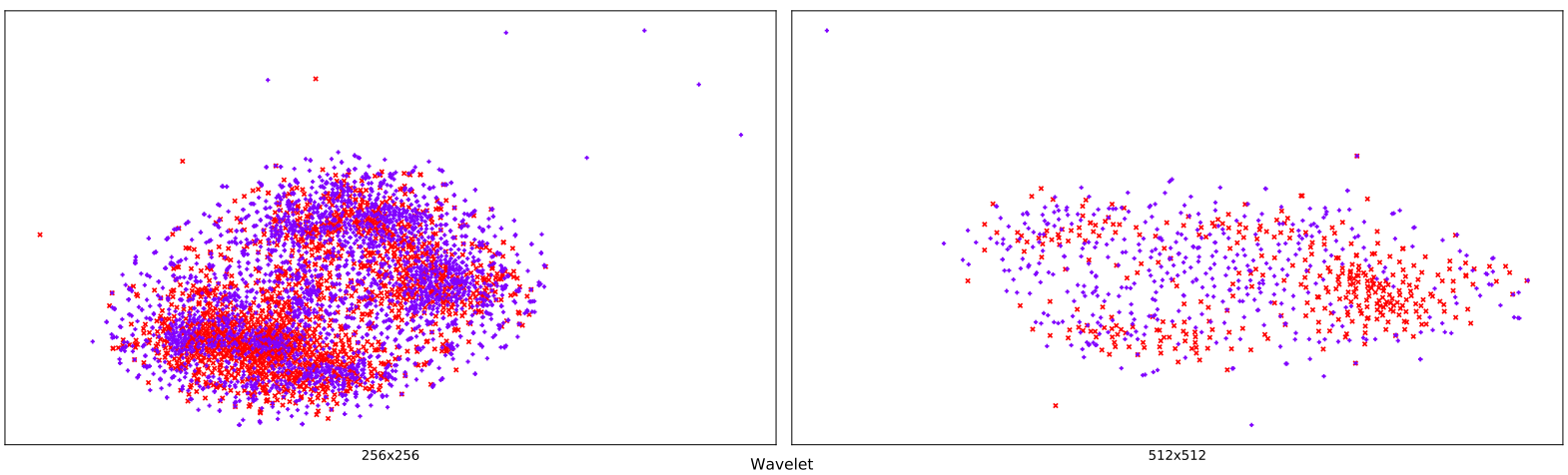
(a) Pre-processing

(b) Feature extraction

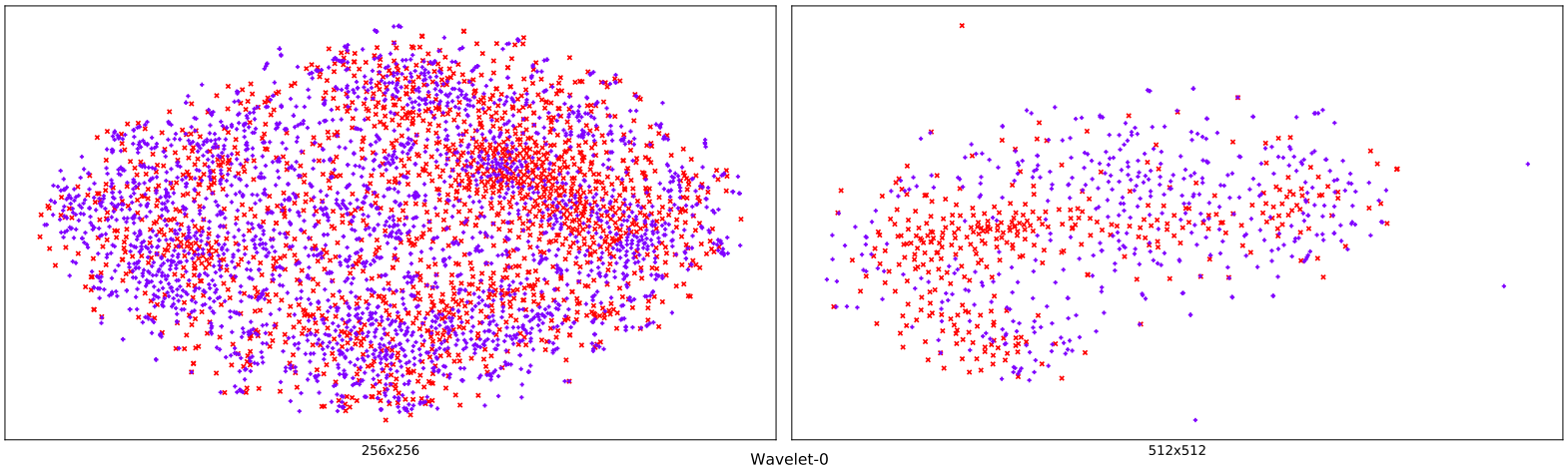
Figure 8: The elapsed time of the image processing steps [msec.].



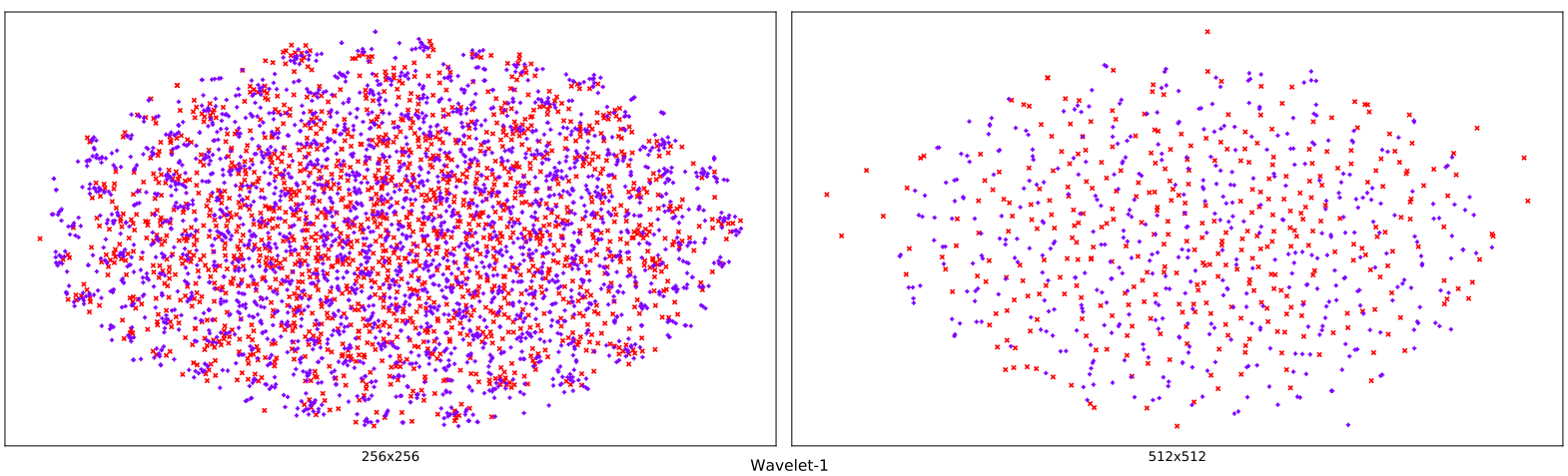
(a) *Image*



(b) Combined *Wavelet*

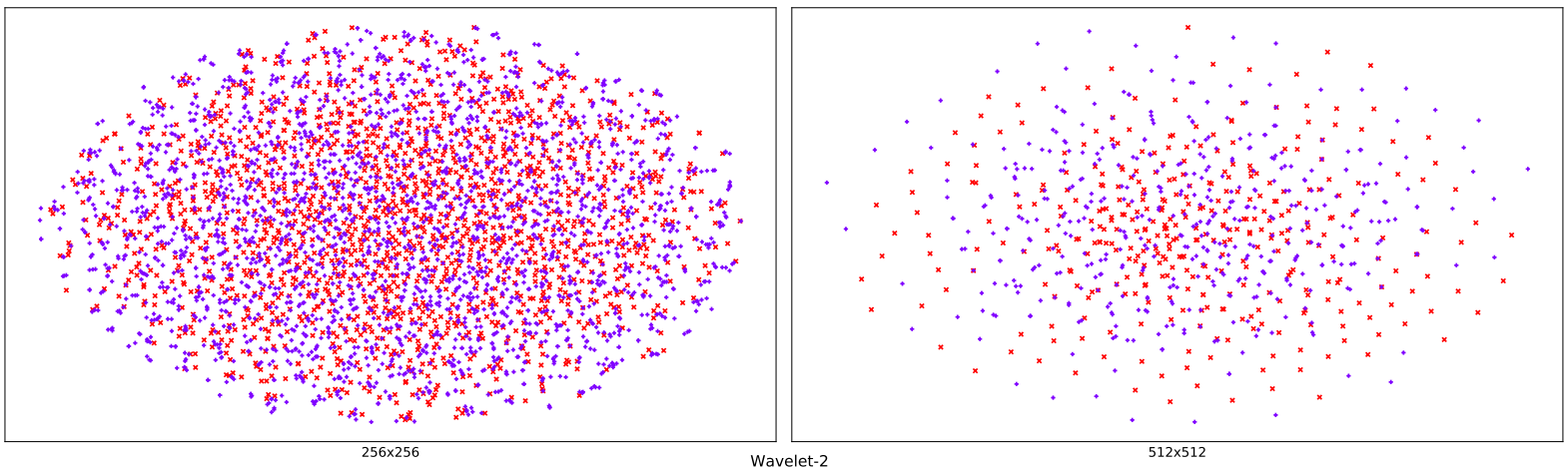


(c) *Wavelet*-0

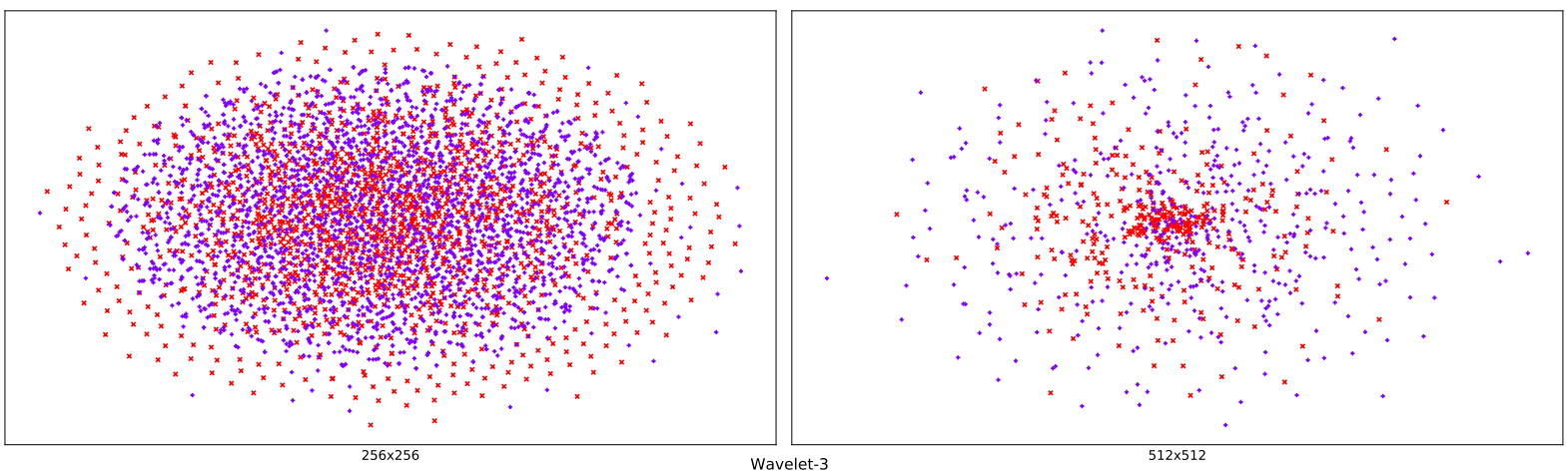


(d) *Wavelet*-1

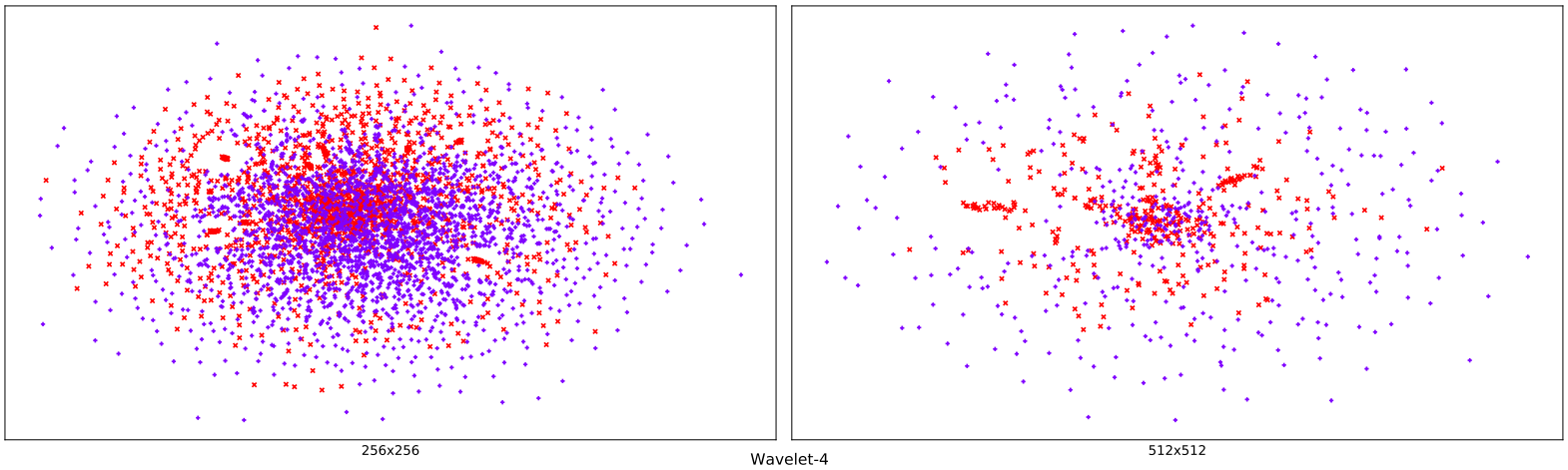
Figure 9: t-SNE 2D embedding of each image feature. Each red **x** and purple + symbol represents the feature of a negative and positive patch respectively.



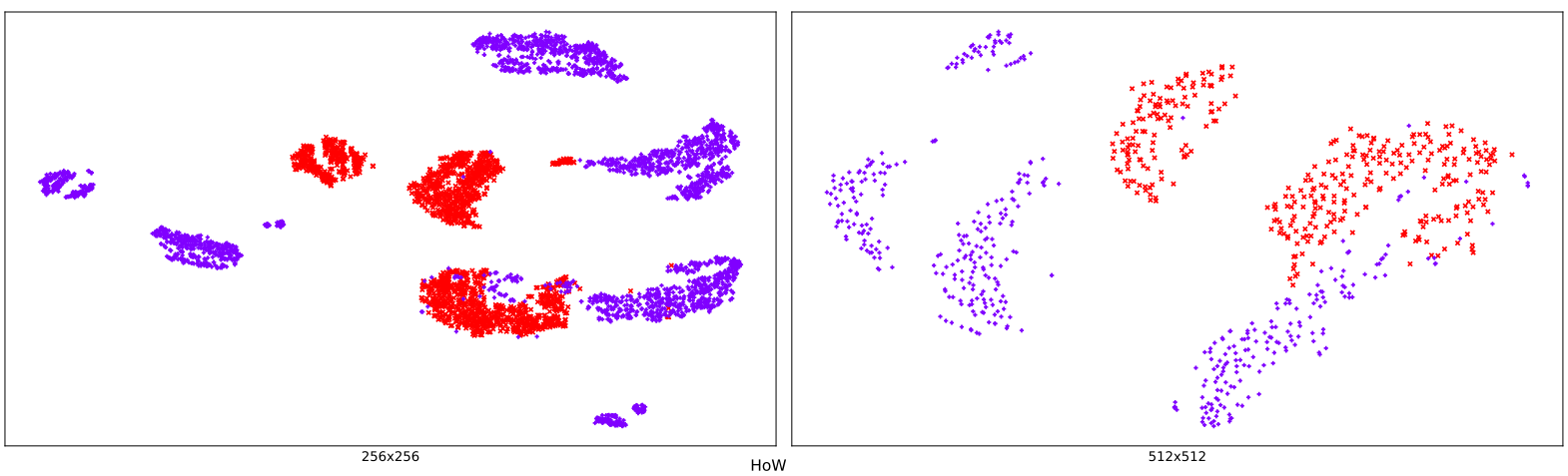
(e) *Wavelet*-2



(f) *Wavelet*-3

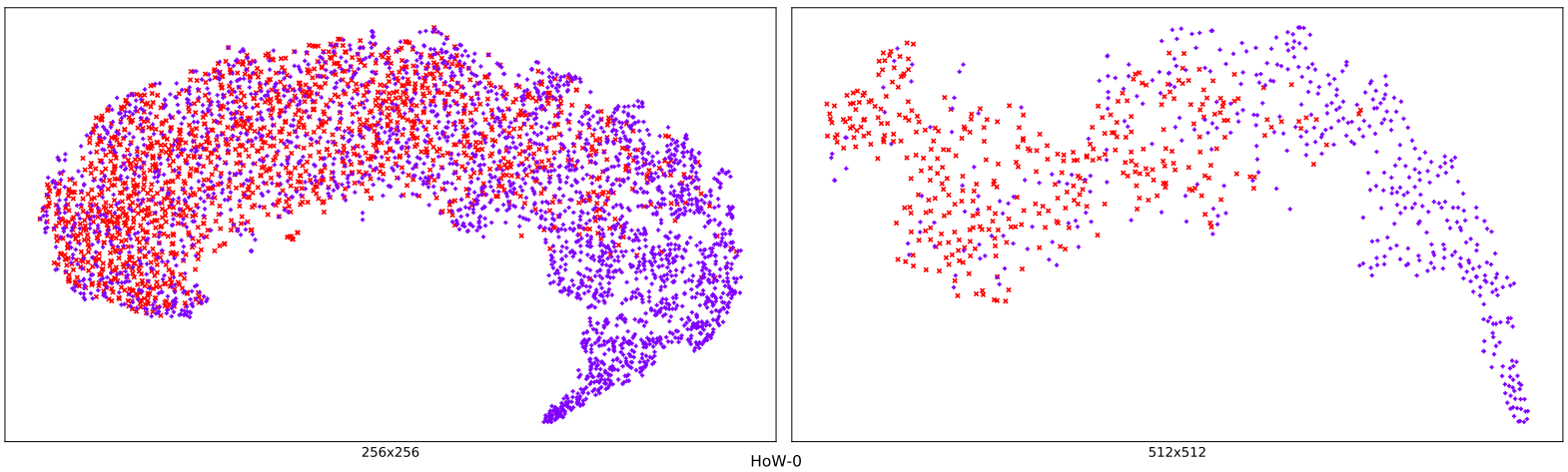


(g) *Wavelet*-4

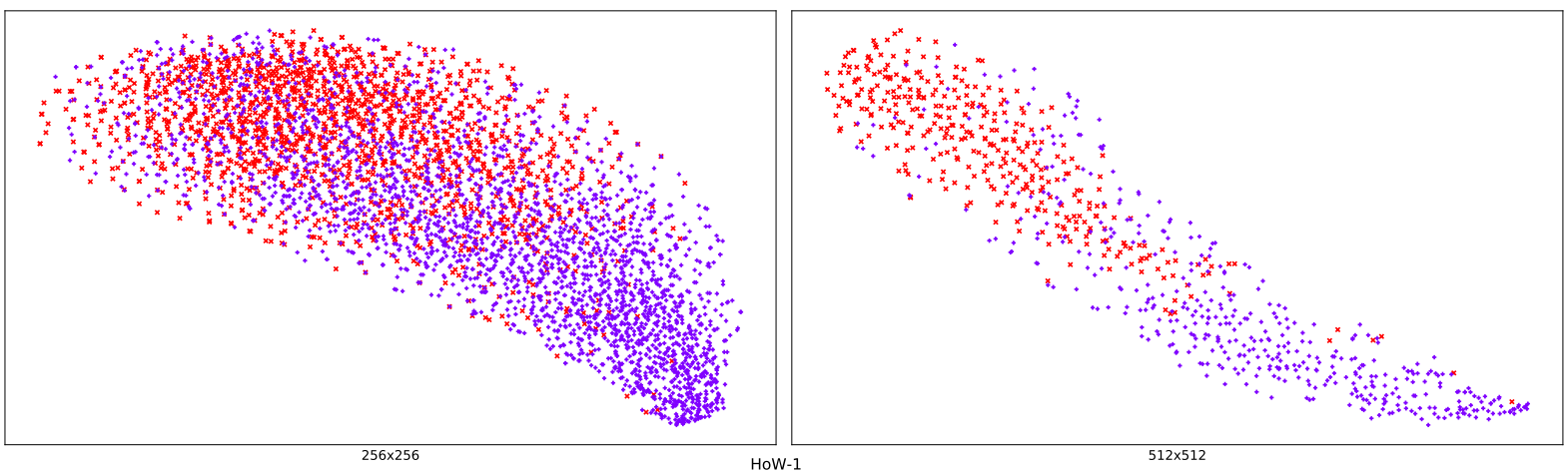


(h) Combined *HoW*

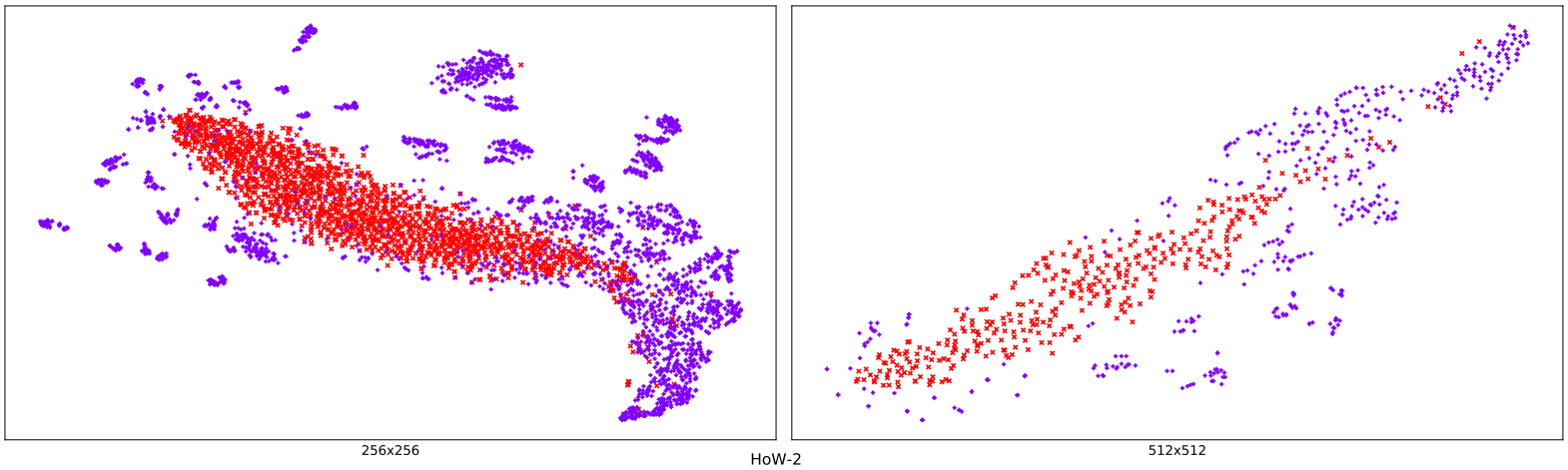
Figure 9: t-SNE 2D embedding of each image feature. Each red **x** and purple + symbol represents the feature of a negative and positive patch respectively.



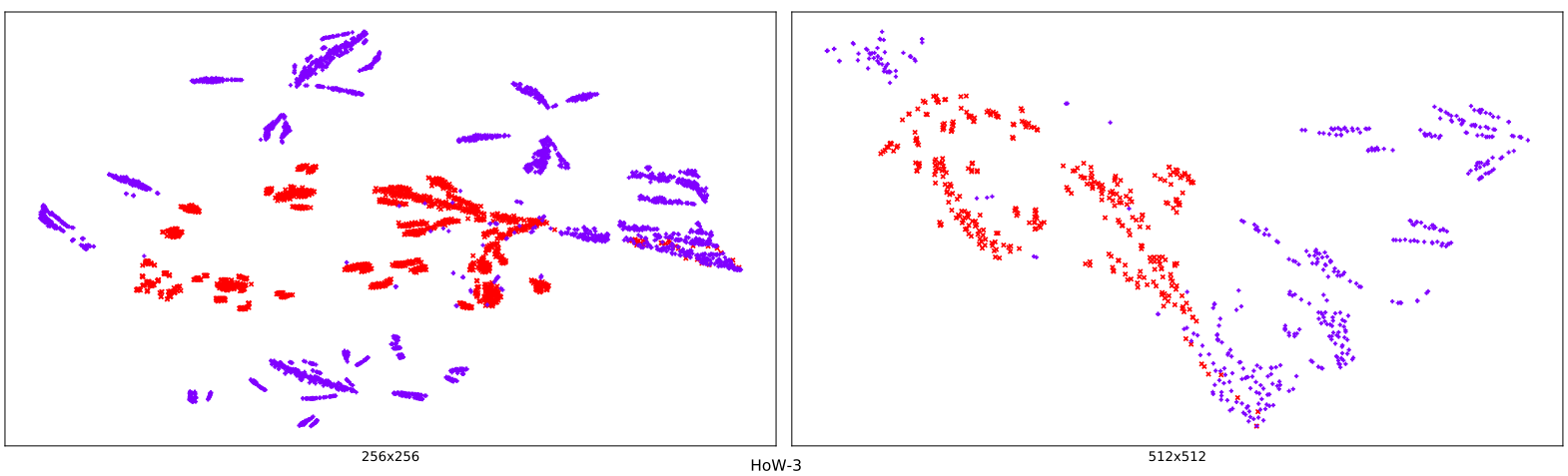
(i) HoW-0



(j) HoW-1

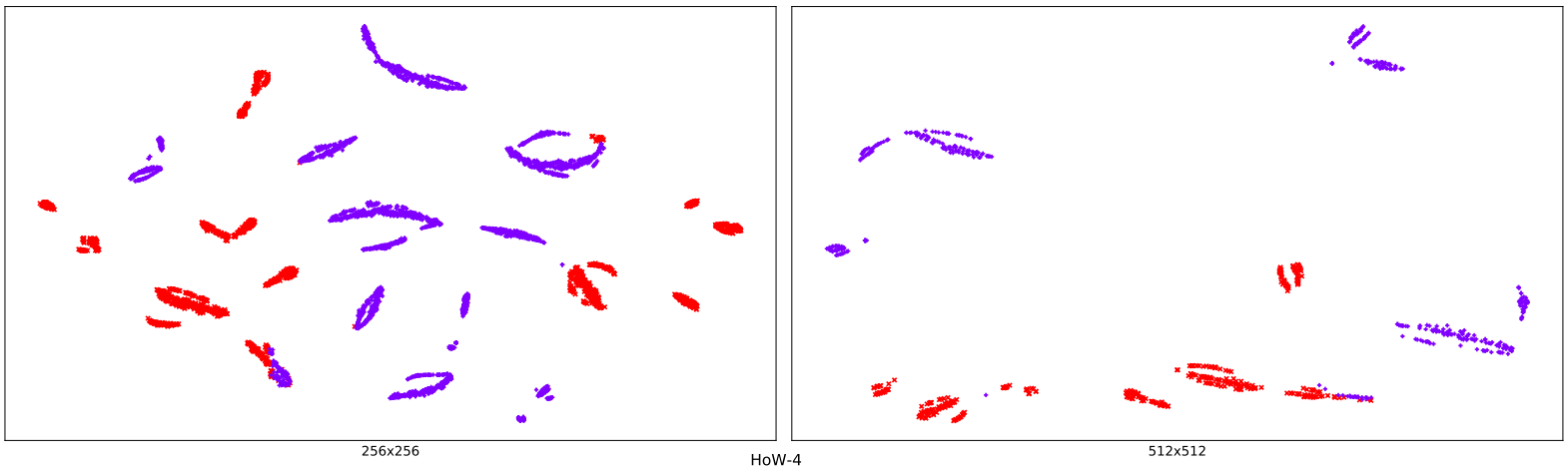


(k) HoW-2



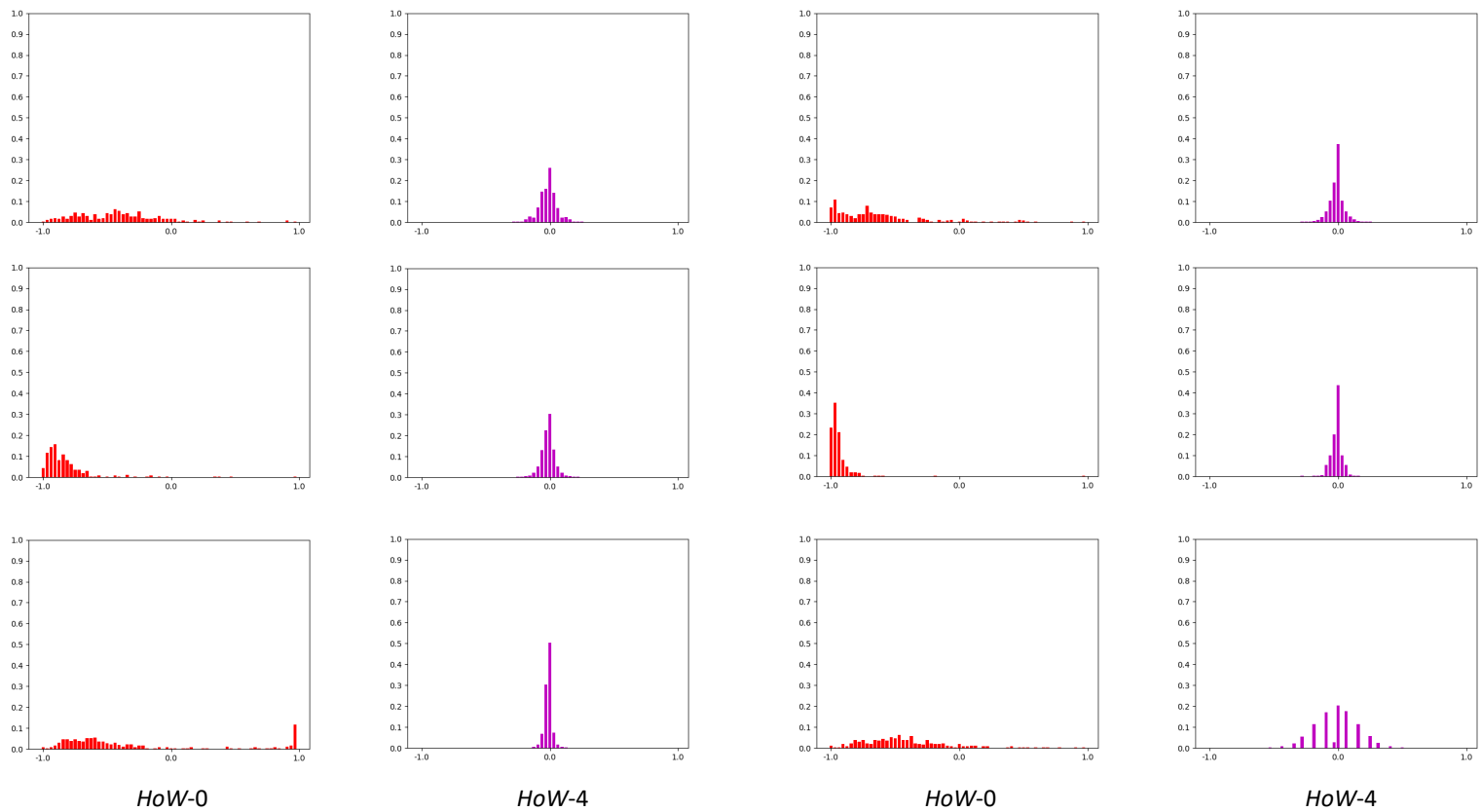
(l) HoW-3

Figure 9: t-SNE 2D embedding of each image feature. Each red x and purple + symbol represents the feature of a negative and positive patch respectively.



(m) HoW-4

Figure 9: t-SNE 2D embedding of each image feature. Each red **x** and purple + symbol represents the feature of a negative and positive patch respectively.



(a) Negative data

(b) Positive data

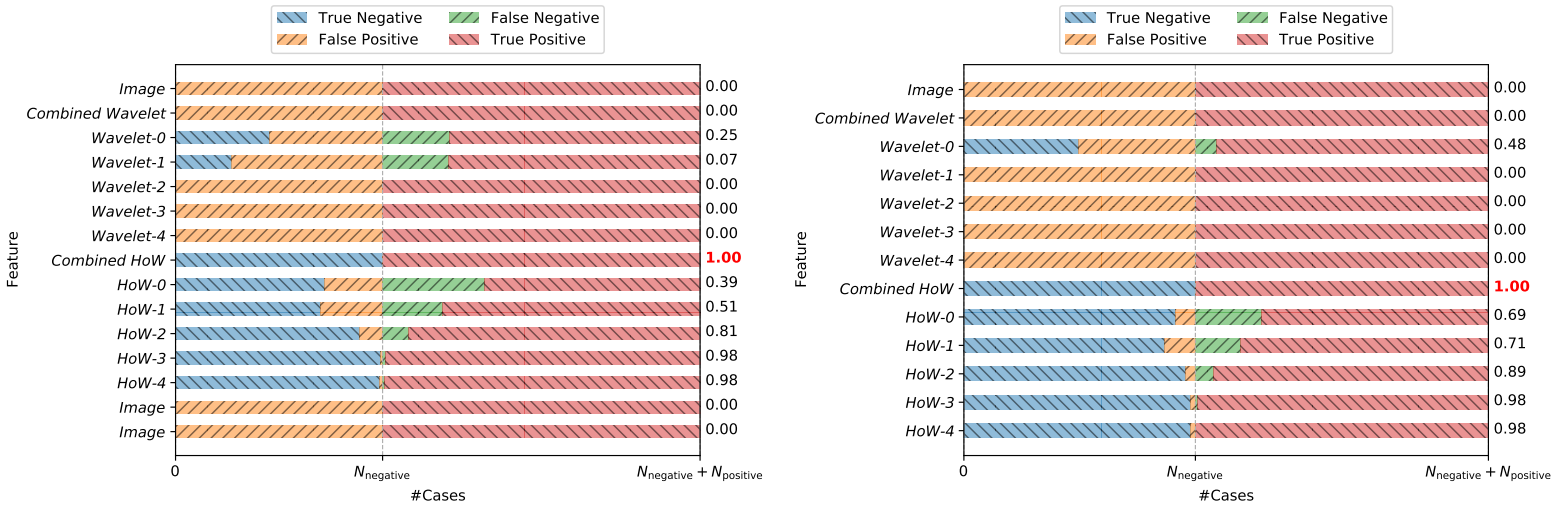
Figure 10: Some HoW-0 and HoW-4 features from the training data of patch size 256×256. The shape of HoW-0 varies even in same class while HoW has similar shape.

Table 5: The elapsed time for training per dataset [sec.].

	256×256	512×512
Image	6.2×10^{3}	1.2×10^{3}
Combined *Wavelet*	7.0×10^{3}	1.2×10^{3}
Wavelet-0	3.2×10	3.5
Wavelet-1	8.5×10	1.0×10
Wavelet-2	3.0×10^{2}	4.0×10
Wavelet-3	1.2×10^{3}	1.6×10^{2}
Wavelet-4	4.6×10^{3}	7.0×10^{2}
Combined HoW	1.0×10	6.7×10^{-1}
HoW-0	5.9	3.4×10^{-1}
HoW-1	5.5	3.2×10^{-1}
HoW-2	5.4	3.6×10^{-1}
HoW-3	3.4	2.9×10^{-1}
HoW-4	3.0	2.7×10^{-1}

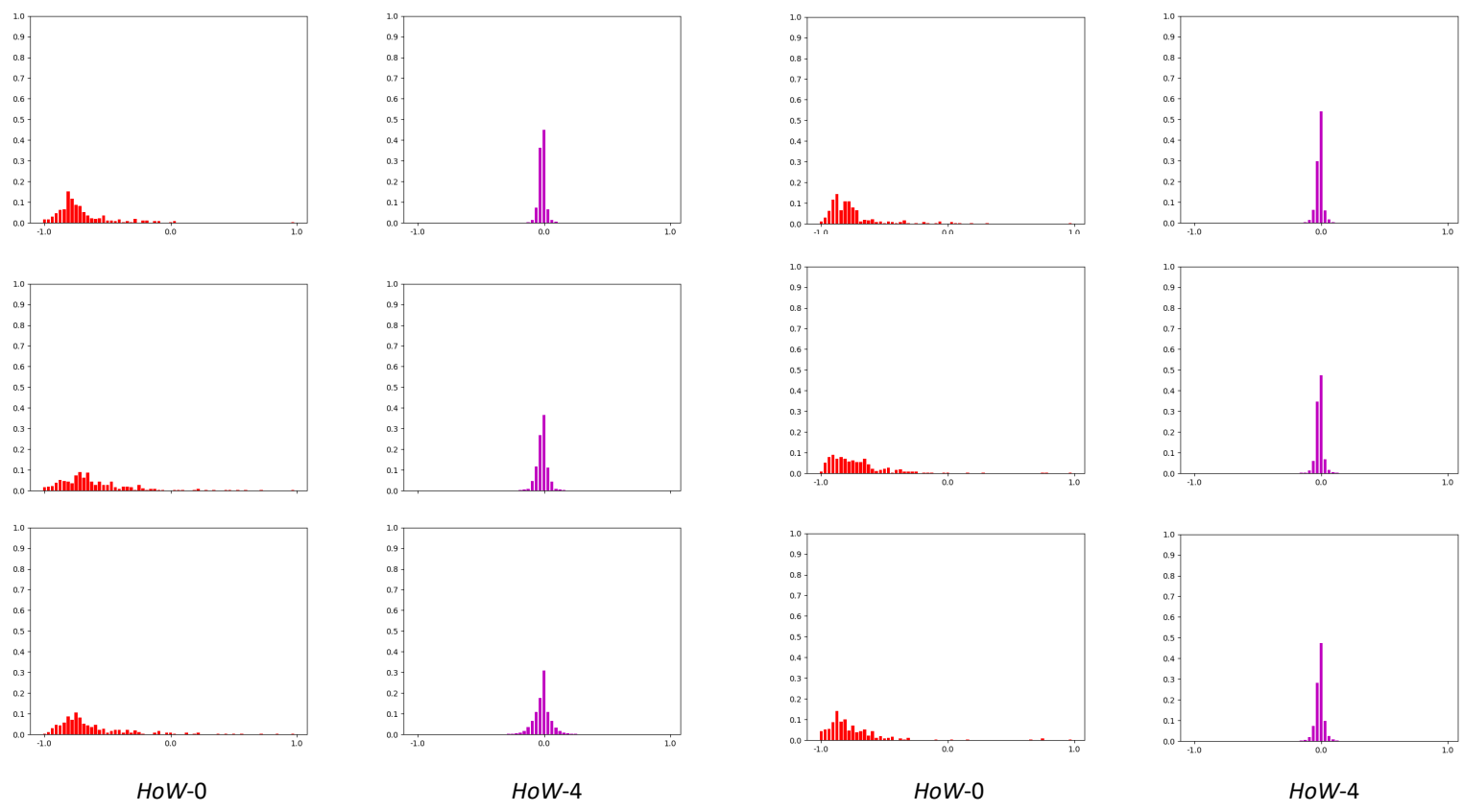
Table 6: The elapsed time for test per patch [msec.].

	256×256	512×512
Image	7.9×10^{2}	1.5×10^{2}
Combined *Wavelet*	8.6×10^{2}	1.5×10^{2}
Wavelet-0	2.7	4.9×10^{-1}
Wavelet-1	9.2	1.6×10^{0}
Wavelet-2	3.5×10	6.5
Wavelet-3	1.4×10^{2}	2.6×10
Wavelet-4	5.5×10^{2}	1.1×10^{2}
Combined HoW	1.3×10^{-1}	1.6×10^{-2}
HoW-0	5.5×10^{-1}	1.9×10^{-2}
HoW-1	4.5×10^{-1}	1.6×10^{-2}
HoW-2	3.2×10^{-1}	1.6×10^{-2}
HoW-3	5.0×10^{-2}	4.0×10^{-3}
HoW-4	2.9×10^{-2}	5.4×10^{-3}



(a) 256×256 (b) 512×512

Figure 11: Visualized confusion matrix and MCC for each feature of different patch size. Horizontal bars represent the number of True Negative, False Positive, False Negative, and True Positive cases respectively. The variables N_{negative} and N_{positive} denote the number of negative and positive data respectively. Numerical values written on the right is the MCC value of each feature. The best MCC score is emphasized by red and bold font.



(a) False negative cases (b) False positive cases

Figure 12: Some HoW-0 and HoW-4 features from the failed cases of patch size 256×256.