Journal of Universal Computer Science, vol. 28, no. 1 (2022), 80-97
submitted: 22/6/2021, accepted: 20/10/2021, appeared: 28/1/2022 CC BY-ND 4.0

Fastener Classification Using One-Shot Learning with Siamese Convolution Networks

Canan Tastimur
(Erzincan Binali Yildirim University, Erzincan, Turkey
https://orcid.org/0000-0002-8429-854X, ctastimur@erzincan.edu.tr)

Erhan Akin
(Firat University, Elazig, Turkey
https://orcid.org/0000-0001-6476-9255, eakin@firat.edu.tr)

Abstract: Deep Learning has been widely used in image-based applications such as object classification, object detection, and object recognition in recent years. Classifying highly similar objects is a very difficult problem. It is difficult to classify datasets in this situation where object similarity between classes and differences between classes are high. In this study, Siamese Convolution Neural Network, which is a similarity measurement-based network, has been practiced to classify 6 types of screws, 5 types of nuts, and 7 types of bolts that are very similar to each other. In addition, this neural network formed with the One-Shot Learning technique is trained. Thanks to the OSL technique, there is no need to use large data sets. Also, there is no need to use large amounts of data from each class. Adding a new class to be classified is also made easier by the use of the OSL technique. The performance results of the proposed method are manifested in detail in the article.

Keywords: Classification, Deep learning, Fastener, Siamese network, One-shot learning
Categories: I.2.0, I.2.1, I.2.6, I.2.8, I.2.10, I.4.0
DOI: 10.3897/jucs.70484

1 Introduction

Fasteners are important for machines to function properly. Computer vision-based detection of fastener defects or malfunctions prevents major problems from occurring. With the development of computer vision techniques, many studies have been conducted on object detection and classification. The fact that fasteners are very similar to each other makes object classification difficult. In deep learning (DL) based object classification applications, large datasets are needed even though there are enough images of each class. Moreover, the convolutional neural network model needs to be retrained and the network parameters need to be adjusted when a new object to be classified is added. These problems do not exist with Siamese networks. Adding a new class to the Siamese network is fairly straightforward. Since there is not much data in our dataset, the use of One-Shot Learning (OSL) technique in this study contributes to this. Studies on Siamese networks and OSL have been explored in the literature.

In this study, object classification has been performed. The objects classified are screws, bolts and nuts. Since the examples of these fasteners are visually very similar to each other and the differences between the classes are small, the classification becomes difficult. Deep learning based methods are used for classification of fasteners.

In addition, Siamese networks are used. Siamese networks measure the similarity between image inputs. In this network, the output of the same convolutional neural network is determined as the result of two image inputs with the same parameter and weight values. OSL has been implemented because the image data in this study is not sufficient for training in the deep learning model. Training with OSL does not require a large number of images from each class. In this study, 6 screws, 7 nuts, and 5 bolts have been classified. The proposed work has been examined as 3-way, 5-way, and 10-way, and the performance results are presented comparatively.

The contributions of own study are summarized below:

- In this study, very similar objects are classified using OSL, Siamese network and DL. From studies using OSL, Siamese network and DL [Agarwal et al. 2020] classified 5 objects, [Hossain et al. 2020] classified 5 objects and [Sabri and Setumin 2021] classified 4 objects, 18 objects were classified in our study.
- While [Agarwal et al. 2020], [Hsiao et al. 2019] and [Sabri and Setumin 2021] performed their studies with only Triplet loss, only L2-distance and only Euclidian loss, respectively at the classification stage, L1-distance and Contrastive loss functions were tested in our study and the experimental results obtained were presented comparatively.
- Although [Sabri and Setumin 2021], [Ullah et al. 2020], [Hsiao et al. 2019] tried the N parameter in N-way OSL on multiple values as in our study, in our study, it was aimed to obtain optimum results by considering parameters such as the number of random examples and number of iteration in addition to the N-way parameter.
- In our study, unlike other studies, Sigmoid, Softplus, and Tanh activation functions and L1-distance and Contrastive loss functions were applied to obtain optimum results in our experimental studies. In addition, while examining the classification performance in our study, the effect of the number of possible classes, the number of random examples, and the number of iteration on the accuracy rate was investigated.

2 Literature Review

The OSL and Siamese networks were used to detect Bangladeshi banknotes with a small dataset [Hossain et al. 2020]. Twenty images were used to recognize five different banknotes. The results were obtained with an accuracy of 97.38% in training and 94.48% in validation. The regularized cross entropy was used as a loss function in the study. The model in [Hossain et al. 2020] was run with 50 epochs. In another study, the similarity between two anomaly sequences is measured using 3D CNN Siamese network for anomaly detection, and OSL is used to solve the problem of lack of data. Once the network is trained, it can be referred to anomaly detection not only for new data but also for new classes. The results are compared with the proposed method in [Ullah et al. 2020], 5-way, 10-way and 14-way. Lungu et al. [Lingu et al. 2020] worked with OSL and Siamese network model to measure the similarity between image pairs and classify new objects quickly. The developed OSL and Siamese network model was tested using Omniglot, Tiny-Imagenet, CIFAR-100 and RoShamBo datasets. The use

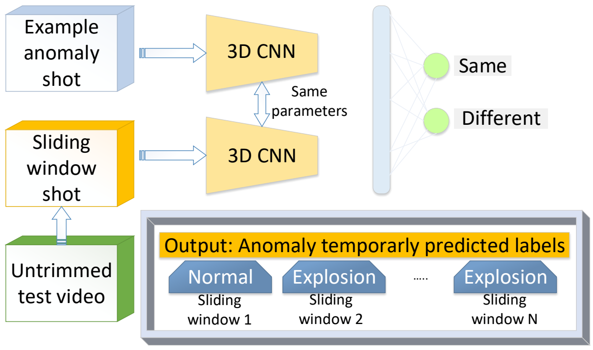

Figure 1: The block scheme of the study in [Ullah et al. 2020]

of the multi-resolution Siamese network contributed to an increase in accuracy of about 10%.

A convolutional Siam network was proposed that learns salient features for relation inference and requires only a limited number of samples for unusual relations [Yuan et al. 2017]. The recommended network gave good results in 1-shot and 10-shot learning situations for domain-specific information extraction. The study in [Agarwal et al. 2020] presented an approach for image-based classification of plastic waste using one-shot learning for the problem of sorting recyclable waste. In this study, Siamese and Triplet Loss convolutional neural networks were used to discriminate between 5 types of plastic waste. An accuracy of 99.74% was achieved in the WaDaBa dataset. A study was conducted on human face recognition by applying OSL to radar sensor data [Pho et al. 2020]. The OSL method facilitated the extraction of features from labeled samples and the application of these features to new samples. Signals from multiple channels were combined to form the input of the OSL model. The accuracy rate of the method produced by combining OSL and Siamese network was 97.6%. In another study, Siamese network and OSL were used for a face drawing recognition study [Sabri and Setumin 2021]. OSL was adopted since each image has only one photo. In the study in [Sabri and Setumin 2021], Siamese networks with the same architecture and weights were used to find out the similarity between the two images. Euclidean distance was used in the similarity calculation in the study. An accuracy of 100% was achieved using 300 epochs, sigmoid activation function and 10-way OSL. The dataset used in the study was CUHK.

A study was conducted to investigate the recent developments in Siamese convolutional neural networks [Eloff et al. 2019]. In the study in [Eloff et al. 2019], a dataset consisting of verbal and visual figures was used. High accuracy was achieved by using pixel distance on the developed Siamese model images. Deep transfer learning was investigated in Siamese networks for animal identification [Van Zyl et al. 2020]. A deep convolutional Siamese network was proposed to achieve high classification accuracy in online signature verification [Vorugunti et al. 2019]. Three datasets MCYT-100, MCYT-330 and SVC -2004-Task2 were used. In [Putra and Setumin 2021], OSL was run in Siamese networks with different activation features for face recognition. Sigmoid, tanh, softmax, softsign and softplus activation functions were tried and the highest performance was obtained with sigmoid activation function and

N-way OSL with 92% accuracy. Siamese convolutional neural network was used to identify a random number of individuals in the camera image [Liu et al. 2018]. Real-time one-shot learning was applied to never-before-seen classes of individuals. In the study in [Sokar et al. 2018], Siamese convolutional neural networks and Support Vector Machine (SVM) were proposed to be used for visual character recognition. The classification accuracy decreases when the available dataset is insufficient. Moreover, the model needs to be re-trained when a new class is to be classified. To avoid these situations, the deep Siamese convolutional neural network extracted salient features. The training was performed once for a group of classes. Different classes can then be recognized without the need to retrain or fine-tune the model. The performance of the network was improved by 12% using the Siamese network.

3 Materials and Method

3.1 One-Shot Learning

It is the process of classification, although there are few examples of each class [O'Mahony et al. 2019; Xiong et al. 2019]. They are used with OSL siamese nets. Because Siamese nets do similarity learning with few data. The first use of Siamese networks was by Browley and LeCun in the early 1990s in the signature verification problem [Bromley et al. 1993]. The concept of OSL is the use of single data for the learning process. It is used in situations where there is no large data set in the traditional approach and constantly changing inputs and outputs. OSL is the learned version of a single data set. OSL technique decides whether two input images taken are the same by comparing them. This situation is illustrated in Figure 2. The use of OSL is recommended when the traditional use of ConvNet is not appropriate [Grigorescu 2018]. Some of these situations have already been mentioned. A small training set is not sufficient to train a robust neural network. The use of ConvNet is not appropriate because the trained feature vectors do not contain important information that can be used for future image recognition. When the number of classes or datasets to be trained increases, the network needs to be retrained, but this is not suitable for using ConvNet because it consumes too much time and resources. For these reasons, OSL can be used.

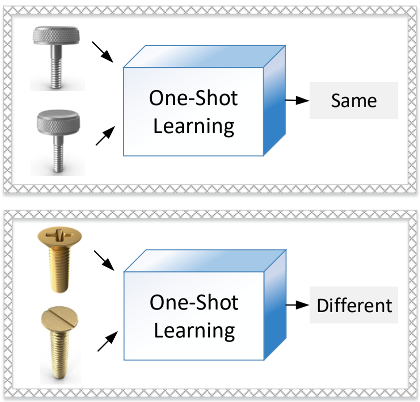

Figure 2: An example of the demonstration of the OSL [Agarwal et al. 2020]

There are several methods to implement the OSL technique: Siamese network, triplet loss integration and contrastive loss for dimensionality reduction.

3.2 Siamese Network for One-Shot Learning

The Siamese network aims to learn how images are encoded in order to measure how different the two images are. These networks consist of two identical neural networks. The parameters are shared in these neural networks. That is, the parameters are the same in both networks [Koch et al. 2015]. This property provides a very important advantage. The predictions generated by neural networks are consistent. That is, when very similar images are used because they are networks with the same weight, the network result shows that these two images belong to the same class. Similarly, if images from different classes are given, the difference will be large, so these two images are assumed to belong to different classes [Torres et al. 2020; Rao et al. 2017]. Siamese networks consider the difference between the outputs that both inputs will produce in the network. Based on that, it decides that these two images belong to the same or different classes. An input given to the neural network is encoded in the network and the output is exhibited. Then the second input given to the network is encoded in the same way and the output is obtained. The difference between these two outputs is considered. If this difference is less than the given threshold, these two images are considered to belong to the same class and if it is large, these two images belong to different classes. This situation is illustrated in Figure 3.

The working principle of the Siamese network is as follows: First, the network takes two inputs and then these inputs are passed through the neural network with the same weights and structure to obtain the feature vectors. The difference value of the obtained feature vectors is calculated using any distance measurement function. If the difference value is less than the specified threshold, the two inputs are similar,

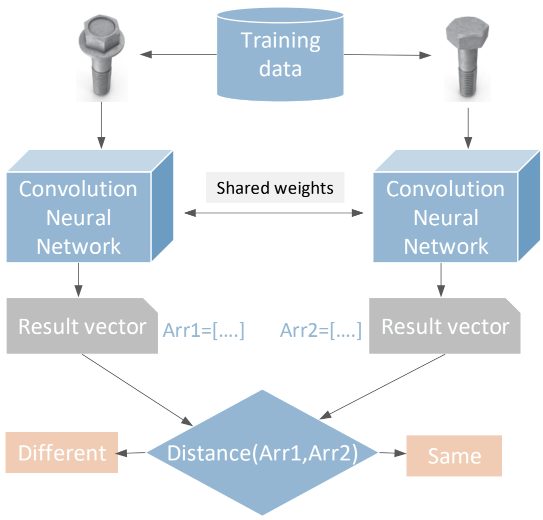

Figure 3: An example demonstration of the Siamese network [Micro Articles 2021]

otherwise they are different from each other. Siamese network uses the same weights to work with two different input vectors to compute output vectors, similar to an artificial neural network. In this network, one of the output vectors is precomputed to create a baseline against which the other output vector is compared. The Siamese network obtains an output vector by encoding the images it receives as input. Since it performs the similarity computation over this vector, the network model must learn how to encode the images to obtain this output vector.

One advantage of using the Siamese network is that it further reduces intra-class imbalance by using only a small amount of data from each class for classification. Another advantage is that it can provide better classification results than the average two correlated supervised models. Moreover, Siamese networks are also useful in learning semantic similarity. The disadvantages of Siamese networks are that they require more training time than conventional networks and the probability value of the classes cannot be generated eventually because the training involves binary learning, it gives the distance to each class, not the probabilities of the prediction.

3.3 Triplet Loss Integration

In Siamese networks, three images are used as input. The first of these images is the base image, the second is the positive image and the third is the negative image. The base image is any image from any class. A positive image is an image that is in the same class as the base image, but different from the base image. A negative image is an image of a different class than the base image. When examining the difference in distance between these images, the difference between the base image and the positive image is expected to be less than the difference between the base image and the negative image [Hsiao et al. 2019].

$$\|f(B) - f(P)\|^2 < \|f(B) - f(N)\|^2 \tag{1}$$

$$\|f(B) - f(P)\|^2 - \|f(B) - f(N)\|^2 < 0 \tag{2}$$

In equation (1), the values B, P, and N represent the base image, the positive image, and the negative image, respectively. The model can be solved by giving all values the value 0. To do this, a value (marg) is inserted into equation (2) and thus equation (3) is obtained. Based on this information, the loss function is expressed in equation (4).

$$\|f(B) - f(P)\|^2 - \|f(B) - f(N)\|^2 + marg < 0 \tag{3}$$

$$L = max\ ((\|f(B) - f(P)\|^2 - \|f(B) - f(N)\|^2 + marg), 0) \tag{4}$$

3.4 Contrastive Loss for Dimensionality Reduction

The term ‘dimensionality reduction’ refers to the process of reducing the dimensions of a feature vector. When converting the input image to a low-dimensional feature vector, it is critical that the one-dimensional arrays formed from the image after passing through the fully connected convolutional layer capture the features in the original image. Traditional neural networks are designed to reduce the loss function to the smallest achievable value. Unlike prediction-based error functions, contrastive loss is a distance-based loss function. It compares three images as contrastive loss and triplet loss. However, instead of comparing the images of contrastive loss simultaneously with triplet loss, it performs the comparisons in pairs one after another.

$$L = \frac{(1-Y)}{2} D_w^2 + \frac{Y}{2} max\,(0, marg - D_w)^2 \tag{5}$$

If the images given as input are equal, the variable Y in equation (5) returns 0; if they are different, it returns 1. The distance between the two output vectors is denoted by the variable D_w. The loss function will be 0 if the distance between the two images is 0, so we cannot train the model, so the marg value is added to the loss function. It is 1 if the images we give as input belong to the same class, and 0 if they belong to different classes. Error functions are calculated based on the distance between the images. This situation is expressed in Figure 4 as base image, negative image and positive image.

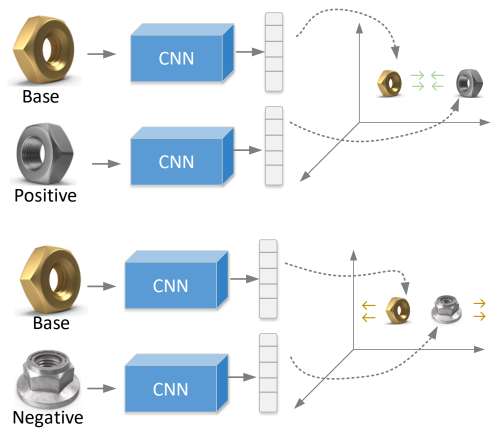

Figure 4: The functioning of the contrastive loss function [HeartBeat 2020]

3.5 The Proposed Method

A new approach is proposed for the classification of highly similar connectors. Siamese networks, OSL and deep learning techniques have been used in the proposed approach. The proposed study has been tried as 3-way, 5-way and 10-way OSL. The number of classes to be classified using one-shot learning technique is described by the N-value in the N-way variable. The number of samples used in training for each class to be classified is one. The success rate of the proposed method decreases as the number of classes to be classified increases. The accuracy rates obtained as a result of the experimental studies are presented comparatively. In this study, the L1 distance is used as the loss function of the Siamese network. Manhattan distance, also known as L1 distance, is used to calculate the distance between i and j points in p-dimensional space. It is equal to the sum of the lengths of the line segment projections between the points on the coordinate axes. It is the sum of the absolute differences between the measures in all dimensions of two points in the base measure. Equation (6) formulates this explanation.

$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{ip} - x_{jp}| \quad (6)$$

$$L1\_Distance = \sum_{i=1}^{n} |y_{true} - y_{prediction}| \quad (7)$$

In equation (7), the y_{true}value represents the actual tag value of the classified object, while the $y_{prediction}$ value represents the estimated tag value of the classified object. It is used in many fields, such as regression analysis and frequency distribution. Hermann Minkowski was the one who first introduced it. The L1 distance between two extracted feature vectors is computed using a matched layer, rather than the regular CNNs used in standard classification. The number and size of filters have been determined here by different experiments. The two outputs of the two image inputs given to this CNN network architecture designed for Siamese networks have been evaluated by the comparison using the L1 distance and the contrastive loss functions, and the similarity score of the two inputs has been measured. In the output of the network, similarity scores between 0 and 1 are recorded, with as many values as the number of classes to classify.

In the constructed Siamese CNN network, the input images are 105x105 pixels in RGB format. Several pre-processing steps are used to convert the input image imported into the network into a two-dimensional grayscale format, which reduces the computational cost in the network. There are 64 10x10 convolutional layers, 2x2 maximum pooling layers, 128 7x7 convolutional layers, 2x2 maximum pooling layers, 128 4x4 convolutional layers, 2x2 maximum pooling layers, 256 4x4 convolutional layers, flatten layer, 256 fully connected layers, and a customized layer in the content of the identical CNN network architecture used in Siamese networks. The accuracy rate of a Siamese neural network with k one-shot tasks and N times true predictions is formulated in equation (8) during the testing procedure.

$$Accuracy\_ratio = (100 * N)/k \quad (8)$$

The performance of the Siamese network with one-shot learning built in this study has also been compared to fundamental methods like 1-Nearest Neighbor and Random Guessing. Since the result produced by the K-NN method is matched with the outcome of the one-shot learning method, the k value is chosen as 1 and the 1-nearest neighbor method is employed. In the K-NN method, L2-distance is utilized in the calculation of the loss function. The weights are randomly initialized in the random guessing technique until all of the training images are correctly classified [Hsiao et al. 2019].

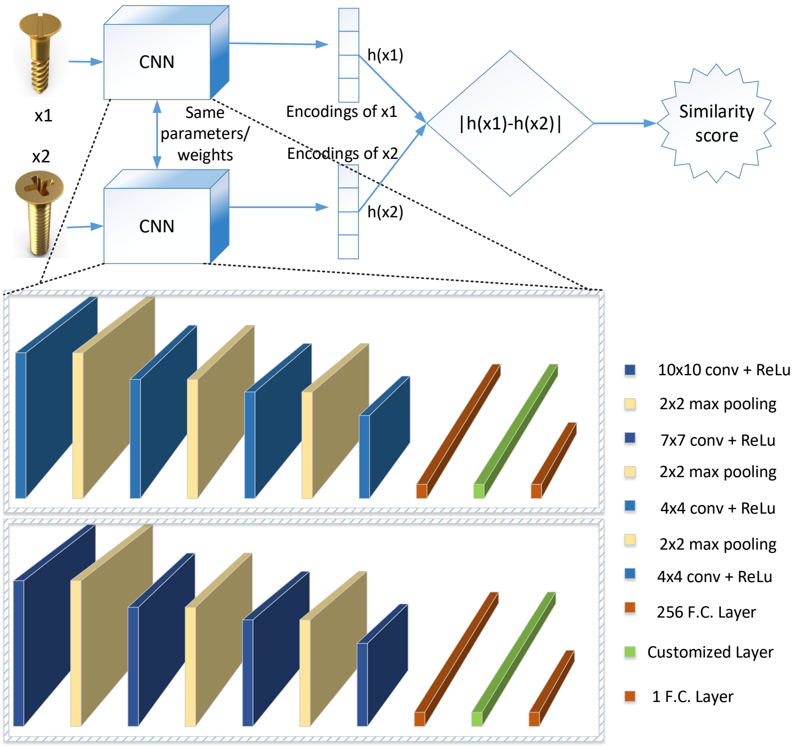

Figure 5: The proposed approach with DL

The N-way value in the testing procedure depends on the accuracy of the random guessing approach.

In addition, the Constructive loss function has been also tested under equal conditions and it is noticed that the best result is reached with the L1 distance function. CNN architecture developed with Siamese networks has been tested with Sigmoid, Softplus, and Tanh activation functions. The hyper-parameters that are 16 batch size, 100 epochs, Adam optimizer and 50 one-shot tasks on validating are set. When partitioning the dataset utilized in this study for training, validation, and testing, K-fold cross-validation has been used. The process of dividing the dataset by 20% has been carried out.

4 The Experimental Results

This research has been carried out using a desktop computer with an Intel(R) Core(TM) i7-8550U CPU running at 1.80GHz, 8GB of DDR3 RAM, and an NVIDIA GeForce MX150 GPU. The fastener dataset created has been used in this study. There are 18 categories in this dataset. Since the similarity between the images in the categories is quite high, it becomes difficult to classify them. In the study, the Siamese CNN network developed with 3-way, 5-way, and 10-way OSL has been tested. Obtained performance results are presented. Performance results of Sigmoid, Softplus, and Tanh are given as the activation function in the proposed network.

Hyper-parameter	**Explanation of hyper-parameter**	**Value**
batch_size	Number of image pairs in each iteration	16
epoch	Number of training steps of the network	100
n_val	Number of one-shot tasks to validate	50
n_way	Number of classes for testing one-shot tasks	3, 5, 10
evaluate_every	Interval value for evaluating one-shot tasks	1
loss_every	Interval value for printing loss on iterations	50
loss_function	The function of the network used to calculate the loss value	L1_Distance
Optimizer	Algorithm that constantly minimizes the error in the predictions	Adam

Table 1: Hyper-parameter values set in this study.

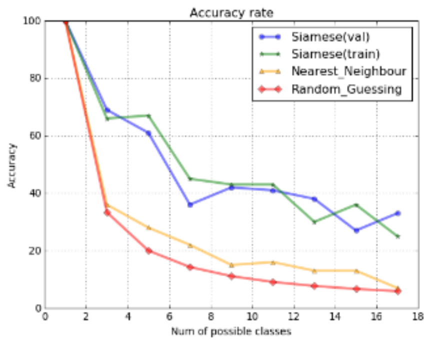

Figure 6: Accuracy rate by the number of possible classes with Sigmoid activation function, L1 distance loss and 3-way OSL

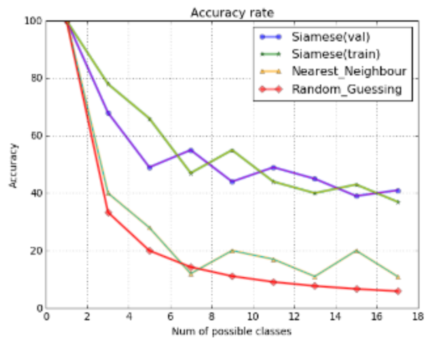

Figure 7: Accuracy rate by the number of possible classes with Softplus activation function, L1 distance loss and 3-way OSL

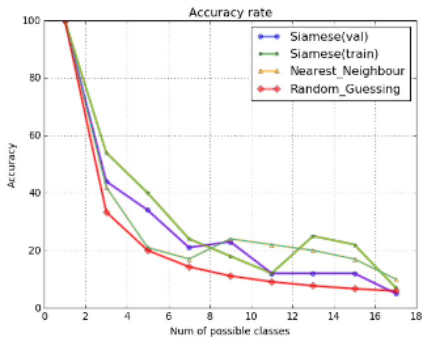

Figure 8: Accuracy rate by the number of possible classes with Tanh activation function, L1 distance loss and 3-way OSL

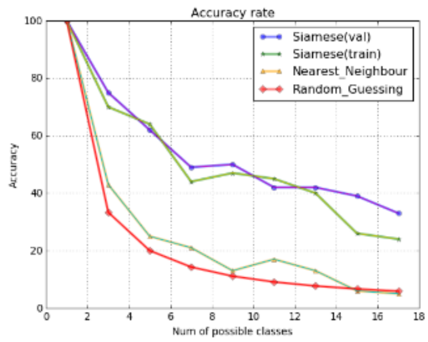

Figure 9: Accuracy rate by the number of possible classes with Sigmoid activation function, L1 distance loss and 5-way OSL

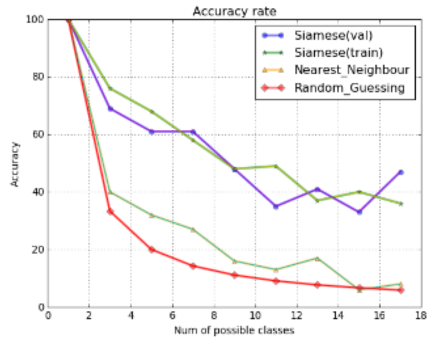

Figure 10: Accuracy rate by the number of possible classes with Softplus activation function, L1 distance loss and 5-way OSL

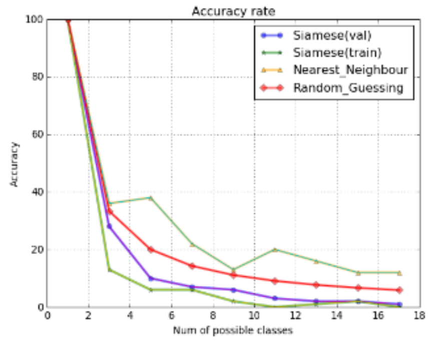

Figure 11: Accuracy rate by the number of possible classes with Softplus activation function, Constrastive loss and 5-way OSL

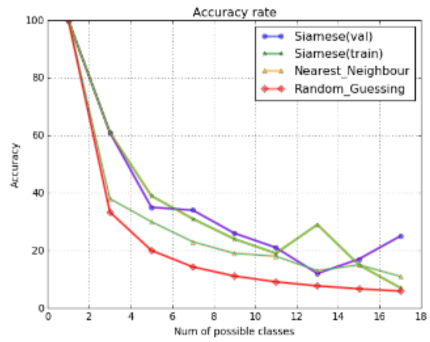

Figure 12: Accuracy rate by the number of possible classes with Tanh activation function, L1 distance loss and 5-way OSL

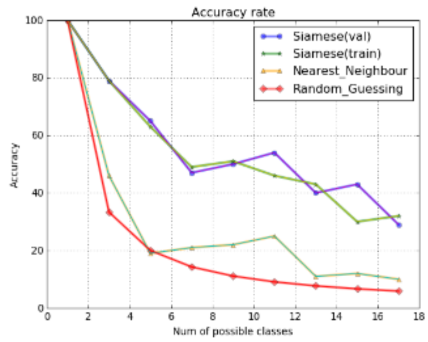

Figure 13: Accuracy rate by the number of possible classes with Sigmoid activation function, L1 distance loss and 10-way OSL

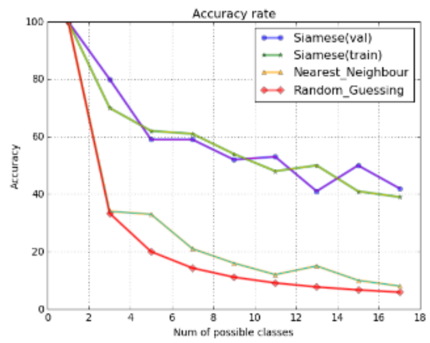

Figure 14: Accuracy rate by the number of possible classes with Softplus activation function, L1 distance loss and 10-way OSL

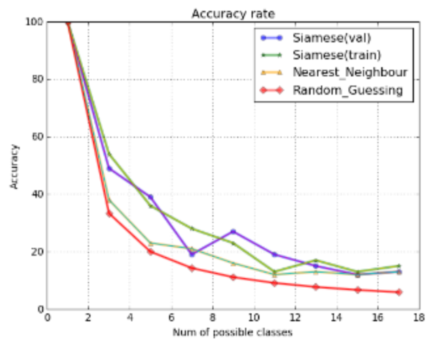

Figure 15: Accuracy rate by the number of possible classes with Tanh activation function, L1 distance loss and 10-way OSL

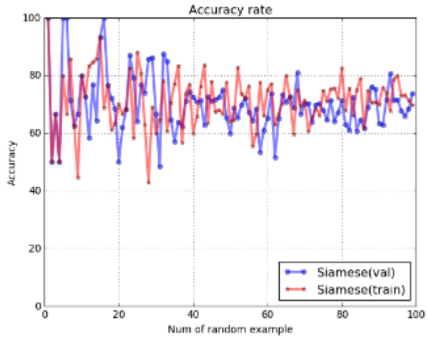

Figure 16: Accuracy rate by the number of random example with Sigmoid activation function, L1 distance loss and 3-way OSL

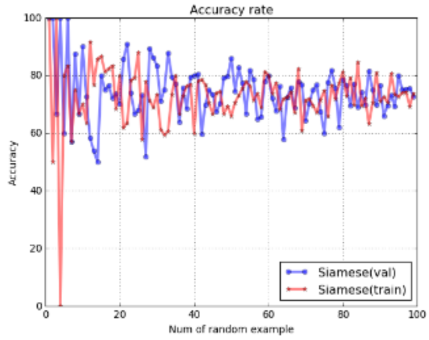

Figure 17: Accuracy rate by the number of random example with Softplus activation function, L1 distance loss and 3-way OSL

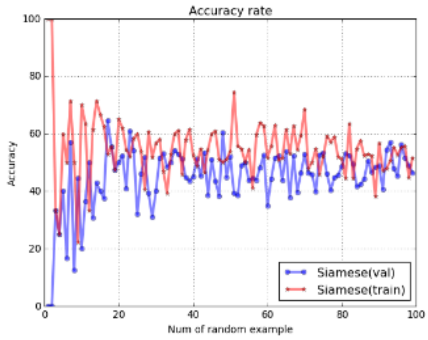

Figure 18: Accuracy rate by the number of random example with Tanh activation function, L1 distance loss and 3-way OSL

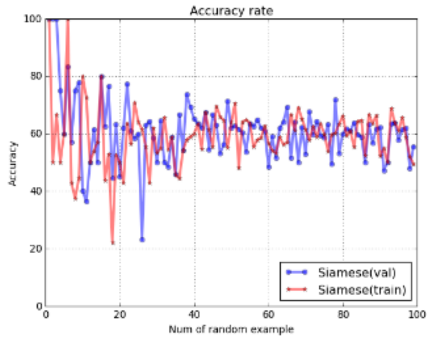

Figure 19: Accuracy rate by the number of random example with Sigmoid activation function, L1 distance loss and 5-way OSL

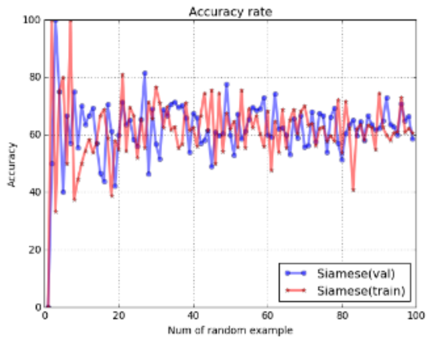

Figure 20: Accuracy rate by the number of random example with Softplus activation function, L1 distance loss and 5-way OSL

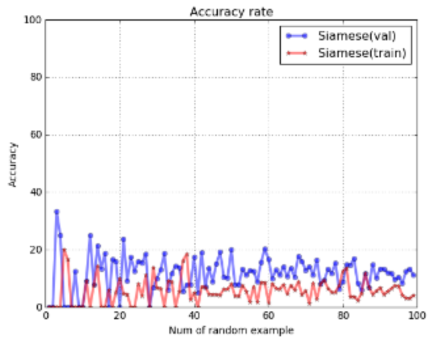

Figure 21: Accuracy rate by the number of random example with Softplus activation function, Constrastive loss and 5-way OSL

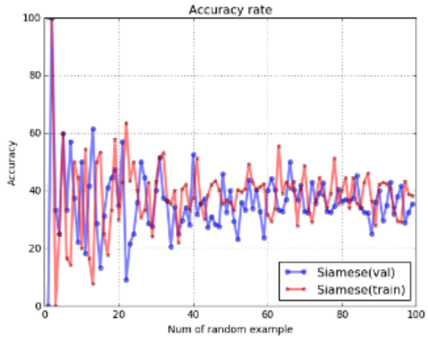

Figure 22: Accuracy rate by the number of random example with Tanh activation function, L1 distance loss and 5-way OSL

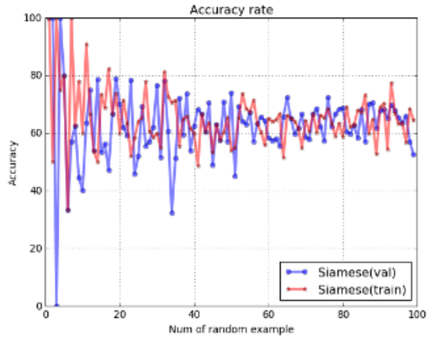

Figure 23: Accuracy rate by the number of random example with Sigmoid activation function, L1 distance loss and 10-way OSL

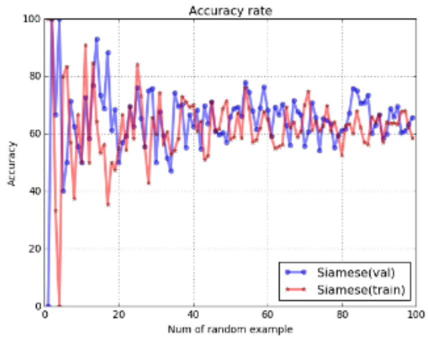

Figure 24: Accuracy rate by the number of random example with Softplus activation function, L1 distance loss and 10-way OSL

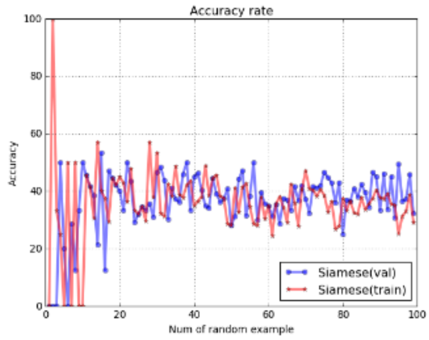

Figure 25: Accuracy rate by the number of random example with Tanh activation function, L1 distance loss and 10-way OSL

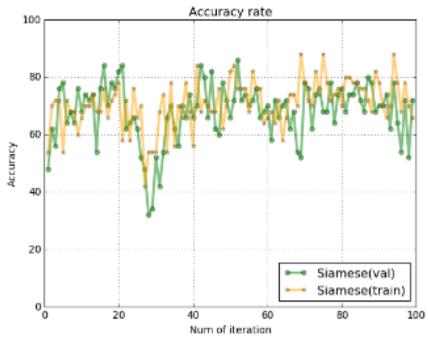

Figure 26: Accuracy rate by the number of iteration with Sigmoid activation function, L1 distance loss and 3-way OSL

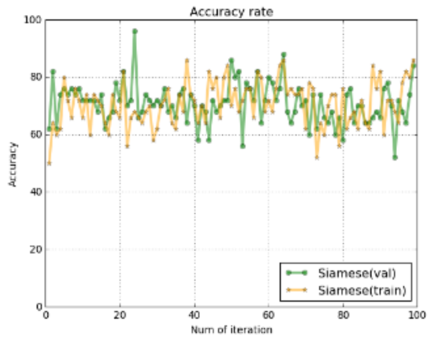

Figure 27: Accuracy rate by the number of iteration with Softplus activation function, L1 distance loss and 3-way OSL

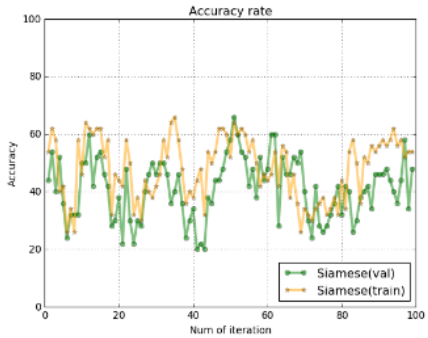

Figure 28: Accuracy rate by the number of iteration with Tanh activation function, L1 distance loss and 3-way OSL

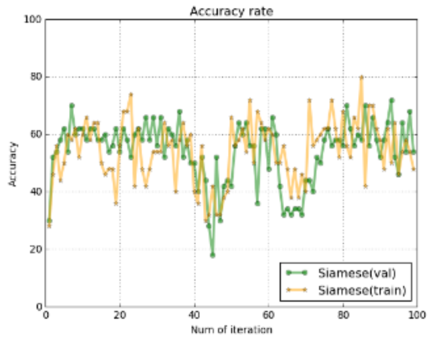

Figure 29: Accuracy rate by the number of iteration with Sigmoid activation function, L1 distance loss and 5-way OSL

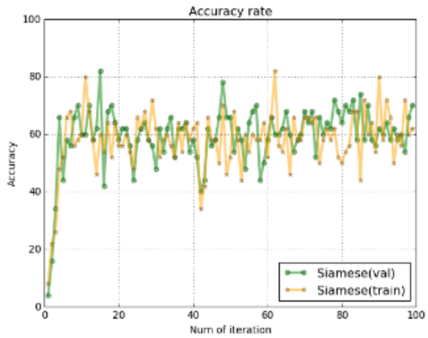

Figure 30: Accuracy rate by the number of iteration with Softplus activation function, L1 distance loss and 5-way OSL

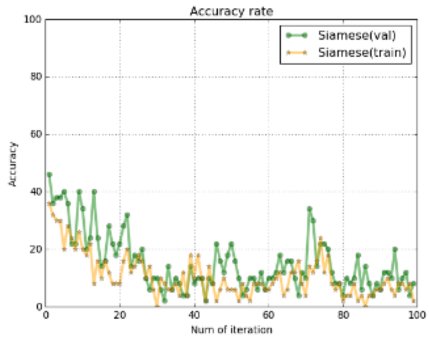

Figure 31: Accuracy rate by the number of iteration with Softplus activation function, Constrastive loss and 5-way OSL

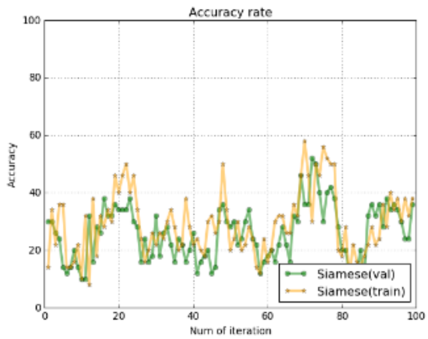

Figure 32: Accuracy rate by the number of iteration with Tanh activation function, L1 distance loss and 5-way OSL

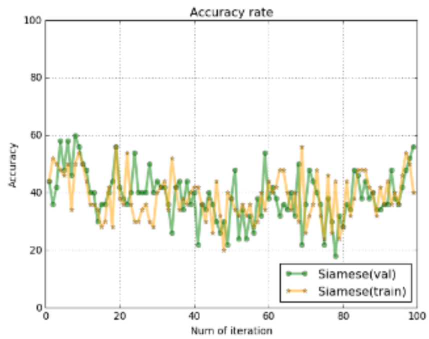

Figure 33: Accuracy rate by the number of iteration with Sigmoid activation function, L1 distance loss and 10-way OSL

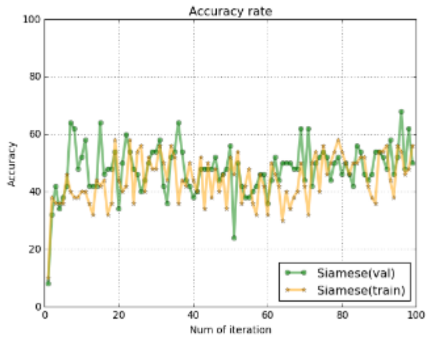

Figure 34: Accuracy rate by the number of iteration with Softplus activation function, L1 distance loss and 10-way OSL

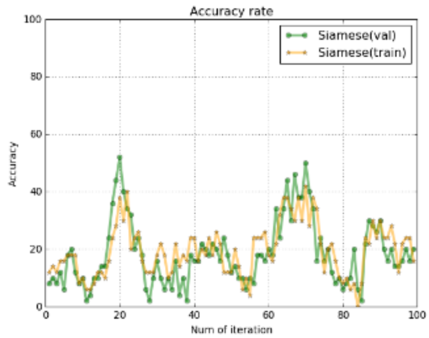

Figure 35: Accuracy rate by the number of iteration with Tanh activation function, L1 distance loss and 10-way OSL

In our study, the effect of many parameters on performance during the use of OSL and Siamese CNN networks has been evaluated. For this, OSL technique has been applied as 3-way, 5-way, and 10-way. In addition, the effect of loss function and activation function on performance has been also investigated. In our study, L1-Distance loss and Contrastive loss functions have been applied. The contribution of both functions to the study has been determined. Finally, the effect of Sigmoid, Tanh and Softplus activation functions on performance has been investigated. In addition to all these, the importance of parameters such as the number of possible classes, the number of random example and the number of iteration on the classification have been examined. If we consider the N-way, loss function and activation function parameters when we want to examine the effect on the number of possible classes classification (from Figure 6 to Figure 15), the increase in the number of classes to be classified in the network causes a decrease in the accuracy of the network. The best accuracy rate has been obtained in Figure 14. With the increase in the number of classes to be classified, it has been observed that the network can mislabel these classes that are quite similar to each other. If we consider the N-way, loss function and activation function parameters when we want to examine the effect of the number of random example on the classification (from Figure 16 to Figure 25), if the increase in the number of sample images to be used randomly in the validation phase of the classification exceeds a certain threshold value. runs at certain ranges. No matter how much the number of samples increases after a certain value, the network performance remains constant within a certain range. The best accuracy rate was obtained in Figure 17. Finally, when the accuracy rates according to the number of iterations were examined, it was determined that the best performance was obtained in Figure 27.

5 Conclusions

Traditional CNN is widely practiced in object classification studies. However, it becomes difficult to classify with traditional CNN when the number of training samples in the dataset is insufficient, when new categories need to be added to the dataset and when the samples in the dataset can change frequently. In these cases, the use of the OSL method is beneficial. Siamese networks are needed during the implementation of the OSL method. Siamese networks consist of two identical networks, and the output values produced against the inputs to these networks are compared with the difference calculation metrics, and thus the similarity score between the two inputs is calculated. In this study, classification of the highly similar screw, bolt, and nut fasteners using deep learning, OSL, and Siamese networks have been performed. A total of 18 classes have been classified with 7 types of bolts, 5 types of nuts, and 6 types of screws. Siamese CNN network formed with OSL has been tested under various scenarios. Accuracy rates of the Siamese CNN network formed according to the determined number of iterations of 3-way, 5-way, and 10-way OSL, the number of random samples used in the validation phase and the number of classes to be classified has been calculated. L1-distance and Contrastive loss distance measurement metrics are tried as loss functions. In addition, the similarity score of the two input images has been produced 3 different activation functions (Sigmoid, Softplus, and Tanh) in the developed Siamese CNN network. When the experimental results obtained are reviewed, the performance ranking among all N-way states is Softplus, Sigmoid, and Tanh activation functions, respectively. When the accuracy rates according to the number of classes have been observed, it is obtained from the 3-way OSL with the Softplus activation function. In this case, the performance of the network decreases as the number of classes increases in N-way OSL. The accuracy rate of the basic comparison methods 1-NN and Random Guessing methods is lower than the performance of Siamese networks in all cases. With the increase in the number of possible classes, the accuracy rates have decreased greatly. Since the OSL method compares the samples in the chosen class, the increase in the number of classes complicates the problem. When the performance of the network is examined according to the number of iterations, as the number of iterations increases, overfitting is prevented, the network performance has been in a constant range after a certain number of iterations. When the performance of the network is checked according to the number of random samples selected in the validation process, the accuracy rate remains within a certain range if the number of samples exceeded a certain threshold value. When the performance of L1-distance and Contrastive loss functions in the network is observed under equal conditions, there has been a reduction of approximately 50% in the network with the application of Contrastive loss. For this reason, the L1 distance function approves to be used in the developed network.

References

[Agarwal et al., 2020] Agarwal, S., Gudi, R. & Saxena, P. (2020). One-Shot learning based classification for segregation of plastic waste. arXiv preprint.

[Bromley et al., 1993] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. Advances in neural information processing systems, 6:737-744.

[Eloff et al., 2019] Eloff, R., Engelbrecht, H. A. & Kamper, H. (2019). Multimodal one-shot learning of speech and images. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8623-8627.

[Grigorescu, 2020] Grigorescu, S. M. (2018). Generative One-Shot Learning (GOL): A semi-parametric approach to one-shot learning in autonomous vision. In 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 7127-7134.

[HeartBeat, 2020] One-shot learning (Part 1/2): Definitions and fundamental techniques, 2020, https://heartbeat.fritz.ai/one-shot-learning-part-1-2-definitions-and-fundamental-techniques-1df944e5836a

[Hossain et al., 2020] Hossain, M. E., Islam, A. & Islam, M. S. (2020). A Proficient Model to Classify Bangladeshi Bank Notes for Automatic Vending Machine Using a Tıny Dataset with One-Shot Learning & Siamese Networks. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-4.

[Hsiao et al., 2019] Hsiao, S. C., Kao, D. Y., Liu, Z. Y. & Tso, R. (2019). Malware image classification using one-shot learning with siamese networks. Procedia Computer Science, 159:1863-1871.

[Koch et al., 2015] Koch, G., Zemel, R. & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In ICML deep learning workshop, 2.

[Liu et al., 2018] Liu, Z., McClung, A., Yeung, H. W., Chung, Y. Y. & Zandavi, S. M. (2018). Top-down person re-identification with Siamese convolutional neural networks. In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-8.

[Lungu et al., 2020] Lungu, I. A., Hu, Y. & Liu, S. C. (2020). Multi-resolution siamese networks for one-shot learning. In 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), pp. 183-187.

[Micro Articles, 2021] One-Shot Learning with Siamese Network, 2021, https://www.samaya.tech/pages/similar/g0q4t9aix1.html?utm_source=similar&utm_medium=pages&utm_campaign=v2

[O'Mahony et al., 2019] O'Mahony, N. et al. (2019). One-Shot Learning for Custom Identification Tasks; A Review. Procedia Manufacturing, 38:186-193.

[Pho et al., 2020] Pho, H. A., Lee, S., Tuyet-Doan, V. N. & Kim, Y. H. (2020). Radar-Based Face Recognition: One-Shot Learning Approach, IEEE Sensors Journal, 21(5):6335-6341.

[Putra and Setumin, 2021] Putra, A. A. R. & Setumin, S. (2021). The Performance of Siamese Neural Network for Face Recognition using Different Activation Functions. In 2021 International Conference of Technology, Science and Administration (ICTSA), pp. 1-5.

[Rao et al., 2017] Rao, D. J., Mittal, S. & Ritika, S. (2017). Siamese neural networks for one-shot detection of railway track switches, arXiv preprint.

[Sabri and Setumin, 2021] Sabri, N. I. A. & Setumin, S. (2021). One-Shot Learning for Facial Sketch Recognition using the Siamese Convolutional Neural Network. In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 307-312.

[Sokar et al., 2018] Sokar, G., Hemayed, E. E. & Rehan, M. (2018). A generic OCR using deep siamese convolution neural networks. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 1238-1244.

[Torres et al., 2020] Torres, L., Ribeiro, B. & Arrais, J. (2020). Exploring a Siamese Neural Network Architecture for Drug Discovery. Deep Drug Discovery and Deployment Project, University of Coimbra, CISUC, DEI.

[Ullah et al., 2020] Ullah, A., Muhammad, K., Haydarov, K., Haq, I. U., Lee, M. & Baik, S. W. (2020). One-Shot Learning for Surveillance Anomaly Recognition using Siamese 3D CNN. In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8.

[Van Zyl et al., 2020] Van Zyl, T. L., Woolway, M. & Engelbrecht, B. (2020). Unique Animal Identification using Deep Transfer Learning For Data Fusion in Siamese Networks. In 2020 IEEE 23rd International Conference on Information Fusion (FUSION), pp. 1-6.

[Vorugunti et al., 2019] Vorugunti, C. S., Mukherjee, P. & Pulabaigari, V. (2019). OSVNet: convolutional siamese network for writer independent online signature verification. In 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1470-1475.

[Yuan et al., 2017] Yuan, J., Guo, H., Jin, Z., Jin, H., Zhang, X. & Luo, J. (2017). One-shot learning for fine-grained relation extraction via convolutional siamese neural network. In 2017 IEEE International Conference on Big Data (Big Data), pp. 2194-2199.

[Xiong et al., 2019] Xiong, P., He, K., Song, A. & Liu, P. X. (2019). A Novel Multi-Modal One-Shot Learning Method for Texture Recognition. IEEE Access, 7:182538-182547.