# Training with single class images and Generalizing for multi-class images using Kasami Orthogonal Classification Layer

Mohamed Saadeldin *[1], Ammar Khairi[2], Amel AbdElraheem[2], Arjun Pakrashi[1], and Brian MacNamee[1]

[1]University College Dublin
[2]University of Khartoum

## Abstract

This work studies and evaluates extending the generalization of neural networks trained on single class data to provide classification output for multi-label data without further training. In this context we evaluate and compare performance of neural networks with conventional classification/output layer against neural networks utilizing the *Kasami Orthogonal Classification Layer* (KOCL) proposed in [5]. KOCL is used as an output layer for classification networks and consists of a fully connected layer same like conventional classification/output layer but it has fixed weights (non-trainable) that are equal to a set of orthogonal Kasami codes. Evaluation was carried out by training (two types of networks) a VGG/WResNet network on the standard MNIST, FashionMNIST, CIFAR10, SVHN single label data-sets then testing it on multi-label images synthetically generating by merging two random images from the same data-set. Results show that neural networks trained on single label images can generalize to multi-label images and provide classification predictions with high accuracy. Moreover when comparing neural networks trained with conventional classification layer to networks trained with KOCL then testing them for multi-label classification under the same setup, the later provided far better multi-label classification accuracy.

*Corresponding Author: mohamed.saadeldin@ucd.ie

## 1 Introduction

The KOCL was proposed as an output layer for classification networks in [5], it consists of a fully connected layer same like conventional classification/output layer but it has fixed weights (non-trainable) that are equal to a set of orthogonal Kasami codes. Each Kasami code among the set is assigned to one of the output neurons of the classification/output layer. Networks trained with KOCL learns to generate a unique latent representation for each data class that has values equivalent to the Kasami code values initially assigned as fixed weights to the output neuron associated with that specific data class.

Another interesting property of Kasami codes is that they can be nicely separated from each other when mixed together, since each code is orthogonal to all other codes. In this work we exploit this property by training neural networks only on single label images then test it with multi-label images without additional multi-label training. Although the operations and transformations performed by the neural network can not be considered linear, herein we assume that latent representations generated by the network for a data point that belongs to more than one class can be approximated as the sum of the individual latent representations learned during training for all the classes present in this multi-label data point.

## 2 Experimental Evaluation

A set of experiments were conducted to evaluate the ability of neural networks trained on single label

Table 1: Conventional layer classification accuracy of two test images merged with various overlap levels

| | **0%** | **10%** | **20%** | **30%** | **40%** | **50%** | **60%** | **70%** | **80%** |
|---|---|---|---|---|---|---|---|---|---|
| **MNIST** | 97.96 | 98.49 | 97.77 | 96.23 | 94.33 | 89.81 | 80.04 | 76.8 | 73.09 |
| **Fashion** | 86.96 | 86.24 | 83.49 | 79.89 | 82.08 | 77.9 | 72.82 | 71.78 | 64.85 |
| **CIFAR10** | 74.11 | 73.07 | 68.43 | 65.56 | 59.92 | 55.25 | 51.49 | 49.33 | 49.01 |
| **SVHN** | 80.35 | 79.4 | 79.84 | 74.73 | 66.28 | 60.24 | 58.22 | 58.19 | 55.77 |

Table 2: KOCL classification accuracy of two test images merged with various overlap levels

| | **0%** | **10%** | **20%** | **30%** | **40%** | **50%** | **60%** | **70%** | **80%** |
|---|---|---|---|---|---|---|---|---|---|
| **MNIST** | 99.53 | 99.61 | 99.52 | 99.23 | 98.19 | 95.69 | 87.65 | 82.18 | 76.49 |
| + % | 1.57 | 1.11 | 1.76 | 2.99 | 3.85 | 5.88 | 7.61 | 5.38 | 3.41 |
| **Fashion** | 88.51 | 91.76 | 88.29 | 86.23 | 86.94 | 81.27 | 78.69 | 78.31 | 69.91 |
| + % | 1.55 | 5.52 | 4.81 | 6.35 | 4.87 | 3.37 | 5.87 | 6.53 | 5.06 |
| **CIFAR10** | 85.65 | 84.12 | 81.21 | 75.18 | 67.34 | 59.22 | 52.87 | 50.79 | 50.11 |
| + % | 11.54 | 11.05 | 12.78 | 9.62 | 7.42 | 3.97 | 1.38 | 1.46 | 1.1 |
| **SVHN** | 88.65 | 85.05 | 87.05 | 84.82 | 75.32 | 67.18 | 61.56 | 63.09 | 62.13 |
| + % | 8.3 | 5.65 | 7.21 | 10.09 | 9.04 | 6.94 | 3.34 | 4.9 | 6.36 |

data to generalize its classification prediction results to multi-label data and compare the test accuracy of conventional classification layer and KOCL in performing this task.

As they are well understood and widely used in the community, in our experiments we use four popular image classification datasets: MNIST [1], Fashion MNIST [7], CIFAR10 [3], and SVHN [4]. Each of these datasets was divided into an 80% training set and a 20% validation set.

A small CNN with 7 convolutional layers was constructed, using 3 VGG like blocks [6] and a single point-wise convolutional layer at the end. Each of the 3 blocks consists of two convolutional layers with kernel size of 3X3, padding of 1, and a stride of 1. Batch normalization is applied after each convolutional layer, followed by ReLU nonlinearity. The spatial dimension is reduced by half at the end of each block using a 2X2 max pooling layer. The 3 VGG blocks are then preceded by a final layer of point-wise convolution where a kernel of size 1X1 is used to reduce the number of output feature maps. Finally a global average pooling is applied to produce the output latent representation vector. This latent representation vector is then passed to either a conventional classification layer or KOCL depending on which setup is being tested.

The adam optimizer [2] with learning rate equal to 0.001 and zero weight decay is used throughout training. A batch size of 200 examples was used and each network was allows to train for 200 epochs. These values were selected after a limited hyper-parameter tuning applied to the network described above when using a conventional classification layer at the end of the network. The same hyper-parameters were used when training the network with KOCL without further tuning or adjustments.

# 3 Conclusion

Neural networks trained on single label data showed very good potential to generalize for multi-label test data without requiring any additional multi-label training. KOCL showed significant performance advantage compared to conventional classification layer where it consistently provided higher testing accuracy of multi-label data in all the experiments conducted in this study. KOCL ability to encourage the network to produce Kasami codes as latent representation for different data classes provide it with a competitive advantage when the network is tested against multi-label images due to the orthogonality properties of Kasami codes.

## References


[1] L. Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[2] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[3] A. Krizhevsky, G. Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[4] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[5] M. Saadeldin and B. Macnamee. An orthogonal classification layer with kasami sequences for discriminative feature learning in neural networks. In *Proceedings of The 33rd IEEE International Conference on Tools with Artificial Intelligence*, page Not available online yet, 2021.

[6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[7] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.