

## 2020 Special Issue

## Deep learning for symbols detection and classification in engineering drawings



Eyad Elyan\*, Laura Jamieson, Adamu Ali-Gombe

School of Computing Science and Digital Media, Robert Gordon University, UK

## ARTICLE INFO

## Article history:

Available online 1 June 2020

## Keywords:

Deep learning  
YOLO  
P&ID  
Engineering drawings  
Symbols recognition  
GANs

## ABSTRACT

Engineering drawings are commonly used in different industries such as Oil and Gas, construction, and other types of engineering. Digitising these drawings is becoming increasingly important. This is mainly due to the need to improve business practices such as inventory, assets management, risk analysis, and other types of applications. However, processing and analysing these drawings is a challenging task. A typical diagram often contains a large number of different types of symbols belonging to various classes and with very little variation among them. Another key challenge is the class-imbalance problem, where some types of symbols largely dominate the data while others are hardly represented in the dataset. In this paper, we propose methods to handle these two challenges. First, we propose an advanced bounding-box detection method for localising and recognising symbols in engineering diagrams. Our method is end-to-end with no user interaction. Thorough experiments on a large collection of diagrams from an industrial partner proved that our methods accurately recognise more than 94% of the symbols. Secondly, we present a method based on Deep Generative Adversarial Neural Network for handling class-imbalance. The proposed GAN model proved to be capable of learning from a small number of training examples. Experiment results showed that the proposed method greatly improved the classification of symbols in engineering drawings.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

Large volumes of un-digitised and paper-based documents are still very common across different domains. Amongst this legacy, engineering drawings are known to be one of the most complex types of documents to process and analyse. They are widely used in different industries such as construction and city planning (i.e. Floor Plan diagrams (Ahmed, Liwicki, Weber, & Dengel, 2012)), Oil and Gas (i.e. P&IDs (Elyan, Garcia, & Jayne, 2018)), Mechanical Engineering (Vaxiviere & Tombre, 1992), AutoCAD Drawing Exchange Format (DXF) (Goh, Mohd. Shukri, & Manao, 2013) and others. Interpreting these drawings requires highly skilled people, and in some cases long hours of work.

In recent years, the digitisation of these drawings is becoming increasingly important. This is partly due to the urgent need to improve business practices such as inventory, assets management, risk analysis, safety checks and other types of applications, and also due to the recent advancements in the domain of machine vision and image understanding. Deep Learning (DL) (Goodfellow, Bengio, & Courville, 2016), in particular, had significantly improved the performance by orders of magnitude in

many domains such as Gaming and AI (Holcomb, Porter, Ault, Mao, & Wang, 2018), Natural Language Processing (Yang et al., 2016), Health (Esteva et al., 2019), and others. One particular domain that has benefited hugely from DL is machine vision (Girshick, 2015). Convolutional Neural Networks (CNNs) (Gu et al., 2018) have made significant progress in recent years in many image-related tasks (Ali-Gombe, Elyan, & Jayne, 2017). It has been successfully applied to several fields such as hand-written recognition (Lecun, Bottou, Bengio, & Haffner, 1998), image classification (Krizhevsky, Sutskever, & Hinton, 2017; Szegedy et al., 2015), Face Recognition & Biometrics (Park & Jain, 2010) and others. Before the CNNs, the improvements in image classification, segmentation, and object detection were marginal and incremental. However, the introduction of CNNs revolutionised this field. For example, Deep Face (Taigman, Yang, Ranzato, & Wolf, 2014), a face recognition system that was first proposed by FaceBook in 2014 achieved an accuracy of 97.35%, beating the state-of-the-art then, by 27%.

Core image processing tasks such as shape and object detection, recognition, and tracking have become much less challenging even under different conditions and in much less controlled environments. Faster Region-based CNN (R-CNN) (Ren, He, Girshick, & Sun, 2015), Single Shot Detectors (SSD) (Liu et al., 2015), Region-based Fully Convolutional Networks (R-FCN) (Dai, Li, He,

\* Corresponding author.

E-mail address: [e.elyan@rgu.ac.uk](mailto:e.elyan@rgu.ac.uk) (E. Elyan).

& Sun, 2016) and You Only Look Once (YOLO) (Redmon, Divvala, Girshick, & Farhadi, 2016) are all relatively recent methods that showed superior performance in the field of object detection, tracking, and classification. These methods and their extensions have significantly advanced this area of research and solved some of the most challenging and inherent vision problems such as occlusions, light conditions, orientation, and others, which were considered major challenges, even for a specific vision task in a more controlled environment (Zhao, Chellappa, Phillips, & Rosenfeld, 2003).

Significant advancement has also been made in the area of Generative Models and was successfully applied in many applications. Among these, Generative Adversarial Networks (GAN) proved to be one of the most established and commonly used methods in generating content. GANs were initially introduced by Ian Goodfellow in 2014 (Goodfellow et al., 2014). In the Methods section, we will discuss our GAN-based method to handle the class imbalance problem. This is another challenging problem that is common across many domains (Vuttipittayamongkol, Elyan, Petrovski, & Jayne, 2018) including engineering drawings, where one or more class of symbols in the diagrams are either underrepresented or overrepresented in the dataset (Vuttipittayamongkol & Elyan, 2020).

Despite this massive progress in the field of image processing and analysis, very little progress has been made in the area of digitising complex engineering drawings, and extracting information from these diagrams is still considered a challenging problem (Arroyo, Fay, Chioua, & Hoernicke, 2014). To date, a major problem of most of the existing solutions is that they still follow a traditional image-processing approach, which requires extensive features extraction and engineering and carefully designed heuristics (Moreno-García, Elyan, & Jayne, 2018). These are often very domain-dependent, sensitive to noise and data distribution, and mostly dedicated to solving part of the problem (i.e. detecting symbols, separating graphics from text, and so on). As can be seen in Fig. 1, not only such an approach difficult to generalise across different scenarios, but also the performance of any machine learning algorithm will hugely depend on the quality and accuracy of the extracted features.

In this paper, we propose an end-to-end framework for processing and analysing complex engineering drawings. We argue that the core task of such a framework is the accurate localisation and recognition of symbols in the drawing that constitute a major part of it and simplifies subsequent tasks (i.e. line and text detection). We show how one of the main inherent problems in classifying engineering symbols, namely class-imbalance can be addressed using Generative Adversarial Neural Networks. Fig. 2 provides a schematic diagram of the work presented in this paper. The main contributions of this work are outlined as follows:

- We propose a novel pipeline for processing and analysing complex engineering drawings. At the core of this pipeline is the accurate detection and recognition of symbols.
- We show that an advanced-bounding-box detection method performs very accurately on challenging engineering diagrams. To the best of our knowledge, Deep Learning models (e.g. YOLO Redmon & Farhadi, 2017, RCNN Girshick, 2015) were never used in such domain at a large scale of symbols with minimal difference. This is mainly due to the complexity of the problem, and the very little variation and noise within symbols of engineering drawings.
- Methods to handle the class-imbalance within engineering drawings are presented and thoroughly evaluated. We present a fine-grained method to train GAN models to generate engineering symbols of different overlapping classes.
- Thorough evaluation using large collection of P&ID diagrams provided by an industry partner in the Oil and Gas sector.

The rest of this paper is organised as follows: Section 2 presents an overview and critical discussion of relevant work. In Section 3, we present our methods, dataset and pre-processing steps carried out. In Section 4 we present our experiments and discuss results. Finally, conclusions and future work are outlined in Section 5.

## 2. Related work

In this section, we discuss relevant literature. First, we discuss literature related to the processing and analysis of engineering drawings. This will be followed by a brief introduction to Generative Adversarial Neural Networks and how it can be applied to handle the class imbalance problem.

### 2.1. Engineering drawings

An engineering drawing is a 2D image that contains different types of shapes, symbols, lines, and text. These drawings are commonly used in different domains and provide a rich representation of complex engineering workflows or situations. The digitisation of these drawings was subject to extensive research from the machine vision research community over the past four decades (Ablameyko & Uchida, 2007; Chhabra, 1997; Cordella & Vento, 2000; Zhang & Lu, 2004). In recent years, and due to the significant progress in machine vision research, computer power and also due to the availability of large volumes of un-digitised data, the demand to have a fully automated framework for digitising these drawings is becoming increasingly important.

Examples of work that aimed at extracting information from engineering documents include analysis of musical notes (Blostein, 1995), mechanical drawings (Kanungo, Haralick, & Dori, 1995), optical character recognition (OCR) (Kulkarni & Barbadekar, 2017; Lu, 1995; Mori, Suen, & Yamamoto, 1992), and extracting information from P&ID drawings (Elyan et al., 2018; Howie, Kunz, Binford, Chen, & Law, 1998; Moreno-García, Elyan, & Jayne, 2017). It can be argued that most of the existing literature followed a traditional image processing approach (Gonzalez & Woods, 2008), which requires some form of feature extraction from the image (Chhabra, 1997), features representations (Zhang & Lu, 2004), and classification to determine the class of objects (i.e. symbols, digits, ...) (Ablameyko & Uchida, 2007).

The key limitation of traditional machine vision methods is that they require extensive features engineering, depend heavily on the quality of extracted features, and often will not generalise well to other unseen examples. A recent extensive review showed that most of the existing literature focused on solving part of the problem rather than providing a fully automated framework for digitising an engineering diagram (Moreno-García et al., 2018). Examples include methods for recognising symbols and lines in a drawing (Boatto et al., 1992), detecting and separating text from symbols and other graphics elements in diagram (Moreno-García et al., 2017), classifying symbols in engineering drawings (Elyan et al., 2018) and so on. This is partly due to the complexity of the problem (i.e. localising every single element in the document), and also due to the limitations of the traditional image processing and analysis methods and the inherent vision problems such as the sensitivity to noise, quality of the image, the orientation of shapes and so on. Consider for example the work in Elyan et al. (2018), the authors used a set of heuristic rules to localise symbols in the drawings, a Random Forest (Elyan & Gaber, 2017) was then used to classify the symbols achieving an average accuracy higher than 95%. Similar work was presented in Moreno-García et al. (2017), where a set of heuristics were also used to detect and separate text from graphics elements. However, such an approach is very dependent on the data distribution, and

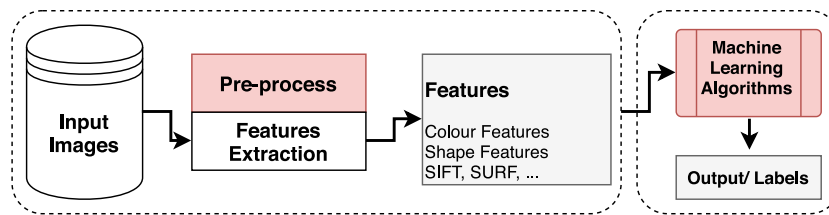


Fig. 1. Traditional frameworks for analysing images/ documents.

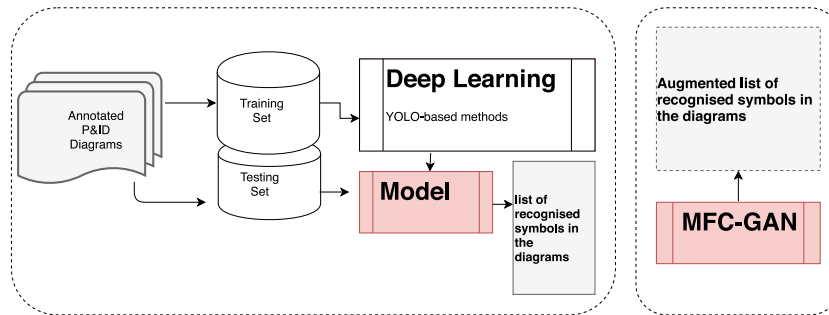


Fig. 2. Schematic diagram of the framework for processing and analysing engineering documents.

a slight variation in the diagrams or in symbols representation might require adapting the existing heuristics rules or creating new ones.

In a closely related area, [Rebello, Capela, and Cardoso \(2010\)](#) presented a study on optical music recognition and classification methods for musical symbols. They suggested that adjoining staff lines, presence of symbols in close proximity to music notes, broken symbols, overlapping symbols and areas with high symbol density all contributed to the complexity of optical music recognition. Four classification methods namely a multi-layer perceptron neural network model, Hidden Markov Model, K-nearest neighbour and Support Vector Machine (SVM) were evaluated on datasets of both synthetic and handwritten music scores. The highest performance was obtained with an SVM model, however all approaches implemented detection then removal of music staff lines and segmented the symbols prior to symbol classification.

[Khan, Ansell, Kuru, and Bilal \(2018\)](#), used video image analysis as part of a flight deck warning system, which combined automated dial reading of flight instruments with domain knowledge. Experiments on a flight simulator and real flight aimed to obtain the position of a white needle on the flight instrument using three image processing approaches: background subtraction, pattern matching and a convolution based approach. Results showed that the convolution method obtained the highest accuracy, highest true positive rate and highest true negative rate.

In recent years, DL-based methods were explored and successfully applied to some tasks that are similar to engineering drawings analysis. Ziran et al. proposed a method, based on Single Shot Detectors (SSD) ([Liu et al., 2015](#)), to detect and recognise furniture objects, doors, and windows in floor plan diagrams ([Pacha, Haji and Calvo-Zaragoza, 2018](#)). The results were encouraging. However, the datasets used were simple with a limited number of furniture objects in each drawing (12). The performance also dropped under the imbalanced class distribution of objects in the images.

Faster R-CNN was used in [Julca-Aguilar and Hirata \(2017\)](#) for the detection and recognition of handwritten characters. Although the work focused mainly on specific elements of the documents (mathematical expressions and flowcharts), promising results over other traditional methods were achieved.

Detection and recognition of musical notes in documents have also benefited from adopting Deep Learning-based methods ([Pacha et al., 2018](#); [Pacha, Haji et al., 2018](#)). R-CNN, R-FCN, and SDD were applied successfully to detect and recognise handwritten music notes ([Pacha, Choi et al., 2018](#)). Results showed an improvement in symbols detection over other traditional structured machine vision methods.

A framework for extracting information from P&ID drawings was presented very recently in [Rahul, Paliwal, Sharma, and Vig \(2019\)](#). The authors used a two-step approach. First, Deep Learning methods were used to localise symbols and text, and then heuristic-based methods were employed to detect other elements of the drawing (i.e. Euclidean metrics for associating tags and symbols with pipelines, probabilistic Hough transform to detect pipelines, etc.) The methods for localising symbols were based on a fully connected convolutional neural network. A dataset of four sheets consisting of 672 flow diagrams was used. Results were an improvement over other traditional methods. However, accuracy was not consistent across all components. Class accuracy ranged from 100% for some components to 64.0% for others (i.e. symbols of a certain class). Moreover, only a limited number of symbols were used in this study (10 different classes of symbols) and the P&ID sheets seem to be of a very good quality which is not often the case in the real world.

To summarise, existing literature shows a clear gap between the current state of machine vision and image understanding -*due to the rapid development in this field-* and the slow and incremental progress in a very important application domain across many industries.

## 2.2. GAN models

Generative Adversarial Networks (GAN) were initially introduced by Ian Goodfellow in 2014 ([Goodfellow et al., 2014](#)). These are considered as generative models that are capable of producing new content. GANs are made of two contesting models (i.e. Neural Networks, CNN's, etc...), the Generator (G), and the Discriminator (D). The discriminator is a classifier that receives input from the training set (authentic content), and from the generator (fake input). During the training process, the discriminator will learn how to distinguish between authentic and fake input

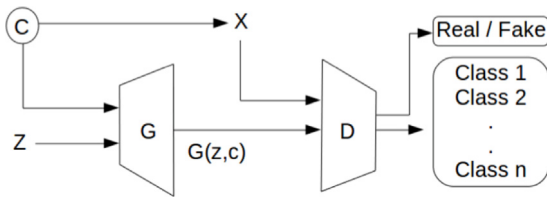


Fig. 3. Generative Adversarial Neural Networks.

samples. On the other hand, the generator is trained to generate samples that capture the underlying characteristics of the original data (replicating original content). Fig. 3 depicts the GAN model.

Adversarial training of both models  $G$ , and  $D$  is carried out using value function as can be seen in Eq. (1).

$$\min_D \max_G V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where  $p_{data}(x)$  is the probability distribution over the real data,  $x$  is a sample from the real training data,  $p_z$  is the probability distribution over the noise vector  $z$ , and  $G(z)$  is the output from the generator function  $G$  (or generated images). GANs are state-of-the-art in terms of the quality of the image generated.

GANs have been successfully applied to different problems including image generation (Ali-Gombe, Elyan, & Jayne, 2019; Ali-Gombe, Elyan, Savoye, & Jayne, 2018), segmentation and speech synthesis. In recent years they were also successfully applied to handle class-imbalance problems (Ali-Gombe & Elyan, 2019; Antoniou, Storkey, & Edwards, 2017). The class imbalance is common across different domains including health, security, and banking (Vuttipittayamongkol et al., 2018). The problem happens when one or more class is either underrepresented or overrepresented in the dataset. In such scenarios, a typical supervised learning algorithm tends to be biased towards the majority class when dealing with imbalanced datasets (Vuttipittayamongkol & Elyan, 2020).

Supervised GANs provide an extension to the original GAN framework by introducing conditional probabilities in the value function. This allows more control over the generated samples and introduces diversity which is needed for augmenting synthetic input data for class-imbalanced datasets. Typical examples include vanilla GAN (Goodfellow et al., 2014), CGAN (Mirza & Osindero, 2014) and AC-GAN (Odena, Olah, & Shlens, AUG 2017). Although the literature shows that these models can be hugely affected by class-imbalance especially in extreme cases (Mariani, Scheidegger, Istrate, Bekas, & Malossi, 2018).

Recent work appeared in Ali-Gombe and Elyan (2019) introduced a new extension of the GAN models. The authors trained the GAN models at a fine-grained level by updating the discriminator objective to not only distinguish between fake and real instances but also to classify the fake instances into different classes (i.e. Fake 1, Fake 2, etc...). Extensive experiments using four different datasets showed superior results over other GAN models. Generated samples proved to be of good quality and were successfully used to augment the dataset and improve the detection rate of minority class instances.

### 3. Methods

Most engineering drawings contain a set of symbols, connectivity information (lines) and some form of annotation (text). However, no public dataset is available for evaluation purposes. In Section 3.1 we introduce our approach for end-to-end symbols recognition from complex engineering drawings. The following subsection will discuss in detail the dataset used for experiments.

This will include data exploration and pre-processing. Finally, Section 3.4 provides details of our proposed method to handle class-imbalance in these drawings.

#### 3.1. Symbols recognition

For locating and recognising symbols in the P&IDs, we propose to use YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) method. This allows us to represent the problem as a set of bounding box coordinates and class probabilities. The method is based on dividing the entire image into  $S \times S$  grid, where each cell predicts  $B$  bounding boxes and confidence scores for those boxes (Redmon et al., 2016). The confidence scores are used to decide if a cell contains a symbol or not. These are represented as a five-dimensional vector ( $x, y, w, h$ , and confidence). Here, ( $x, y$ ) represents the centre of the bounding box, while the width and height are predicted relative to the whole image. The prediction from a grid is presented as  $S \times S \times (B * 5 + C)$ , where  $S$  is the size of the grid,  $B$  is the bounding box and  $C$  is the class probabilities (i.e. probability of the symbol being gate valve, sensor, etc...). Fig. 4 depicts this setting.

The YOLO model was chosen for two main reasons. First, it is a simple framework, which allows simultaneous predictions of multiple bounding boxes and class probabilities using a single convolutional neural network. Second, compared to other models, YOLO is considered extremely fast. For testing P&IDs that may contain on average 180 instances of various engineering symbols, this is very important in a practical context.

#### 3.2. Dataset - P&ID diagrams

For experiments in this paper, we chose to work with Piping and Instrumentation Diagrams (P&IDs) Fig. 5. A collection of 172 P&ID sheets were obtained from an Oil and Gas industrial partner for evaluation purposes. These diagrams contain different types of symbols, lines, and text (Fig. 5).

Additionally, the P&IDs are of different qualities, which makes the dataset suitable for evaluation purposes. The P&ID diagrams can be defined as schematic diagrams representing the different components of the process and the connectivity information. It is a representation of equipment (often depicted as symbols) and process flow (depicted as different types of lines) (Elyan et al., 2018).

Such diagrams are available across many industries in the form of paper or scanned documents. Interpreting and analysing these documents requires expert knowledge, and is often time-consuming (Arroyo, Hoang, & Fay, 2015; Moreno-García et al., 2018). Moreover, a misinterpretation of such documents can be very costly. For example, if a pipe needs to be replaced in an Oil and Gas installation, then an engineer needs to check the corresponding P&ID diagram, identify the valves that must be closed before carrying on the task to ensure safety. In other words, accurate interpretation of these drawings is paramount.

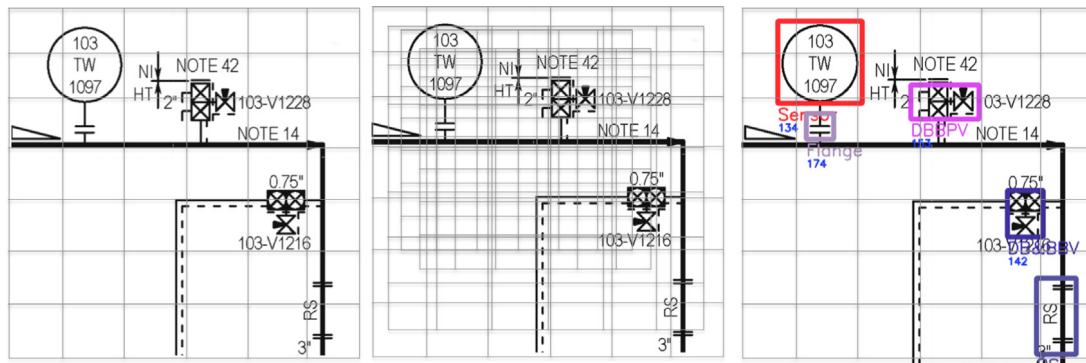
#### 3.3. Data exploration & pre-processing

The original P&IDs sheets are large images,  $7500 \times 5250$  pixels. To speed up the training process we divided the sheet into  $6 \times 4$  grid, resulting in 24 sub-images (patches) with relatively much smaller sizes compared to the original sheets ( $1250 \times 1300$ ).

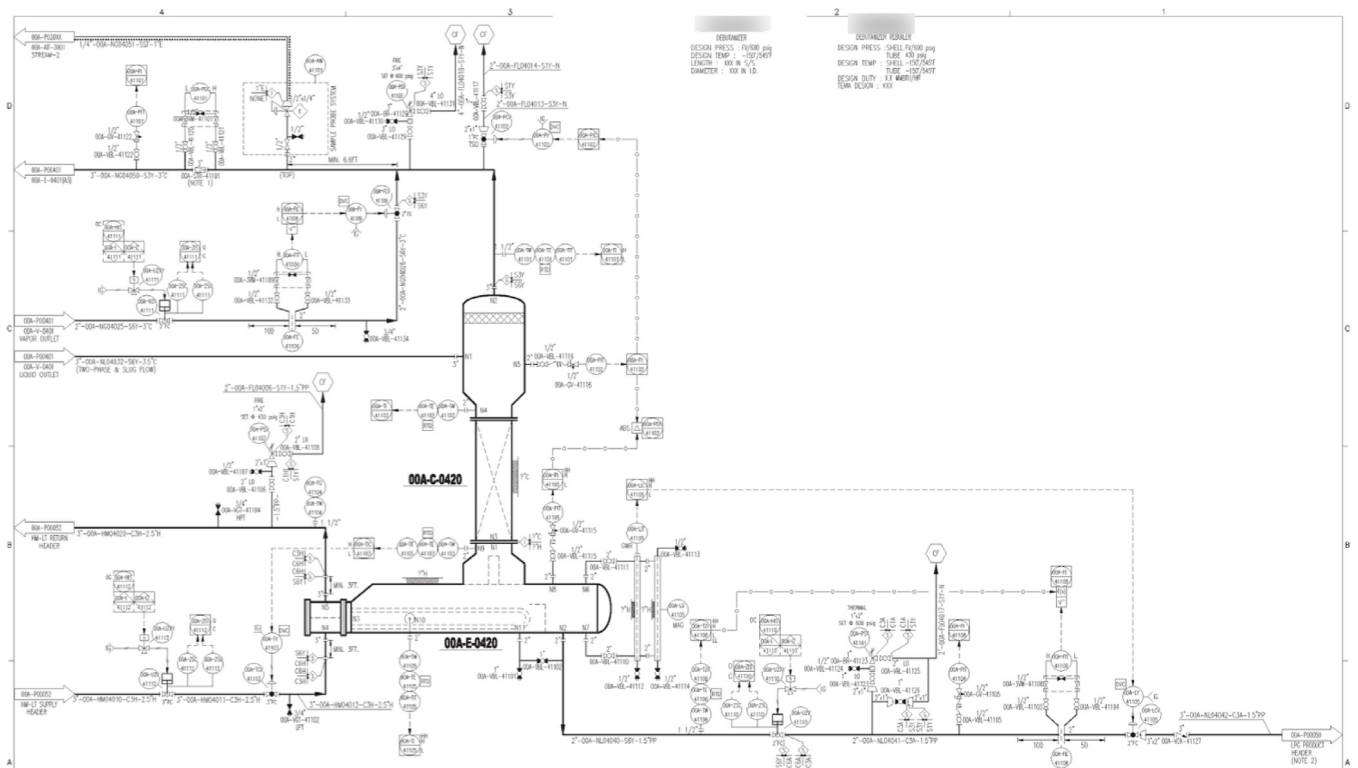
Training a Deep Learning model requires fully annotated images/ diagrams. To do so, we have used the Sloth tool<sup>1</sup> to annotate the collection of P&ID diagrams. In total 29 different symbols were annotated in the whole dataset (Fig. 7). The annotation

<sup>1</sup> <https://sloth.readthedocs.io/en/latest>.





**Fig. 4.** The method divides the P&ID Diagram into a grid, following the YOLO model (Redmon et al., 2016), and predicts the class probabilities of the bounding boxes. The figure shows the symbols sensor, flange, DBBPV, DB&BV and RS.



**Fig. 5.** Part of a P&ID Diagram.

process is simple and involves customising the sloth tool to record the corresponding symbols names (class) and its location in the diagram.

The resulting annotation of data is captured in a file representing the 29 symbols. Data recorded included the x,y coordinates of the centre of the bounding boxes, width and height of the bounding box enclosing symbols. In total, 13,327 symbols belonging to the 29 different classes were annotated. The dataset is hugely imbalanced as can be seen in Fig. 6.

Only 25 symbols of these were used in the experiments. These are shown in Fig. 7. Five symbols that were extremely under-represented in the whole dataset (i.e. only one or two instances of these symbols appear in the training and testing sets) were excluded from the first experiment.

### 3.4. MFC-GANs

To handle the class-imbalance in the dataset of engineering symbols (at the classification level), we are proposing to use a

method similar to the MFC-GAN model presented in Ali-Gombe and Elyan (2019). This model is chosen due to the very minor and in some cases subtle difference between different classes of symbols. MFC-GAN model allows us to train the discriminator to classify not only real symbols but also fake symbols, which provides more fine-grained discrimination between instances.

For this work, the discriminator network is designed to have four convolution layers with strides of two and batch normalisation is used between layers. All convolution layers are activated using Leaky ReLu with alpha set to 0.2, and Sigmoid function is used in the final layer as the activation function.

The discriminator layers are shared with a classifier model that outputs  $2 \times N$  soft-max, where  $N$  is the number of classes. We also designed the generator to have one linear layer and five transpose convolution layers with strides of two in each layer. Batch normalisation was also used between adjacent layers and all layers were activated using Leaky ReLu apart from the final layer which is sigmoid activated.

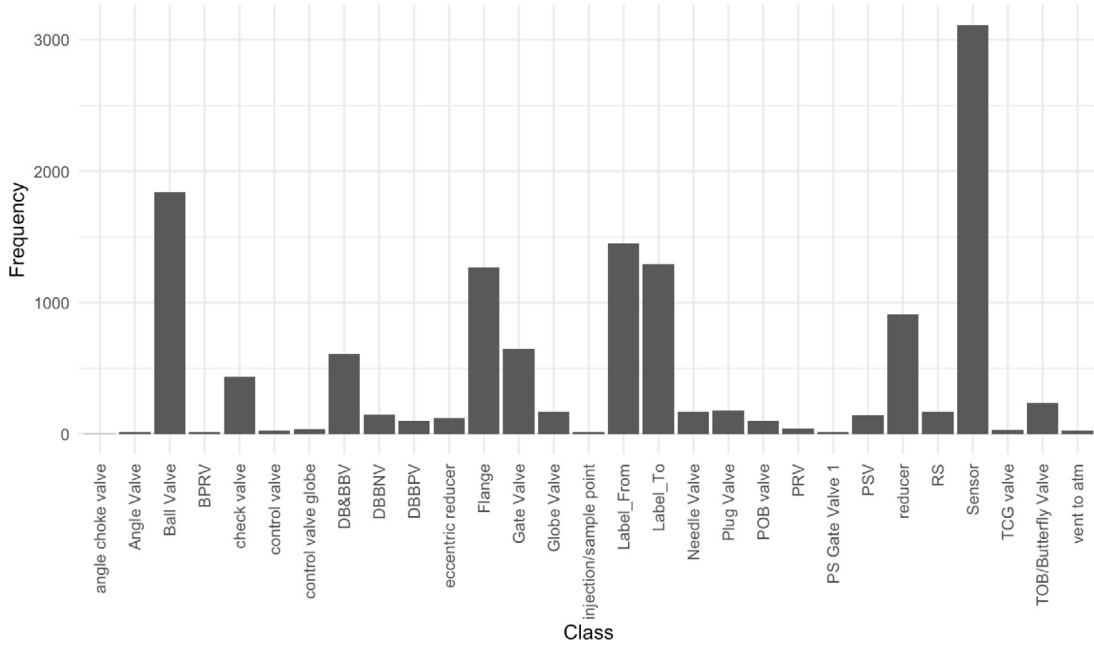


Fig. 6. Class distribution of symbols in the whole dataset.

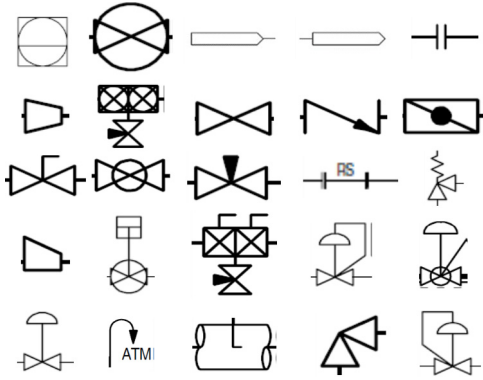


Fig. 7. Symbols used in the training and testing sets.

Similar to most GAN models the generator's input is a noise vector of size 100 and combined with symbol label encoding (see Ali-Gombe & Elyan, 2019 for details). This label encoding is used to control the class-specific generation, which is essential for our experiment.

The generator output is a  $64 \times 64$  greyscale symbols image. For our experiments (following sections) we used a batch size of 100 and a learning rate of 0.001 which was experimentally chosen. Spectral normalisation was used in both the generator and the discriminator. The proposed model will be trained using Eqs. (2) and (3), (4).

$$\mathcal{L}_s = \mathbb{E}[\log P(S = \text{real}|X_{\text{real}})] + \mathbb{E}[\log P(S = \text{fake}|X_{\text{fake}})] \quad (2)$$

$$\mathcal{L}_{cd} = \mathbb{E}[\log P(C = c|X_{\text{real}})] + \mathbb{E}[\log P(C' = c'|X_{\text{fake}})] \quad (3)$$

$$\mathcal{L}_{cg} = \mathbb{E}[\log P(C = c|X_{\text{real}})] + \mathbb{E}[\log P(C = c|X_{\text{fake}})] \quad (4)$$

where  $\mathcal{L}_s$  is used to estimate the sampling loss, which represents the probability of the sample being real or fake.  $\mathcal{L}_{cd}$  and  $\mathcal{L}_{cg}$  are used to estimate the classification losses over the generator and the discriminator.  $X_{\text{real}}$  represents the training data and  $X_{\text{fake}}$  is the set of generated images.

## 4. Experiment & results

Two experiments were carried out. The first experiment was designed to evaluate an end-to-end solution for recognising symbols in engineering drawings. We are assuming here, that locating and recognising these symbols will simplify subsequent tasks in a framework for analysing the whole drawings (i.e. detecting text, pipelines, etc...). This is simply because the majority of these types of drawings are made of symbols. The second experiment is separate and is focused on handling the class-imbalance problem using GAN-based methods.

### 4.1. Symbols recognition

In our dataset, the P&ID sheets were approximately 7500 x 5250 pixels in size. To use such image size in training data is computationally expensive and therefore each P&ID was split into 24 patches by dividing the original P&ID width by 6 and the height by 4. This gave a patch size of approximately 1250 x 1300 pixels. The annotation data for each patch was obtained using the annotations for the whole P&ID as discussed in the previous section.

For the training phase, we excluded symbols that overlapped multiple patches. After extensive experiments, it was decided to use the 3rd version of the YOLO framework which proved to be improving the detection rate of small objects compared to the first and second YOLO models (Redmon et al., 2016; Redmon & Farhadi, 2017). It is worth pointing out that the sizes of the various engineering symbols in our dataset are relatively small compared to the image size.

The YOLO architecture was customised for the purpose of this experiment. First the number of classes in each of the three YOLO layers was set to 25, and the number of filters was changed accordingly and was set to  $3 \times (\text{Class}_{no} + 5)$ , where the  $\text{Class}_{no}$  denotes the number of classes in the dataset.

The dataset was split approximately 90%:10% into training (155 P&IDs) and test (16 P&IDs) sets. A pre-trained Network was used and retrained using our dataset and all layers were fine-tuned. Darknet implementation of the YOLO was used in this

experiment.<sup>2</sup> During the training process the network input size of  $416 \times 416$  was adjusted after every ten batches; adjusting the input size during training was reported to improve object detection across different object scales (Redmon & Farhadi, 2017). The network was trained with a learning rate of 0.001 and training is stopped when the model was trained on 10,000 batches, (batch size of 64).

At testing time, the model input was adjusted from  $416 \times 416$  to  $2400 \times 2400$ . In this way, we were able to test on the original P&ID images and simplify symbol detection across a whole P&ID diagram in one step as opposed to combining detections from the P&ID patches. For evaluation, symbols were compared against the ground truth and the Intersection over Union *IOU* was set experimentally to 0.5. A simple front end was developed using Python Libraries and OpenCV<sup>3</sup> for visualisation and manual error analysis purposes.

#### 4.1.1. Results

The training accuracy achieved was  $\sim 96\%$ . On the testing set, 1352 symbols out of 1424 were correctly located and recognised with a testing accuracy equal to  $\sim 94.9\%$ . A heatmap of the confusion matrix for the testing set is presented in Fig. 8. It can be seen and as expected that majority class instances were accurately detected and recognised. In other words, symbols with enough examples in the training set were accurately recognised.

A typical output from the proposed methods where different symbols are highlighted in different colours is shown Fig. 9. Recognised symbols here were numbered and the predicted labels were recorded for further comparison against the ground truth. These symbols include inlets/outlets that are denoted by *label\_to*, *label\_from*, sensors, ball valve, reducers, gate valves, globe valves, and others.

Table 1 provides more details about the number of instances of each symbol in the training and testing set named as *No of Training Symbols* and *No of Testing Symbols* respectively. Furthermore, it also shows the number of correctly recognised symbols in the testing set (*Correctly Recognised*) and the testing accuracy per class (*Class Accuracy*).

Results show that the majority of instances were accurately detected and recognised (1352 of 1424). Fig. 10 shows different symbols from various P&ID diagrams. Notice here that symbols are accurately detected and recognised regardless of its orientation. For example, reducers, gate valves, check valves, and others appear in different orientations (Fig. 10). Similarly, sensors are accurately detected and recognised regardless of the text overlap with these instances. This clearly shows that unlike traditional methods the proposed method is robust to these inherent vision challenges (at least in this context).

As can be seen in Table 2, in total 72 instances of the P&ID symbols were either unrecognised at all (missed), or incorrectly classified as different symbols. Of these, 8 instances of symbols were incorrectly classified (Table 2). Additionally, 64 symbols were completely missed. This can be largely attributed to the nature of the drawings, wherein these cases symbols will have text and annotation almost covering its entirety. This is evident if we look at the *IOU* in Table 2 which is zero across all these missed symbols.

Further visual analysis of the results presented in Table 2 showed that some symbols were incorrectly labelled. In particular, the instance of the symbol *Ball Valve*, although the model predicted the 'wrong' class symbol, visualising the results showed that the model actually predicted the right class for these symbols despite the wrong label. This is illustrated in Fig. 11. Here, we use

**Table 1**

Results of the proposed methods for symbols recognition (Testing set).

Symbol	No of training symbols	No of testing symbols	Correctly recognised (Testing)	Class accuracy
Sensor	2810	302	297	98%
Ball valve	1629	213	212	99%
Label_From	1347	103	103	100%
Label_To	1178	113	113	100%
Flange	1110	158	121	77%
Reducer	821	91	90	99%
DB&BBV	542	67	66	98%
Gate valve	535	110	104	94%
Check valve	396	42	42	100%
TOB/Butterfly valve	178	59	58	98%
Plug valve	173	8	8	100%
Globe valve	161	7	7	100%
Needle valve	160	10	10	100%
RS	143	26	24	92%
PSV	118	25	22	88%
Eccentric reducer	98	23	22	96%
POB valve	84	16	16	100%
DBBPV	83	15	15	100%
PRV	32	8	8	100%
Control valve globe	30	6	6	100%
Control valve	22	5	5	100%
Vent to atm	19	8	2	25%
Injection/sample point	13	2	1	50%
Angle valve	11	2	0	0.0%
BPRV	11	5	0	0.0%

**Table 2**

Unrecognised and misclassified symbols in engineering drawings.

Actual class	No of instances	Predicted class	IOU
Ball valve	1	Reducer	0.81
BPRV	5	PRV	0.91
Eccentric reducer	1	Reducer	0.72
Reducer	1	Eccentric reducer	0.90
Angle valve	2	–	0.00
Flange	37	–	0.00
Gate valve	6	–	0.00
Sensor	5	–	0.00
TOB/Butterfly valve	1	–	0.00
Vent to Atm	6	–	0.00
Injection/sample point	1	–	0.00
PSV	3	–	0.00
RS	2	–	0.00
DB&BBV	1	–	0.00

a simple front end to visualise the recognised symbols alongside an item number that we assign for each of them. This has greatly facilitated the analysis and visualisation of the results.

Similarly, consider the symbol of class *BPRV* which was classified as *PRV* in all five instances in the testing set. First, it is worth noting that the number of training instances of this symbol is extremely low (11). Additionally, the symbol is very much similar to the *PRV* class. However, it is anticipated that more training examples of this symbol will certainly improve its detection rate, as it is the case with most majority class symbols (i.e. Sensor, Ball Valve, Reducer, Gate Valve, Check Valve, Globe Valve and so on). Fig. 12 shows samples of the *BPRV* symbols which were incorrectly classified alongside the actual *PRV* symbol.

In summary, it can also be seen from the results presented in Table 1 that the recognition rate of the symbols (*vent to atm*, *Angle Valve*, and *BPRV*) was quite low. This is mainly due to the limited numbers of training samples that represent these symbols. But overall, and excluding these three symbols, the average class accuracy of the remaining 22 symbols in the dataset is over 92% which is very encouraging results for such a challenging problem.

<sup>2</sup> <https://github.com/AlexeyAB/darknet>, A. A.B., Darknet,(2019).

<sup>3</sup> <https://opencv.org/>.

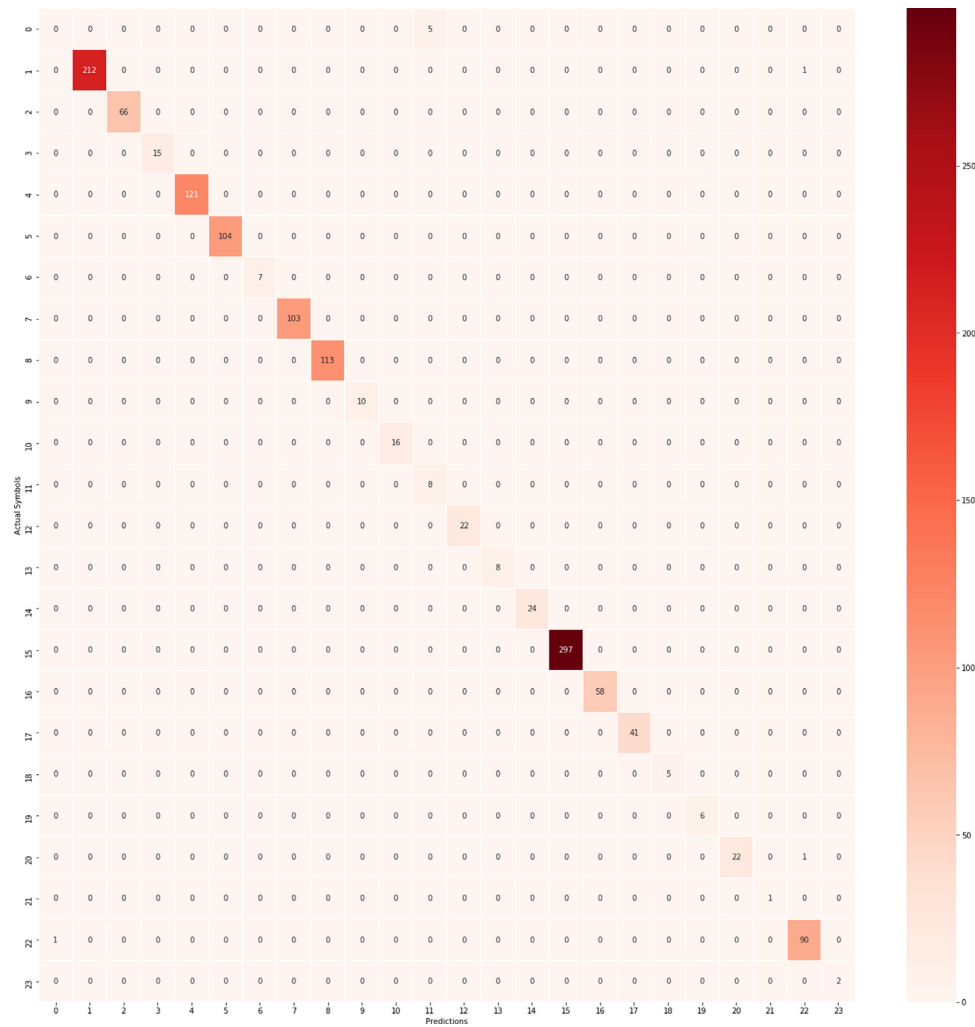


Fig. 8. Heatmap of the Confusion matrix of the 25 symbols predictions (Testing Set).

#### 4.2. MFC-GAN for class-imbalance

In this experiment we aim at evaluating a GAN-based model to handle the class-imbalance problem in the dataset. This is not a recognition task as in the first experiment but rather a classification problem. The experiment aims first at generating more symbols using MFC-GAN model. Then these synthesised samples will be used to augment the training set aiming at improving classification results.

The dataset used in this experiment is almost the same one used in Experiment 1. All symbols were resized to  $64 \times 64$  grey-scaled images. The problem is formulated as a supervised learning task where the aim is to learn a function  $f(x)$  that maps an instance  $x_i$  of a particular engineering symbol to the corresponding class  $y_i$ . In this case,  $y_i \in Y$  where  $Y$  is a discrete set of classes representing the 29 symbols in the dataset. As discussed earlier the dataset is hugely imbalanced, and some of the instance that were dropped in Experiment 1 populates less than  $\sim 0.01\%$  such as angle choke valve.

The experiment was carried out in two stages, a GAN training stage and a classification stage. First, we trained MFC-GAN using all the samples in the dataset. The MFC-GAN model was conditioned to generate engineering symbols in extreme cases of class imbalance. To do so, we considered the least represented symbols in the whole dataset. These are Angle Choke Valve, Angle Valve, Injection Sample Point, Back Pressure Regulating Valve, PS Gate Valve, Control Valve, Through Conduit Gate Valve, Control

Valve Globe and Pressure Regulating Valve. These symbols have 2, 13, 15, 17, 17, 27, 31, 36 and 42 instances respectively in the training set. The model was trained only once on this dataset and the samples were generated after training was completed. During training, the minority classes were resampled to encourage learning of minority instances structure.

The trained MFC-GAN model was then used to generate symbols of minority class instances (the least represented in the dataset, nine symbols). The original dataset was split 70% for training and 30% for testing set. Synthetic datasets were then added to the training set. For each minority class, 5000 more synthetic samples were added. This enabled us to rebalance the dataset by increasing the presence of the least represented symbols.

In order to evaluate the quality of the generated symbols, we build a classification model to compare performance before and after adding the generated symbols to the training set.

The classification model chosen is a CNN with 4 layers. The first three layers are convolution layers with 32, 64, 128 outputs. These layers have a kernel size of  $3 \times 3$ ,  $2 \times 2$  and max-pooling in-between them. The fourth layer is a fully connected layer with 256 units that feeds in to a 29-way SoftMax output representing the 29 symbol classes. The CNN was trained using SGD with a batch size of 64 and a learning rate of 0.001. Classification results were recorded using common metrics, namely true positive rate, balanced accuracy, G-mean and F1-Score.



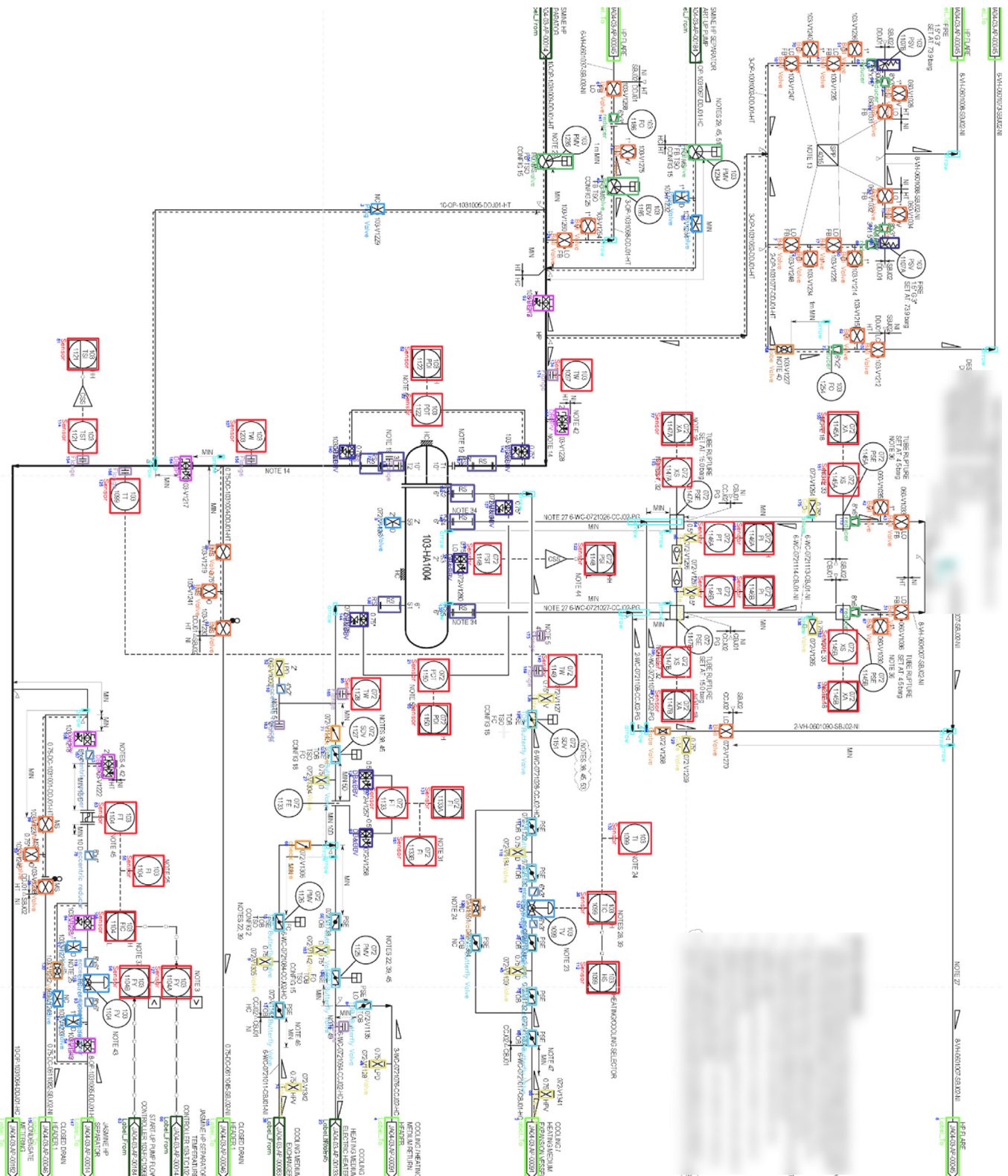


Fig. 9. A P&ID diagram with various recognised symbols (Testing Set).

#### 4.2.1. Results

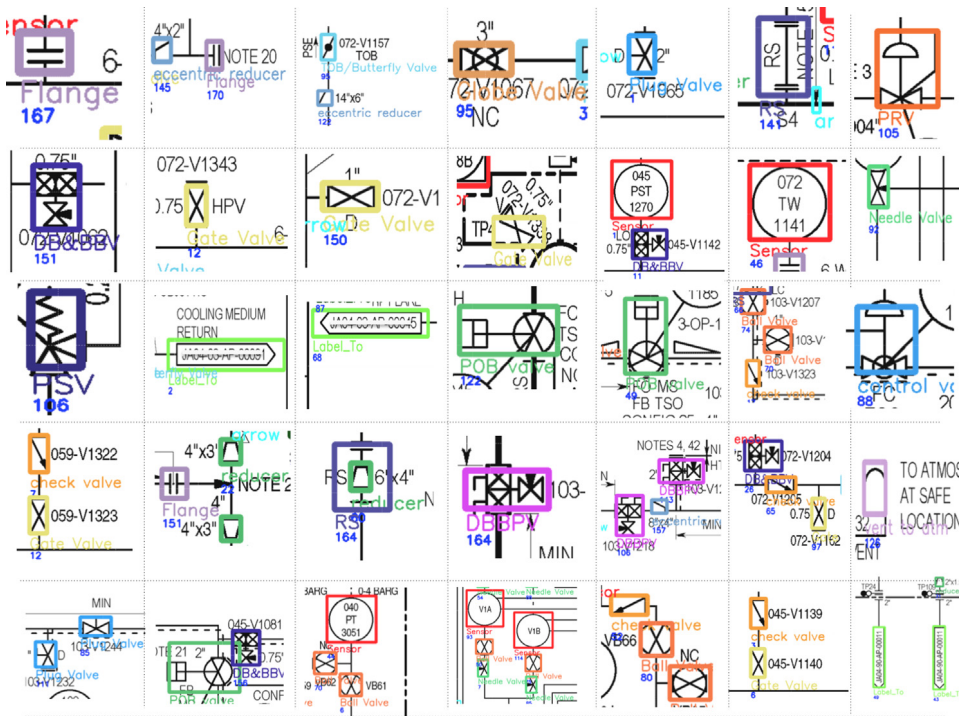
Fig. 13 compares the generated samples from MFC-GAN model with the original symbols from the diagram. We also report the symbols classification results in Table 3.

MFC-GAN generated far superior and more realistic samples. Visual inspection revealed distinct symbols features and the

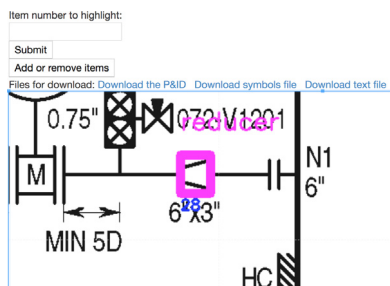
required categories were generated in each instance. Moreover, MFC-GAN high-quality samples had a positive effect on the performance of the classifier. For example, the G-Mean and sensitivity improved from 0 to 100% on angle choke valve as can be seen in Table 3 with just two instances of the classes. This result is consistent in seven of the nine minority classes. However, we

**Table 3**  
CNN performance on symbols classification.

Metric	Model	Angle choke valve	Angle valve	BPRV	Control valve	Control valve globe	Injection sample point	PRV	PS gate valve	TCG valve
Sensitivity	Baseline	0.00	0.50	0.60	0.88	1.00	0.80	1.00	1.00	0.89
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	<b>0.80</b>	0.88	<b>1.00</b>	<b>0.88</b>	0.77	<b>1.00</b>	<b>0.91</b>
Specificity	Baseline	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Precision	Baseline	0.00	1.00	1.00	1.00	0.85	1.00	0.72	1.00	1.00
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	0.67	<b>1.00</b>	0.92	<b>1.00</b>	<b>0.91</b>	0.83	<b>1.00</b>
F1-score	Baseline	0.00	0.67	0.75	0.93	0.92	0.89	0.84	1.00	0.94
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	0.73	<b>0.93</b>	<b>0.96</b>	<b>0.93</b>	0.83	0.91	<b>0.95</b>
Accuracy	Baseline	0.50	0.75	0.80	0.94	1.00	0.90	1.00	1.00	0.95
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	<b>0.90</b>	0.94	<b>1.00</b>	<b>0.94</b>	0.89	<b>1.00</b>	<b>0.96</b>
G-mean	Baseline	0.00	0.71	0.77	0.94	1.00	0.89	1.00	1.00	0.94
	MFC-GAN	<b>1.00</b>	<b>1.00</b>	<b>0.89</b>	0.94	<b>1.00</b>	<b>0.93</b>	0.88	<b>1.00</b>	<b>0.95</b>



**Fig. 10.** Examples of detected symbols.



**Fig. 11.** Incorrectly labelled symbol.

observed that the model did not improve the baseline in the other two classes control valve and PRV classes. A closer look at Fig. 13 revealed a high similarity between symbols. There is extreme similarity between angle valve (fifth symbol from the top) with control valve globe (eighth symbols from the top) and PRV (seventh symbol from the top) and BPRV (second symbol

from the top). Although symbols were distinctly generated, the similarity of symbols dwindled the classification results in these classes. The low precision in BPRV and control valve globe classes from Table 3 further solidifies this observation.

MFC-GAN models proved in this experiment to be able to generate minority class instances that are extremely under-represented in the dataset. The quality of these samples was evaluated subjectively by inspecting the resulting samples, and objectively by measuring a classifier performance before and after adding the generated samples to the training sets. Results show clearly that performance improved across several common evaluation metrics. However, it has to be said that MFC-GAN is only one method that can be used to handle the class imbalance problem. Other possible methods can also be explored and utilised. Class-imbalance is a very well researched problem, and there is a wide range of methods that ranges from simple data augmentation, sampling to more advanced methods such as GAN (Vuttipittayamongkol & Elyan, 2020). For an extensive review of different possible methods, the reader is referred to Anon (2017).

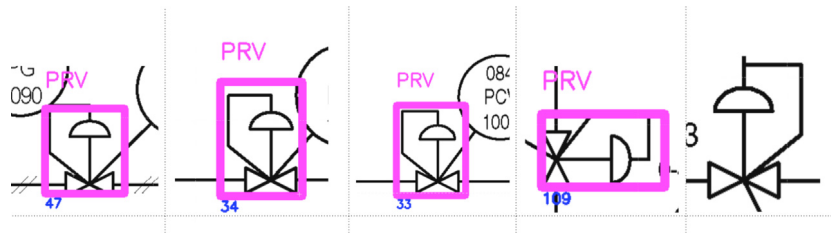


Fig. 12. Incorrectly classified BPRV symbols (first four instances from left) as PRV symbol (fifth instance).

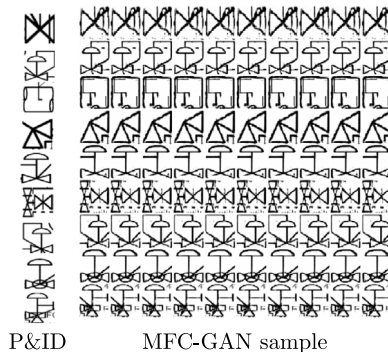


Fig. 13. Comparing original P&ID samples with MFC-GAN generated samples.

## 5. Conclusion & future direction

In this paper, we proposed an end-to-end framework for processing and analysing complex engineering drawings. Thorough experiments using a large collection of P&ID sheets from an industrial partner showed that our method accurately recognises more than 94% of the symbols in the drawings. Advanced-bounding-box detection methods proved in our experiments that they perform accurately in such challenging tasks by recognising symbols of 25 different classes, despite the very little differences between some of these symbols. Additionally, we proposed a GAN-based model to handle class-imbalance in the symbols dataset. Our experiments demonstrated that our method was capable of generating plausible engineering symbols and also proved to be improving classification accuracy when augmenting the training set with this synthesised data. Experimental results show that the proposed GAN model can learn from a smaller number of training examples.

A future direction of this work will focus on utilising Generative Adversarial Neural Networks to generating symbols in a diagram context. In other words, generate part of the engineering diagram, and not only the symbols. This will greatly help in saving efforts needed for manual data annotation. Additionally, future work will include building a unified framework based on the proposed methods to allow full processing and analysis of engineering diagrams such as P&ID. We hypothesise that the work presented in this paper will greatly simplify subsequent tasks such as text localisation and line detection.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by Scotland Data Lab Innovation Centre, UK, Oil and Gas Innovation Centre, UK and DNV GL, Norway.

## References

- Abelameyko, S. V., & Uchida, S. (2007). Recognition of engineering drawing entities: Review of approaches. *International Journal of Image and Graphics*, 07(04), 709–733. <http://dx.doi.org/10.1142/S0219467807002878>, arXiv:<http://www.worldscientific.com/doi/pdf/10.1142/S0219467807002878>.
- Ahmed, S., Liwicki, M., Weber, M., & Dengel, A. (2012). Automatic room detection and room labeling from architectural floor plans. In *2012 10th IAPR international workshop on document analysis systems* (pp. 339–343). <http://dx.doi.org/10.1109/DAS.2012.22>.
- Ali-Gombe, A., & Elyan, E. (2019). Mfc-gan: Class-imbalanced dataset classification using multiple fake class generative adversarial network. *Neurocomputing*, 361, 212–221.
- Ali-Gombe, A., Elyan, E., & Jayne, C. (2017). Fish classification in context of noisy images. In G. Boracchi, L. Iliadis, C. Jayne, & A. Likas (Eds.), *Engineering applications of neural networks* (pp. 216–226). Cham: Springer International Publishing.
- Ali-Gombe, A., Elyan, E., & Jayne, C. (2019). Multiple fake classes gan for data augmentation in face image dataset. In *2019 international joint conference on neural networks (IJCNN)* (pp. 1–8). <http://dx.doi.org/10.1109/IJCNN.2019.8851953>.
- Ali-Gombe, A., Elyan, E., Savoye, Y., & Jayne, C. (2018). Few-shot classifier gan. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1–8). <http://dx.doi.org/10.1109/IJCNN.2018.8489387>.
- Anon (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, 220–239.
- Antoniou, A., Storkey, A., & Edwards, H. (2017). Data augmentation generative adversarial networks. arXiv preprint [arXiv:1711.04340](https://arxiv.org/abs/1711.04340).
- Arroyo, E., Fay, A., Chioua, M., & Hoernicke, M. (2014). Integrating plant and process information as a basis for automated plant diagnosis tasks. In *Proceedings of the 2014 IEEE emerging technology and factory automation (ETFA)* (pp. 1–8). <http://dx.doi.org/10.1109/ETFA.2014.7005098>.
- Arroyo, E., Hoang, X. L., & Fay, A. (2015). Automatic detection and recognition of structural and connectivity objects in svg-coded engineering documents. In *2015 IEEE 20th conference on emerging technologies factory automation (ETFA)* (pp. 1–8). <http://dx.doi.org/10.1109/ETFA.2015.7301510>.
- Blostein, D. (1995). General diagram-recognition methodologies. In: *Proceedings of the 1st international conference on graphics recognition (GREC'95)* (pp. 200–212).
- Boatto, L., Consorti, V., Del Buono, M., Eramo, V., Esposito, A., Melcarne, F., et al. (1992). Detection and separation of symbols connected to graphics in line drawings. In *Proceedings, 11th IAPR international conference on pattern recognition. Vol.II. Conference B: Pattern recognition methodology and systems* (pp. 545–548). <http://dx.doi.org/10.1109/ICPR.1992.201837>.
- Chhabra, A. K. (1997). Graphics recognition algorithms and systems. In *Proceedings of the 2nd international conference on graphics recognition (GREC'97)* (pp. 244–252). [http://dx.doi.org/10.1007/3-540-64381-8\\_40](http://dx.doi.org/10.1007/3-540-64381-8_40).
- Cordella, L. P., & Vento, M. (2000). Symbol recognition in documents: A collection of techniques? *International Journal on Document Analysis and Recognition*, 3(2), 73–88. <http://dx.doi.org/10.1007/s100320000036>.
- Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: object detection via region-based fully convolutional networks. CoRR abs/1605.06409. [arXiv:1605.06409](https://arxiv.org/abs/1605.06409).
- Elyan, E., & Gaber, M. M. (2017). A genetic algorithm approach to optimising random forests applied to class engineered data. *Information Sciences*, 384, 220–234. <http://dx.doi.org/10.1016/j.ins.2016.08.007>, URL <http://www.sciencedirect.com/science/article/pii/S0020025516305783>.
- Elyan, E., Garcia, C. M., & Jayne, C. (2018). Symbols classification in engineering drawings. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1–8). <http://dx.doi.org/10.1109/IJCNN.2018.8489087>.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <http://dx.doi.org/10.1038/s41591-018-0316-z>.
- Girshick, R. (2015). Fast r-cnn. In *ICCV '15, Proceedings of the 2015 IEEE international conference on computer vision (ICCV)* (pp. 1440–1448). Washington, DC, USA: IEEE Computer Society, <http://dx.doi.org/10.1109/ICCV.2015.169>.



- Goh, K. N., Mohd. Shukri, S. R., & Manao, R. B. H. (2013). Automatic assessment for engineering drawing. In H. B. Zaman, P. Robinson, P. Olivier, T. K. Shih, & S. Velastin (Eds.), *Advances in visual informatics* (pp. 497–507). Cham: Springer International Publishing.
- Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, URL <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press, <http://www.deeplearningbook.org>.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377. <http://dx.doi.org/10.1016/j.patcog.2017.10.013>, URL <http://www.sciencedirect.com/science/article/pii/S0031320317304120>.
- Holcomb, S. D., Porter, W. K., Ault, S. V., Mao, G., & Wang, J. (2018). Overview on deepmind and its alphago zero ai. In *ICBDE '18, Proceedings of the 2018 international conference on big data and education* (pp. 67–71). New York, NY, USA: ACM, <http://dx.doi.org/10.1145/3206157.3206174>, <http://doi.acm.org/10.1145/3206157.3206174>.
- Howie, C., Kunz, J., Binford, T., Chen, T., & Law, K. (1998). Computer interpretation of process and instrumentation drawings. *Advances in Engineering Software*, 29(7), 563–570. [http://dx.doi.org/10.1016/S0965-9978\(98\)00022-2](http://dx.doi.org/10.1016/S0965-9978(98)00022-2), URL <http://www.sciencedirect.com/science/article/pii/S0965997898000222>.
- Julca-Aguilar, F. D., & Hirata, N. S. T. (2017). Symbol detection in online handwritten graphics using faster R-CNN. CoRR abs/1712.04833. arXiv:1712.04833.
- Kanungo, T., Haralick, R. M., & Dori, D. (1995). Understanding engineering drawings: A survey. In *Proceedings of the 1st international conference on graphics recognition (GREC'95)* (pp. 119–130).
- Khan, W., Ansell, D., Kuru, K., & Bilal, M. (2018). Flight guardian: Autonomous flight safety improvement by monitoring aircraft cockpit instruments. *Journal of Aerospace Information Systems*, 15(4), 203–214. <http://dx.doi.org/10.2514/1.1010570>, arXiv:https://doi.org/10.2514/1.1010570.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <http://dx.doi.org/10.1145/3065386>.
- Kulkarni, C. R., & Barbadekar, A. B. (2017). Text detection and recognition: A review. *International Research Journal of Engineering and Technology (IRJET)*, 4(6), 179–185.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <http://dx.doi.org/10.1109/5.726791>.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C., et al. (2015). SSD: single shot multibox detector. CoRR abs/1512.02325. arXiv:1512.02325.
- Lu, Y. (1995). Machine printed character segmentation - an overview. *Pattern Recognition*, 28(1), 67–80. [http://dx.doi.org/10.1016/0031-3203\(94\)00068-W](http://dx.doi.org/10.1016/0031-3203(94)00068-W).
- Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., & Malossi, C. (2018). Bagan: Data augmentation with balancing gan. arXiv preprint arXiv:1803.09655.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
- Moreno-García, C. F., Elyan, E., & Jayne, C. (2017). Heuristics-based detection to improve text / graphics segmentation in complex engineering drawings. In *Engineering applications of neural networks, Vol. CCIS 744* (pp. 87–98).
- Moreno-García, C. F., Elyan, E., & Jayne, C. (2018). New trends on digitisation of complex engineering drawings. *Neural Computing and Applications*, <http://dx.doi.org/10.1007/s00521-018-3583-1>.
- Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7), 1029–1058. <http://dx.doi.org/10.1109/5.156468>.
- Odena, A., Olah, C., & Shlens, J. (AUG 2017). Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, page 2642–2651 70 (AUG 2017) (pp. 2642–2651).
- Pacha, A., Choi, K., Coüasnon, B., Ricquebourg, Y., Zanibbi, R., & Eidenberger, H. (2018). Handwritten music object detection: Open issues and baseline results. In *2018 13th IAPR international workshop on document analysis systems (DAS)* (pp. 163–168). <http://dx.doi.org/10.1109/DAS.2018.51>.
- Pacha, A., Haji, J., & Calvo-Zaragoza, J. (2018). A baseline for general music object detection with deep learning. *Applied Sciences*, 8(9), <http://dx.doi.org/10.3390/app8091488>, URL <http://www.mdpi.com/2076-3417/8/9/1488>.
- Park, U., & Jain, A. K. (2010). Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3), 406–415. <http://dx.doi.org/10.1109/TIFS.2010.2049842>.
- Rahul, R., Paliwal, S., Sharma, M., & Vig, L. (2019). Automatic information extraction from piping and instrumentation diagrams. CoRR abs/1901.11383. arXiv:1901.11383.
- Rebelo, A., Capela, G., & Cardoso, J. S. (2010). Optical recognition of music symbols: A comparative study. *International Journal on Document Analysis and Recognition*, 13(1), 19–31. <http://dx.doi.org/10.1007/s10032-009-0100-1>.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 779–788). <http://dx.doi.org/10.1109/CVPR.2016.91>.
- Redmon, J., & Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 6517–6525). <http://dx.doi.org/10.1109/CVPR.2017.690>.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *NIPS'15, Proceedings of the 28th international conference on neural information processing systems - volume 1* (pp. 91–99). Cambridge, MA, USA: MIT Press, URL <http://dl.acm.org/citation.cfm?id=2969239.2969250>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *2015 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1–9). <http://dx.doi.org/10.1109/CVPR.2015.7298594>.
- Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE conference on computer vision and pattern recognition* (pp. 1701–1708). <http://dx.doi.org/10.1109/CVPR.2014.220>.
- Vaxiviere, P., & Tombre, K. (1992). Celesttin: cad conversion of mechanical drawings. *Computer*, 25(7), 46–54. <http://dx.doi.org/10.1109/2.144439>.
- Vuttipittayamongkol, P., & Elyan, E. (2020). Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. *Information Sciences*, 509, 47–70. <http://dx.doi.org/10.1016/j.ins.2019.08.062>, URL <http://www.sciencedirect.com/science/article/pii/S0020025519308114>.
- Vuttipittayamongkol, P., Elyan, E., Petrovski, A., & Jayne, C. (2018). Overlap-based undersampling for improving imbalanced data classification. In H. Yin, D. Camacho, P. Novais, & A. J. Tallón-Ballesteros (Eds.), *Intelligent data engineering and automated learning - IDEAL 2018* (pp. 689–697). Cham: Springer International Publishing.
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1480–1489). San Diego, California: Association for Computational Linguistics, <http://dx.doi.org/10.18653/v1/N16-1174>, URL <https://www.aclweb.org/anthology/N16-1174>.
- Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1), 1–19. <http://dx.doi.org/10.1016/j.patcog.2003.07.008>.
- Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399–458. <http://dx.doi.org/10.1145/954339.954342>, <http://doi.acm.org/10.1145/954339.954342>.