A study of Teacher-Student anomaly detection in images

1st Chien Nguyen Hong
FPT Software
Quy Nhon, Vietnam
ChienNH2@fsoft.com.vn

2nd Nho Nguyen
FPT Software
Quy Nhon, Vietnam
NhoN1@fsoft.com.vn

***Abstract*—We study a simple teacher-student learning approach for the challenging problem of unsupervised anomaly detection. Teacher network is the pre-trained Resnet18, and the student network is proposed by us. Student networks are only trained on anomaly-free images to reconstruct the output of the teacher network. Anomalies are detected when the outputs of the student networks differ from that of the teacher network. Our experiments demonstrate our student network is good for this task.**

***Index Terms*—Anomaly detection, teacher-student, deep learning**

I. Introduction

The goal of anomaly detection is to identify observations in a dataset that significantly deviate from the remaining observations. Since anomalies are rare and naturally diverse, it is not feasible to obtain a labeled dataset representative of all possible anomalies. A successful approach for anomaly detection is to learn a model of the normal class, under the assumption that the training data consists entirely of normal observations. If an observation deviates from that learned model, it is classified as an anomaly.

We take the idea from the paper [1]. In this paper, we review the teacher-student learning, and propose a simpler approach for solving the anomaly detection in images. In some data types, this method achieves high accuracy.

Our code is available at https://github.com/nhchien93/student-teacher-abnormaly-detection.

II. Related work

We summarize some approach for unsupervised task.

A. Auto Encoder model

Auto Encode (AE) contains an encoder net and a decoder net. This network is trained on free-anomaly images. The objective is to reconstruct anomaly-free images to anomaly-free images. The reconstruction loss functions could be $mse, mae, ssim$.

B. Generative model

Variational Autoencoder (VAE) is an auto-encoder whose encoding distribution is regularised during the training in order to ensure that its latent space has good properties allowing us to generate new data. VAE uses KL loss, and this is a different point from traditional AE.

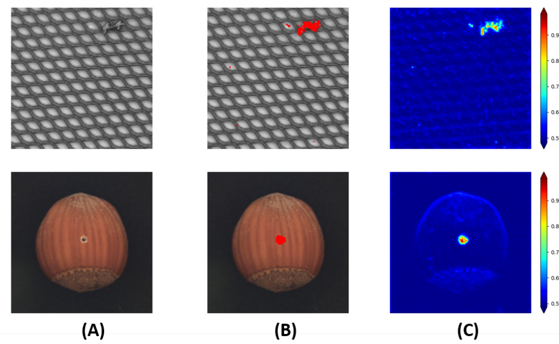

Fig. 1. Some anomaly detection result from our work. (A) Original images, (B) images with highlight anomaly region and (C) anomaly score map.

Generative adversarial network contains two networks. The generative network learns to map from a latent space to a data distribution of interest, while the discriminative network distinguishes candidates produced by the generator from the true data distribution. Some experiments demonstrate that GAN is a good choice for the anomaly detection task.

C. One class classification model

Deep features are first extracted from a pre-trained deep model for given training images. Classification is done using a one-class classification method such as one-class SVM, k-nearest neighbor using extracted features. A small fraction of data points are allowed to lie on the other side of the decision boundary: those data points are considered outliers.

III. Proposed Methods

We propose to apply teacher-student learning.

A. Teacher-student learning

The Teacher-student framework contains teacher and student networks. The Teacher network is a pre-trained model which is trained on a large data set. The Student net is trained by using knowledge distillation loss on a new data set. The objective is to gain the cognition of the teacher on anomaly-free data. On abnormal data, the cognitive of teacher net and student net will be different. Then the abnormal data point could be identified.

We can assume that the input image $I \in \mathbb{R}^{3\times n\times n}$, T is the teacher net, the $S_i, i = 1, \ldots, s$, are the student nets and deep features $T(I), S_i(I) \in \mathbb{R}^{d\times k\times k}$.

B. Knowledge distillation loss

Knowledge is transferred from the teacher model to the student by minimizing a MSE loss function, aimed at matching feature outs of teacher net as well as ground-truth labels.

$$loss = \frac{1}{d \times k \times k} \sum_{i=1}^{s} (T(I) - (S_i(I)))^2 \in \mathbb{R}. \tag{1}$$

C. Scoring function

We use $l2{-}distance$ to calculate distance between teacher's feature and student's features

$$distanceMatrix = \frac{1}{s \times d} \sum_{i=1}^{s} \sum_{j=1}^{d} (T(I) - (S_i(I)))^2 \in \mathbb{R}^{k\times k}. \tag{2}$$

Then we normalize $distanceMatrix$ in the range $[0, 1]$.

$$distanceMatrix = \frac{distanceMatrix}{max - min}, \tag{3}$$

where max and min are the maximum and minimum value of $distanceMatrix$.

D. Evaluation metric

Receiver operating Characteristics (RoC) is a plot of False positive rate (FPR) against True positive Rate (TPR) of binary classifiers. FPR is defined as

$$FPR = \frac{FP}{FP + TN}, \tag{4}$$

where FP is False positive and TN is True negative. FPR corresponds to the proportion of negative data points mistakenly predicted positive to all negative data points. TPR, also called *Recall*, is defined as

$$TPR = \frac{TP}{TP + FN}, \tag{5}$$

where TP is True positive and FN is False negative. TPR corresponds to the proportion of positive data points that are correctly predicted positive to all positive data points. At a special value called *threshold* we have one FPR - TPR pair values. So the closer the RoC is to top-left border, the better the quality of predictions by the prediction model (Fig. 2).

In our work, the evaluation is on the pixel level, the positive class is the abnormal, the negative class is the normal. We use Area under RoC Curve (AuC) metric to evaluate models. AuC is a measure of how well a binary classifier can perform predictions of labels. The AuC of a classifier is equal to the probability the classifier will rank a randomly chosen positive record higher than a randomly chosen negative one. If the AuC value is equal to 1, we have a perfect classifier.

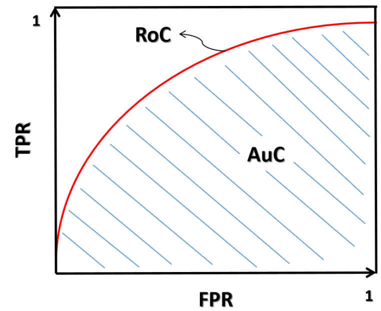

Fig. 2. RoC curve and AuC.

E. Anomaly visualization

We use $distanceMatrix$ to highlight pixel that have value higher than a given threshold p.

IV. Experiment

A. Data set description

We applied our proposed anomaly detection approach to MVTec Anomaly Detection (MVTec AD) [2]. The data set contains over 5000 high-resolution images divided into fifteen object and texture categories such as grid, bottle, leather,... Each category consists of a set of defect-free training data and a test set of images with various kinds of defects as well as images without defect. All anomalies were precisely annotated pixel by pixel. Some object images without anomaly with their anomaly and ground-trust were shown in Fig. 3.

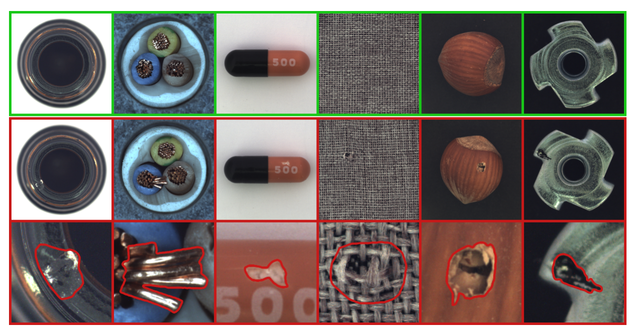

Fig. 3. The MVTect data set [2].

B. Teacher network

In this study, we used the RestNet18 pre-trained model without last five layers to extract feature from training images. The architecture of the model was shown in Fig. 4.

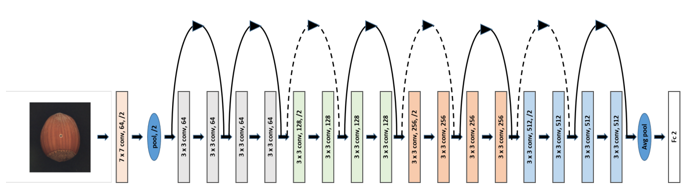

Fig. 4. The ResNet18 CNN architecture.

RestNet18 is a convolutional neural network that is trained on more than a million images from the ImageNet database [3]. There are 18 layers present in its architecture. This neural network is very useful and efficient in images classification.

C. Student network

For student networks, we proposed an ensemble containing $s = 3$ networks with the same architecture. Each student is a simple deep convolutional neural networks (CNNs) with only convolutional and max pooling layers. The architecture of the network is shown in Fig. 5.

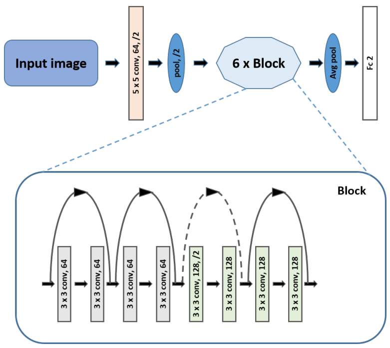

Fig. 5. The Student network architecture.

We build the student network based on Fig. 5. The network takes input $(3, 512, 512)$, after through the network we obtain features $(64, 128, 128)$.

D. Training details

We train on anomaly-free images for 100 epochs with batch size 1. For network optimization we use Adam with initial learning rate 10^{-3}. Input images are randomly horizontal flipped with probability of 0.5, resized to $(512, 512)$. It takes about thirty minutes to train one student network.

E. Evaluation and conclusion

We use AuC metric to evaluate the performance of the model, the results shown on TABLE I

TABLE I
EVALUATION REPORT

	Category	AUC
Texture	Carpet	0.86
	Grid	**0.96**
	Leather	0.91
	Tile	0.88
	Wood	0.89
Object	Bottle	0.74
	Cable	0.60
	Capsule	0.73
	Hazelnut	0.92
	Metal nut	0.77
	Pill	0.80
	Screw	**0.96**
	Toothbrush	0.91
	Transistor	0.57
	Zipper	0.91

The Teacher-Student framework has high performance on the texture data. However, our algorithm is not good with *Transistor* and *Cable* data categories.

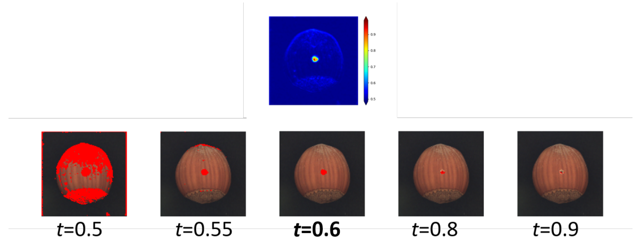

Fig. 6. Anomaly visualization with some thresholds t. *Red-highlight* is the detected anomaly regions.

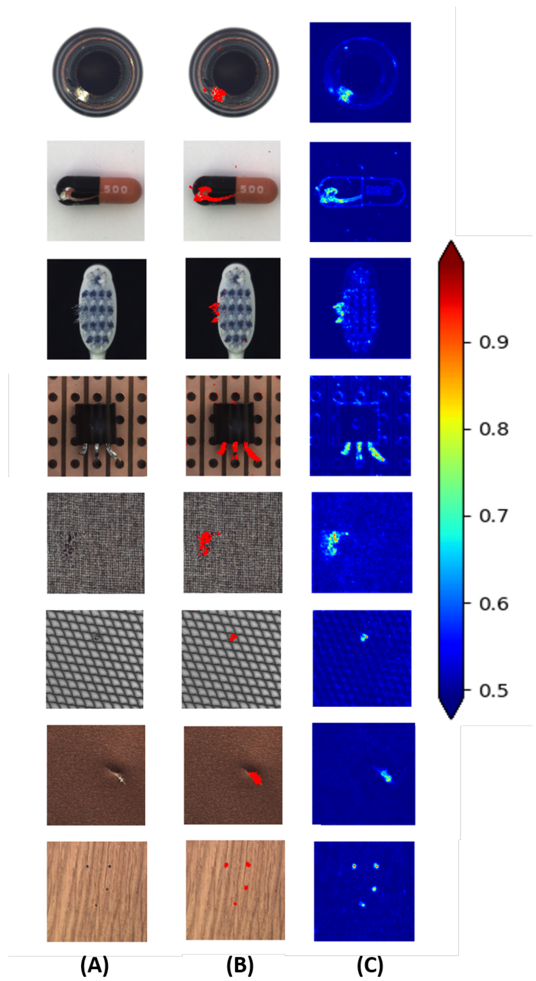

Fig. 7. Anomaly detection result from our method on MVTec AD data set. **Left column:** Input images containing defects. **Center column:** Images with highlight defect regions. **Right column:** Anomaly score map

ACKNOWLEDGMENT

We would like to express our very great appreciation to Dr. Hieu, Dr. Nguyen for their reviewing and constructive suggestions during the development of this research. We would also like to thank Mr. Tam for his advice that keeps motivated. Finally, we would like to thank QAI-team for their support and encouragement throughout our research.

REFERENCES

[1] Paul Bergmann, Michael Fauser, David Sattlegger, Carsten Steger. Uninformed Students: Student–Teacher Anomaly Detection with Discriminative Latent Embeddings; in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), March 2020

[2] Paul Bergmann, Michael Fauser, David Sattlegger, Carsten Steger. MVTec AD - A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection; in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition; in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Dec 2015