

Hazardous Object Detection in Bulk Material Transport using Video Stream Processing

Vanessa Meulenberg, Kamal Moloukbashi
Al-Kahwati, Johan Öhman, Wolfgang Birk
& Rune Nilsen

Hazardous Object Detection in Bulk Material Transport using Video Stream Processing

Vanessa Meulenberg
Predge AB
Luleå, Sweden

Kamal Moloukbashi Al-Kahwati
Predge AB
Luleå, Sweden

Johan Öhman
Predge AB
Luleå, Sweden

Wolfgang Birk
Predge AB
Luleå, Sweden

Rune Nilsen
LKAB
Narvik, Norway

ABSTRACT

Belt conveyor systems are a primary means of bulk material transport in industrial applications due to their high bulk capacity and limited need for human involvement. Abnormal objects on the belt conveyor can be hazardous to the operation of the belt conveyor systems and/or downstream equipment. The dependability of production on a well-operating system in combination with the high degree of automation and limited inspection accessibility, establishes the need for a continuous and fully automated monitoring solution. In this paper, a monitoring solution comprising a camera, object detection and classification model, and decision support is presented and discussed. The detection and classification model is comprised of two steps: a classical brightness and contour detection algorithm using colour channel weighing, and a subsequent processing by a Convolutional Neural Network (CNN). The CNN performs a classification of the detections as True Positives (TP) or False Positives (FP). Further, the object size is estimated providing a measure for the risk imposed by the object. The solution makes use of an off-the-shelf industrial network camera that communicates with an edge computing device close to the installation site. The edge device is further connected to a SaaS solution for predictive maintenance and decision support where results (classified detections) are visualized in a dashboard. There, operators can assess classified detections as TP or FP, which provides a ground truth for subsequent retraining of the solution. Moreover, these actionable insights enable a warning and stopping mechanism that can be implemented when the operators trust the solution. The solution is implemented and tested at LKAB Narvik and operational since 2021. Initially, the solution was trained using artificially introduced objects and manually labelled video frames, followed by a validation phase to assess the performance of the solution. The solution exceeds targeted performance while having a low false positive rate.

Keywords

Pattern recognition, image analysis, convolutional neural networks, labelling, autonomous decision making, object detection.

1. INTRODUCTION

For the transportation of large quantities of bulk materials in mining industries, belt conveyor systems play a crucial role due to their high capacity and limited need for human involvement. At LKAB Narvik, where the experimental work of this study was carried out, a dense network of conveyor belts is used to load ships with ore ar-

riving by train from the LKAB mines in Kiruna, Northern Sweden. The conveyor systems are critical logistic elements in the production chain and are often exposed to harsh and adverse environmental conditions. Dust, humidity, excessive loading, and freezing temperatures are just a few examples of the conditions in which conveyors must operate reliably [1].

Occasionally, hazardous objects can end up on the conveyor belts, such as silo chute plates, large chunks of ice, concrete blocks, tools and other undesired objects. Such objects can be damaging to the operation of the belt conveyor systems and/or downstream equipment. For example, metal plates with sharp edges can rip the conveyor belt, and concrete blocks can get stuck in enclosed conveyor belts or sieves, causing material build-up and damage. Such damages can occur anywhere in the production line, where there is a high degree of automation and limited possibilities to perform frequent inspections. Moreover, objects on the belts causing unplanned stops and disruptions in the operations can lead to a delay in material delivery to clients and increased costs such as contamination fees, increased harbour fees, etc. A monitoring solution to detect and warn for abnormal objects would therefore be beneficial for safe and reliable operation of a conveyor system and can be a complementary part of a larger predictive maintenance solution for belt conveyor systems, as discussed by Al-Kahwati et al. [2][3].

Object detection has a well-established background, and classical object detection methods, such as brightness and contour detection algorithms, are suitable for basic applications. Furthermore, segmenting images into their respective red, green, and blue color channels can aid in detecting objects that exhibit lower intensity, yet are distinctly colored. This technique leverages the color information in images to improve object detection accuracy [4][5]. However, in diverse and unpredictable operational environments, more advanced detection methods are needed to increase the detection rate while keeping the false detection rate low. The harsh conditions in which conveyor belts operate create unwanted features that are detected by the camera, such as dust, flares, and so on. These features can be referred to as noise and may trigger an object detector, but they may not warrant the stopping of operations. Convolutional neural networks (CNNs) are a widely used deep learning algorithm for image classification [6], and have proven to work reliably in environments with variability, such as the automotive sector and self-driving car applications. The most significant advantage of CNNs is their ability to automatically detect image features without human supervision [7]. Therefore, they are a valuable tool for dis-

tinguishing between true positives (TPs) and false positives (FPs) as detected by a primitive object detector.

In this paper, the authors present an object detection methodology implemented in a Software-as-a-Service (SaaS) platform, utilizing edge devices for local video stream processing. The proposed method consists of a two-step object detection model. Firstly, a classical brightness and contour detection algorithm is employed that uses colour channel weighing to identify objects of different colours, exceeding a predetermined size threshold. The second step involves subsequent processing by a CNN to classify the detections as TP or FP. The solution is implemented on an edge device at the asset site and is connected to a camera installed above the belt conveyor system.

2. METHODOLOGY

An Axis P1455-LE IP camera was installed above one of the conveyor belt systems in the LKAB Narvik operation chain and connected to the local area network (LAN). Furthermore, a computer was installed at the asset site and connected to the same LAN to ensure swift communication between the devices. The architecture can be seen in Figure 1. The frames captured by the camera are stored in a queue until a worker thread is available for the processing of said frame. When a thread becomes available, the frame is pre-processed in a manner to extract the relevant features for the first classification step - including the colour and size. Should the first classification step lead to a detection, the frame is further processed in a second step to determine whether the detection is TP or FP. In the case of a TP, a warning is issued to the operators and the frame is visualized in a SaaS platform. Here, the operators can decide if an action such as stopping the conveyor is necessary so that the object can be removed and thus provide a ground truth assessment. The ground truth information is stored for further training of the classification models.

2.1 Frame pre-processing

In order to address the issue of detecting on empty conveyor belt frames, an intensity thresholding approach was employed, whereby frames exhibiting a much darker intensity level than those containing ore were eliminated. Prior to the first classification step, frames were pre-processed by cropping them to a size of 935 x 580 pixels, following which a mask was applied to a region of interest (ROI). To further refine the frames and remove unwanted noise and graininess, a non-linear median filtering technique was applied.

2.2 Initial classification step

The initial step of object classification is designed to be less stringent, with a view to reducing the possibility of missing objects on the belt. Specifically, the approach is tailored to permit a greater number of false positives to pass through. The first classification step works as follows: a key frame was generated by averaging multiple frames during regular conveyor operation. Each frame was subsequently converted to grayscale, and the red, green, and blue (RGB) color channels were extracted. The original, grayscale and RGB channel images can be seen in the example of Figure 2. Subsequently, the differential frames obtained by subtracting the key frames from the grayscale and RGB frames were analyzed for contours. The border following method proposed by Suzuki and Abe [8] was used for detecting the contours of the differential frames, as

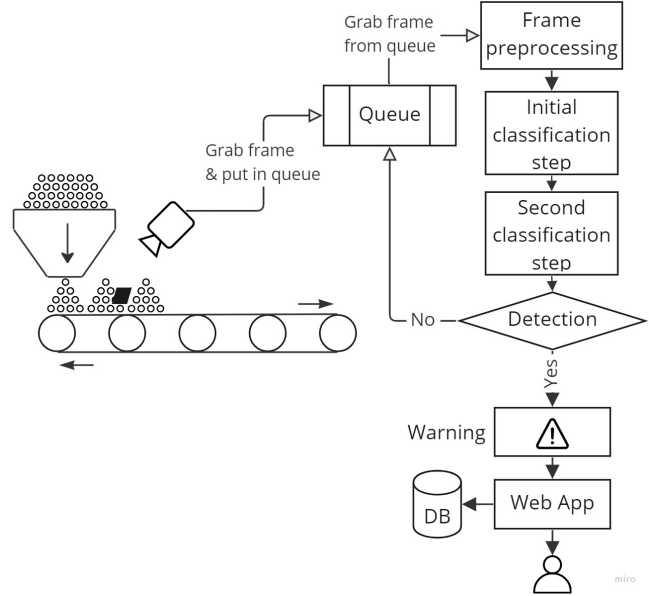


Figure 1: Architecture of the two-step object detector as constructed on the edge device.

shown in Figure 3.

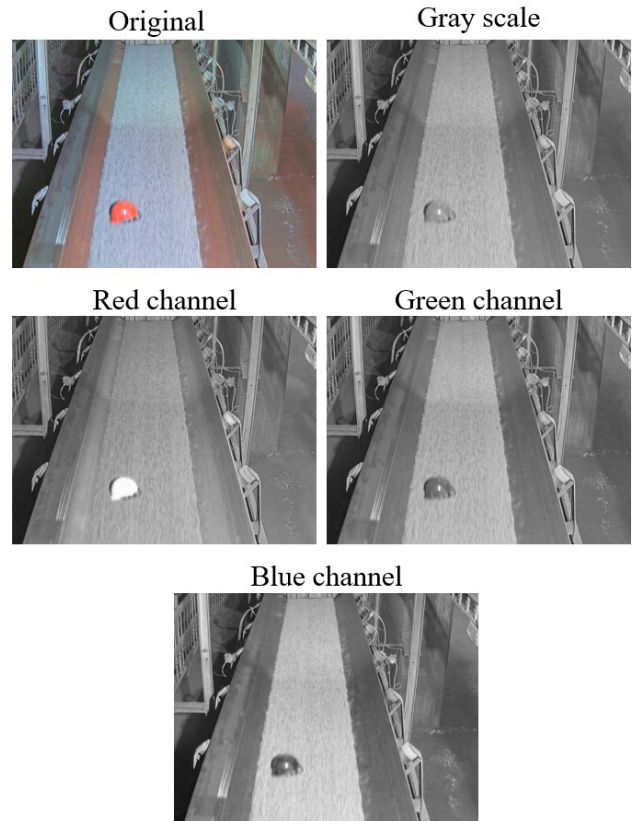


Figure 2: Original, grayscale, red, green and blue images of a safety helmet.

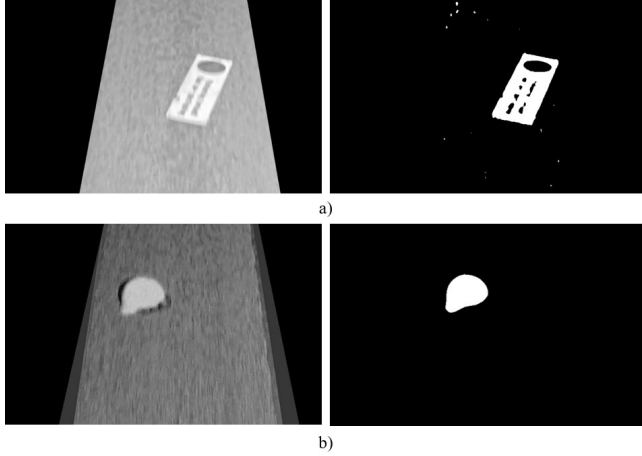


Figure 3: On the left the cropped, images with ROI mask applied is show and on the right the differential frame from the key frame is shown of a) a metal sign from a gray scale image and b) a safety helmet from the red channel image.

Through consultations with the operators, it was agreed that a minimum size threshold of 20 x 20 cm or 500 cm² would be used to classify detected objects. It is known that the spacing between the left and right idlers is 200 cm, thus, the allowable pixel area threshold with respect to the location along the belt can be defined as follows:

$$A_T = 9.606y + 656, \tag{1}$$

where A_T is the threshold area in px² and y is the location of the center point of the detected object (y is zero at the top of the image). The area of a rectangular bounding box which enclosed the detected contours were calculated and compared to the threshold calculated in equation 1.

2.3 Second classification step

The second classification step was used to distinguish whether the detections made during the less aggressive first classification step are TP or FP. This was done using a image classification CNN. Full-sized images with the ROI mask in place were rescaled to 480 x 480 pixels before being used for training. The model was trained on 739 images, manually labeled as unacceptable, and 916 manually labeled as acceptable. The acceptable database contained images of normal operation, flares, small lumps (lumps that did not make the size threshold as calculated in equation 1) as can be seen in Figure 4a. The unacceptable dataset contained images with artificially introduced objects, blocks of ice and concrete lumps (see examples in Figure 4b).

The CNN architecture can be seen in Figure 5. To compensate for the relatively small dataset, a data augmentation layer was added to the model, where a random horizontal flip was applied. The images were then rescaled to have pixel values between 0 and 1. A sequential model consisting of three convolution blocks with 32, 64 and 128 filters, respectively, was used followed by a max pooling layer after each of the convolution layers. To prevent over-fitting,

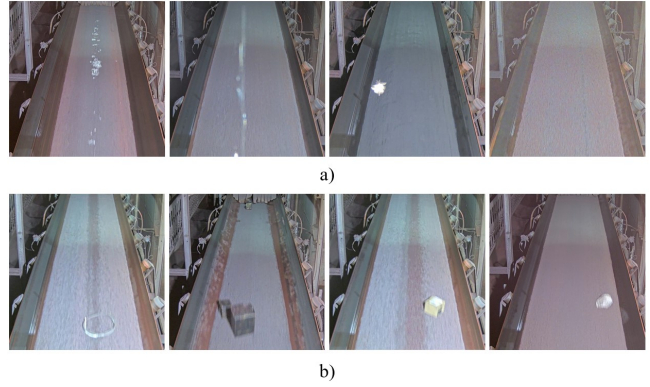


Figure 4: Example images of the training dataset classified as a) acceptable (normal operation) and b) unacceptable (objects, etc. on the belt).

a dropout regularization layer was added to the network. A flatten layer was then added, followed by a fully connected dense layer with 256 units activated by the ReLU activation function.

The model was compiled using the Adam optimizer, which is an effective stochastic gradient descent (SGD) method based on adaptive first- and second-order moments, as defined by Kingma and Ba [9]. The model losses were computed using the sparse categorical cross-entropy loss function. The model was trained using 50 epochs with a training-validation split of 20% and a batch size of 32.

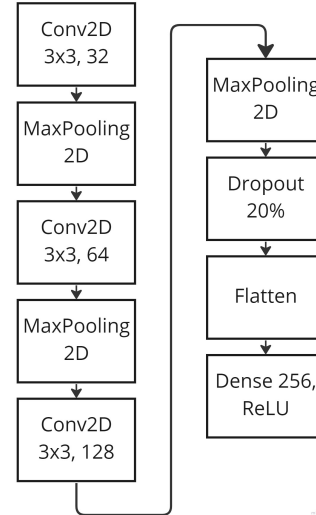


Figure 5: Details and layers of the convolutional neural network.

2.4 Testing and Validation

The initial assessment of the first classification step involved the introduction of artificial objects onto the belt. Objects of varying color, brightness, and size were placed on the belt and frames were collected. These frames along with frames from normal operations in different conditions where the stream yielded frames that are normal, empty, dusty, with flares and reflections, flooding, etc. were used to test the first classification. In total 1,655 frames were used

Table 1: Results of the first classification step

Metric	Grayscale	Red	Green	Blue
Accuracy	0.72	0.60	0.71	0.59
PPV	0.90	0.56	0.90	0.55
Recall	0.41	0.51	0.40	0.37
Processing time	34 ms/image			

for testing. The resulting detections in different color channels and grayscale were analyzed. The number of TPs, FPs, true negatives (TNs) and false negatives (FNs) recorded for the frames. Example frames can be seen in Figure 4.

The training set for the CNN consisted of the images from the first classification step divided into two classes, acceptable (FPs and TNs from the first classification step) and unacceptable (TPs and FN from the first classification step), where the unacceptable class represents frames containing unwanted objects, and the acceptable class represents normal operation. The CNN validation set comprised 20% of all images, and an additional 300 images were withheld for further testing. Model validity was assessed by evaluating various metrics, including accuracy, positive predictive value (PPV) and the recall (see equations 2-4). Additionally the inference time as well as the training and validation accuracy and losses were analysed.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

$$PPV = \frac{TP}{TP + FP}, \quad (3)$$

$$Recall = \frac{TP}{TP + FN}, \quad (4)$$

3. RESULTS AND DISCUSSION

3.1 First classification step

The results from the first classification step are shown in Table 1. The accuracy of the object detection model for grayscale images was found to be highest, at around 0.72. Notably, gray and green channels exhibited similar TP and FP values as can be seen in Figure 6. Dark objects such as the boxes shown in Figure 4b were not detected in any channels, likely due to their non-reflective nature which results in their loss during differential frame creation. PPV scores for grayscale and green channel frames were found to be significantly higher than those for red- and blue channels, which yielded more FPs due to greater intensity of red and blue tints in iron ore. The true positive rate, or recall of the red channel was found to have the highest score (0.51), followed by gray, green, and blue channels. The red channel model yielded the highest number of TPs, likely due to the presence of red objects such as the helmet.

These results indicate that although the red channel yields more detections, it also results in more FPs. Considering the two-step nature of the object detection model, it is advantageous to utilize RGB channels in conjunction with grayscale images to ensure detection of all objects and prevent belt damage. Without the second step classification, grayscale images would yield the most desirable results as false warnings could lead to unnecessary belt stoppage. Finally, the processing time per frame for this model was found to

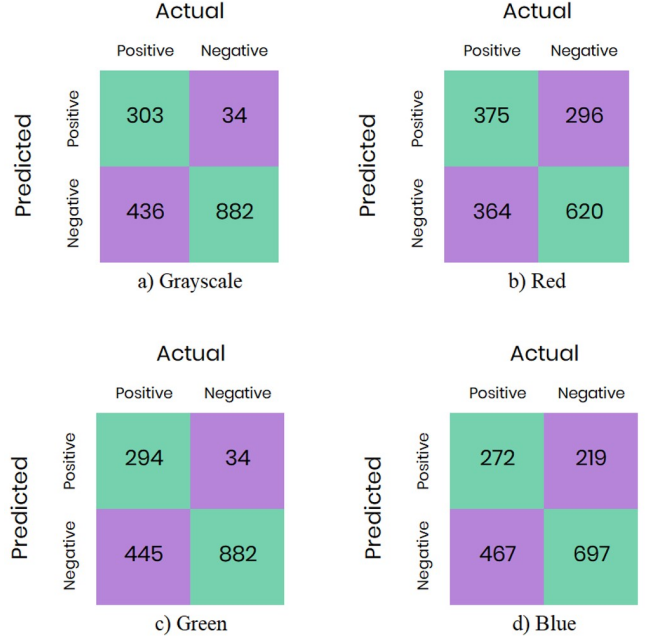


Figure 6: Confusion matrix for the first classification step of the a) gray images, b) red channel images, c) green channel images and d) blue channel images

be 34 milliseconds, which is deemed sufficient for detecting objects in real-time on the live stream.

3.2 Second classification step

The performance of the CNN model has been evaluated and is presented in Figure 7. The validation accuracy of the model is notably high, indicating that it performs well on unseen data. The validation accuracy exhibit a linear increase with each training step. The training and validation losses closely follow each other, indicating that overfitting has been avoided.

The evaluation metrics of the CNN model are summarized in Table 2. The model achieved outstanding results, with an accuracy of 0.97. Remarkably, the model did not produce any FPs from the validation set. The recall rate of the model was 0.99, suggesting that it accurately classified a high proportion of validation frames as acceptable, such as flares, smaller lumps, dusty frames, etc. Frames containing objects such as concrete lumps, metal plates, safety gear were accurately classified as unacceptable by the model. Finally, the CNN model was found to have a fast inference time of 8.1 milliseconds per image, making it highly efficient for practical use. The findings presented in this study provide a promising initial framework for understanding operations in the conveyor belt environment. However, it is crucial to acknowledge the restricted scope of the dataset utilized in this research, which may not comprehensively reflect all operational events, particularly those that are unanticipated. These unforeseen circumstances frequently occur in the surroundings of the conveyor belt, thus highlighting the necessity for further data collection and retraining of the model.

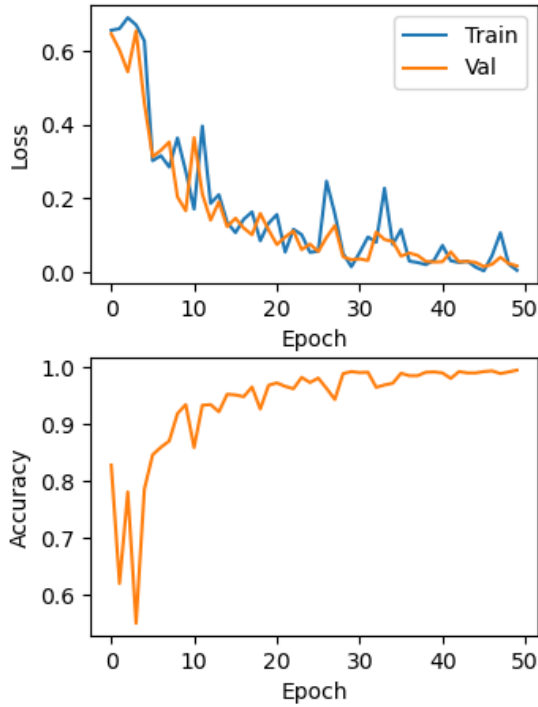


Figure 7: Training and validation accuracy and loss per training step.

Table 2: Metrics of the CNN

Metric	Value
Accuracy	0.97
PPV	0.95
Recall	0.99
Processing time	8.1 ms/image

4. CONCLUSIONS AND RECOMMENDATIONS

In this paper, an object detection model was presented for identifying hazardous materials in conveyor transport systems using data collected from a video stream. The model comprised a two-step classification algorithm, with the first step being a primitive brightness and contour detection approach that used different colour channels and grayscale images, while the second step was a CNN model that classified the detections as either acceptable or unacceptable.

Although the primitive approach detected many test objects placed on the belt, it also produced several false positives such as flares due to its intensity-based methodology. The use of colour channels assisted in the detection of coloured objects, but dark objects were not detected because of their low intensity. Further development is needed to enhance the detection of dark and less reflective objects.

The second classification step, the CNN model, produced excellent results, effectively eliminating false positives and accurately detecting hazardous objects. Since the solution is a two-step model, the first step can be more aggressive, knowing that false positives can

be eliminated by the second step. This will enable operators to stop the belt when alerted in the SaaS dashboard, reducing unplanned downtime and costs and preventing damage and contamination.

While the results are promising, future work can consider several recommendations. For example, the CNN model may be used without the first classification step, increasing the likelihood of detecting dark objects, simplifying the process, and reducing the required computing power. As the system continues to operate, more data can be collected, and with operator feedback and ground truth assessment, the CNN can be retrained and refined for further unknown situations that the conveyor belt system may present.

ACKNOWLEDGMENTS

The authors of this paper would like to thank Tord Arnøy and the operators at LKAB Narvik for their contributions, insights and help with testing.

REFERENCES

- [1] M. Zhang, H. Shi, Y. Zhang, Y. Yu, and M. Zhou, "Deep learning-based damage detection of mining conveyor belt," *Measurement*, vol. 175, p. 109130, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0263224121001561>
- [2] K. Al-Kahwati, W. Birk, E. F. Nilsfors, and R. Nilsen, "Experiences of a digital twin based predictive maintenance solution for belt conveyor systems," *PHM Society European Conference*, vol. 7, pp. 1–8, 2022. [Online]. Available: <https://doi.org/10.36001/phme.2022.v7i1.3355>
- [3] K. Al-Kahwati, E. Saari, W. Birk, and K. Atta, "Condition monitoring of rollers in belt conveyor systems," in *2021 5th International Conference on Control and Fault-Tolerant Systems (SysTol)*, 2021, pp. 341–347.
- [4] G. Alvarado-Robles, R. A. Osornio-Ríos, F. J. Solís-Muñoz, and L. A. Morales-Hernández, "An approach for shadow detection in aerial images based on multi-channel statistics," *IEEE Access*, vol. 9, pp. 34 240–34 250, 2021.
- [5] Maison, T. Lestari, and A. Luthfi, "Retinal blood vessel segmentation using gaussian filter," *Journal of Physics: Conference Series*, vol. 1376, no. 1, p. 012023, nov 2019. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1376/1/012023>
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: <https://doi.org/10.1038/nature14539>
- [7] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, 2021.
- [8] S. Suzuki and K. be, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp.

32–46, 1985. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0734189X85900167>

[9] D. P. Kingma and J. L. Ba, “Adam: A method for stochas-

tic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.