Towards Artifact Rejection in Microscopic Urinalysis

Tarun Dutt
IIIT Bangalore
Bengaluru, KA 560100, India
Tarun.Dutt@iiitb.org

G.N. Srinivasa Prasanna
IIIT Bangalore
Bengaluru, KA 560100, India
gnsprasanna@iiitb.ac.in

Tathagato Rai Dastidar
SigTuple Technologies,
Bengaluru, KA 560102, India
trd@sigtuple.com

Ananth Shreekumar
IIIT Bangalore
Bengaluru, KA 560100, India
Ananth.Shreekumar@iiitb.org

Abstract

Real-world deep learning systems often have to deal with unseen categories during prediction. The closed set nature of these systems force it to choose only from known categories. In building robust automated urine sample analyzers, multiple artifacts present in urine microscopy images need to be identified and rejected by these systems during test time, while correctly classifying clinically significant objects. We tackle this problem by looking at techniques that adapt existing deep neural network architectures to support identification and rejection of artifacts belonging to unseen/unknown classes. We propose two methods that show **6%**-**10%** better performance in rejecting artifacts compared to existing work in open set recognition and out-of-distribution object detection, for convolutional neural network (CNN) models trained on a microscopic urine sample image data-set.

1 Introduction

Clinicians prescribe urine analysis for various diagnosis. Urinalysis involves a microscopic examination of a urine sample to detect and classify clinically significant objects in the sample, leading to a diagnosis. Objects found include red blood cells (RBCs) or white blood cells (WBCs), epithelial-casts, crystals, microorganisms like bacteria, yeast, and artifacts [16]. Artifacts in urine samples occur due to a variety of reasons, ranging from fibers and dust, to blurring of a sample by an out-of-focus object, and lens artifacts [1]. Objects are detected and extracted as image patches from a large field-of-view (FOV) image captured through a microscope lens. [1] use an unsupervised thresholding method with a U-net to extract clinically significant objects. Sometimes, artifacts are extracted as well due to close resemblance to known class objects. A learning based automated classifier must be able to identify and reject artifact objects coming from an unbounded set of unknown classes, while correctly classifying clinically significant objects from known classes. The accuracy of such a system relies on an exhaustive knowledge of all potential object classes and is constrained by the nature of the training dataset. We propose and evaluate two methods against Openmax [2] from the literature on open set recognition [11, 12, 13] and ODIN [8] from work on Out-Of-Distribution (OOD) [5, 7] object detection. We report results for models trained on the extracted microscopic urine sample images using these techniques.

2 Methodologies

Openmax Bendale and Boult [2] replace the softmax layer with Openmax. A deep neural network is first trained with the Softmax layer by minimizing cross-entropy loss. Using the Nearest Class

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

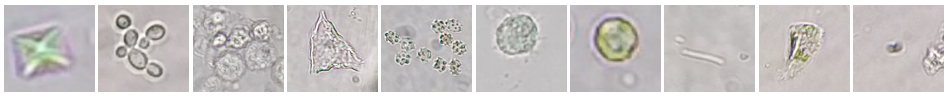

Figure 1: Crystal, yeast, WBC clumps, epithelial cast, RBC clumps, WBC, RBC, bacteria and two artifact objects present in urine samples

Mean described in [9, 10], each class is represented as a mean activation vector (MAV) computed using the mean of the activation vectors (only for the correctly classified training samples) in the penultimate layer of the network. A Weibull distribution is fit for each class using the distances of training samples from their corresponding class MAVs. The activation vector's values are then re-calibrated according to the Weibull, and then used to compute a pseudo-activation for an unknown class. Class probabilities for known and unknown classes are computed by using softmax on the re-calibrated activation vector. Samples are rejected based on a threshold over the softmax values.

ODIN This method increases the difference between the maximum softmax scores of in-distribution and OOD samples by (i) calibrating softmax scores by scaling the logits by a large constant (referred to as temperature) and (ii) pre-processing the input by perturbing it with the loss gradient. ODIN [8] demonstrated that at high temperature values, the softmax score for the predicted class is proportional to the relative difference between the largest un-normalized output (logit) and remaining outputs (logits). Temperature scaling pushes the softmax scores of in- and OOD images further apart when compared to plain softmax. Also, perturbing an input image through gradient ascent w.r.t to the score of its predicted label was shown [8] to further push apart the softmax scores of in- and OOD images.

Nearest Neighbor (NN) Using a CNN model, the feature vector before the softmax layer, known as logits, are extracted for all correctly classified training samples from known classes as well as artifact samples in the training set. Each known class' cluster and artifact cluster is further broken into one or more clusters. For a test sample, the cluster center neighbor that is closest is identified using the L2 norm and labeled as the class associated with that cluster center. With a single cluster for each known class, 88% of artifact samples in the test set were closest to bacteria, crystal, epithelial cast, and yeast classes. This suggests ubiquitous presence of artifacts through multiple known class clusters in feature space, making it harder to discriminate between known class samples and artifacts.

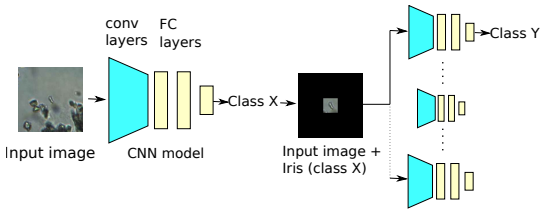

Figure 2: The proposed two stage classifier with iris. Given a test sample, labels from both classifiers are used to determine the final classification label. If X and Y differ in labels, it is rejected as an artifact/unknown class sample, else it is classified as the class corresponding to that label

Iris We propose a two-stage classifier, with the second stage acting as a re-verification of the first stage classification, by localizing objects in the images, using a mechanism called the Iris. Multiple objects may be present in each image, but labels are based on the object at the center. A first stage CNN model is trained on samples from known classes. Most objects in the samples occur within a small fixed size range. A square bounding box (Iris) size is obtained for each known class using the first stage classifier, by linearly iterating over sizes and maximizing the number of rejections while not more than 5% of positive class samples classified correctly are now misclassified or rejected due to the Iris. The second stage contains binary models, one for each known class. Images are pre-processed with the Iris size corresponding to the class label from the first stage. Each binary model is trained as one-vs-rest using images with Iris. The mechanism to handle a test sample is outlined in Figure. 2. In most samples, artifact objects are surrounded by other objects, which could belong to a known class. A first stage classifier trained on known classes is biased to surrounding objects than an object never encountered before, affecting final classification confidence on the central

object. Using the Iris in the second stage eliminates some surrounding objects. With a very large number of training samples, in principle, the CNN model should learn to localize the object based on the label, during training.

3 Experiments and Results

Dataset and Training The urine microscopic image data-set has 25600 training images and 8510 test images of known class and artifact objects extracted from FOVs [1]. There are 8 clinically significant classes of objects, shown in Figure. 1. Images have equal dimensions and are labeled based on the object at the center. The train and test sets have 9089 and 1000 artifact samples from unknown classes, respectively. The dimensions of images in the training set vary from 28×28 pixels for crystal image samples to 2880×2880 for epithelial-cast samples, while all test images are of size 512×512. All images are pre-processed with center cropping or padding to 224×224. The original architecture of VGG-16 [14] is modified, with an extra convolutional layer with 64 filters in the first block and 128 filters in the second block, two fully connected (FC) layers with 512 neurons, and 8 labels in the ouput layer. The model is trained from scratch on eight known classes, with 0.7 dropout[15] between the FC layers, using a categorical cross-entropy loss [4]. When tested on images from known classes, the model achieved an accuracy of 91%. Artifact test samples are run through the model to check performance of softmax with a threshold, similar to openmax. The techniques in [2, 8] are applied to the model described. A model trained and tested on known class and artifact samples drops to 74% in accuracy due to the spread of artifact samples, as outlined in the nearest neighbor (NN) section, leading to poor generalization. For the Iris, four binary models are trained with convolutional layers similar to the first-stage model, but with FC layers having 4096 neurons with 2 labels in the output layer. Image augmentation is used to balance data for training the binary models. The iris sizes obtained using the first-stage model are 40×40, 70×70, 70×70 and 220×220 for bacteria, crystal, yeast, and epithelial cast classes respectively. Including binary models for other classes affected accuracy on known classes without significantly increasing the rejection of artifact samples. All models were implemented over a Keras [3] framework in a Python environment on a machine with Intel Xeon, 324GB RAM, and two NVIDIA V100 32GB GPUs.

Evaluation The Normalized accuracy (NA) uses accuracy on known samples (AKS) and accuracy in rejecting unknown samples or artifacts (AUS). It is defined as $NA = \lambda_r AKS + (1 - \lambda_r)AUS$, where $\lambda_r, 0 < \lambda_r < 1$, regulates trade-off of mistakes between known and unknown samples [6]. λ_r is chosen as the ratio of known class samples to the total number of samples in the training set. We define an efficiency metric, as the ratio of number of test artifact samples rejected to the number of known class test samples rejected or misclassified, for a fixed AKS. For AKS=85%, the number of test samples rejected or misclassified from known class samples is $(91\% - 85\%) * 7510 \approx 450$.

Table 1: Results from the comparison of all methods

Method	AUS (%)(AKS=85%)	NA (%)($\lambda_r = 0.64$)	Efficiency (AKS=85%)
Softmax	43.0	69.88	0.954
Openmax [2]	49.1	72.04	1.089
ODIN [8]	46.2	70.96	1.025
NN	56.0	74.60	1.242
Iris	**60.0**	**76.0**	**1.331**

4 Conclusion

We demonstrate two methods on a microscopic urine sample data-set, that is robust to artifacts. Using the Iris improves rejection rate of artifacts by 10% more than techniques explored from the literature, for the same accuracy among known classes. In general, an Iris can handle samples belonging to an unbounded set of unknown classes in multi-class classification scenarios. Future work would involve (1) extending the Iris method to incorporate Openmax and experiment with temperature and input perturbation in both stages of the classifier, (2) using convex hull approximations for defining clusters in feature space as a pre-processing step to perform nearest neighbor distance checks.

Acknowledgments

We are grateful for support from Sigtuple Technologies Pvt. Ltd., and the Machine Intelligence and Robotics Laborotary (MINRO) at the International Institute of Information Technology, Bangalore.

References

[1] Abdul Aziz, Harshit Pande, Bharath Cheluvaraju, and Tathagato Rai Dastidar. Improved extraction of objects from urine microscopy images with unsupervised thresholding and supervised u-net techniques. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2230–2238, 2018.

[2] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016.

[3] François Chollet et al. Keras, 2015.

[4] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[5] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.

[6] Pedro R Mendes Júnior, Roberto M de Souza, Rafael de O Werneck, Bernardo V Stein, Daniel V Pazinato, Waldir R de Almeida, Otávio AB Penatti, Ricardo da S Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 106(3):359–386, 2017.

[7] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. *arXiv preprint arXiv:1711.09325*, 2017.

[8] Shiyu Liang, Yixuan Li, and R Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.

[9] Thomas Mensink, Jakob Verbeek, Florent Perronnin, and Gabriela Csurka. Distance-based image classification: Generalizing to new classes at near-zero cost. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2624–2637, 2013.

[10] Marko Ristin, Matthieu Guillaumin, Juergen Gall, and Luc Van Gool. Incremental learning of ncm forests for large-scale image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3654–3661, 2014.

[11] Ethan M Rudd, Lalit P Jain, Walter J Scheirer, and Terrance E Boult. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):762–768, 2017.

[12] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7): 1757–1772, 2012.

[13] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2317–2324, 2014.

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

[15] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[16] S. K. Strasinger and L. M. S. Di. *Urinalysis and body fluids 5th Edition*. F. A. Davis Company, 2008.