

Enhanced Symbol Recognition based on Advanced Data Augmentation for Engineering Diagrams

Ong Kai Bin¹, Yew Kwang Hooi², Said Jadid Abdul Kadir³, Haruhiro Fujita⁴, Luqman Hakim Rosli⁵

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia^{1,2,3,5}
Faculty of Management and Information Sciences, Niigata University of International and Information Studies, Niigata City, Japan⁴

Abstract—Symbol recognition has generated research interest for image analytics of engineering diagrams. Techniques including structural, syntactic, statistical, Convolution Neural Network (CNN) were studied to identify gaps of research. Despite popularity, CNN requires huge learning dataset, which often involves costly procurement. To address this, combination between CycleGAN and CNN is proposed. CycleGAN generates more learning dataset synthetically, thus yielding opportunity to improve accuracy of symbol recognition. In the domain of for engineering symbols, standard CNN model is developed and used in experimental testing. Different ratios of training dataset were tested in multiple experiments using Piping and Instrument Diagram (P&IDs) drawings. Result of highest accuracy for symbol recognition is up to 92.85% against baseline and other method. The results determined that gradual reduction of training samples, the effectiveness of recognition accuracy performance after using proposed method was remained substantially stable.

Keywords—Symbol recognition; symbol spotting; engineering drawing; convolution neural network (CNN); CycleGAN; piping and instrument diagram (P&ID)

I. INTRODUCTION

Symbol recognition is a subset of ordinary pattern recognition which focusing on identifying, detecting, and recognizing components in technical drawings. Pattern recognition is a complicated process that requires to analyze data input, feature extraction, classification, and post processing. Therefore, various functions are needed for pattern recognizer. Recognizing familiar patterns automatically is an essential pattern. However, recognition does not work accurately when identifying and classifying unfamiliar objects. Insufficient data input is one of the factors.

An application of symbol recognition is for analysis of engineering drawing. Engineering drawings are frequently used in many fields such as Oil and Gas, manufacturing, construction, and engineering. In this study, Piping and Instrument Diagrams (P&ID) are selected. P&ID are schematic diagrams representing different components and flows in a manufacturing process design for a physical plant. These diagrams aid analysis of operability and safety of a process design.

More samples for engineering drawings may produce impactful benefits due to the application for plant safety. The technique is focusing on Oil and Gas field simultaneously can be generalized for other fields of engineering drawing too.

The motivation to study this technique is due to various factors:

- Current research & development trend focuses on smarter digital diagram for better image analytics. Application of the technology includes determining root cause and inferring risk of a safety deviation.
- Digitisation of schematic diagrams and image analytics are instrumental for Digital Twin (DT) in Cyber-Physical System (CPS) designs. DT provides richer information and more potentials than digital engineering drawing application. These include prospective planning, analysis of existing systems or process-parallel monitoring. In all cases, DT offers exceptional ability to create simulations in which various development and testing work can be carried out.

Identifying components inside a digital drawing is necessary in analysis. However, this is difficult because of the drawing's layout complexity. Wrong identification of any component can lead to a faulty analysis, thus posing risk to safety and operability.

There are three main types of pattern recognition mechanisms used to classify input data. Those types are statistical, structural (syntactic), neural. Statistical methods simply collect historical data and identify new patterns based on observations and analysis of that data. The structural technique is also known as the syntactic method since it is based on primitive sub-patterns. Machines do direct computing in the pattern recognition approaches covered so far. Mathematical and statistical techniques are used in direct calculations. Last but not least, neural approach applies biological concepts into technology for recognizing patterns. The result of this effort was the invention of artificial neural networks. A neural network is an information processing system. However, deep learning is becoming more popular as a result of its superior accuracy while training with huge volumes of data. A neural network-based deep learning method is used. Neural approach is less efficient at processing tasks than deep learning systems, which provide excellent efficiency and performance for tasks. Recently, Convolution Neural Network (CNN) is a class of deep neural networks with the most applications.

CNN has shortcomings for analysis of engineering drawings, in that the model used has not been sufficiently trained for engineering drawing domain. The challenge

concerns on huge quantity amount of data samples are required for training to obtain more accurate classification and recognition in training process. Some of the samples are difficult to collect for engineering drawings especially in P&ID diagrams [1]. The accuracy of recognition is important for analysis of engineering design issues such as operability and safety. Nonetheless, having to work hard to manually collect and correct a huge number of sample images for training remains a significant limitation[2]. Methods that rely on artificial training data are recommended. Therefore, a data augmentation technique is proposed called CycleGAN along with CNN to enhance the accuracy of symbol recognition.

The rest of this paper is organized as follows. Section II reviews the associated literatures with the research topic on pattern recognition approaches. Section III illustrates the proposed framework on enhancing the accuracy of recognition rate. Section IV presents the experimental results. Section V interprets the significant findings along with discussion. Section VI concludes the paper and mentions the future works in this study.

II. RELATED WORK

A. Symbol Recognition Trend

Symbol recognition and spotting is an innovative computer vision technology. This technology aims to replicate portions of the human visual system’s complexing, able to allow computers to recognize and interpret within images or videos. Engineering benefits mankind and technology symbolizes the future. Over the recent years, numerous challenges of computer vision were gradually emerged. Surprisingly, many

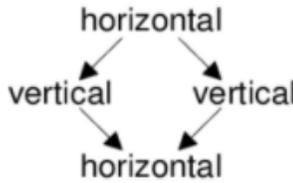
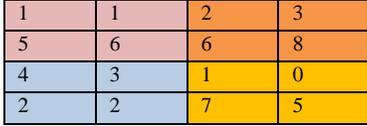
great efforts and solution have been cracked during the evolution timeline.

Table I summarizes the research trend for symbol spotting in the recent years. Symbol Spotting is an active topic in graphical symbol recognition and document analysis. In fact, some of the research papers are focusing on the sub-sections of the issues which are including symbol recognition, symbol detection, and primitive extraction. Those papers are altering from traditional symbol recognition towards the new conceptual of symbol recognition gradually. Symbol spotting has been enhanced in recent years persistently.

There is a demand to design an engineering drawings symbol recognition method compared to few approaches in Table I. Based on the review of history, the state-of-the-art approaches are keep replacing with better algorithms.

Statistical approach has been used precedent for pattern recognition since it is simple to manage. This approach mainly based on statistics and probabilities. Each pattern is obtained in terms of feature collection. The effectiveness of quality pattern depends on the set of feature extraction or measurement. It includes the pre-processing, and the segmentation is not required compared to local pattern representation. The decision boundary is established in the feature extraction via the analysis of probability distribution. Thus, different patterns allocate to the classes respectively. However, this approach has fault tolerant to image distortion since they tend to filter out small change in details.

TABLE I. EVOLUTION OF SYMBOL RECOGNITION TREND

Category approach	Pattern Recognition	Survey
Statistical: statistical features (geometric moments, R-signatures, etc.) + similarity measures	Number of segments: 4 Number of horizontal segments: 2 Number of vertical segments: 2 Number of diagonal segments: 0	Santosh et al. [3-5], 2015, 2016, 2018
Structural: graph-based matching + structural descriptors (visual primitives/ their spatial relations)		
Syntactic: inductive learning + programming spatial predicates (primitives’ relations)		
Semiautomatic and heuristic-based method in Convolutional Neural Network (CNN): Hough transform + connected component analysis + criterion	Pixel Values 	Elyan et al. [6], 2018
Clutter-tolerant cross-correlation for template matching		Rezvanifar et al. [7], 2019
Deep Learning Network in Convolutional Neural Network (CNN)		Yu et al. [8], 2019
Spatial Transformer Network (STN) in Convolutional Neural Network (CNN)		Y.Zhang[9], 2019
RPN method in R-CNN		Yun et al.[10], 2020
GNL method for both the small- and large-symbol networks		Kim et al.[11], 2021

In graphical documents, many objects can be graphically described, especially symbols [12]. Through structural recognition approach for symbols, the patterns can be represented as graphs. The characteristics of patterns are formed based on structure. Recognizing symbol is normally applied with structural recognition but this approach also can implement with other objectives such as learning, data structuring, indexing and others. It is also known as syntactic recognition occasionally since both are using the same formal language theory. Structural depends on the grammars to differentiate between data from several groups based on morphological interrelationships contained within the data. The sub patterns and the interactions between them that make up the data are represented by structural features, also known as primitives. However, recognize the primitives occur many troubles which primarily must focus on segmentation of noisy patterns and gram-mar inference from training data.

Syntactic recognition uses the structure of patterns and the syntax of language to construct a shape. The patterns are depended on the sentences in a language. The patterns are considered as sentences in a language, the primitives are viewed as the language's alphabet, and the sentences are formed using a grammar [12]. Thus, few numbers of primitives and grammatical rules are able to express complex patterns.

Structural and syntactic techniques are much closer to human comprehension. These techniques outperform the statistical in the case of complex symbols. However, the accuracy is obstructed the performance of the vectorization process, and matching based on graph representation which requires high computation cost.

Convolutional neural network (CNN) is a type of artificial neural network. This technique is the most popular to be applied in the recent year [8, 9, 13]. The employment of CNN is not restricted in image classification, recognition, and detection but also in Natural Language Processing (NLP). Therefore, the extensiveness of CNN application is broadly. The CNN achievements are also extravagance; produced many earth-shattering results in computer vision.

B. CNN Architecture in Image Processing

Most of the CNN architectures are based on supervised learning. A large amount of data is needed as a training sample to obtain more accurate classification and recognition in training process. Through enough training on data, CNN can overcome the limitations of the conventional methods such as inter-object interruption, morphological change, and noise problems [8]. However, some samples are difficult to collect. For example, the specific symbols in P&IDs drawings. It is extremely difficult to collect samples due to the limitations of the conditions.

CNN architecture model can be separated into two components which containing features extraction and classification. Feature extraction is performed by the convolution and pooling layers. Detect meaningful features in an image is a complicated task. The convolution layers have to learn such sophisticated features to illustrate the pattern shapes via pixels. Nan S. [14] experimented several feature extraction

algorithms at CNN model. FCN-CRF achieved highest average accuracy for image segmentation and average intersection-over-union compared with Support Vector Machine (SVM), K-means and FCN algorithms.

The few novelties [1, 6, 8] resulted the limited amount of training data for engineering drawing lower the accuracy of recognition. In additional, the lack of datasets occurs the class imbalance problem in classification.

C. Review Data Augmentation

Although more data samples can enhance the machine learning models, in fact collecting more data is not a best solution which channelling into a research topic. Data augmentation is a technique which providing more quantities of synthetic samples. This allows the algorithm to recognize and detect the specific components in an image precisely. Lan Goodfellow [15] proposed a framework called Generative Adversarial Network (GAN). The benefit of GAN can generate more synthetic samples to overcome the shortage and elevate the accuracy of object spotting. GAN is a type of machine learning model which consists of two neural networks compete with each other in order to provide more realistic image in the prediction.

GAN has been implemented in many applications previously for synthesizing more quality images and adding more training data in several studies. Shrivastava et al. [16] developed a GAN-based refiner network and enhanced the realism of simulated eye I mages by developing a GAN-based refiner network, resulting a 21% improvement in performance of an eye gaze estimation algorithm. In the field of medicine analyses, the shortage of image data is also commonly happened due to the lack of available images. Therefore, GAN play a role to synthesize realistic training data for liver lesion images [17], and brain MR images [18].

However, GAN training is highly unsteady because the discriminator and generator training needs to be delicately balanced. A common failure from mode collapse can be happened if the discriminator is too fierce or overwhelms the generator early during training, which results in convergence to a bad local optimum. Therefore, a new technique needs to be discovered instead of replacing GAN.

D. CycleGAN

CycleGAN is a subset of GAN technique which generating the automatic training of image-to-image translation models with unpaired examples. Some related works have been explored recently at the below.

Liu et. al. [19] applied stratified CycleGAN in medical images that generated graded variation in image quality. They resulted quality synthetic images using CycleGAN method. Zhang et. al. [20] presented a novel about road extraction method in generative adversarial network using Aerial images, require few samples and resulted better performance compared with several state-of-the-art techniques in term of detection accuracy. Park et. al. [21] implemented Dense-Net based framework in CycleGAN have eliminated data imbalance issue and deliver a better detection accuracy for wildfires images. Liu et.al. [22] equipped CycleGAN strategy to address ghosting problem. This work synthesizes video

context information and captures interframe stability better. Thus, CycleGAN is one of the impacts on the accuracy of recognition and spotting. The accuracy of the recognition can be influenced by producing more variance images.

III. METHODOLOGY

A. Symbols Dataset – Piping and Instrument Drawings (P&IDs)

Engineering drawings consists of several types of fields. Piping & Instrument Drawings is appointed to be our scope of study. The total number of collections from Piping & Instrument Drawings (P&IDs) is restricted into 7 sheets only due to limited public datasets available. However, this can highlight the problem then figure out a better solution.

In these experiments, only 7 types of symbols are extracted from the P&IDs. These types of symbols which including (a) Check valve, (b) Gate valve, (c) Gate_NC valve, (d) Globe valve, (e) Globe_NC valve, (f) Concentric reducer, (g) Weldcap. These symbols are shown in Fig. 1.

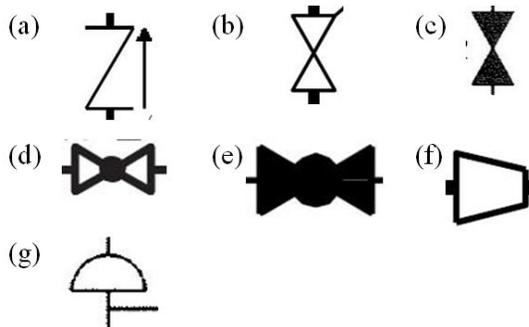


Fig. 1. Types of Symbols Extracted from P&IDs.

In additional, a total number collection of extracted symbols from P&IDs are 1,873. The following Fig. 2 summarized the number of symbols distributed for each class.

Each P&ID sheet displays different qualities of components. Qualities is one of the factors affects the accuracy for spotting. More data information with various qualities can provides the confidence to recognize the symbols precisely. Therefore, the limited symbols from P&IDs will conduct with advanced data augmentation technique to deliver more various qualities images.

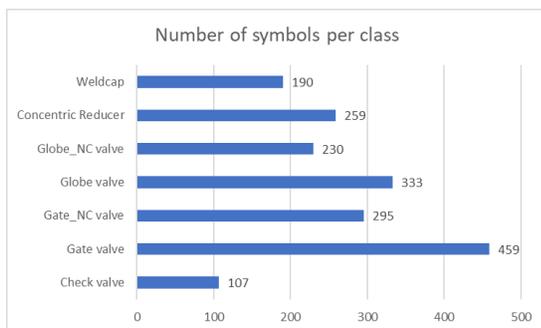


Fig. 2. Class Distribution of Symbols in Whole Dataset.

B. Symbol Recognition – CNN Algorithm

Symbols require an algorithm tool for recognizing. Therefore, a basic Convolution Neural Network (CNN) algorithm to be implemented. The implementation of CNN able to predict several class probabilities and bounding boxes simultaneously.

The CNN architecture involves several convolutional layers, max pooling layers and fully connected layers. A tensor which representing the data structure is required to proceed through several convolutional layers. Then, it is converted to a vector and then transmitted through a dense layer. The overview of CNN architecture is displayed in Fig. 3.

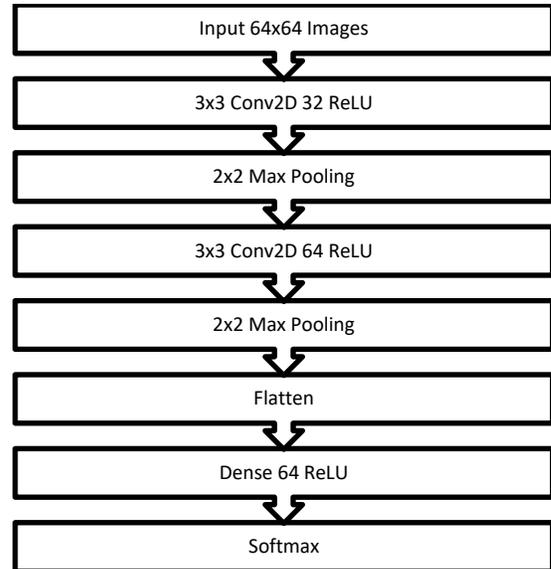


Fig. 3. Basic of CNN Architecture.

Input Images: All the extracted symbols from P&ID drawings and synthetics symbols are converted into 64x64 pixels.

$M \times M$ Conv2D N : In this basic CNN architecture, 2D convolution layer is used, this layer creates a convolution kernel or filter over the input data and perform an elementwise multiplication to produce a tensor of output pixel. Kernel is a convolution matrix or filter that used for features extraction in image processing including blurring, sharpening, embossing, edge detection, and others. $M \times M$ represents the size of filters, and N represents the numbers of filters that the convolution layer learns.

ReLU: Stands for Rectified Linear Unit. Commonly used as an activation function of the CNN layers and the fully connected layers because it is simple, fast, and empirically converge quickly and reliably when training a deep network.

$M \times M$ Max Pooling: Extract each patch from input featured maps. Generally, a filter of size 2x2 with a stride of 2 is implemented. The pooling is downsampling the feature map then spotlight the conspicuous feature in each patch.

Flatten: Also known as fully-connected layer for connecting the final classification model. Flatten the last output of feature map once the model has learnt the features. However, the flatten layer is used to convert the data into 1-dimensional array (1D) for the next layer's input.

Dense: A simple layer for neurons. Each neuron receives an input from the previous layer's neurons. This is used for classifying or predicting image based on the output from previous layers via shallow neural network.

Softmax: Commonly a final output layer in neural network. It is an activation function for generalization of the Sigmoid function but in multiple dimensions. It is used for performing multiclass classification and object recognition. Thus, it normalizes the output as probability distribution to each class.

C. Pre-processing

Prior to any work, the whole dataset containing different number of symbols for each class. All symbols were cropped into a size of 64 x 64 pixels.

According to previous related works, few novelties described the limitation of images. However, there is no findings specify an exact volume. Therefore, 3 methods were conducted in this study including Baseline CNN, CycleGAN + CNN, and Y.Zhang [9] algorithm as for comparison. Each method conducts several experiments with all the extracted symbols from P&IDs drawings. These symbols are classifying into three categories of datasets which including training dataset, validation dataset, and testing dataset. In these experiments, different ratios for training dataset are implemented while validation dataset and testing dataset remain unchanged. The highest ratio of training dataset is up to 4 and lowest ratio is 0.5. The scheme for classifying datasets and the description of each method is shown below:

In Fig. 4, the scheme illustrates the division for training dataset, validation dataset, and testing dataset. The ratio in whole dataset is 1:1:1. The training dataset is further using to implement on the synthetic models which described in Method 2 and Method 3 in the following paragraph. The synthetic images are generated via a model by applying the training dataset with different ratios.

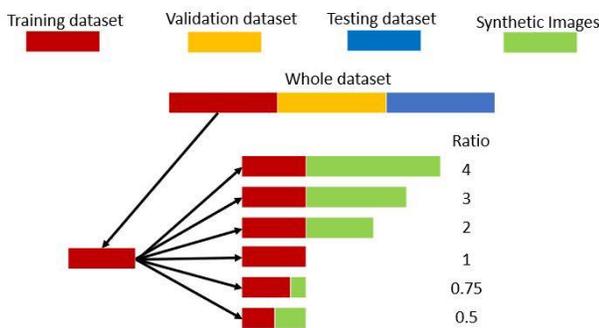


Fig. 4. Dataset Division.

Method 1: Baseline CNN

In baseline CNN method, only extracted symbols of P&IDs to be implemented. The standard CNN algorithm is

used to determine the symbol recognition rate against Method 2 and 3. The numbers of extracted P&ID symbols are calculated as ratio. TABLE II shows the ratio of validation dataset and testing dataset were fixed into 1 while the ratio of training dataset was keep reducing to validate the diversity accuracy of symbol recognition.

TABLE II. BASELINE CNN DATASETS

Baseline (Datasets)		
Training	Validation	Testing
N/A	1	1
N/A	1	1
N/A	1	1
1 (100% training dataset)	1	1
0.75 (75% training dataset)	1	1
0.5 (50% training dataset)	1	1

Method 2: CycleGAN + CNN

Same as Method 1, all extracted P&ID symbols split into training dataset, validation dataset, and testing dataset. However, this method involved synthetic images generated via CycleGAN model to enlarge variation symbols. The synthetic images are added into training dataset for increasing and equalizing the ratio. Then, CNN is used to recognize and classifying the symbols.

In this TABLE III, different ratio numbers of training dataset were tested to discover the difference accuracy recognition of P&IDs symbols.

TABLE III. CYCLEGAN+CYCLEGAN DATASETS

CycleGAN (Datasets)		
Training	Validation	Testing
4 (100% training dataset + 300% synthetic samples)	1	1
3 (100% training dataset + 200% synthetic samples)	1	1
2 (100% training dataset + 100% synthetic samples)	1	1
N/A	1	1
0.75 (75% training dataset)	1	1
0.5 (50% training dataset)	1	1

Method 3: Y.Zhang[9] algorithm

In Method 3, Y.Zhang[9] algorithm was implemented as for the validation against other methods. Since the P&ID symbols used are different from the origin paper, the pattern of the symbols was customized to tally these extracted symbols of P&IDs while the parameters of algorithm remained unchanged in [9]. The feature of this algorithm is to augment synthetic images with various scales, rotations, and random noises. These experiments were applied the ratio of training dataset same as Method 2 in TABLE IV.

TABLE IV. Y.ZHANG [9] ALGORITHM DATASETS

Y.Zhang [9] algorithm (Datasets)		
Training	Validation	Testing
4 (100% training dataset + 300% synthetic samples)	1	1
3 (100% training dataset + 200% synthetic samples)	1	1
2 (100% training dataset + 100% synthetic samples)	1	1
N/A	1	1
0.75 (75% training dataset)	1	1
0.5 (50% training dataset)	1	1

D. Proposed Methodology

Data augmentation is necessary because the limited symbols unable to provide the confidence during recognizing. We proposed to employ CycleGAN [23] with Convolution Neural Network(CNN).

In Fig. 5, the overview of proposed framework is displayed. CycleGAN is an unsupervised deep learning method which conducting a bidirectional translation between the two source domains which are domain X and target domain Y. Generally, the images collection from the source domain and target domain are not requirement that they are associated in any manner. The characteristics of CycleGAN implements two generator networks and a discriminator network for appraisal. Generator (G) and Discriminator (D) networks compete with one another. D is a classifier which trying to distinguish the samples between the synthetic images that generated via generator (fake) and the actual distribution (real). On the other hand, role of G attempts to fool the discriminator by producing synthetic output image. The input of the generator is along with source domain image x from Domain X and its output is a synthetic image. In additional, the inputs of a discriminator D are the synthetic output and an unpaired random image from the target image y of domain Y.

Even though CycleGAN allows unpaired random image but we apply a set of symbol images from a random class for consistent comparison.

Each generator in CycleGAN model involves an encoder, a transformer, and a decoder. Role of the generator ensures the features of images are extracted and converted into transformer (latent space). Transformer uses an attention mechanism to transfer the sequence to decoder. Then, a new feature vector of image is converted and reconstructed as output image in decoder.

In cycleGAN, the discriminator model is implemented as a PatchGAN model [23] which aims at classifying images as real or synthetic. The discrimination starts undergoing an appraisal with identifying the image belongs to real or synthetic. For discrimination, we apply adversarial losses [23] into both mapping functions to match up between the output generated images and real images from domain Y . For the mapping function $G: X \rightarrow Y$ and the discriminator Domain Y, where Generator(G) seeks to generate images $G(x)$ which looking closely images of domain Y, while discriminator D_Y is applied to distinguish between generated images of $G(x)$ and real images y of domain Y. G aims to minimize this training objective but D that tries to maximize it. Therefore, adversarial loss function is used to inverse mapping the first generator resulting $\min_G \max_{D_Y} L_{GAN}(G, D_Y, X, Y)$. A same adversarial loss for the mapping function on second generator (F). According to Figure 5, $F: Y \rightarrow X$ and its discriminator D_X as well resulting $\min_F \max_{D_X} L_{GAN}(F, D_X, X, Y)$.

Adversarial training learns mappings between G and F to generate the outputs closely to target domains Y and X respectively. Nevertheless, adversarial loss alone is uncertain to map the learned function onto an individual input x_i to a desired output y_i . but stimulate the tolerance of the cycle consistency. The cycle consistent argued the learned mapping functions to be further narrowed the space of the possible mapping functions for each image x of domain X. Thus, the image proceeds with translation cycle function to return image.

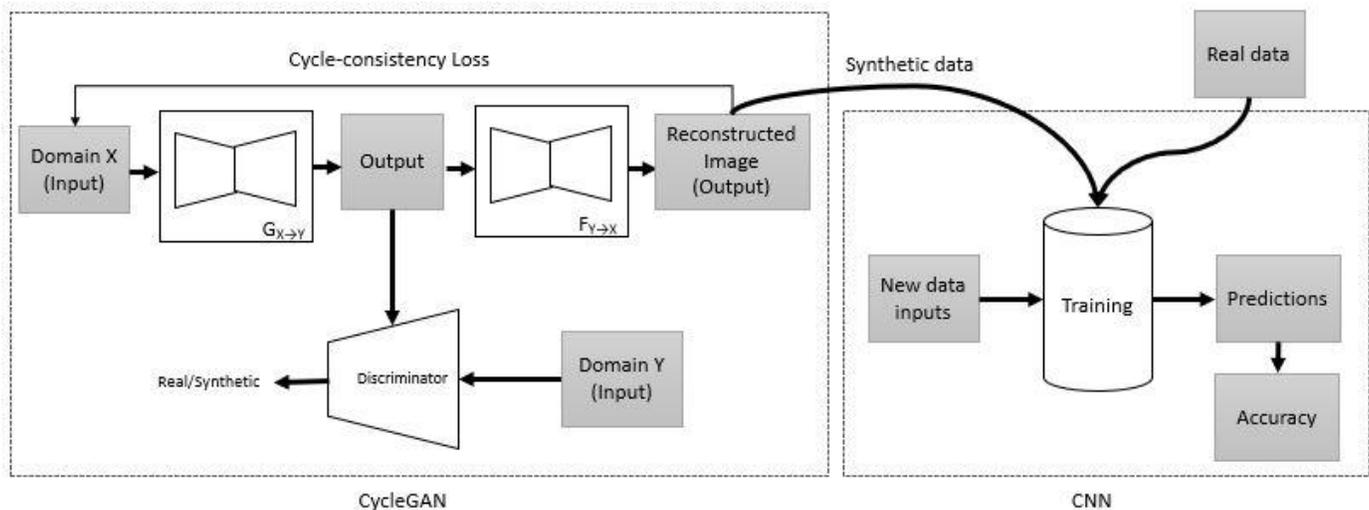


Fig. 5. Proposed Framework: CycleGAN (Generate Synthetic Samples) + CNN (Symbols Recognition).

x back to the original image intimately which resulting as $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. This referred to forward cycle consistency. Besides, every image y from domain Y , G and F should repeat with backward cycle consistency resulting $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. This function is impelled by a cycle consistency loss.

Algorithm 1: CycleGAN Pseudocode

```

1: procedure TRAINCYCLEGAN
2: Select samples  $\{x_i\}$  from domain  $X$ 
3: Select samples  $\{y_i\}$  from domain  $Y$ 
4: Compute generator  $G: X \rightarrow Y$ 
5: Compute the discriminator loss on Domain  $Y$ :

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)]$$


$$+ \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (1)$$

6: Compute generator  $F: Y \rightarrow X$ 
7: Compute Cycle-consistency loss:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1]$$


$$+ \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (2)$$


```

The generated synthetic samples via CycleGAN model to be transferred into second stage for CNN recognition predictions. In CNN stage, mixing synthetic samples and real training data to increase the numbers of dataset for training model. This adding more various patterns and finetune the training model to achieve more accurate performance.

E. CycleGAN + CNN Hyperparameters

The hyperparameter setups for CycleGAN [23] and CNN model were displayed in TABLE V and TABLE VI. These setups were standard in the Tensorflow implementation.

TABLE V. HYPERPARAMETERS FOR CYCLEGAN

Hyperparameters CycleGAN	
Learning rate	0.002
Epochs	50
Beta	0.5
Adversarial loss	L2 Distance
Cycle loss	10
Identity loss	5

TABLE VI. HYPERPARAMETERS FOR CNN

Hyperparameters CNN	
Learning rate	0.01
Batch size	32
Epochs	60

IV. EXPERIMENTS AND RESULTS

A. Evaluation Metrics

The evaluation metrics is implemented with the accuracy of recognition and confusion matrix.

According to the accuracy of recognition metric, Equation 3 is shown below:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

In this equation contains four important terminologies. TP represents the number of positives, TN represents the number of negatives, FP denotes the number of false positives, FN denotes the number of false negatives. The formula of Equation 3 is the total number of samples are divided by the sum of true positives and true negatives gives an overall assessment of the classifier performs across every classes.

Furthermore, another evaluation metric is confusion matrix, also called error matrix. It is a summarized table that is used to evaluate a classification model's performance. The number of positives and negatives predictions are totaled and broken down by class using count values. In confusion matrix table, predicted classifications are indicated by rows while true positives are indicated in every classes respectively by columns.

B. Experiments Results

The experiment results are displayed in TABLE VII. This table showed:

According to TABLE VIII, the results summarized the highest average accuracy is up to 92.85% which belonging to CycleGAN + CNN method. Compared to baseline CNN method with a ratio of 1:1:1, there are no synthesis images or symbol reduction in the dataset, the average accuracy is only 90.75%.

Furthermore, it is undeniable that the average accuracy is decreasing when reducing the number of training symbols in the Baseline CNN method. However, the CycleGAN+CNN method still has higher average accuracy at ratios of 0.5 and 0.75 as a training dataset compared to the Baseline and Y.Zhang [9] algorithm. In contrast, the CycleGAN model was able to generate a large number of synthetic images, but a ratio of 3 and a ratio of 4 average accuracy showed a significant drop in results as the training dataset size increased.

In Y.Zhang [9] method, the average accuracy in all ratios of training dataset are significantly lower compared to Baseline CNN and CycleGAN + CNN methods. These results showed that not all methods that can generate synthetic images improve accuracy due to several factors.

In Fig. 6, a clustered chart displayed the accuracy over the different ratio of training dataset based on the values from TABLE VIII. Various colors represent different methods. CycleGAN + CNN method achieved the peak accuracy values against another methods with all different ratios of training dataset.

TABLE VII. EXPERIMENTAL RESULTS - RECOGNITION ACCURACIES

Dataset Ratio			Baseline CNN (Without synthetic samples)					CycleGAN + CNN					Y.Zhang [9]				
Training	Validation	Testing	Test 1	Test 2	Test 3	Test 4	Test 5	Test 1	Test 2	Test 3	Test 4	Test 5	Test 1	Test 2	Test 3	Test 4	Test 5
4	1	1	N/A	N/A	N/A	N/A	N/A	90.98%	90.02%	90.66%	91.47%	89.21%	73.27%	72.79%	74.07%	75.04%	73.11%
3	1	1	N/A	N/A	N/A	N/A	N/A	91.30%	92.59%	91.79%	92.59%	91.14%	73.91%	73.91%	74.40%	73.59%	74.72%
2	1	1	N/A	N/A	N/A	N/A	N/A	92.92%	93.56%	92.75%	92.59%	92.43%	76.49%	75.52%	79.55%	77.62%	76.17%
1	1	1	91.46%	90.49%	90.98%	89.04%	91.78%	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.75	1	1	89.21%	88.72%	90.66%	90.98%	89.69%	89.69%	91.30%	90.66%	92.11%	90.98%	83.25%	86.63%	87.28%	84.22%	87.18%
0.5	1	1	88.56%	88.24%	87.27%	83.89%	87.60%	88.08%	87.92%	89.21%	86.47%	91.30%	77.30%	77.13%	73.43%	75.52%	76.49%

TABLE VIII. AVERAGE AND STANDARD DEVIATION OF RECOGNITION ACCURACIES FOR FIVE TIMES REPEATS

Dataset Ratio			Baseline CNN (Without synthetic samples)		CycleGAN + CNN		Y.Zhang [9]	
Training	Validation	Testing	Avg. Accuracy (%)	Avg. Std (%)	Avg. Accuracy (%)	Avg. Std (%)	Avg. Accuracy (%)	Avg. Std (%)
4	1	1	N/A	N/A	90.47%	0.88%	73.66%	0.91%
3	1	1	N/A	N/A	91.88%	0.69%	74.11%	0.45%
2	1	1	N/A	N/A	92.85%	0.44%	77.07%	1.58%
1	1	1	90.75%	1.07%	N/A	N/A	N/A	N/A
0.75	1	1	89.85%	1.10%	90.95%	0.89%	85.71%	1.85%
0.5	1	1	87.11%	1.87%	88.60%	1.80%	75.97%	1.58%

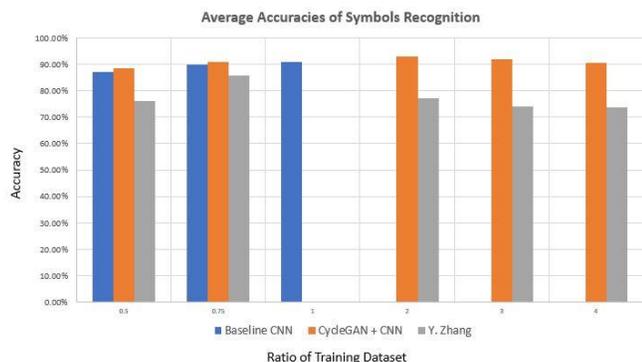


Fig. 6. Cluster Chart for Average Accuracies of Symbol Recognition.

The confusion matrix between Baseline method and CycleGAN method were shown in TABLE IX and TABLE X.

According to TABLE IX and TABLE X, each alphabet represents a type of symbol which is displayed in Section III. Both tables highlighted the difference of symbol recognition. As per comparison, (A) Check valve, (B) Gate valve, and (F) Concentric reducer showed the significant gaps. One of the reasons could be the variation issues. It requires more variation symbols to match with the P&ID symbols in testing dataset. In fact, CycleGAN model generate more variation synthetic images allow CNN architecture to recognize precisely especially these three types of symbols.

TABLE IX. CONFUSION MATRIX FOR THE BASELINE CNN MODEL

	A	B	C	D	E	F	G
A	19	1	0	0	0	4	11
B	0	137	13	0	0	0	3
C	0	0	96	1	1	0	0
D	0	0	0	111	0	0	0
E	0	0	0	0	76	0	0
F	1	0	0	0	0	55	30
G	3	0	0	0	0	0	59

TABLE X. CONFUSION MATRIX FOR THE CYCLEGAN + CNN MODEL

	A	B	C	D	E	F	G
A	23	0	0	0	0	3	9
B	0	146	5	0	0	0	2
C	0	0	97	0	0	0	0
D	0	1	0	110	0	0	0
E	0	0	0	0	76	0	0
F	1	0	0	0	0	70	16
G	2	0	0	0	0	1	59

V. DISCUSSION

According to Section IV, several experiments were tested on different methods. Our proposed framework resulted the combination of CycleGAN and CNN performed effectiveness on accuracy of recognition. Highest average accuracy up to 92.85% was achieved. The result showed CycleGAN the potential when trained on the synthetic samples. It performed better than the algorithm used in Y.Zhang [9]. Using CycleGAN model, generated synthetic samples increased the dataset for training. This provides CNN model better classifying and recognizing with sufficient training. Nevertheless, the confusion matrix of our proposed framework displayed a greater total number of true positives which recognized the specific symbols accurately.

Another interesting aspect from the results, different ratios of training dataset were conducted during experiments testing. Significantly, the accuracy rate of recognition gradually lower down when applying ratio 0.5:0.75 as the training data samples. The work proved the superior of synthetic samples from proposed method. Added CycleGAN synthetic samples to increase back to 100% or ratio 1 of the training dataset. It performed great contribution to accuracy rate even achieved higher than the standard 100% of training dataset from baseline CNN method without adding any synthetic samples. In addition, the incremental ratio of synthetic samples is scaled up to 400%. The best accuracy rate was performed with ratio 2 while ratio 3:4 start declined the accuracy of recognition.

VI. CONCLUSION AND FUTURE WORKS

In this study, the accuracy enhancement for recognizing symbols of engineering drawings required a data augmentation technique. We proposed CycleGAN + CNN method for synthetic symbols to enlarge the quantities of symbols in Piping & Instrument Diagrams (P&IDs). Our work addresses the lack of labelled data in deep networks by utilizing CycleGAN to generate synthetic images to supplement training data. We achieved the highest average accuracy of this method as high as 92.85%, which is a significant enhancement over CNN recognition alone. However, excessiveness or insufficient images can reduce the accuracy of recognition rate due to some factors such as overfitting and underfitting.

Future work is schemed to apply spatial transformations in CycleGAN architecture. If there are too many changes, it is difficult for the generator to learn the basic style efficiently. Additionally, spotting all types of symbols in an engineering drawing. An advanced object detection algorithm will be studied especially design for symbols. This will allow to detect symbols precisely regardless the pixel of engineering drawings and the size of symbols.

ACKNOWLEDGMENT

This research/paper was fully supported by Universiti Teknologi PETRONAS, under the Yayasan Universiti Teknologi PETRONAS (YUTP) Fundamental Research Grant Scheme (YUTP-FRG/015LC0-280).

REFERENCES

- [1] E. Elyan, L. Jamieson, and A. Ali-Gombe, "Deep learning for symbols detection and classification in engineering drawings," Elsevier, vol. 129, pp. 91-102, 1 June 2020.
- [2] C. F. E. Moreno-García, E.; Jayne, C., "New trends on digitisation of complex engineering drawings," *Neural Comput. Appl.*, vol. 31, pp. 1695-1712, 2019.
- [3] S. K., "Complex and composite graphical symbol recognition and retrieval: a quick review," in In: *International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R'16)*, 2016: Springer, Singapore, pp. pp 3-15.
- [4] W. L. Santosh K., "Graphical symbol recognition," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015.
- [5] S. K., "Document image analysis," in In: *Document Image Analysis*, 2018: Springer, Singapore, pp. pp 1-15.
- [6] E. Elyan, Garia, C.M., Jayne,C., "Symbols Classification in Engineering Drawings," presented at the International joint conference on neural networks 2018 (IJCNN), Rio de Janeiro, Brazil, 8-13 July 2018, 2018.
- [7] A. Rezvanifar, M. Cote, and A. Branzan Albu, "Symbol spotting for architectural drawings: state-of-the-art and new industry-driven developments," *IPSN Transactions on Computer Vision and Applications*, vol. 11, no. 1, p. 2, 2019/05/10 2019, doi: 10.1186/s41074-019-0055-1.
- [8] E. Yu, Cha, J.M., Lee, T., Kim, J., Mun, D., "Features Recognition from Piping and Instrumentation Diagrams in Image Format Using a Deep Learning Network " 2019.
- [9] Y. Zhang, "CNN-based Symbol Recognition and Detection in Piping Drawings," 16-Aug-2019.
- [10] D.-Y. Yun, S.-K. Seo, U. Zahid, and C.-J. Lee, "Deep Neural Network for Automatic Image Recognition of Engineering Diagrams," *Applied Sciences*, vol. 10, no. 11, p. 4005, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/11/4005>.
- [11] H. Kim et al., "Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams," *Expert Syst. Appl.*, vol. 183, p. 115337, 2021.
- [12] M. Delalandre, É. Trupin, and J.-M. Ogier, "Symbols Recognition System for Graphic Documents Combining Global Structural Approaches and Using a XML Representation of Data," Berlin, Heidelberg, 2004: Springer Berlin Heidelberg, in *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 425-433.
- [13] W. Fang, F. Zhang, V.-S. Sheng, and Y. Ding, "A Method for Improving CNN-Based Image Recognition Using DCGAN," *Computers, Materials & Continua*, vol. 57, no. 1, pp. 167-178, 2018. [Online]. Available: <http://www.techscience.com/cmcc/v57n1/22963>.
- [14] N. Shuping, "Feature Extraction and Segmentation Processing of Images Based on Convolutional Neural Networks," *Optical Memory and Neural Networks*, vol. 30, no. 1, pp. 67-73, 2021/01/01 2021, doi: 10.3103/S1060992X21010069.
- [15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, Warde-Farley, and D. e. al., "Generative adversarial nets.," *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014.
- [16] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2107-2116, 2017.
- [17] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321-331, 2018.
- [18] C. Han et al., "Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images," *arXiv preprint arXiv:1902.09856*, 2019.
- [19] J. Liu, J. Li, T. Liu, and J. Tam, "Graded Image Generation Using Stratified CycleGAN," *Cham*, 2020: Springer International Publishing, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 760-769.

- [20] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, and L. Jiao, "Aerial Image Road Extraction Based on an Improved Generative Adversarial Network," *Remote Sensing*, vol. 11, no. 8, p. 930, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/8/930>.
- [21] M. Park, D. Q. Tran, D. Jung, and S. Park, "Wildfire-Detection Method Using DenseNet and CycleGAN Data Augmentation-Based Remote Camera Imagery," *Remote Sensing*, vol. 12, no. 22, p. 3715, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/22/3715>.
- [22] S. Liu, H. Wu, S. Luo, and Z. Sun, "Stable Video Style Transfer Based on Partial Convolution with Depth-Aware Supervision," *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.
- [23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired imaged-to-image translation using cycle-consistent adversarial networks," In *IEEE International Conference on Computer Vision (ICCV)*, 2017.