IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version August 21, 2024.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

ViGLAD: Vision Graph Neural Networks for Logical Anomaly Detection

FIRAS ZOGHLAMI[1], DENA BAZAZIAN[1], GIOVANNI MASALA[2], MARIO GIANNI[3] and ASIYA KHAN[1]

[1]School of Engineering, Computing and Mathematics, University of Plymouth, United Kingdom.
[2]School of Computing, University of Kent, United Kingdom.
[3]Department of Computer Science, University of Liverpool, United Kingdom.

Corresponding author: Firas Zoghlami (e-mail: firas.zoghlami@plymouth.ac.uk), ORCID: 0000-0001-9609-2568.

ABSTRACT Quality inspection is an industrial field with a growing interest in anomaly detection research. An anomaly in an image can either be structural or logical. While structural anomalies lie on the image objects, challenging logical anomalies are hidden in the global relations between the image components. The proposed approach, Vision Graph based Logical Anomaly Detection (ViGLAD), uses the graph representation of an image for logical anomaly detection. Defining an image as a structure of nodes and edges leverages new possibilities for detecting hidden logical anomalies by introducing vision graph autoencoders. Our experiments on public datasets show that using vision graphs enhances the performance of state-of-the-art teacher-student-autoencoder neural networks in logical anomaly detection while achieving robust results in structural anomaly detection.

INDEX TERMS Logical anomaly detection, graph neural networks, vision graphs.

I. INTRODUCTION

ANOMALY detection is the task of recognizing a deviation in the test data based on a learned data distribution during training [1], [2]. Anomalies are the rare deviations that can occur, which can for instance be related in image data to the structure of certain objects, called structural anomalies, or to the global structure of the image, called logical anomalies.

While a structural anomaly is local to one object on the image, a logical anomaly is usually hidden behind the relation between multiple parts of the image, hence its global aspect. Logical anomalies cannot be fully detected with the traditional anomaly detection approaches based on local features extracted from convolution layers or embedded in sequences extracted from transformer blocks. For this reason, logical anomaly detection methods have been following other approaches focusing on extracting both the local and global features of the image or on the relation between its segmented objects.

Image anomaly detection research has been focusing on detecting structural anomalies on objects found in the MvTec anomaly detection dataset [3]. An anomaly detection dataset has two classes, namely the normal class and the anomalous class and is usually unbalanced [4]. State-of-the-art approaches, such as Glass [5] and EfficientAD [6] are unsupervised and achieve very high performance in this benchmark. Recent research has been shifting to logical anomalies found in the MvTec Logical Constraints [7] dataset. This task represents a challenge for most of the methods mentioned above. For instance, EfficientAD [6] achieves an image area under receiver operating characteristic curve (AUC) value of 55.26 on the "screw bag" subset of this logical anomaly detection dataset.

The nature of the logical anomalies affecting the global relations of image parts is similar to anomalies that can be found in graph data. Anomaly detection for graph data is an active research field that has multiple applications in financial, security and biological fields. Until now, graph anomaly detection has been considered as a separate research field from image anomaly detection. The idea and aim of our research is to combine graph and image anomaly detection by introducing Vision Graph Logical Anomaly Detection (ViGLAD). Our work proposes a novel approach in the field of logical anomaly detection in image data based on the graph representation of an image. Graph representations for images has been considered in previous works, however, only in the context of image and point cloud semantic segmentation [8] and recently for image classification and object detection [9]. Our idea is to use this graph representation and the recently introduced vision graph blocks [10] for logical anomaly detection in image data. Vision graph blocks transform and ex-

change information among the graph nodes of image patches. We show in this work that representing an image as graph and learning how to reconstruct its graph representation with vision graph convolutions and deconvolutions allows the detection of logical anomalies in images by learning local and global logical relations of its objects. We summarize the contribution of this work as follows:

- We introduce a novel approach that uses both convolution layers and vision graph blocks for logical anomaly detection.
- We propose a novel vision graph autoencoder architecture based on vision graph blocks using novel vision graph deconvolutions inspired by vision graph convolutions, which have been used for graph information processing.
- We combine graph autoencoders with image convolution neural networks to establish a robust logical and structural anomaly detection method that outperforms state-of-the-art approaches on logical anomaly detection datasets such as MvTec Logical Constraints [7] and on structural datasets such as VisA [11].

This Section I has introduced the considered research challenge and the contribution idea. Section II describes the research state in the novel field of logical anomaly detection and the recent development of vision graph neural networks. Section III explains the ViGLAD method and Section IV presents our experimental work to evaluate its performance in both logical and structural anomaly detection tasks. In Section VI, we discuss the results of the evaluation and describe the limitations of our approach to present the potential for future work that can be based on ViGLAD in Section VII.

II. RELATED WORK

The datasets used in the training phase of anomaly detection methods are highly unbalanced, containing usually no or few anomalous data, in contrast to training datasets for image classification or object detection and segmentation. For this reason, researchers are more focused on unsupervised learning approaches [12] rather than supervised learning methods [13], [14]. Based on the availability of anomaly data during training, the supervision approach for anomaly detection can vary. While supervised image classification or object detection can be used for anomaly detection with balanced datasets, unsupervised anomaly detection is more researched because of the usually unbalanced nature of datasets in practice. In fact, if examples from the anomaly class are available and labeled on image level, binary image classification can be used to predict if an image is normal or anomalous. Object detection techniques can also be adopted for anomaly detection, if example anomaly images are available for training and are labeled on object level by defining the areas or pixels representing the anomaly. The output of such models will be a bounding box or a segmentation map of the anomaly area. The unsupervised context is based on the assumption that the training set exclusively contains normal data. In practice one cannot exclude the possibility that some anomaly data have contaminated the training set. This setup is called fully unsupervised anomaly detection [15]. Recently, few-shot anomaly detection [16] approaches have been emerging, especially with the use of unified models that are trained on different classes and able to detect anomalies within a new class of objects.

Based on the nature of the test data, other settings can be considered. If the anomaly type has not been seen during training, then supervised approaches are developed to fit the open-set setup [13], [14]. This is also considered in the Out-Of-Distribution task, where the unseen anomaly data is from a novel class that is not considered during training [17], [18]. However if the unseen test anomaly is from a known class but from another domain, then the task becomes anomaly detection under domain or distribution shift [19].

A. LOGICAL ANOMALY DETECTION

The focus of the anomaly detection task has been developing from detecting structural anomalies on individual objects in an image, such as scratches or deformations, to also consider logical anomalies in the global context of the image, such as wrong number of objects or wrong spatial order. This development originating from practical examples has been introduced to research with novel logical anomaly datasets, such as the MvTec Logical Constraints dataset [7].

Logical anomaly detection methods can be categorized in two groups. The first one focuses on detecting anomalies in the relations between the components of an image defined in a prior segmentation step. In [20], the authors propose an approach that uses a histogram matching and an entropy loss based segmentation to define the image components composing both a component and a class memory bank parallel to a patch memory bank used to compute the anomaly score. In [21], the authors also perform an image segmentation with the help of a pre-trained model to calculate a component memory bank during training. In a similar approach, the authors in [22] use the segmentation of multiple scales in a decoder to define a foreground and a background component in each stage and define if the foreground is an anomaly in the context of the background of each scale.

The second category of logical anomaly detection techniques consider both local and global features of the image. In [23], the authors introduce a framework to extract local features and their corresponding global features through a local-global bottleneck. In a second stage, local and global feature estimation networks based on the transformer architecture are trained on normal data and are used in the inference stage to compute the anomaly score. In another work, the authors in [24] define logical anomalies as unpicturable anomalies that have to be detected based on local and global features extracted by the teacher-student-auto-encoder architecture of the EfficientAD approach introduced in [6].

In summary, logical anomaly detection is an active research field with a growing attention. To the best of our knowledge, unsupervised approaches are achieving a performance level that is not consistent over all types of logical anomalies.

This can be discovered in the performance difference over the subsets of the MvTec Logical Constraints dataset [7], especially in "screw bag" and "breakfast box".

B. VISION GRAPH NEURAL NETWORKS

Hypergraph theory, introduced by Berge in 1987, models complex problems in operational research and combinatorial optimization. It extends traditional graph theory by representing multi-way relationships, which are essential in fields like psychology, biology, and artificial intelligence. Hypergraphs are particularly effective for applications involving network modeling, data structures, process scheduling, and computational systems due to their ability to capture more general types of relationships beyond binary ones. In image analysis, hypergraphs provide a nuanced representation of interactions between image segments, enhancing the effectiveness in advanced image processing tasks [25]. Graph Neural Networks (GNNs), introduced by [26] in 2009, are neural networks specifically designed for graph-structured data. Unlike traditional neural networks, which handle fixed-size input vectors or sequences, GNNs can process graphs directly, making them ideal for applications involving relationships and interactions between entities. Their key innovation is the ability to propagate information along graph edges, enabling the learning of node representations that reflect the structure and features of their neighborhoods. GNNs have been also extended to graph convolution networks (GCN) in [27] to enable effective learning on non-Euclidean domains and to GNNs with attention mechanisms to enhance their ability to capture more complex relationships in graphs [28] and be robust against corrupt test data [29].

GNNs can predict molecular properties, aiding in biological and chemical computation by analyzing molecular interactions [30] and they can predict social impacts and links in social networks and in traffic networks, they accurately forecast traffic conditions. In neuroscience, they help study conditions like bipolar disorder and diabetic optic neuropathy. Additionally, they are used to enhance text categorization in natural language processing, to improve image and text classification, to predict drug side effects and to develop recommender systems [31]. Graph Neural Networks are also widely used for anomaly detection in social networks like BlogCatalog and Flickr, as well as academic paper citation networks such as ACM [32], in traffic networks security detecting attacks and threats [33] or in trust networks to identify suspicious users in trading networks [34].

Recently, Vision Graph [10] and Vision Hypergraph [9] networks have been introduced for image classification and object detection tasks, where an image is represented as a graph or a hypergraph and fed to an isotropic and pyramid networks to identify low-level features and their high level dependencies. With the novel vision graph and hypergraph blocks, these networks have been able to achieve state-of-the-art results in both tasks. Analogically, a hypergraph neural network architecture for electron micrograph classification has been introduced by [35]. This architecture encodes visual hypergraphs to capture structural and feature information, facilitating the learning of relation structure-aware embeddings. It identifies discrete visual elements and their dependencies, optimizing the representation of scale-variant elements for an improved classification performance.

Our work is inspired by the use of graph structures for anomaly detection in general. However, we use the graph representation of an image for a novel task, namely to capture logical anomalies in images with the use of the recently introduced vision graph blocks.

III. APPROACH

The proposed approach, shown in Figure 1 and described in Section III-C, is composed of two main branches: a vision graph auto-encoder branch and an image convolution teacher-student branch. The vision graph is built with a vision graph constructor based on the input image as described in Section III-A and then processed through a vision graph auto-encoder described in Section III-B. The outputs of both branches is combined in a later step to build a local and global anomaly map. During inference, the maximum value of this map is combined with the Mahalanobis distance between the extracted features of the test image and a representation of the training images to compute the anomaly score.

A. VISION GRAPH CONSTRUCTOR

The first component of our proposed approach is responsible of computing the graph representation of the image and is composed of a tockenization step followed by a nearest neighbours graph optimization as explained below. We set the image graphs to be both finite, unidirected and connected with no isolated vertex or nodes [25]. Considering an input image $I \in \mathbb{R}^{H\times W\times 3}$, the graph constructor G outputs a vision graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$\mathcal{G} = G(I). \tag{1}$$

The graph $\mathcal{G}$ is constructed as follows: We first resize the image I to $I' \in \mathbb{R}^{H'\times W'\times 3}$. Following the structure in Figure 2 according to [36], we embed the image into $N = \frac{H'}{4} \times \frac{W'}{4}$ patches $p_i \in \mathbb{R}^{D\times D}$ with $D = 4$, building a feature matrix $P = [p_1, p_2, \ldots, p_N]$ representing a nodes set $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ with the each feature vector p_i associated with the node v_i. For each node v_i, the set of its k nearest neighbours $\mathcal{N}_k(p_i)$ is defined based on the distance between the feature vectors. The edge matrix $\mathcal{E} \in \mathbb{R}^{N\times N}$ is built as

$$e_{ij} = \begin{cases} 1 & \text{if } v_j \in \mathcal{N}_k(v_i), \\ 0 & \text{else.} \end{cases} \tag{2}$$

B. VISION GRAPH AUTO-ENCODER

The proposed vision graph auto-encoder is composed of an encoder built with Convolution Vision Graph (ConvViG) blocks and a decoder built with Deconvolution Vision Graph (DeconvViG) blocks. In a ConvViG block, we start with a

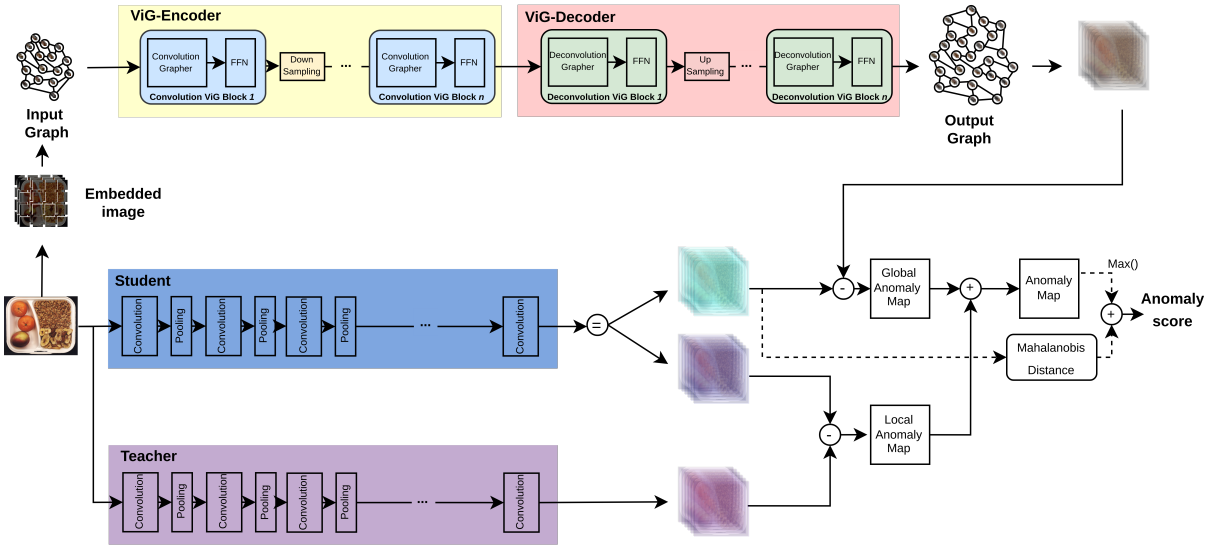


FIGURE 1: Overall structure of the Vision Graph Logical Anomaly Detection (ViGLAD) approach. From left to right, it is composed of a graph constructor including an image embedding step, a feature extraction phase composed of a vision graph auto-encoder built with ConViG and DeconViG blocks, a PDN [6] based student and teacher. The features extracted are extracted to build an anomaly maps used in addition to the Mahalanobis distance of some extracted features in order to compute the anomaly score in inference.

graph convolution step. For each node feature vector p_i and its k neighbours $q_j \in \mathcal{N}_k(p_i)$, we compute the output p'_i of the graph convolution step H as

$$p'_i = H(g_{conv}(p_i)) \tag{3}$$

with the basic graph convolution function g_{conv} defined as in [37]. The graph convolution function H is composed of a multi-head representation of h heads and a fully connection layer FC_{conv} with weights $W^i_{FC_{conv}}$

$$p'_i = [head^1(g_{conv}(p_i))W^1_{FC_{conv}}, \ldots, head^h(g_{conv}(p_i))W^h_{FC_{conv}}]. \tag{4}$$

This results in a new feature matrix $P' = [p'_1, p'_2, \ldots, p'_N]$. We summarize the graph convolution step as *GraphConv*

$$P' = GraphConv(P) \tag{5}$$

The *GraphConv* operation is wrapped by two fully connected layers FC_{gIn} and FC_{gOut} and an activation function σ_g. The output Y of this operation is computed as

$$Y = FC_{gOut}\left[\sigma_g\left[GraphConv\left(FC_{gIn}(P)\right)\right]\right] + P \tag{6}$$

The second component of the ConvViG Block is a Feed-Forward Network (FFN) with two fully connected layers FC_{FFN1} and FC_{FFN2} separated by an activation function σ_{FFN}. The output Z of the ConvViG Block can be computed as

$$Z = FC_{FFN2}\left[\sigma_{FFN}\left[FC_{FFN1}(Y)\right]\right] + Y. \tag{7}$$

In this work, we introduce the inverse operation of the graph convolution, which we integrate in the decoder part of the vision graph auto-encoder. The structure of the Deconvolution Vision Graph (DeconvViG) block is similar to the ConvViG block using a basic transpose convolution function g_{deconv} instead of the basic convolution function g_{conv}. Given the feature vector p^l_i of a node i in a layer l, the output feature feature vector $p^{l+1}_i = g_{deconv}(p^l_i)$ is computed as

$$p^{l+1}_i = \sigma_{deconv}\left[norm\left[TrConv2D(concat[p^l_i, p'^l_i])\right]\right] \tag{8}$$

with *norm* being a batch normalization operation [38], *TrConv2D* being a transposed convolution operation [39] and *concat* being a concatenation of both feature vectors. The neighbourhood distance vector p'^l_i is computed

$$p'^l_i = max[q^l_j - p^l_i | q^l_j \in \mathcal{N}_k(p^l_i)]. \tag{9}$$

with $\mathcal{N}_k(p^l_i)$ being the k neighbours set of the feature vector p^l_i.

Based on the ConvViG and DeconvViG blocks, we build a vision graph auto-encoder A_g. Given the input image I and the constructed graph $\mathcal{G}$, the output of the vision graph auto-encoder composed of an encoder E of depth d_{enc} and a decoder D of depth d_{dec} is

$$A_g(\mathcal{G}) = D[E(\mathcal{G})]. \tag{10}$$

The output E_i in each layer i of the encoder is computed as

$$E_i = downsampling(ConvViG(E_{i-1})), \tag{11}$$
$$E_0 = downsampling(ConViG(\mathcal{G})). \tag{12}$$

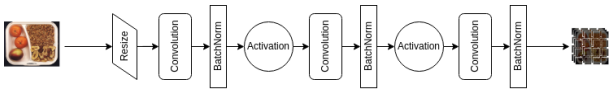


FIGURE 2: Image embedding procedure according to [36]. The image input is first resized then transformed with three blocks of convolution and batch normalization. The non-linear Gaussian Error Linear Units activation function (GeLu) [40] is added after the first and second block. The output is a four channel feature map divided in N patches of size 4×4.

For the decoder, the output D_j of each layer j is computed as

$$D_j = upsampling(DeconvViG(D_{j-1})), \tag{13}$$
$$D_0 = upsampling(DeconViG(E_{d_{enc}})). \tag{14}$$

The *downsampling* and *upsamplinbg* operations are based on convolution and transposed convolution [39] layers followed by a batch normalization layer [38].

C. VISION GRAPH LOGICAL ANOMALY DETECTION

The architecture of the introduced approach, Vision Graph Logical Anomaly Detection (ViGLAD) is inspired from the EfficientAD unsupervised anomaly detection method [6]. ViGLAD is mainly composed of a teacher network T, a student network S based on the Patch Description Network (PDN) architecture proposed in [6] and our proposed graph auto-encoder A_g. The teacher and student networks have the same layer structure. The teacher is trained in a first step based on a knowledge distillation from a WideResNet-101 backbone [41] pretrained on the ImageNet dataset [42] for a classification task.

The student network is trained to imitate the distilled teacher network frozen during training by minimizing the L_{ST} loss function

$$L_{ST} = \frac{1}{CWH}\left[\sum_c \|T(I)_c - S(I)_c\|_F^2 + \sum_c \|S(I_r)_c\|_F^2\right] \tag{15}$$

with C being the channel number of the output features and I_r being a random image from the ImageNet dataset [42] used in the knowledge distillation phase. All losses introduced in this section are computed based on the $\|.\|_F^2$ Frobenius norm [43]. With this loss, the student will be able to predict the output of the teacher for normal images and fails to predict it for images with structural anomalies.

The vision graph auto-encoder is also trained to imitate the teacher network frozen during training by minimizing the L_{A_gT} loss function

$$L_{A_gT} = \frac{1}{CWH}\sum_c \|T(I)_c - A_g(\mathcal{G})_c\|_F^2 . \tag{16}$$

With this loss, the vision graph auto-encoder will be able to predict the teacher output except for fine-grained structural anomalies and will focus on the global structure of the image. In order to extract the logical anomalies, a third term L_{SA_g} is added to the loss function describing the distance between the second half of to the student output S' and the output of the vision graph auto-encoder

$$L_{SA_g} = \frac{1}{CWH}\sum_c \|A_g(\mathcal{G})_c - S'(I)_c\|_F^2 . \tag{17}$$

The total loss function of ViGLAD during training $L = L_{ST} + L_{A_gT} + L_{SA_g}$ is intended to train the network to detect both structural and especially logical anomalies. The intention behind using the vision graph auto-encoder instead of the original auto-encoder of EfficientAD [6] is to increase the capability of the network to learn features depending on logical relations from the graph representation in order to detect logical anomalies.

During inference, the anomaly score is calculated based on the difference between the output of the teacher and first half of the student networks, called local anomaly map, as well as the distance between the output of the vision graph auto-encoder and the second half of the student network, called global anomaly map. After an average pooling step over the channels, both anomaly maps are merged and the maximum is defined as the reconstruction anomaly score A_r. In order to further highlight the logical anomalies, we add a second anomaly score component, called feature anomaly score A_f computed as

$$A_f = M_G[y'_{S'(I)}] \tag{18}$$

with M_G being the Mahalanobis distance under Gaussian distribution of the features computed in the second part of the student output. M_G is computed according to [24] as:

$$M_G[y'_{S'(I)}] = \sqrt{(y'_{S'(I)} - \mu)^T \Sigma^{-1} (y'_{S'(I)} - \mu)} \tag{19}$$

with $y'_{S'(I)}$ being the average pooling result of the output features from the student second half. μ and Σ are the mean and covariance matrix of the features produced by the student second half with a set of images from the training set. The total anomaly score of ViGLAD is $A = A_r + A_f$.

IV. EXPERIMENTAL WORK

A. DATASETS

The proposed approach ViGLAD is designed to detect logical anomalies in image data in an unsupervised setup. We evaluate its performance based on the public logical anomaly datasets MvTec Logical Constraints [7], CAD-SD [44] and Digit Anatomy [45]. The structure of the selected datasets, as shown in Table 1, is composed of a training set containing only normal data, a validation set containing a smaller number of normal data and a test set composed of normal data and anomaly data. We split the anomaly data in the test set into logical and structural anomalies, in order to describe the performance of our approach for logical anomaly detection in comparison to structural anomaly detection. We build the Digit Anatomy dataset [45] by randomly choosing examples from the MNIST dataset [46] and introducing misorder and

TABLE 1: Structure of the considered datasets. Each dataset is compose one or multiple subsets. The number of images for training, validation and testing set of each subset is presented. Only logical anomalies are considered in testing set.

Dataset	Subset	Train normal	Validate normal	Test normal	Test anomaly
MvTec Logical Constraints [7]	Screw bag	360	60	122	137
	Breakfast box	351	62	102	83
	Juice bottle	335	54	94	142
	Pushpins	372	69	138	91
	Splicing connectors	360	60	119	108
VisA [11]	Candle	810	90	100	100
	Capsule	488	54	60	100
	Cashew	405	45	50	100
	Chewing gum	408	45	50	100
	Fryum	405	45	50	100
	Macaroni1	810	90	100	100
	Macaroni2	810	90	100	100
	Pcb1	814	90	100	100
	Pcb2	811	90	100	100
	Pcb3	816	90	100	100
	Pipe fryum	405	45	50	100
CAD-SD [44]	Screw	400	72	139	85
Digit Anatomy [45]	Digits	360	60	110	120

flipping of digits as logical anomalies, while considering missing and novel digits as structural anomalies.

In addition to the logical anomaly detection datasets, we execute an evaluation on the structural anomaly dataset VisA [11], in order to compare our proposed method to state-of-the-art approaches and evaluate its generalization capability to other types of anomalies with a different context than the objects found in the logical anomaly detection datasets.

B. BASELINES

In this section, we present the results of the experimental work with the target to describe the performance of our proposed approach in comparison to state-of-the-art methods in logical and structural anomaly detection. We select as baselines methods based on both feature embedding and reconstruction. From feature embedding methods, we select PatchCore [41], Padim [47] and Fastflow [48]. These methods are based on a pretrained feature extracting backbone that feed different architectures to transform and cluster these features in order to build a representation of the normal data, where anomaly data during inference cannot fit in. These approaches have been successful in different anomaly detection tasks in the last years. For reconstruction methods, we select EfficientAD [6] and PUAD [24], since they contain both an auto-encoder for reconstruction and can be a direct baseline to compare to our method. With this direct comparison, we can directly interpret the effect of using graph representation of images and the effect of convolution and deconvolution vision graph blocks.

TABLE 2: Image AUC results for experiments on MvTec Logical Constraints [7], CAD-SD [44] and DigitAnatomy [45] for EfficientAD (EffAD) [6], PUAD [24], PatchCore [41], Padim [47] and Fastflow [48] and ViGLAD (ours). The architectures of EfficientAD [6], PUAD [24] and ViGLAD (ours) are based on a medium PDN [6] the Mahalanobis distance is computed based on the second half of the student features (average of 5 runs).

Dataset	Subset	EffAD [6]	PUAD [24]	PatchCore [41]	Padim [47]	FastFlow [48]	ViGLAD (ours)
MvTec Logical Constraints [7]	Screw bag	55.26	79.47	58.08	48.80	67.78	**81.83**
	Breakfast box	84.49	89.44	57.04	45.51	58.07	**93.95**
	Juice bottle	99.97	**99.98**	35.93	54.30	55.52	99.68
	Pushpins	**98.94**	96.61	49.43	48.86	64.46	87.30
	Splicing connectors	96.21	**96.38**	57.74	49.53	76.29	92.43
CAD-SD [44]	Screw	99.63	99.38	73.29	48.37	66.70	**99.91**
Digit Anatomy [45]	Digit	93.78	96.08	88.66	94.39	86.27	**96.11**

C. IMPLEMENTATION

For the experiments of this article, we build ViGLAD with a medium PDN for the teacher and student networks as described in [6]. For the embedding of the input images, we resize the input images to $(224, 224, 3)$ and build the vision graph with dilated knn ($k = 12$). In the encoder, as well as in the decoder, we use 12 vision graph blocks with a downsampling and an upsampling step after the second, fourth, tenth and twelfth vision blocks of the encoder and decoder respectively. The number of output channels of the teacher and auto-encoder is $c = 384$ and $c = 768$ for the student. For graph convolution, we use the max-pooling graph feature aggregator [37], that we have also adopted for the deconvolution step. We conduct our experiments on an RTX 4070 Nvidia GPU with 8 GB of memory hardware for 100.000 epochs with a batch size of 1, a starting learning rate of 10^{-4} and an Adam optimizer. We set the normalization quantiles introduced in [6] to $q_{start} = 0.9$ and $q_{end} = 0.995$ for all experiments, except for Digit Anatomy [45], where we set the set the normalization quantiles to $q_{start} = 0.9$ and $q_{end} = 0.999$.

D. RESULTS

Our evaluation on the logical anomaly datasets reveals two key results that can be seen in Table 2. First, the methods that achieve state-of-the-art performance for structural anomaly detection are not able to detect logical anomalies. In contrast, methods that also consider global features and their relations in their loss functions achieve good results detecting both types of anomalies, even though they are based on the same feature extracting backbones. Second, incorporating the

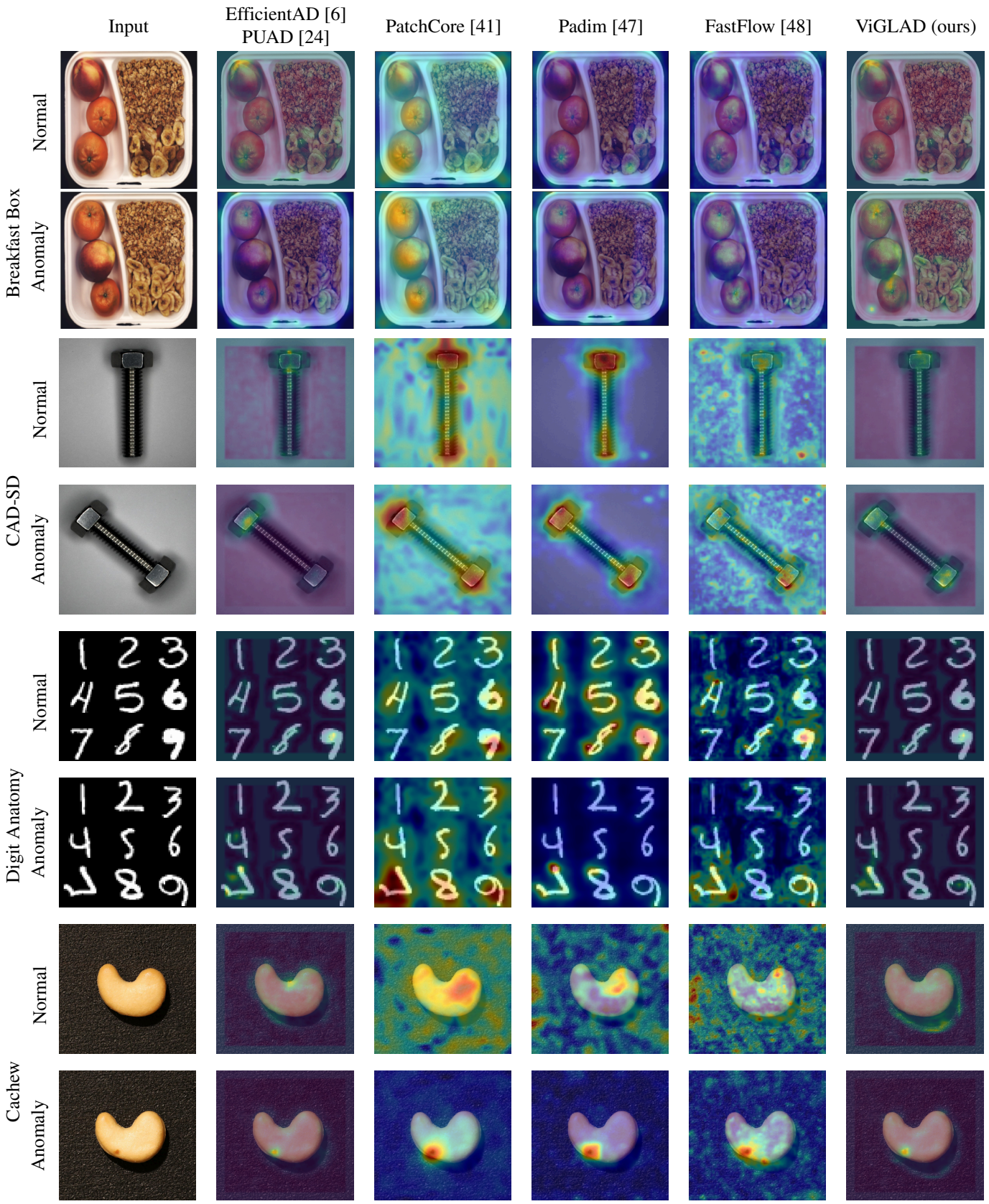


FIGURE 3: Example qualitative results for normal images and images with logical anomalies in the "breakfast box" subset from the MvTec Logical Constraints dataset [7], in CAD-SD [44], in DigitAnatomy [45] and in the "cashew" subset from the visA dataset [11]. Since EfficientAD [6] and PUAD [24] generate the same anomaly map and differ only in the anomaly score calculation, we combine the results for both methods in this figure.

TABLE 3: Image AUC results for experiments on visA [11] subsets for EfficientAD (EffAD) [6], PUAD [24], PatchCore [41], Padim [47] and Fastflow [48] and ViGLAD (ours). The architectures of EfficientAD [6], PUAD [24] and ViGLAD (ours) are based on a medium PDN [6] the Mahalanobis distance is computed based on the second half of the student features (average of 5 runs).

Dataset	Subset	EffAD [6]	PUAD [24]	PatchCore [41]	Padim [47]	FastFlow [48]	ViGLAD (ours)
VisA [11]	Candle	98.92	98.99	95.84	85.40	96.84	**99.08**
	Capsule	85.93	81.75	46.80	44.93	71.80	**92.91**
	Cashew	98.26	98.87	55.27	51.11	84.71	**99.01**
	Chewing gum	**99.96**	99.90	30.47	30.48	99.91	99.70
	Fryum	**98.82**	98.74	30.72	53.59	91.91	98.41
	Macaroni1	**99.75**	99.61	55.48	55.91	90.88	99.06
	Macaroni2	**97.70**	94.42	72.52	51.47	49.75	92.58
	Pcb1	**99.98**	99.96	43.72	74.80	79.64	99.89
	Pcb2	99.84	**99.95**	45.76	56.56	74.43	99.78
	Pcb3	99.02	**99.03**	98.79	37.27	68.95	98.97
	Pipe fryum	**99.94**	99.60	57.92	94.39	79.27	99.65

graph representation of an image and using graph convolution and deconvolution lead to better results for logical anomaly detection, especially in subsets that are challenging for logical anomaly detection methods without the graph representation of the input image. This can be observed in the "screw bag" and "breakfast box" subsets, where our approach achieve 81.83 and 93.95 image AUC compared to the second best method PUAD [24] that achieves 79.47 and 89.44 image AUC. Regarding the remaining subsets, our approach achieves comparable results and its performance does not deteriorate, even in the "juice bottle" subset, whose logical anomalies are very close to its structural anomalies.

On the other hand, the results of the evaluation on the structural anomaly detection dataset VisA [11] from Table 3 show that our method also achieve state-of-the-art results and outperform the baseline methods on multiple subsets. In average, our approach outperforms the best result of the baseline methods achieved by EfficientAD [6] with an average image AUC of 98.09 compared to 98.01.

V. ABLATION STUDY

In order to analyze the proposed architecture of our approach, we conduct an ablation study by defining the impact of different setups on the overall performance on the "screw bag" subset, which is the most challenging in the MvTec logical constraints dataset [7]. First, we analyze the graph representation with the use of vision graph convolution and deconvolution blocks in different components of ViGLAD. For this purpose, we test three different implementations: using vision graph blocks in the teacher and student networks (GPDN-AE), using vision graph blocks in the auto-encoder (PDN-ViGAE) and using vision graph blocks in the teacher, student and auto-encoder (GPDN-ViGAE). The results presented in Figure 4 show that using the Mahalanobis distance for the teacher or student output as part of the anomaly score plays an important role in highlighting the global features extracted from convolution layers in the student next to the global features extracted from the vision graph blocks in the vision graph auto-encoder. As a result, only building the auto-encoder based on the vision graph blocks (PDN-ViGAE) in combination with using the Mahalanobis distance as part of the anomaly score lead to the best results for both logical and structural anomaly detection. the proposed approach ViGLAD in the evaluation section above is built as a PDN-ViGAE with the Mahalanobis distance included in the anomaly score.

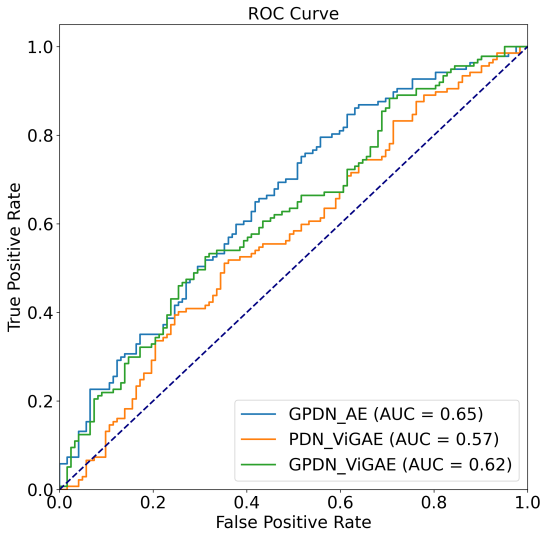


FIGURE 4: Ablation study vision graph blocks use on the "screw bag" logical subset.

TABLE 4: Ablation study PDN [6] size. The architecture of the network is ViG-AE with Mahalanobis distance computed based on the second half of the student features.

PDN size	small	medium
Logical	78.07	**81.14**
Structural	**91.69**	90.68
All	84.88	**85.91**

Second, we study the influence of the PDN [6] size in the student and teacher networks. Table 4 shows that using a deeper PDN [6] leads to better results on logical anomalies. However, for structural anomalies, using a small PDN [6] slightly outperforms the medium size. Since our focus in this paper lies in detecting logical anomalies, we use medium size PDN [6] in our approach that achieves a better performance in average.

Table 5 describes the results of using different features for the computation of the Mahalanobis distance used in the anomaly score to highlight unpicturable (logical) anomalies.

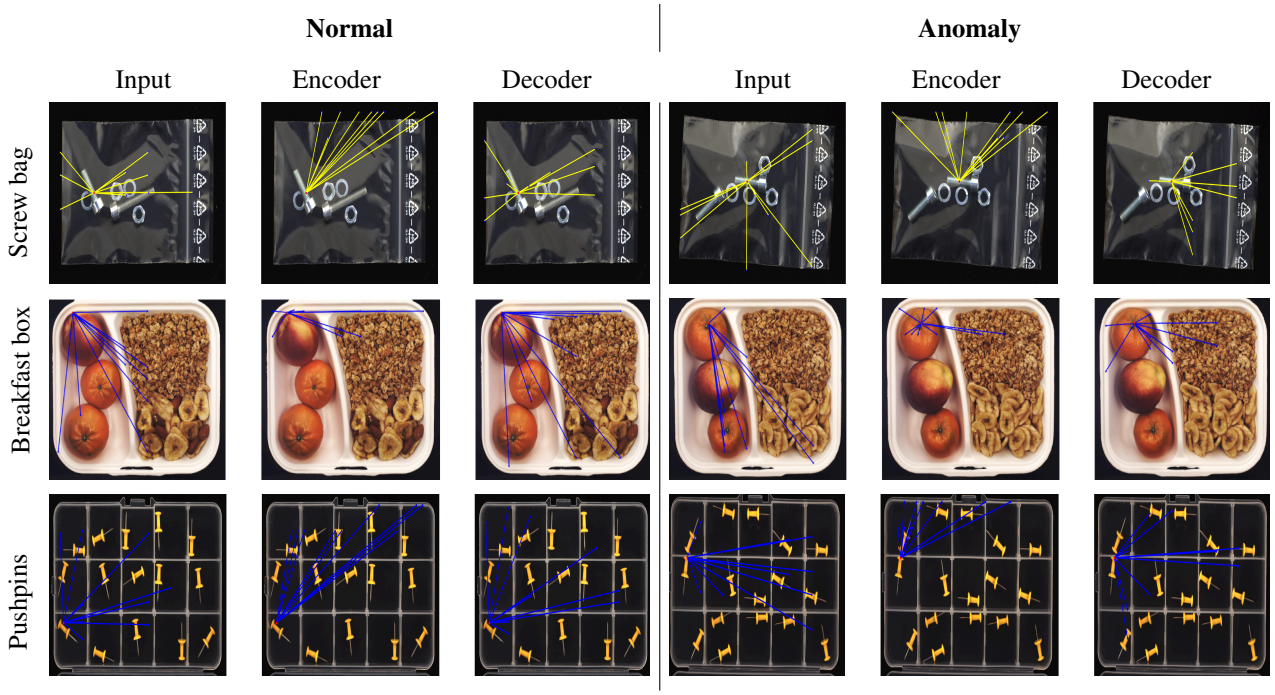


FIGURE 5: The graph representation of the node with the highest anomaly score. The first column represents the input graph before the first vision graph block. The second column represents the output graph of the encoder. The third column represents the output graph of the decoder.

TABLE 5: Ablation study about the features used for the computation of the Mahalanobis distance for the unpicturable anomaly score. The experiments have been conducted on the "screw bag" sub-dataset from MvTec logical constraints dataset [7]. The size of the PDN [6] has been set to medium.

Features	Teacher	Student former	Student second
Logical	64.83	64.15	**81.83**
Structural	89.27	87.30	**90.68**
All	77.05	75.72	**85.91**

TABLE 6: Ablation study about the number of k-neighbours used for constructing the graph of each image. The experiments have been conducted on the "screw bag" sub-dataset from MvTec logical constraints dataset [7]. The size of the PDN [6] has been set to medium.

k-neighbours	k=3	k=9	k=12	k=15
Logical	77.62	**81.14**	**81.83**	79.10
Structural	93.35	90.68	91.64	**94.50**
All	85.48	85.91	**86.73**	**86.80**

Using the second half of the student output leads to the best results on the considered dataset since they are responsible for training the student to also imitate the graph vision auto-encoder in learning global features and their higher relations.

For the last part of the ablation study, we focus on the graph representation of the image and its features throughout the vision graph auto-encoder. For this purpose, we test ViGLAD with different k values for the dilated knn based graph construction step. The results summarized in Table 6 show that using higher k values lead to better results on logical anomalies. However the performance stagnates with values higher than 9. Since the k value determines how many nodes are connected to each other, we opt for the smallest value being $k = 9$, in order to limit the complexity of the image graphs.

VI. DISCUSSION AND LIMITATIONS

The results presented in the previous section confirm our hypothesis, stating that representing an image as graph and learning how to reconstruct its graph representation with graph convolutions and deconvolutions enable the capability to learn local and global logical relations of the objects of an image. This allows detecting logical anomalies, which can affect these local and global relations. The graph representation and processing have been used as a second branch parallel to an image convolution branch that has been proven to be efficient in detecting structural anomalies. This combination has been shown in our work to be efficient in detecting both logical and structural anomalies and achieve state-of-the-art results in general image anomaly detection. The conducted ablation study has shown that the graph representation in both branches of ViGLAD enhances its performance in comparison to the baseline EfficientAD [6] architecture that is solely based on image convolution. However, it is still to be studied if using the proposed vision graph convolution and deconvolution blocks can be transferred to other architectures, and if it enhances their capability to detect logical anomalies

without decreasing their performance in detecting other types of anomalies.

Figure 5 shows some example results of our approach and the development of the graph representation of the image. We observe for normal images the capability of the vision graph auto-encoder to reproduce the graph even for the node with the highest anomaly score. For anomaly images, the vision graph auto-encoder is not capable to reconstruct the input graph.

The performance gap experienced by our approach in the subset "pushpins" from the MvTec logical constraints [7] can be explained by the presence of multiple objects in the image in comparison to other subsets. In this case, we experience an under representation of the possible global relations in one image between the different objects. This means that the low complexity of the constructed graph could be not able to represent all possible global relations in an image with a high number of objects. This can be explored in further research in order to find the optimal graph representation for an image independent from its number of objects.

VII. CONCLUSION

Logical anomalies are challenging to be detected in images because they are unpicturable, meaning that they cannot be directly seen on the individual objects of a scene but have to be interpreted from the global relations between the objects of the scene. State-of-the-art anomaly detection methods, which have been originally developed to detect structural anomalies related to individual objects, are not able to achieve good performance in detecting logical anomalies in images. On the other side, anomaly detection in graph data have been focusing on detecting the relations between different nodes of the graph. This has been the motivation to introduce graph representation which has been recently introduced to image classification also to logical anomaly detection in images. Our proposed approach proposes the use of graph representation and its feature extracting and representation in baseline anomaly detection methods. This have resulted in an enhancement of their ability to detect logical anomalies without decreasing their performance in detecting structural anomalies. Our approach achieves state-of-the-art performance in detecting both types of anomalies with a remarkable advantage in benchmark datasets. This work lies the basis for further exploring the advantages and limitations of this novel framework for logical anomaly detection.

REFERENCES

[1] Srikanth Thudumu, Philip Branch, Jiong Jin, and Jugdutt Singh. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, 7, 07 2020.

[2] Chi-Hua Chen, Chuanlei Zhang, Jiangtao Liu, Wei Chen, Jinyuan Shi, Minda Yao, Xiaoning Yan, Nenghua Xu, and Dufeng Chen. Unsupervised anomaly detection based on deep autoencoding and clustering. *Security and Communication Networks*, 2021:7389943, 2021.

[3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The mvtec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4):1038–1059, April 2021.

[4] Mirko Nardi, Lorenzo Valerio, and Andrea Passarella. Centralised vs decentralised anomaly detection: when local and imbalanced data are beneficial. volume 154 of *Proceedings of Machine Learning Research*, pages 7–20. PMLR, 17 Sep 2021.

[5] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A Unified Anomaly Synthesis Strategy with Gradient Ascent for Industrial Anomaly Detection and Localization, 2024.

[6] Kilian Batzner, Lars Heckler, and Rebecca König. EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies, 2024.

[7] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization. *Int. J. Comput. Vis.*, 130(4):947–969, 2022.

[8] Dena Bazazian and Dhananjay Nahata. DCG-net: Dynamic capsule graph convolutional network for point clouds. *IEEE Access*, 8:188056–188067, 2020.

[9] Yan Han, Peihao Wang, Souvik Kundu, Ying Ding, and Zhangyang Wang. Vision HGNN: An Image is More than a Graph of Nodes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19878–19888, October 2023.

[10] Kai Han, Yunhe Wang, Jianyuan Guo, Yehui Tang, and Enhua Wu. Vision GNN: An Image is Worth Graph of Nodes. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2024. Curran Associates Inc.

[11] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 392–408, Cham, 2022. Springer Nature Switzerland.

[12] Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly Detection with Conditioned Denoising Diffusion Models, 2023.

[13] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching Both Gray and Black Swans: Open-set Supervised Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[14] Jiawen Zhu, Choubo Ding, Yu Tian, and Guansong Pang. Anomaly Heterogeneity Learning for Open-set Supervised Anomaly Detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024.

[15] Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22883–22892, June 2024.

[16] Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. FastRecon: Few-shot Industrial Anomaly Detection via Fast Feature Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 17481–17490, October 2023.

[17] Reza Averly and Wei-Lun Chao. Unified Out-Of-Distribution Detection: A Model-Specific Perspective. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1453–1463, 2023.

[18] Alvaro Gonzalez-Jimenez, Simone Lionetti, Dena Bazazian, Philippe Gottfrois, Fabian Gröger, Marc Pouly, and Alexander Navarini. Hyperbolic Metric Learning for Visual Outlier Detection. *Springer, Computer Vision – ECCV 2024*, 2024.

[19] Tri Cao, Jiawen Zhu, and Guansong Pang. Anomaly Detection Under Distribution Shift. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6511–6523, October 2023.

[20] Soopil Kim, Sion An, Philip Chikontwe, Myeongkyun Kang, Ehsan Adeli, Kilian M. Pohl, and Sang Hyun Park. Few Shot Part Segmentation Reveals Compositional Logic for Industrial Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):8591–8599, Mar 2024.

[21] Tongkun Liu, Bing Li, Xiao Du, Bingke Jiang, Xiao Jin, Liuyi Jin, and Zhuo Zhao. Component-aware anomaly detection framework for adjustable and logical industrial visual inspection. *Advanced Engineering Informatics*, 58:102161, 2023.

[22] Shyam Nandan Rai, Fabio Cermelli, Dario Fontanel, Carlo Masone, and Barbara Caputo. Unmasking Anomalies in Road-Scene Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4037–4046, October 2023.

[23] Haiming Yao, Wenyong Yu, Wei Luo, Zhenfeng Qiang, Donghao Luo, and Xiaotian Zhang. Learning Global-Local Correspondence With Semantic

Bottleneck for Logical Anomaly Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3589–3605, 2024.

[24] Shota Sugawara and Ryuji Imamura. PUAD: Frustratingly Simple Method for Robust Anomaly Detection, 2024.

[25] Alain Bretto, Hocine Cherifi, and Driss Aboutajdine. Hypergraph imaging: an overview. *Pattern Recognition*, 35(3):651–658, Jan 2002.

[26] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[27] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[28] P Velickovic, A Casanova, P Liograve;, G Cucurull, A Romero, and Y Bengio. Graph Attention Networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[29] Boris Knyazev, Graham W Taylor, and Mohamed Amer. Understanding Attention and Generalization in Graph Neural Networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[30] Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin He. Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19. AAAI Press, 2019.

[31] Lilapati Waikhom and Ripon Patgiri. A Survey of Graph Neural Networks in Various Learning Paradigms: Methods, Applications, and Challenges. *Artif. Intell. Rev.*, 56(7):6295–6364, nov 2022.

[32] Kunpeng Zhang, Guangyue Lu, Yuxin Li, and Cai Xu. A Graph Autoencoder-based Anomaly Detection Method for Attributed Networks. In *2023 5th International Conference on Natural Language Processing (ICNLP)*, pages 330–337, 2023.

[33] Patrice Kisanga, Isaac Woungang, Issa Traore, and Glaucio H. S. Carvalho. Network Anomaly Detection Using a Graph Neural Network. In *2023 International Conference on Computing, Networking and Communications (ICNC)*, pages 61–65, 2023.

[34] Tong Zhao, Tianwen Jiang, Neil Shah, and Meng Jiang. A Synergistic Approach for Graph Anomaly Detection With Pattern Mining and Feature Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2393–2405, 2022.

[35] Rajat Kumar Sarkar, Sagar Srinivas Sakhinana Sakhinana, Sreeja Gangasani, and Venkataramana Runkana. Vision HGNN: An electron-micrograph is worth hypergraph of hypernodes. In *PML4DC Workshop, International Conference on Learning Representations(ICLR)*, 2023.

[36] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. PVTv2: Improved Baselines with Pyramid Vision Transformer. *Computational Visual Media*, 8:415–424, 2022.

[37] G. Li, Matthias Müller, Ali K. Thabet, and Bernard Ghanem. DeepGCNs: Can GCNs Go As Deep As CNNs? *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9266–9275, 2019.

[38] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *CoRR*, abs/1502.03167, 2015.

[39] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2018.

[40] Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415, 2016.

[41] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards Total Recall in Industrial Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14318–14328, June 2022.

[42] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[43] Roland Herzog, Frederik Köhne, Leonie Kreis, and Anton Schiela. Frobenius-Type Norms and Inner Products of Matrices and Linear Maps with Applications to Neural Network Training, 2023.

[44] Kengo Ishida, Yuki Takena, Yoshiki Nota, Rinpei Mochizuki, Itaru Matsumura, and Gosuke Ohashi. SA-PatchCore: Anomaly Detection in Dataset With Co-Occurrence Relationships Using Self-Attention. *IEEE Access*, 11:3232–3240, 2023.

[45] Tiange Xiang, Yixiao Zhang, Yongyi Lu, Alan L. Yuille, Chaoyi Zhang, Weidong Cai, and Zongwei Zhou. SQUID: Deep Feature In-Painting for Unsupervised Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23890–23901, June 2023.

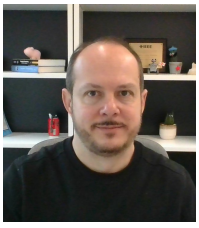
[46] Li Deng. The MNIST Database of Handwritten Digit Images for Machine Learning Research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[47] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges. Virtual Event, January 10-15, 2021, Proceedings, Part IV*, volume 12664 of *Lecture Notes in Computer Science*, pages 475–489. Springer, 2020.

[48] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows. *CoRR*, abs/2111.07677, 2021.

FIRAS ZOGHLAMI is currently pursuing his Ph.D studies in the school of Engineering, Computing and Mathematics at the university of Plymouth, UK. His Ph.D research is entitled: Anomaly Detection based Post Gripping Perception for Logistics Robotics. He is also leading the logistics robotics innovation development cluster at BMW Group. He has received his Master of Science degree in electrical engineering and information technology at the Technical University Munich. He has worked as a research associate in the department of Applied Sciences and Mechatronics at the University of Applied Sciences Munich. He has authored, co-authored and published multiple research articles in the Journal of Computing and Information Science in Engineering, in the Proceedings of the Design Society and IEEE conferences. ORCID: 0000-0001-9609-2568.

DR DENA BAZAZIAN is a lecturer in machine vision and robotics at the University of Plymouth. Previously, she was a senior research associate at the Visual Information Laboratory of the University of Bristol. Prior to that, she was a research scientist at CTTC (Centre Tecnològic de Telecomunicacions de Catalunya) and a postdoctoral researcher at the Computer Vision Center (CVC), Universitat Autònoma de Barcelona (UAB) where she accomplished her PhD in 2018. She had long-term research visits at NAVER LABS Europe in Grenoble, France in 2019 and at the Media Integration and Communication Center (MICC), University of Florence, Italy in 2017. She was working with the Image Processing Group (GPI) at Universitat Politècnica de Catalunya (UPC) between 2013 and 2015. Dena Bazazian was one of the main organisers of the series of Deep Learning for Geometric Computing (DLGC) workshops at CVPR2024-20, ICCV2021, Women in Computer Vision (WiCV) Workshops at CVPR2018 and ECCV2018, Robust Reading Challenge on Omnidirectional Video at ICDAR2017. ORCID: 0000-0002-1229-4494.



DR GIOVANNI L. MASALA received the Ph.D. degree in Physics, at the University of Cagliari (Italy). He is currently a Senior Lecturer in Computer Science and the Leader of the Cognitive Robotics Laboratory, at the University of Kent, Canterbury, U.K. He has produced several influential scientific publications in international journals and conference proceedings. His research interests include artificial intelligence (AI) and robotics, human machine interaction, social robots for eldercare, and machine vision. He is involved with numerous international research grants and recently led the U.K. partnership of the EU Interreg "AGE Independently" (AGE'IN) Project and other InnovateUK grants. He became a Senior Member of the IEEE Computational Intelligence Society in 2020. He has been part of program committees and chaired several international conferences (e.g. ECAI 2024, ISNCC, ISC2) and he is an Associate Editor at Frontiers, in Robotics and AI–computational intelligence, at the International Journal Robot Learning and at International Journal of Environmental Research and Public Health. Giovanni L. Masala is a senior IEEE member. ORCID: 0000-0001-6734-9424.

DR MARIO GIANNI is currently a Senior Lecturer (Associate Professor) in Robotics at the Department of Computer Science, University of Liverpool, UK. Previously, he was an Associate Professor of Robotics at the School of Engineering, Computing and Mathematics, University of Plymouth, UK. Dr Gianni has been PI and Academic Leader of the EPSRC Project NCNR EP/R02572X/1 – FPF 18-11411 and the Innovate UK project KTP011575. In addition, he was co-PI of the AGE IN project Interreg 2 Seas Mers Zeeën, European Regional Development Fund. Dr Gianni's main work focuses on designing, developing and deploying autonomous robotic systems collaborating with humans in monitoring and intervention operations in extreme environments. He has published more than 50 peer-reviewed papers in this research field and served as peer reviewer and committee member in several International Journals and Conferences in Robotics, including SSRR, Autonomous Robots, IROS, ICRA and the Journal of Field Robotics. Dr Gianni has also participated in collaboration with the Italian National Fire Corp to the search and rescue interventions in the earthquakes of Amatrice and Mirandola in Italy assessing damage to historical buildings and cultural artifacts using heterogeneous multi-robot systems. ORCID: 0000-0001-5410-2377.

DR ASIYA KHAN is an Associate Professor of Multimedia Communication & Intelligent Control and Associate Dean of Education at the Faculty of Science and Engineering at the University of Plymouth. She received her BEng (Hons) in Electrical & Electronic Engineering from University of Glasgow, MSc in Communications, Control & Digital Signal Processing from Strathclyde University and PhD from University of Plymouth. She is a Chartered Engineer, Fellow of IET, Senior Member of IEEE and a Senior Fellow of the Higher Education Academy. She has published over 70 journal and conference papers; 4 book chapters and several abstracts. She received a 'best paper award' in 2009 and a 'best extended abstract award' in 2022. She is currently leading the RAEng Diversity Impact programme in her School supporting students from disabled and neurodiverse engineering backgrounds. Further, she is managing work-packages on 3 EU ERDF projects, 1 by EPSRC and two from the British Council, one of them involves strengthening STEM Education in Pakistan. Her current research interests include prediction and control of video quality using artificial intelligence, machine learning, cloud computing, fuzzy logic, applying computer vision techniques and deep learning in pedestrian recognition, disease identification in cotton crops and damage recognition in wind turbine blades. She is a reviewer for IEEE, Elsevier, IET, and Springer journals. She is a member of IEEE and IET. ORCID: 0000-0003-3620-3048.