2021

Class distribution-aware adaptive margins and cluster embedding for classification of fruit and vegetables at supermarket self-checkouts

Khurram Hameed

Douglas Chai

Alexander Rassau

10.1016/j.neucom.2021.07.040

Hameed, K., Chai, D., & Rassau, A. (2021). Class distribution-aware adaptive margins and cluster embedding for classification of fruit and vegetables at supermarket self-checkouts. *Neurocomputing*, *461*, 292-309. https://doi.org/10.1016/j.neucom.2021.07.040

Neurocomputing 461 (2021) 292–309

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Class distribution-aware adaptive margins and cluster embedding for classification of fruit and vegetables at supermarket self-checkouts ☆

Khurram Hameed *, Douglas Chai, Alexander Rassau

School of Engineering, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027, Perth, Australia

ARTICLE INFO

Article history:
Received 4 March 2021
Revised 21 May 2021
Accepted 15 July 2021
Available online 20 July 2021
Communicated by Zidong Wang

Keywords:
Distribution-aware clustering
Adaptive classification margins
Deep feature embedding
Fruit and vegetables classification

ABSTRACT

The complex task of vision based fruit and vegetables classification at a supermarket self-checkout poses significant challenges. These challenges include the highly variable physical features of fruit and vegetables i.e. colour, texture shape and size which are dependent upon ripeness and storage conditions in a supermarket as well as general product variation. Supermarket environments are also significantly variable with respect to lighting conditions. Attempting to build an exhaustive dataset to capture all these variations, for example a dataset of a fruit consisting of all possible colour variations, is nearly impossible. Moreover, some fruit and vegetable classes have significant similar physical features e.g. the colour and texture of cabbage and lettuce. Current state-of-the-art classification techniques such as those based on Deep Convolutional Neural Networks (DCNNs) are highly prone to errors resulting from the inter-class similarities and intra-class variations of fruit and vegetable images. The deep features of highly variable classes can invade the features of neighbouring similar classes in a learned feature space of the DCNN, resulting in confused classification hyper-planes. To overcome these limitations of current classification techniques we have proposed a class distribution-aware adaptive margins approach with cluster embedding for classification of fruit and vegetables. We have tested the proposed technique for cluster-based feature embedding and classification effectiveness. It is observed that introduction of adaptive classification margins proportional to the class distribution can achieve significant improvements in clustering and classification effectiveness. The proposed technique is tested for both clustering and classification, and promising results have been obtained.

1. Introduction

Classification of fruit and vegetables is a complex task that is based on a set of highly variable attributes i.e. texture, colour, shape and size. These attributes are used as features and their variability poses significant challenges for a classification task. Due to this variance, it is very impractical to obtain an exhaustive dataset to train a classifier for fruit and vegetables classification. Such features also provide a perfect environment for creation of imbalanced and incomplete datasets under the open-set protocol [62]. Training on an imbalanced dataset can constraint performance of a classifier to readily available samples [31,32]. For example, a fruit and vegetables dataset will usually have more images of a fruit with normal colour, texture, shape and size than a fruit with irregular colour, shape or size. In the past several years, Deep Learning (DL) approaches have achieved state-of-the-art classification performance, but are prone to errors when trained using imbalanced and complex datasets. Deep Convolutional Neural Networks (DCNNs) have achieved a performance boost for different applications e.g. object recognition [1,19], and face recognition and verification [51,71,75,60]. The non-linear hyper-dimensional features extracted by the layered architecture of the DCNN have the capability to learn the lower and higher level visual details of an image. Incremental depth enhancement [41,61] is used as the most common technique to deal with complex datasets in the state-of-the-art DCNNs, although smaller stride sizes [59] and non-linear activations [26] are also used. The strong learning capability of the DCNNs can also create issues of overfitting and memorisation of the complete training datasets. Large scale dataset [56], CNN connection dropout [39], data augmentation [61], feature regularisation [64,26] and stochastic pooling [72] are amongst the recent techniques proposed to deal with this issue.

☆ This work is supported by Edith Cowan University (ECU), Australia and Higher Education Commission (HEC) Pakistan, The Islamia University of Bahawalpur (IUB) Pakistan (5–1/HRD/UESTPI(Batch-V)/1182/2017/HEC). The authors would like to thank ECU Australia, HEC and IUB Pakistan for PhD grant of first and corresponding author of this paper.

* Corresponding author.
E-mail addresses: k.hameed@ecu.edu.au (K. Hameed), d.chai@ecu.edu.au (D. Chai), a.rassau@ecu.edu.au (A. Rassau).

https://doi.org/10.1016/j.neucom.2021.07.040
0925-2312/

Much research has been reported recently on the methods to overcome the issue of imbalanced and complex datasets [32,11,38,18,46]. The most common techniques include synthetic sampling and sensitive cost estimation. Class priors are balanced by synthetically over or under sampling the classes. For instance, a synthetic sampling technique has been used in [49], where foreground and background images are re-sampled for object classification. Inverse class frequencies have been used as a cost function for semantic segmentation with an ensemble of Support Vector Machines (SVMs) in [9]. An inverse class frequency is used as a scaling factor for the loss function in [47], where a significant improvement is reported for an imbalanced class semantic segmentation task. Similarly, a relative [55] and median [21] class frequency has been used to scale the loss function for face attributes recognition and complex scene understanding, respectively. The loss function is regularised for edge detection in [58], where an equal scaling factor is used for positive and negative class edge examples. An improvement of this work is proposed in [54], where an adaptive scaling factor has been applied at pixel-level for observed segmentation loss. Feature transformation and embedding is also used as a technique to strengthen DCNNs with more discriminative features. The intuition behind this approach is to maximise the inter-class separability and intra-class compactness to enhance the distribution margins between the features in an imaginary space with d dimensions. However, this is a complex task for applications with significant intra-class variance e.g. fruit and vegetables classification. Recently, much research has been reported on feature transformation and embedding to enhance the intra-class compactness, hence inter-class separability [27,57]. However, most of the techniques proposed such as [27,57] have an inherent limitation based on the significant time complexity. Both techniques work on tuples of training data and can reach to a non-linear time complexity, for example $\mathcal{O}(N^3)$ in the case of [57], where N is number of training samples.

In our research, we have explored vision-based classification of fruit and vegetables at a supermarket self-checkout as a complex application of computer vision. This application also has additional complexities due to variable lighting conditions and backgrounds in addition to the feature variability. The schematic and design of a proposed supermarket self-checkout kiosk along with preliminary efforts on vision-based fruit and vegetables classification is reported in [29,30,2]. However, more sophisticated vision-based classification techniques need to be explored for this complex application [28]. Moreover, considering our commercial application, a significant accuracy is required to classify the fruit and vegetables to a known class in order to demonstrate the variability of this approach. There are significant limitations in the state-of-the-art techniques discussed above i.e. are these techniques effective for fruit and vegetables classification in a supermarket, where the imbalance and complexity of datasets is barely considered? Moreover, classification of fruit and vegetables requires a DCNN to estimate the complex hyperplane due to the large number of fruit and vegetable varieties. The above discussed techniques have been studied for shallow and small CNNs [22], but their implementation to DCNNs has not previously been studied in detail. There are also inherent limitations from synthetic sampling and sensitive cost estimation. For instance, over and under sampling techniques can introduce unnecessary noise and or loss of significant information, respectively. The meta heuristic techniques applied for cost sensitivity analysis have been applied to DCNNs however, no noticeable performance increase has been reported [35,36,10].

To overcome the limitations of the current state-of-the-art techniques, we have proposed cluster-based feature embedding with adaptive classification margins, which can maintain a large inter-class separability and intra-class compactness. This cluster-based embedding is motivated by the observation that the fruit and vegetables classes have significant intra-class variability and inter-class similarities. There are fair chances that the instances from the classes with significant variability can invade the classes with a reasonable similarity. The invasion of instances offer an extra level of complexity for a classifier, where the estimated hyperplanes will get confused. This condition is true for many fruit and vegetable classes that have significant similarity in physical features e.g. lettuce and cabbage, orange and mandarin. An example of Gaussian kernel density distribution of colour channels in the RGB and YC_bC_r colour spaces of fruit and vegetable images with similar physical features is depicted in Fig. 1 illustrating the overlapping of the vision-based features. A set of 100 images per class has been used to extract the sample colour features and to understand the concept of feature overlapping. We have used Apple, Cabbage, Carrot, Lettuce, Mandarin, Orange and Tomato as fruit and vegetables with significant feature similarity for representation and presentation of our proposed concept, more details on the dataset are described in Section 4. We have selected these classes based on three pairs of classes with significant feature similarities i.e. Cabbage and Lettuce, Mandarin and Orange, and Apple and Tomato. We have also used images of Carrots as an independent class for better understanding and representation of the proposed approach. Implementation of the proposed technique with DCNNs can be beneficial to classify the considered fruit and vegetables with significantly similar physical features. This technique can also be considered as robust towards fruits and vegetables with deformation, damage and extra ordinary features, hence an open-set classification protocol. A comparison of the effectiveness achieved for the proposed and the state-of-the-art techniques for classification of fruit and vegetables in a supermarket environment is presented in Table 1. The comparison is emphasised on the application of the proposed technique in a real-life supermarket considering the segmentation techniques, number of classes, features representation and classification techniques. It is noted that inter-class feature similarity and intra-class variance is scarcely considered along with limited datasets for experiments. The earliest approach for vision-based classification of fruit and vegetables in a supermarket is reported in [8]. A dataset of 150 produce items consisting of 5000 images including fruit and vegetables has been developed where a combination of colour and texture features is used for classification achieving an accuracy of 96%. However, details regarding class-level dataset distribution are missing, if considered as a uniform distribution it can be concluded that the dataset is significantly small and can result in the overfitting. A similar limitation can be noted in [53,20,50] where there are a small number of images per class, moreover the inter-class similarities are not considered for classification. A visual produce verification has been performed in [7] using the HSI based colour histogram where no fresh produce (i.e. fruit and vegetables) is considered. Similarly, no fresh produce is considered for vision-based retail in [70] where produce item are detected and classified using You Only Look Once (YOLO) CNN. A custom supermarket dataset based fruit and vegetables classification is performed in [28] where an AdaBoost CNN optimisation technique is used for classification. However, it can be noted that less effective classification results are obtained for fruit and vegetables with similar physical features. The techniques presented in [50,23] can also be considered prone to misclassification of the fruit and vegetables with similar physical features. Considering these limitations a significant corollary can be concluded that the state-of-the-art techniques are limited for classification of fruit and vegetates with similar physical features. We have emphasised on the classification of fruit and vegetables with significant similar physical features where a reasonable number of samples per class are considered in our

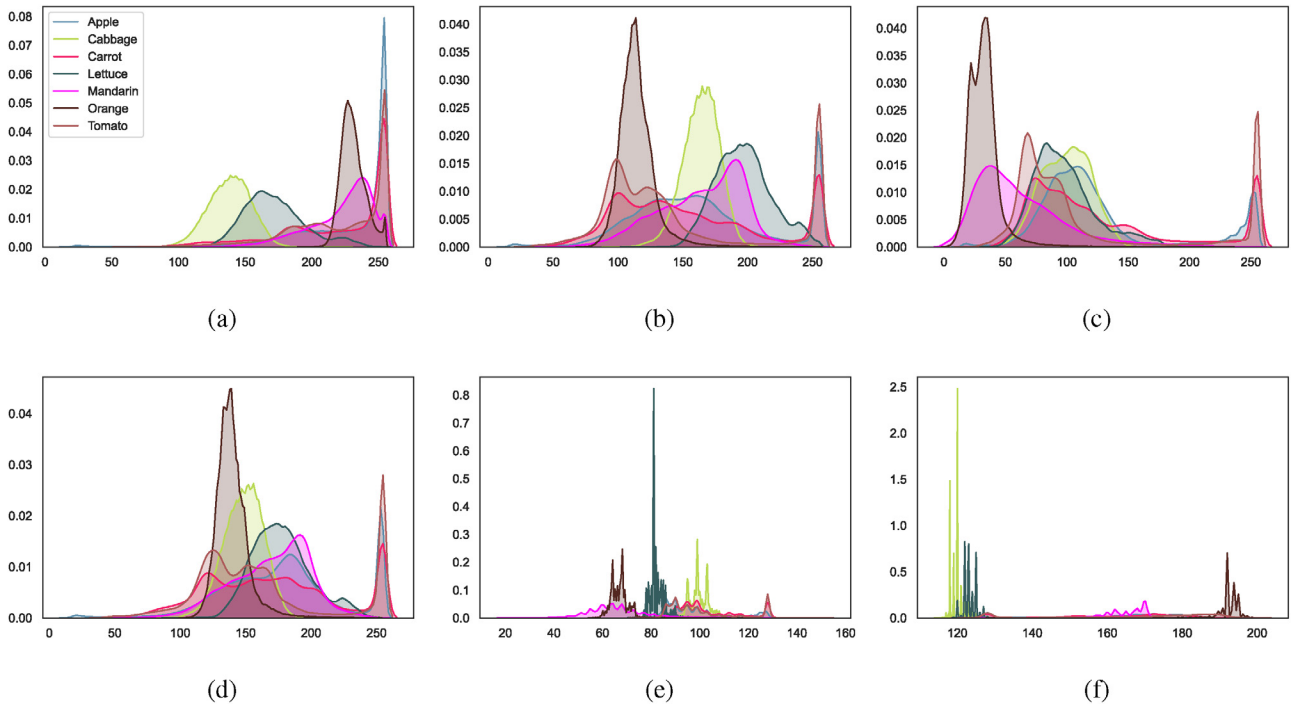

Fig. 1. Example Gaussian kernel density distribution of fruit and vegetables in RGB and YC_bC_r colour spaces: (a) R-channel, (b) G-channel, (c) B-channel, (d) Y-channel, (e) C_b-channel, and (f) C_r-channel.

Table 1
A comparison of the state-of-the-art fruit and vegetable classification techniques in a supermarket environment.

Year	Ref.	Classes	Segmentation	Features	Classification	Accuracy (%)
1996	[8]	150	Threshold based	HSI histogram and mask based texture	Euclidean KNN	96.00
2010	[53]	15	Normalised background	GCH, CCV and Unser's	SVM	95.00
2011	[7]	20	Pixel-level blob	HSI colour and edge contours	Barcode verification	70.00
2013	[20]	15	K-means clustering	Improved sum and difference histogram	SVM	96.90
2016	[70]	24	YOLO mask	YOLO mask shape	CaffeNet	66.40
2018	[50]	15	Resizing	CNN features	Fruit-Alex CNN	97.50
2018	[23]	10	Resizing	CNN features	MobileNet	97.00
2019	[24]	9	Manual cropping	Multi-CNN features	Inception net	97.70
2019	[52]	10	YOLOv3 mask	CNN features	RetinaNet	80.20
2020	[30]	15	Fixed background	CNN features	Custom CNNs	91.30
2020	[12]	12	YOLO mask	CNN features	Faster-RCNN	97.90
This work		**15**	**Fixed background**	**ResNet50 features**	**Distribution-aware feature embedding**	**99.00**

experiments to avoid overfitting; promising results have been achieved reported in Table 1.

The rest of the paper is organised as follows. The state-of-the-art techniques used for intra-class compactness and inter-class separability enhancement are discussed in Section 2. A detailed discussion on the proposed cluster-based embedding with adaptive classification margins is presented in Section 3. The experimental implementation inline with the tests performed to verify the effectiveness of the proposed approach is described in Section 4. The effectiveness of the proposed technique based on the results obtained is analysed in Section 5. An overall discussion on the proposed technique and future directions is provided in Section 6.

2. Related work

To the best of our knowledge, only a few efforts have been reported to deal with complex and imbalanced dataset classification with DCNNs [35,36,22,10,74,68,17,66,67,63,48]. A complementary Neural Network is used as an under-sampling technique for class imbalance improvement in [35]. A re-sampling is also performed based on the Synthetic Minority Over-sampling Technique (SMOTE) [11] technique to re-balance the training dataset. However, the SMOTE technique is not constrained with respect to the neighbouring samples considered i.e. the neighbouring samples can be from other overlapped classes, which can introduce significant noise. A DL on complex and imbalanced datasets has been proposed as an extension of conventional imbalanced learning techniques in [36]. A cost sensitivity loss function has been derived based on the complexity and skewness of the dataset for classification with a CNN. A similar approach has been proposed in [10], where a joint objective function is proposed on a binary classification problem. A Cost Sensitive Multi Layer Perception (CSMLP) method is proposed considering a single cost parameter based on the skewness of the samples in two classes. However, this cost parameter is significantly sensitive to the number of classes, level of imbalance, and overlap among features of different classes. This technique can also be considered as DL based extension of conventional techniques proposed in [6,37]. An alternative approach is proposed in [66], where a Mean Squared False Error (MSFE) loss has been defined on a binary classification. The proposed loss func-

tion is tested on eight imbalanced datasets and a significant performance gain is reported. A stacked auto encoder technique is used to deal with the overlapping classification problem. The learned features are concatenated to achieve a better representation using different properties of a dataset. A Sigmoid and tanh functions are compared for DL of complex datasets, where the former is reported as robust while learning the features of a complex dataset. An adaptive sample batch weighting technique has been proposed in [60] based on the gradient direction. A more sophisticated and unbiased validation dataset, is however, required for training the classifier. A multidimensional dataset distribution skewness is studied in [68], where the balance in one dimension does not guarantee the balance in all dimensions for features of a class in a hyperplane. The learned knowledge of balance features in a dimension is transferred to another dimension with skewed features. This technique, however, results in significant time complexity for complex and large datasets. There are also other significant limitations of these approaches, e.g. no class structure or data distribution is considered for all these approaches.

Significant efforts have been reported to enhance the loss function to deal with complex datasets for classification and recognition tasks. The Softmax loss is used in the state-of-the-art vision-based classification and recognition techniques i.e. face recognition and image classification. However, the Softmax loss is limited to the discriminative features for classification. Recently, many enhancements of Softmax loss have been proposed to deal with the less discriminative features [44,43,65]. These enhancements are based on the concept of forcing a large margin between the classification hyper-surfaces. More recent techniques have used the triplet loss [57] for classification of the complex dataset. A range-based loss is defined in [73], where the maximum intra-class feature difference (range) is used to adapt the inter-class classification margins. Combined Softmax, margin loss and centre loss are also used for classification of complex datasets [69,15]. However, all these techniques are limited to inter-class margin enhancement while ignoring the intra-class data distribution structure and any neighbouring feature invasion within the complex classes. A Class-level Rectification Loss (CRL) has been defined in [17] where the classes with small numbers of samples, imbalanced features and skewed features are identified at batch level. The features of such classes are then regularised to normalise the cross entropy loss for DL. This is a preliminary effort towards the class structure consideration however, feature regularisation and loss normalisation only for minority classes cannot be equally effective for a dataset like fruit and vegetables. In a fruit and vegetable dataset all classes are equally prone to imbalanced class distribution, for example samples with extraordinary features i.e. highly irregular shape. Multiple classes in a fruit and vegetables dataset can also have a significantly overlapping features. Considering this limitation we have proposed a multi-class cluster embedding to enhance the intra-class compactness and adaptive hyperplane margins to accommodate the inter-class separability. The proposed technique is considered more aware of distribution structure of dataset classes at global level as compared to the previously proposed approaches.

3. Methodology

DCNNs have significant limitations when learning non-linear hyperplanes to differentiate between complex and imbalanced classes in fruit and vegetable datasets. A simple example has been demonstrated in Fig. 2, where we have considered a set of 200 images per class for representation of learned features by a pre-trained ResNet50 [33]. The learned features have been embedded into a normalised unit sphere using t-distributed Stochastic Neighbor Embedding (t-SNE) [45] based feature embedding. A ResNet50 [33] pre-trained on the ImageNet dataset [14] has been used to learn these features for seven classes. It can be observed that there is significant inter-class overlap in the learned features especially for the fruit and vegetables with similar physical features e.g. texture and colour. This overlap confirms our concept of inter-class similarity and intra-class variations, where the significant variation in the features of a class can invade similar neighbouring classes.

Considering this limitation of DCNNs, we have proposed a multi-class cluster-based embedding approach with adaptive classification margins to deal with the complex dataset of fruit and vegetables. A block diagram of the proposed approach is depicted in Fig. 3. A simple assumption of minimum inter-class margins under the d-dimensional hyperplane has been considered where the minimal inter-class margins should be greater than the maximum intra-class variations. This assumptions has been considered carefully based on the state-of-the-art techniques for complex feature embedding discussed in Section 2. To overcome the limitation of a complex and imbalanced dataset we have used the vector projection based similarity of feature vectors extracted with a DCNN for clustering. The training set $\mathcal{Z} = \{s_i, c_i\}_{i=1}^{L}$, consists of a sample s_i with a corresponding class label $c_i \in \{1, \ldots, \mathcal{P}\}$, where $\mathcal{P}$ represents the number of classes. Our goal is to assign a cluster θ_i^c for each sample s_i in $\mathcal{Z}$, where the maximum number of clusters are represented by $\mathcal{K}$ described as:

$$\theta_1^c, \ldots, \theta_{\mathcal{K}}^c = \underset{\theta_1^c, \ldots, \theta_{\mathcal{K}}^c}{\arg\max} \sum_{k=1}^{\mathcal{K}} \sum_{i \in \theta_k^c} \mathcal{V}(s_i)^{\mathbf{T}} \bullet \mu_k^c, \tag{1}$$

$$\mu_k^c = \frac{1}{|\theta_k^c|} \sum_{i \in \theta_k^c} \mathcal{V}(s_i), \tag{2}$$

where μ_k^c represents the centre of the cluster k for a class c and CNN based features of sample s_i are represented as $\mathcal{V}(s_i)$. These supervised class label based cluster centroids are kept consistent for the training and testing process. A fixed number of clusters equal to the number of classes have been used for our experiments where the proposed technique allows us to achieve significant classification effectiveness as compared to the state-of-the-art differential labelling techniques e.g. pointwise mutual information, Chi-Squared and Euclidean norm based label estimation. These learned features are embedded to a d-dimensional space such that $\|\mathcal{V}(s_i)\|_2 = 1$. This constraint on the features helps to achieve significant benefits toward the highly variant features i.e. lighting conditions and any reasonable colour difference in the samples from a class. The cluster size can be expressed as $l_c = |\theta_k^c|$ where an assumption has been made that the cluster sizes are equal for all classes. This assumption will reduce the complexity of feature embedding to $\mathcal{K}$ clusters. To deal with the variable class distribution and complex dataset with imbalanced features due to significant feature variation of fruit and vegetables we have used a product of vectors projection in a d-dimensional Euclidean space as a similarity measure for clustering. This similarity measure for clustering is significantly helpful to deal with global data distribution-aware clustering, which is achieved by dealing with the complex and over lapping features from multiple classes in a coherent manner i.e. product of vector projection. The concept here is to use the Euclidean difference of cluster centroids and the projections of feature vectors on a unit sphere $\mathcal{E}$ for the similarity measure, which can be achieved as:

$$\mathcal{E}(\mathcal{V}(x_i), \mu_k^c) = \cos(\mathcal{V}(s_i), \mu_k^c) = \frac{\langle \mathcal{V}(s_i), \mu_k^c \rangle}{\|\mathcal{V}(s_i)\| \|\mu_k^c\|}. \tag{3}$$

This clustering technique helps in a standard characterisation of the complex features of a fruit and vegetables dataset with overlapping features. In the training process of $\mathcal{J}$ iterations a sample

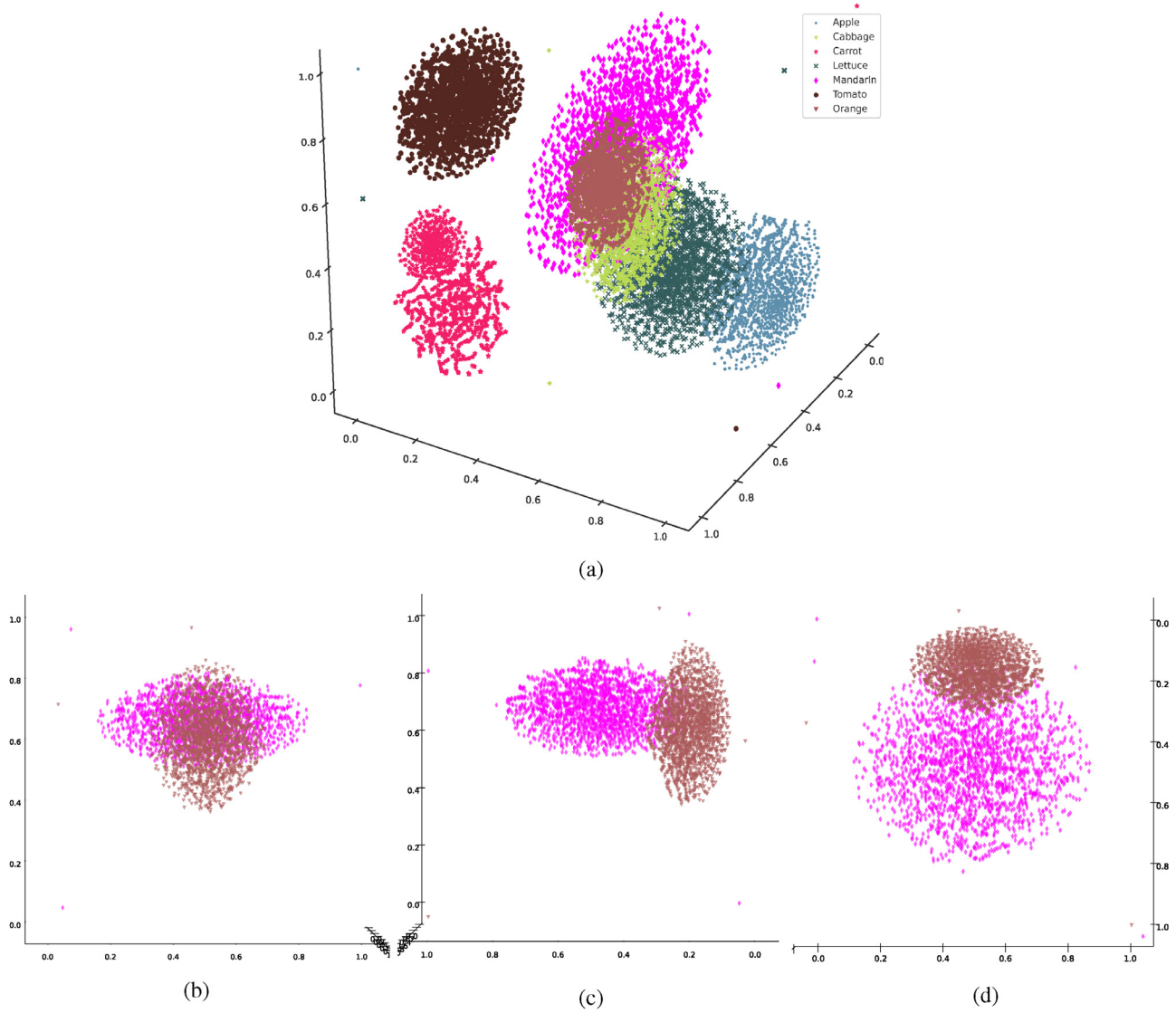

Fig. 2. A t-SNE [45] based 3D embedding to represent: (a) inter-class similarity and intra-class variation of pre-trained ResNet50 features of seven classes, and the respective (b) X-Y, (c) Y-Z, and (d) Z-X axes views of Mandarin and Orange featu.res overlapping.

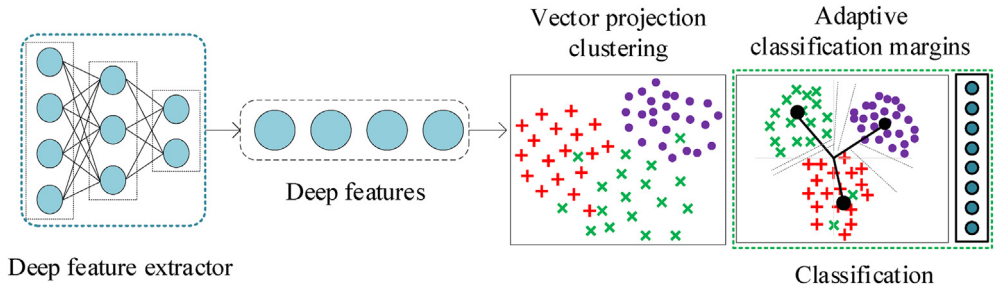

Fig. 3. A block schematic of the proposed multi-class cluster-based embedding and adaptive classification margin technique.

s_i is mapped to a cluster θ_i based on the similarity measure between s_i and $\mathcal{K}$ clusters depending upon the angle α between centroids and the feature vector projection onto a unit sphere. The intra-class compactness can be ensured by correctly mapping the high dimensional features i.e. $\mathcal{V}(s_i)$ with respect to μ_k^c. An extension of Softmax loss has been derived based on concept presented in [42,13] to enhance the intra-class compactness with the vector projection based clustering, described as:

$$L_c = -\frac{1}{\mathcal{P}}\sum_{i=1}^{\mathcal{P}}\log\left(\frac{e^{\mathcal{V}(s_i)^{\mathbf{T}}\bullet\mu_i}}{\sum_{\mathcal{K}}e^{\mathcal{V}(s_i)^{\mathbf{T}}\bullet\mu_i}}\right)$$
$$= -\frac{1}{\mathcal{P}}\sum_{i=1}^{\mathcal{P}}\log\left(\frac{e^{\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos\alpha_i}}{e^{\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos\alpha_i}+\sum_{j\neq i}^{\mathcal{J}}e^{\|\mathcal{V}(s_j)^{\mathbf{T}}\|.\|\mu_j\|\cos\alpha_j}}\right). \quad (4)$$

A projection based similarity of inter and intra class features is obtained by evaluating the query feature vector w.r.t. centroids of all classes. As discussed before we have normalised the class cen-

troid and query feature vector to a unit sphere i.e. L_2 norm is equal to 1. Considering the significantly larger number of classes of fruit and vegetables we have considered a sphere of radius $r \geqslant 1$. This consideration will allow us to deal with a scalable embedded feature space to deal with a large number of classes. The adaptive radius of the feature embedding space can be described as:

$$L_c = -\frac{1}{\mathcal{P}}\sum_{i=1}^{\mathcal{P}} \log\left(\frac{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos\alpha_i}}{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos\alpha_i} + \sum_{j\neq i}^{\mathcal{J}} e^{r\|\mathcal{V}(s_j)^{\mathbf{T}}\|.\|\mu_j\|\cos\alpha_j}}\right). \tag{5}$$

This cluster based embedding will distribute the features on an imaginary sphere of radius r based on the similarity between the centroids and the query feature vector. To improve the intra-class compactness in a progressive way with the corresponding training iterations, each class centroid is updated after a particular number of iterations. This centroid update will reduce the size of each cluster significantly and hence reduce the inter-class overlapping. The progressive update is intended to shift the centroid to the centre of the cluster, resulting in a reduced standard deviation as a metric of cluster size. Moreover, considering significant physical feature variations, centroid updating can help us to deal with the outliers. The centroid updating also allows us to consider the significant inter-class feature variations including both physical and environmental features. However, the inter-class distribution should also be considered to deal with the complex, imbalanced and open-set classification problem. We have used a linear penalty to enforce margin between classification boundaries. This linear penalty has been considered as a phase between the angles of embedded clusters on a feature space in terms of a sphere with a radius of r. This addition of phase is considered as an offset between the class centroids. This offset acts as an additive margin φ as described in (6). This additive margin shifts the angles of the classes w.r.t. to the centre of the feature space. This phase angular shift is equivalent to an enhanced distance between the centres of classes in a geodesic space. An illustration of the process of intra-class compactness and inter-class distribution enhancement based on the proposed technique is depicted in Fig. 4.

$$L_c = -\frac{1}{\mathcal{P}}\sum_{i=1}^{\mathcal{P}} \log\left(\frac{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos(\alpha_i+\varphi)}}{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\|\cos(\alpha_i+\varphi)} + \sum_{j\neq i}^{\mathcal{J}} e^{r\|\mathcal{V}(s_j)^{\mathbf{T}}\|.\|\mu_j\|\cos\alpha_j}}\right). \tag{6}$$

The loss function inherently depends upon the local similarity of the features as described in (6) based on the preliminary approach towards the local similarity consideration described in [34]. However, we have considered a global distribution-aware similarity measure to cluster the DCNN features. Moreover, we have designed an adaptive phase shift based angular margin technique that can translate well w.r.t. to our proposed similarity approach described in (3). This clustering similarity along with feature normalisation to a unit sphere helped us to achieve significant invariance to environmental conditions, scale and rotation, which is one of the most desirable properties for a complex and imbalanced dataset classifier. Considering the angle α between the centroids and query feature vector we have estimated an upper bound of the angular shift φ, where a lower bound is considered as zero. We have also considered a rule of thumb that the minimum inter-class difference should be greater than the maximum intra-class variations i.e. cluster size for each class. This intuition leads us to a simple and effective rule to set the adaptive value of φ. Considering the projection of class centroids and query feature vectors to a hyper dimensional sphere feature space depicted in Fig. 4 the class distribution can be represented on angle α in feature space. As an extreme case, we assumed that each of the classes collapses to single point, where they achieve maximum intra-class compactness. In this case, classes can be represented as single point i.e. the centroids in (2) with intra-class similarity in (3) equal to 1. Hence, the maximum inter-class margin will be $\varphi = 2\pi/\mathcal{P}$. However, it is very unlikely to achieve maximum intra-class compactness. For a complex and imbalanced dataset, each cluster occupies a proportion of the embedding hyperspace proportional to the standard deviation of the class. We have estimated an upper bound on the additive angular margin in order to choose an adaptive margin to enhance the inter-class distribution. The adaptive angular penalty for a class with centroid μ_c and standard deviation σ_c is defined as:

$$\varphi_c = |\sigma_c/h|, \tag{7}$$

where h represents the class-level distribution scaling factor. Using different values of h we can control the distribution awareness of our proposed technique for each class in an experiment where an equal scaling factor is used for all classes in a particular experiment. We have used a value of h = 1, 2, 4 and 8 corresponding to full to $1/8th$ of a class distribution considered, respectively. The embedding effectiveness achieved for different values of h is depicted in Fig. 8 where class distribution can be derived as:

$$\sigma_c = \sum_{i\in\theta_c}\left[\frac{\mathcal{V}(s_i) - \mu_c^k}{l_c}\right]^{\frac{1}{2}}. \tag{8}$$

The consideration of standard deviation σ_i also help us to achieve a class distribution aware embedding and penalisation of inter-class margins. As an effect of this penalisation each class is compacted internally, hence classification margins are improved. A graphical representation of this is presented in Fig. 5. We suppose that the centroids and query feature vectors are initially orthogonal i.e. $\alpha \leqslant 90$, which are converged as an effect of adaptive

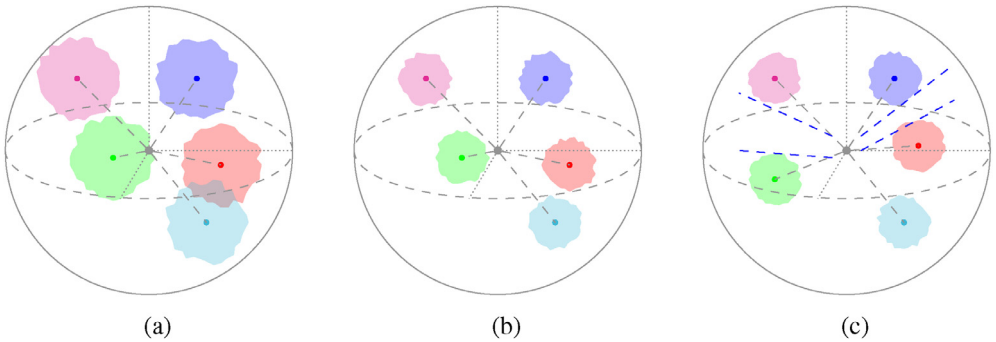

Fig. 4. An illustration of methodology: (a) pre-trained ResNet50 features, (b) vector projection based cluster embedding to enhance intra-class compactness, and (c) cluster-level classification margin adaptation to improve the inter-class distribution.

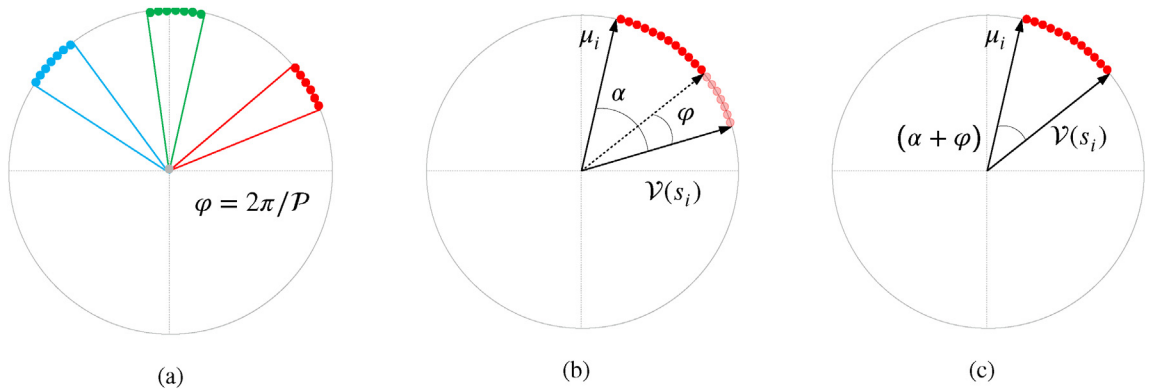

Fig. 5. A graphical depiction of adaptive angular margin: (a) a representation of maximum inter-class margin depicted for three classes, (b) an effect of additive angular margin based on the standard deviation of the individual class, and (c) intra-class compactness achieved by adaptive angular margin.

angular penalty φ_c. Moreover, this adaptive angular margin has better geometric attributes with respect to a class-level feature distribution, which is referred to as cluster distribution in our work. As stated previously highly variant features of the fruit and vegetables can result in highly distributed clusters hence, it is more desirable to deal with the embedding of each class in an adaptive manner. A detailed description of the class distribution-aware adaptive angular margins based cluster embedding is presented in Algorithm 1.

Algorithm 1. Class distribution-aware adaptive margins and cluster embedding

Result: Distribution-aware clusters
Feature space radius r, Distribution scaling factor h, Ground truth labels $c_i \in \{1, \dots, \mathcal{P}\}$;
for $j \leftarrow 1$ *to* m **do**
  $\mu_k^c = \mathcal{V}(s_i)$; ▷ Initialising cluster centroid with class-level deep features, m = Maximum number of iterations
  if m % *1000* **then**
    $\mu_k^c = \frac{1}{|\theta_k^c|} \sum_{i \in \theta_k^c} \mathcal{V}(s_i)$; ▷ Centroid updation after every 1000 iterations
  end
  $\cos(\mathcal{V}(s_i), \mu_k^c) = \frac{\langle \mathcal{V}(s_i), \mu_k^c \rangle}{\|\mathcal{V}(s_i)\| \|\mu_k^c\|}$; ▷ Deep feature and class centroid projection similarity estimation
  $\underset{\theta_1^c, \dots, \theta_{\mathcal{K}}^c}{\operatorname{argmax}} \sum_{k=1}^{\mathcal{K}} \sum_{i \in \theta_k^c} \mathcal{V}(s_i)^{\mathbf{T}} \bullet \mu_k^c$; ▷ Feature projection similarity based clustering
  $\varphi_c = \sum_{i \in \theta_c} \left[\frac{\mathcal{V}(s_i) - \mu_c^k}{l_c} \right]^{\frac{1}{2}} / h$; ▷ Class distribution based additive angular margin estimation
  $-\frac{1}{\mathcal{P}} \sum_{i=1}^{\mathcal{P}} \log \left(\frac{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\| \cos(\alpha_i + \varphi)}}{e^{r\|\mathcal{V}(s_i)^{\mathbf{T}}\|.\|\mu_i\| \cos(\alpha_i + \varphi)} + \sum_{j \neq i}^{\mathcal{J}} e^{r\|\mathcal{V}(s_j)^{\mathbf{T}}\|.\|\mu_j\| \cos \alpha_j}} \right)$; ▷ Loss function reduction based on the estimated angular margin
end

4. Experimental implementation

An extensive and detailed experimental validation has been performed to evaluate the proposed technique. We have used a dataset of 15 fruit and vegetable classes for our experiments. This dataset includes Onion (Brown onion), Carrot, Cauliflower, Cucumber (Continental cucumber), Potato (Creme potato), Cabbage (Drumhead cabbage), Granny Smith (Grannay Smith apple), Lettuce (Iceberg lettuce), Banana (Lady finger banana), Mandarin, Orange (Navel orange), Pear (Packham pear), Apple (Pink lady apple), Strawberry, and Tomato consisting of 2000 images per class. The images have a base resolution of 4384 × 3288 pixels with a fixed background where the sensor distance and lighting conditions have been considered and controlled during the image capturing process. In particular, a horizontal and vertical distance of 19.5 cm and 8.0 cm, respectively have been maintained where a uniform room ambient lighting is used for all images. Moreover, fixed background reduces the efforts required for pre-processing for example detection and segmentation. A low cost embedded system based (ArduCam MTF9001) High Definition (HD) sensor has been used for the imaging process. A useful discussion on the dataset development and the environmental conditions can also be found in our recent work [30]. Example dataset images used for the proposed technique are shown in Fig. 6. The images are resized to a resolution of 224 × 224 pixels as an input to ResNet50. Using this lower resolution can help to achieve lower time complexity and robustness to implement the proposed approach on platforms with lower computational power.

The dataset has been apportioned randomly into disjoint training, validation and testing sets by a ratio of 80%, 10% and 10%, respectively. A similar dataset distribution techniques is used for experiments with 2, 7, and 15 classes. This assumption can help us to gain better understanding of the proposed approach and a consistent experimental setup. A ResNet50 [33] pre-trained on the ImageNet dataset [14] has been used as a backbone DCNN fea-

Fig. 6. Example dataset images of Granny Smith apple, onion, pear, potato and tomato.

ture extractor. This selection has been made based on an assumption to test our proposed method using the recent state-of-the-art DL networks. The proposed technique has been implemented based on the concept presented in [44] with a learning rate between 0.1 to 0.001 which is sequentially reduced after a particular number of iterations. We have used the pre-trained weight as an initialisation for the ResNet50 and have been kept this the same for all experiments. The convergence of our technique is entirely dependent upon the proposed clustering and adaptive classification margins, hence this initialisation has negligible effect on the overall performance. To test our concept we have also transfer learned the ResNet50 on a set of 200 images for seven classes with significant similarities. Considering the significant size and number of parameters we have transfer learned only the last five layers of ResNet50 while the rest of the layers were frozen. This assumption is made based on the similar lower-level abstractions of ImageNet and our custom dataset. Significant benefits can be achieved to deal with overfitting and time complexity issues by updating the lower number of layers. An assumption has been made that the initialisation with pre-trained weights on ImageNet is less crucial for the proposed technique due to embedding and updating of centroids based on the similarity measure and adaptive margin after a particular number of iterations. The training dataset has been divided into batches, where each batch has 20 random images of each class. This technique of training data distribution in batches help us to perform the inter and intra class clustering simultaneously in a coherent and consistent manner while achieving a global distribution-aware clustering. The initial learning rate of 0.1 is used where considering the performance plateau, the learning rate was divided by 10 at 16 K and 18 K iterations. The proposed network was trained for up to 24 K iterations, where cluster centroids are updated ever 1 K iterations. A weight decay and momentum value was considered as 0.9 and $5e^{-4}$, respectively. A disjoint validation dataset (10% of dataset) is used to avoid overfitting where an L_2 weight regularisation is also performed with a fixed weight decay rate. L_2 regularisation has an inherent capability to deal with large number of features as a continuous function for complexity management by reducing the weight proportional to the training iterations. A significant large weight decay rate is considered to penalise the CNN for uniform weight distribution across the features, hence avoiding feature estimation loss. The reduced CNN weights proportional to training iterations can achieve a better overfitting goal for larger numbers of iterations. Moreover, the additive angular margin based loss function defined in (4), (5) and (6) also applies strict penalties to coverage as compared to the state-of-the-art CNN loss function e.g. Softmax loss. This property is achieved by forcing the features into a concise feature space, hence reducing the overlapping. These strict convergence penalties also enhance the capability to deal with overfitting. A combination of discrete leaning rate decay [25] and Adam optimisation [16] is used to optimise the CNN training process. A small learning rate for higher number of iterations avoid the loss function bouncing at false global minimums and can result in a faster convergence. Adam optimisation as a combination of RMSprop and Stochastic Gradient Descent (SGD) with a momentum offers significant benefits in our proposed technique. A parameter-level moment optimisation is obtained with the inherent moving gradient average of Adam optimisation. A diagonal gradient rescaling invariance is also obtained with less computational and memory requirements. Considering the discrete learning rate decay we have used an exponential decay rate of first (β_1) and second (β_2) momentum as 0.9 and 0.999, respectively where a division normalisation (ϵ) of $1e^{-8}$ is used to initialise the Adam optimisation. Only the feature embedding part of ResNet50 (162 MB) was used for testing and training process i.e. the proposed Softmax layers were used along with the clustering based embedding. For testing purposes we extracted features from the 1st fully connected layer of the ResNet50. The extracted features are then used to determine the similarity measure w.r.t to the cluster centroids. The use of the fully connected layer based feature can help us to achieve a scalability w.r.t. to number of classes in future. The feature scaling i.e. the radius of the embedded feature space r is set to 32. A Python (3.7) and Tensorflow (2.1.0) based implementation has been performed where data analysis and visualisation is supported by Sklearn2. Real-time computer vision and image processing support is obtained through the Opencv-python (4.1.1.26) library. The network is trained and tested on a 12 GB Tesla K80 (4992 cores) with 32 GB installed memory.

Centroid Updating. We have considered $\mathcal{P}$ = 2, 7, and 15 clusters representing classes of fruit and vegetables in our experiments. The centroids for these clusters have been initialised on the features of the first training batch as described in (2). However, it is worth noting that the feature representation is updated in a gradual manner during the training process. We update the centroids of the clusters to achieve a true distribution of features in an embedding space after a particular number if iterations. A running index of the clustering process is maintained and is updated to maintain the true distribution of features i.e. the centroids of classes as defined in (2) is re-evaluated after a particular number of iterations. It is considered that the clustering process has negligible computational cost as compared to the feature extraction and neural network based classification margin estimation.

Visualisation. We have used seven classes of fruit and vegetables to represent the concept of intra-class variations and inter-class similarities. This consideration has helped us to emphasise on the classes with significantly similar physical features. Two different kinds of visualisations have been used to represent our concept and compare the experimental results. The ResNet50 learned features has been represented in Fig. 2. The learned features are normalised to an imaginary three dimensional space to represent the feature overlapping and distribution in an arbitrary feature space based on the technique in [45]. A set of 200 samples have been used for this representation where a perplexity of 7 has been used with a learning rate of 10.0 and minimum gradient norm is set to $1e^{-7}$. An experimental observation has been performed to select the perplexity, where a large learning rate ranging from 10.0 to 1000.0 is recommended for t-SNE based visualisation. This large value of learning rate is crucial to avoid the local minimum however, using significantly large learning rates for small samples can force the features into a uniform equidistant distribution,

which is less suitable to understand the feature overlapping. An exaggeration technique defined in [40] has been used to control the embedded inter-cluster distribution. An exaggeration value of 12.0 has been used for our implementation. An enhanced features dependence estimation and cluster assignment is also reported for this technique [40]. A Euclidean distance based features similarity is used with the Barnes Hut [5] optimisation technique for feature embedding where an effective time complexity of $\mathcal{O}(n \log n)$ can be achieved. A maximum number of 1000 iterations has been initialised, however, the loss function is observed for a maximum number of 300 iterations in case of performance plateau based on minimum gradient norm. The cluster-level embedding with the adaptive classification margins has been used for classification of fruit and vegetables. We have used a Silhouette score based visualisation to analyse the clustering and classification of seven considered classes of fruit and vegetables. A complete inter and intra cluster feature-level analysis of our test dataset has been performed, where the higher values of the Silhouette score represent an effective inter and intra class clustering, hence classification. The Silhouette score uses Euclidean distance to estimates the efficacy of our angular margin based clustering and the effect of angular margin in an imaginary geodesic feature space. The proposed technique is used to enforce compactness and inter class margins simultaneously where the Silhouette score as a metric helps to analyse the classification margins of a complex dataset.

5. Results

The proposed cluster-level embedding approach with adaptive margins is evaluated not only for clustering effectiveness but also for classification of a fruit and vegetables dataset. We have tested and evaluated our proposed concept in a progressive manner i.e. with different numbers of classes. Initially, we considered two pairs of Cabbage and Lettuce, and Mandarin and Orange to illustrate our concept for two classes with significantly similar physical features. Our main tests are performed for seven classes of fruit and vegetables with significant inter-class similarities where results have been evaluated for cluster embedding and classification. These classes have been considered carefully to meet the requirements of inter-class similarity and intra-class distribution. The selection has also helped us to achieve a dataset with significant complexity and imbalance. Finally, we have tested the technique for the dataset of 15 classes described in Section 4. We have used the Silhouette score analysis to analyse the clustering achieved, where we have used a combination of metrics for classification effectiveness. The complexity of the dataset is taken into account for classification where we have used Receiver Operating Characteristic (ROC) curves to analyse the classification effectiveness with different values of h.

Cluster Embedding. A Silhouette score analysis has been presented for different numbers of classes where significant support for our proposed technique can be observed. The Silhouette score is analysed for DCNN based features and the proposed technique with different number of classes and adaptive angular margins proportional to the class distribution. An initial analysis performed for two pairs of Cabbage and Lettuce (Pair 1), and Mandarin and Orange (Pair 2) is presented in Fig. 7. These pairs have been considered due to significant inter-class similarities where likelihood of inter-class overlapping is high as depicted in Fig. 2. The pre-trained ResNet50 based features of each class pair is extracted and are used for adaptive margin based clustering where $h = 1$ i.e. considering the complete class distribution. Significant intuitions can be established based on analysis of these two class pairs, where it can be observed that the state-of-the-art DCNNs deals with the highly similar classes in a consistent manner. The sample Euclidean distance between the class centroids is small with an average Silhouette score of 0.32 and 0.30 obtained for Pair 1 and Pair 2, respectively. A marginal difference of 0.02 Silhouette scores can be explained as a result of the limited capability of the sate-of-the-art DCNNs to deal with significantly similar classes. A significantly improved inter-class distribution and classification result has been obtained with the proposed techniques in the case of the considered two class pairs. An average Silhouette score for Pair 1 is 0.89, while Pair 2 has achieved a 0.92 score, which can validate the effectiveness of the proposed techniques to deal with different similar classes robustly.

We have further extended our analysis for a seven class dataset, where a significant overlapping among the classes can be observed in Fig. 8(a) for the ResNet50 features as previously illustrated in Fig. 2. An average Silhouette score of 0.37 is obtained where it can be observed that a significant number of samples has been wrongly clustered, hence resulting in wrong classifications. The Silhouette score estimates the distance of samples w.r.t. to each class centroid, where higher value indicates the samples are at significant distance from the neighbouring clusters. We have used Euclidean distance based metrics for estimation of the Silhouette score to relate the angular margin and geodesic distance for our proposed technique. An accumulation of samples can be observed below the average Silhouette score in Fig. 8(a), which can be translated to an overlapping set of features with thin classification margins among the classes. A wrong classification can be observed by proportion of samples with a negative Silhouette coefficient. For our analysis we have used the same size of cluster, i.e. equal numbers of samples per class, however, inconsistent width of the clusters indicates a merging of samples from multiple classes. We have also estimated the Silhouette score for different values of adaptive angular margins, where class distribution is considered as standard deviation in (7). We have used h = 1, 1, 2, 4, and 8 for Fig. 8(b)–(f), respectively. It can be observed that the best results, with a Silhouette score of 0.86, has been obtained for a $h = 1$ on the transfer learned ResNet50 based features in Fig. 8(b), where, a consistent cluster-level distribution can be observed. As we had assumed, the features obtained by pre-trained or transfer learned ResNet50 have a negligible effect on the overall convergence of the clusters and classification accuracy. A comparison of the results presented in Fig. 8(b) and (c) strengthen our concept, where a clustering is performed for the same number of clusters and adaptive angular margin for the pre-trained and transfer learned ResNet50. A Silhouette score difference of 0.04 has been observed, where clustering on the pre-trained ResNet50 features without transfer learning obtain a score of 0.82. Moreover, no samples have been reported for wrong classification with a negative Silhouette score. An effect of global class distribution on the adaptive angular margin has been depicted in Fig. 8(d)–(f). The class structure considered as standard deviation has a significant effect on overall performance, where reducing the adaptive angular margin by a factor of 2 causes a clustering loss of on average, a 0.12 Silhouette score. An average Silhouette score of 0.74, 0.61, and 0.46 has been achieved for $h = 2, 4$, and 8, respectively on pre-trained ResNet50 features.

The analysis is further extended to a dataset of 15 classes of fruit and vegetables as our base-line application of classification at a supermarket self-checkout. A Silhouette score analysis is presented in Fig. 9 for the 15 classes. Considering our results presented in Fig. 8, we have used a transfer learned ResNet50 as a backbone for feature extraction with $h = 1$. The average Silhouette score of 0.34 and 0.87 has been obtained for ResNet50 features and the proposed approach, respectively. Based on the obtained results the proposed approach can be considered scalable and effective for larger number of classes where significant classification accuracy

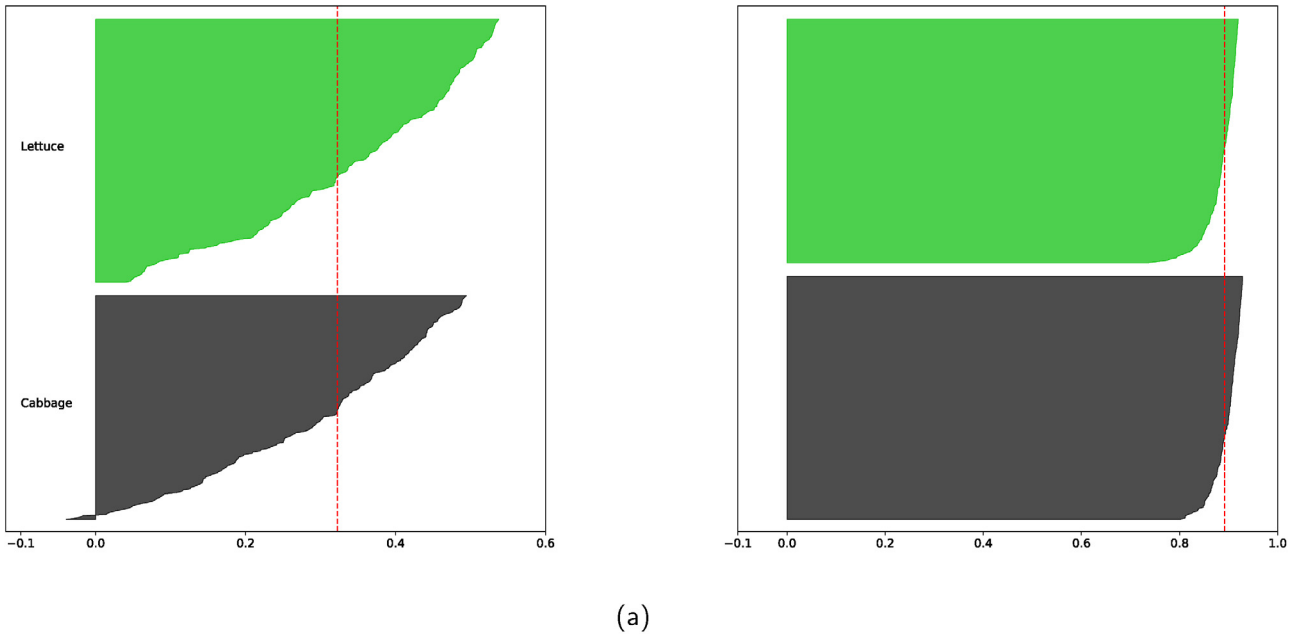

Fig. 7. A two class pair analysis of the proposed approach: (a) Cabbage and Lettuce (Pair 1), and (b) Mandarin and Orange (Pair 2).

can also be achieved, as presented in Fig. 11. A clustering comparison of K-means and K-means^{++} has been performed for pre-trained ResNet50 features of fruit and vegetables. An average Silhouette score of 0.544 and 0.606 has been achieved for K-means and K-means^{++}, respectively where results are shown in Fig. 9(c) and (d). A fixed number of clusters i.e. 15 has been set for this implementation where centroids are initialised randomly for K-means clustering. A uniform centroid distribution technique based on [4] is used for K-means^{++} centroid initialisation. A final clustering is achieved using 10 different seeds for random centroids initialisation for both techniques using a maximum number of 300 iterations. A Euclidean norm based relative tolerance of 1 e^{-4} is used for centroids differences to estimate convergence on two progressive iterations. A set of 20 training dataset images per class is used to estimate Euclidean norm based class labels. Significant results have been obtained where compact clustering margins and wrong cluster assignment can be observed. Moreover, K-means and K-means^{++} also have significant time complexity and are prone to larger and more variable dataset distributions. The proposed technique has used centroid updating to deal with the complex varying data distributions where significant outlier consideration can also be achieved by updating centroids after a particular number of iterations. It can be noted that the additive angular margin and vector projection based features similarity has significantly improved the clustering and classification for the proposed technique.

An experiment has been performed to validate the proposed technique for an imbalanced dataset of fruit and vegetables with similar physical features. A dataset of Cabbage and Lettuce consisting of 200 (20%) and 800 (80%) images, respectively is used for this

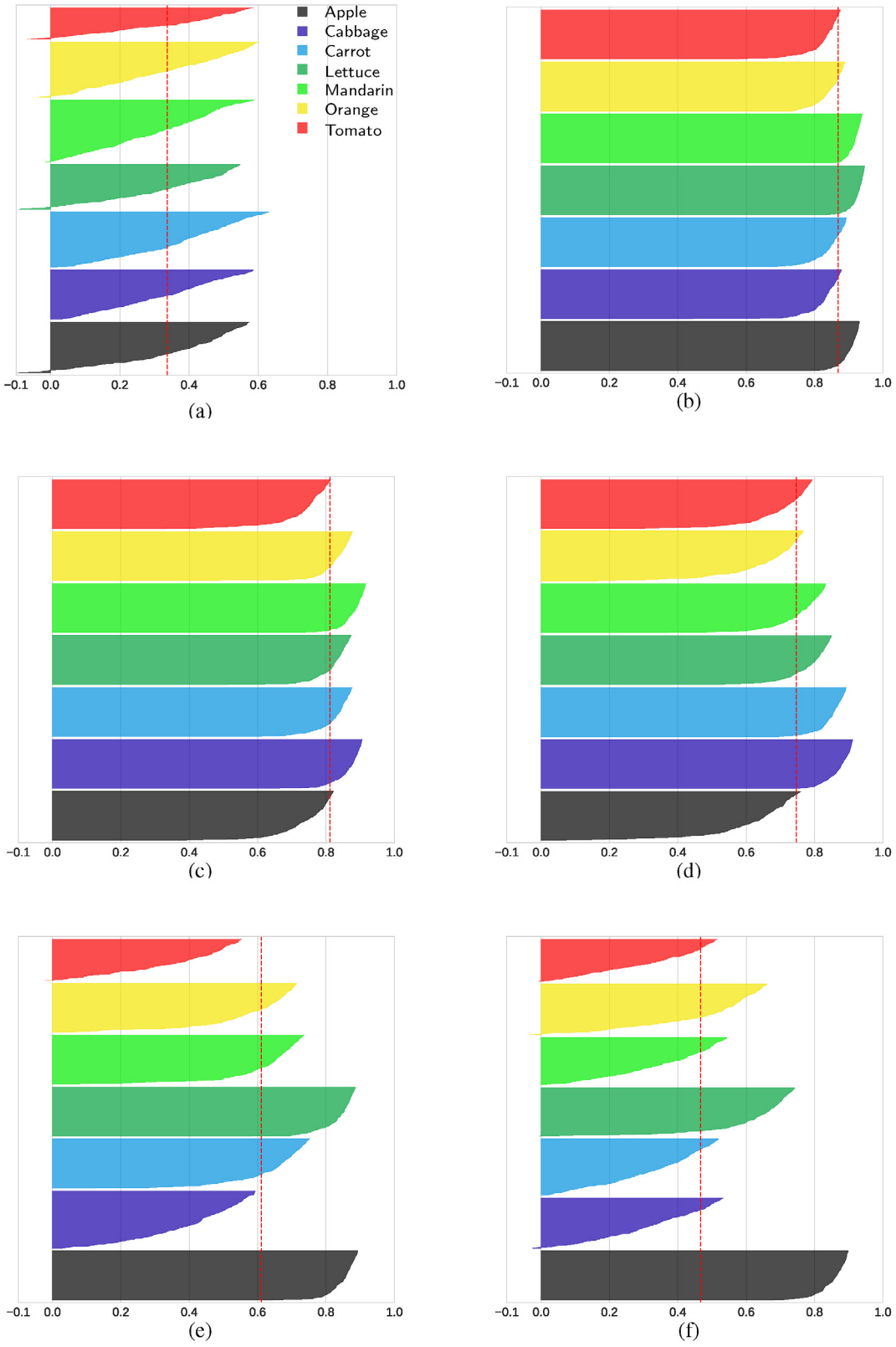

Fig. 8. Silhouette score analysis: (a) Learned pre-trained ResNet50 features, (b) $h = 1$ (transfer learned ResNet50), (c) $h = 1$, (d) $h = 2$, (e) $h = 4$, and (f) $h = 8$.

experiment. Considering the physical features similarity the classes have been selected carefully for this experiment. The proposed technique has been tested for h = 1, 2, 4, 8 where Silhouette score analysis is presented in Fig. 12. The proposed approach has been tested for both pre-trained and transfer learned ResNet50, where a significant clustering effectiveness has been achieved. The result obtained can be considered consistent for balanced and imbalanced data distributions. An average Silhouette score of 0.776 and 0.626 has been achieved for transfer learned and pre-trained ResNet50 features, respectively with $h = 1$. An effect of

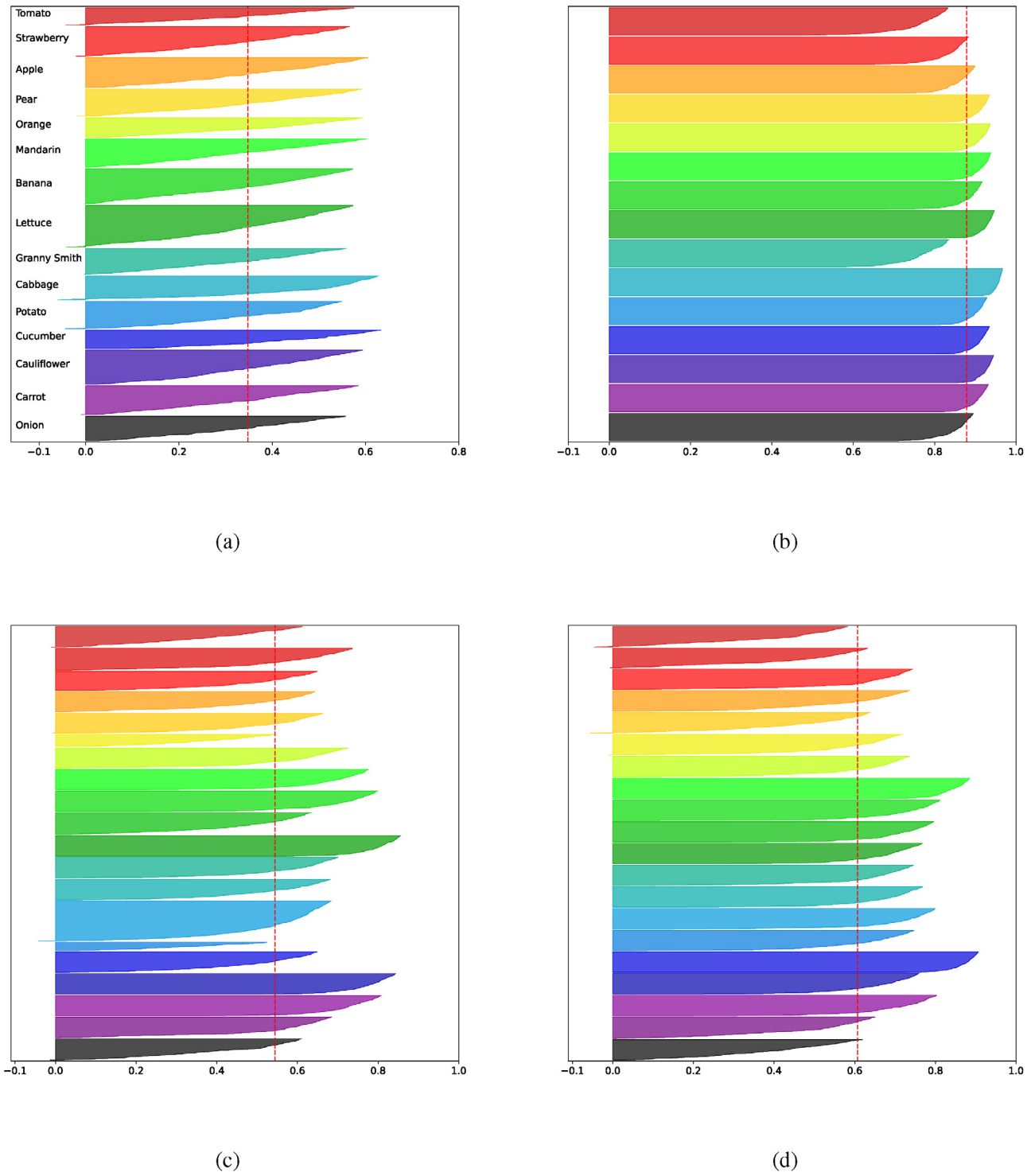

Fig. 9. Implementation of distribution-aware clustering and adaptive margins on 15 classes of fruit and vegetables: (a) ResNet50 features, (b) transfer learned ResNet50 ($h = 1$), (c) K-means clustering (ResNet50 features), and (d) K-means^{++} (ResNet50 features).

class distribution on clustering effectiveness is noted where considering the smaller proportion of class distribution results in compact and overlapping inter-class clusters. An average Silhouette score of 0.595, 0.459 and 0.369 has been obtained for h = 2, 4 and 8, respectively where wrong cluster assignment can also be noted in Fig. 12. Moreover, the clustering effectiveness of the proposed approach is supposed prone to the significant imbalance of the dataset distribution where a difference of 0.114 has been noted in the average Silhouette score (Figs. 7 and 12). This difference can be translated as reduced inter-class clusters distribution margins which can lead to the cluster overlapping, hence wrong classification. However, no wrong clustering assignment has been noted while considering complete class distribution i.e. $h = 1$ for both pre-tranined and transfer learned ResNet50 features. As the proposed approach is significantly dependent upon the deep features extractor (ResNet50 in our case) the reduced inter-class distribution can be related to the capability of ResNet50 to deal with the imbalanced dataset. Considering the complexity of the fruit and

vegetables dataset, it is supposed that a more sophisticated feature similarity measure for feature embedding can significantly improve the technique for imbalanced class distribution. A dynamic class distribution measure consideration w.r.t. the training iterations can also help to improve the proposed approach for imbalanced data distributions.

Classification. The proposed technique is analysed for classification of fruit and vegetables. The classification effectiveness is analysed for our main experiments with seven classes which is further extended to our 15 class dataset of fruit and vegetables. We have tested for different values of h to compare the classification effectiveness of the proposed technique considering different levels of class feature distribution. A ROC curve based comparison of the proposed technique is presented in Fig. 10, where a zoomed plot along with Area Under the Curve (AUC) is presented in Fig. 10(b). The classification results obtained are significantly consistent to the clustering results presented Fig. 8, where the proposed technique used with the transfer learned ResNet50 based features has outperformed. To analyse the statistical significance of the classification results achieved with pre-trained and transfer learned deep features, a p-vale hypothesis testing [3] has been performed. Considering the small difference of AUC for pre-trained and transfer learned deep features a null hypothesis (H_0) of consistent classification results has been considered where a significance level (α) of 0.10 and 0.05 i.e. 10% and 5%, respectively is used. A two-tailed p-value of 0.053 has been obtained where a strong statistical significance can be noted for $\alpha = 0.10$. Moreover, a marginal significance is evident for the more rigorous significance level condition i.e. $\alpha = 0.05$. Considering the p-value observations it can be concluded that the transfer learned deep feature based class distribution-aware features embedding has outperformed for fruit and vegetables classification. Moreover, an improved classification effectiveness can be achieved by transfer learning on larger numbers of images for deep features. Reducing the adaptive angular margin i.e. $h = 2, 4$, and 8, reduces the classification margins in the embedding feature space, hence causing discrepancy among classes. These reduced classification margins caused the overlapping of classes with significantly similar features, resulting in a reduced classification accuracy. A detailed comparison of the conventional classification metrics, i.e. Accuracy (ACC), Error Rate (ER), Positive Predictive Value (PPV), True Negative Rate (TNR), and F1 score (F1) is presented in Table 2. This analysis is presented for both pre-trained and transfer learned ResNet50 features with $h = 1$ i.e. the two leading variants of classifier according to ROC and AUC analysis. The proposed approach used with the transfer learned ResNet50 has outperformed, however the results with the pre-trained ResNet50 are also comparable. These comparable results strengthen the idea that the convergence of the proposed technique is independent of deep feature extractor and initialisation of the weights. Hence, the proposed distribution aware cluster embedding allows us to use the proposed approach with any readily available state-of-the-art DL networks. The usage of these DL networks also extends the capability of the proposed approach to even larger numbers of classes. Considering these findings, we have applied the proposed technique for classification of 15 classes of fruit and vegetables with significant inter-class similarities and intra-class distribution. A confusion matrix based comparison of classification effectiveness for a transfer learned ResNet50 and

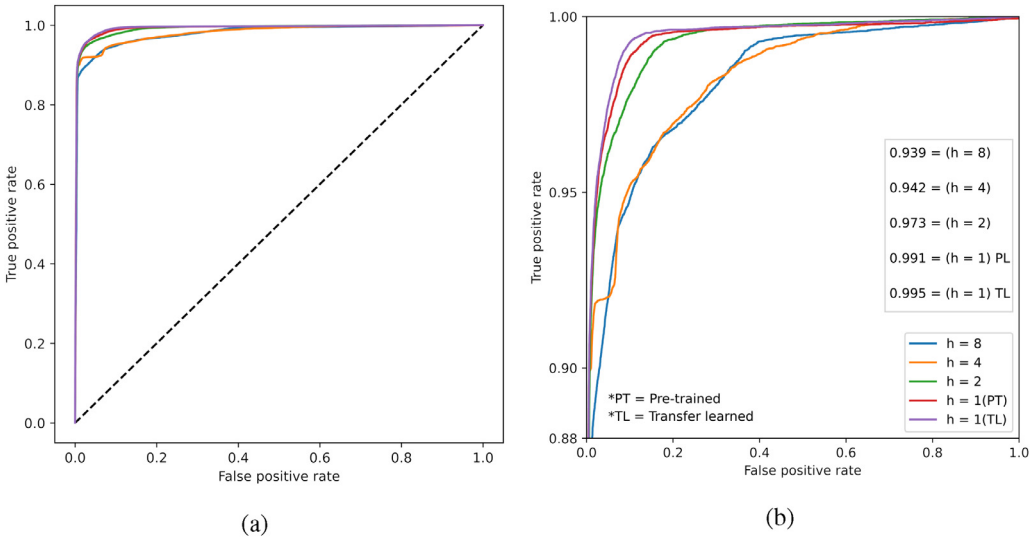

Fig. 10. ROC curves and AUC of different adaptive angular margins: (a) ROC curves of different adaptive angular margins, and (b) a zoomed in ROC curve with AUC for better understanding.

Table 2
Conventional classification metric comparison for pre-trained and transfer learned ResNet50 $h = 1$.

	Pre-trained ResNet50					Transfer learned ResNet50				
Fruit/Veg	ACC(%)	ER(%)	PPV(%)	TNR(%)	F1	ACC(%)	ER(%)	PPV(%)	TNR(%)	F1
Apple	98.00	2.00	93.88	99.00	0.929	99.14	0.86	100.00	100.00	0.969
Cabbage	97.43	2.57	90.20	98.33	0.911	99.14	0.86	97.96	99.67	0.970
Carrot	98.57	1.43	95.92	99.33	0.949	99.71	0.29	98.04	99.67	0.990
Lettuce	97.43	2.57	91.84	98.67	0.909	99.14	0.86	96.08	99.33	0.970
Mandarin	98.00	2.00	92.16	98.67	0.931	98.57	1.43	95.92	99.33	0.949
Orange	97.71	2.29	92.00	98.67	0.920	98.29	1.71	92.31	98.67	0.941
Tomato	98.00	2.00	92.16	98.67	0.931	98.57	1.43	94.12	99.00	0.950

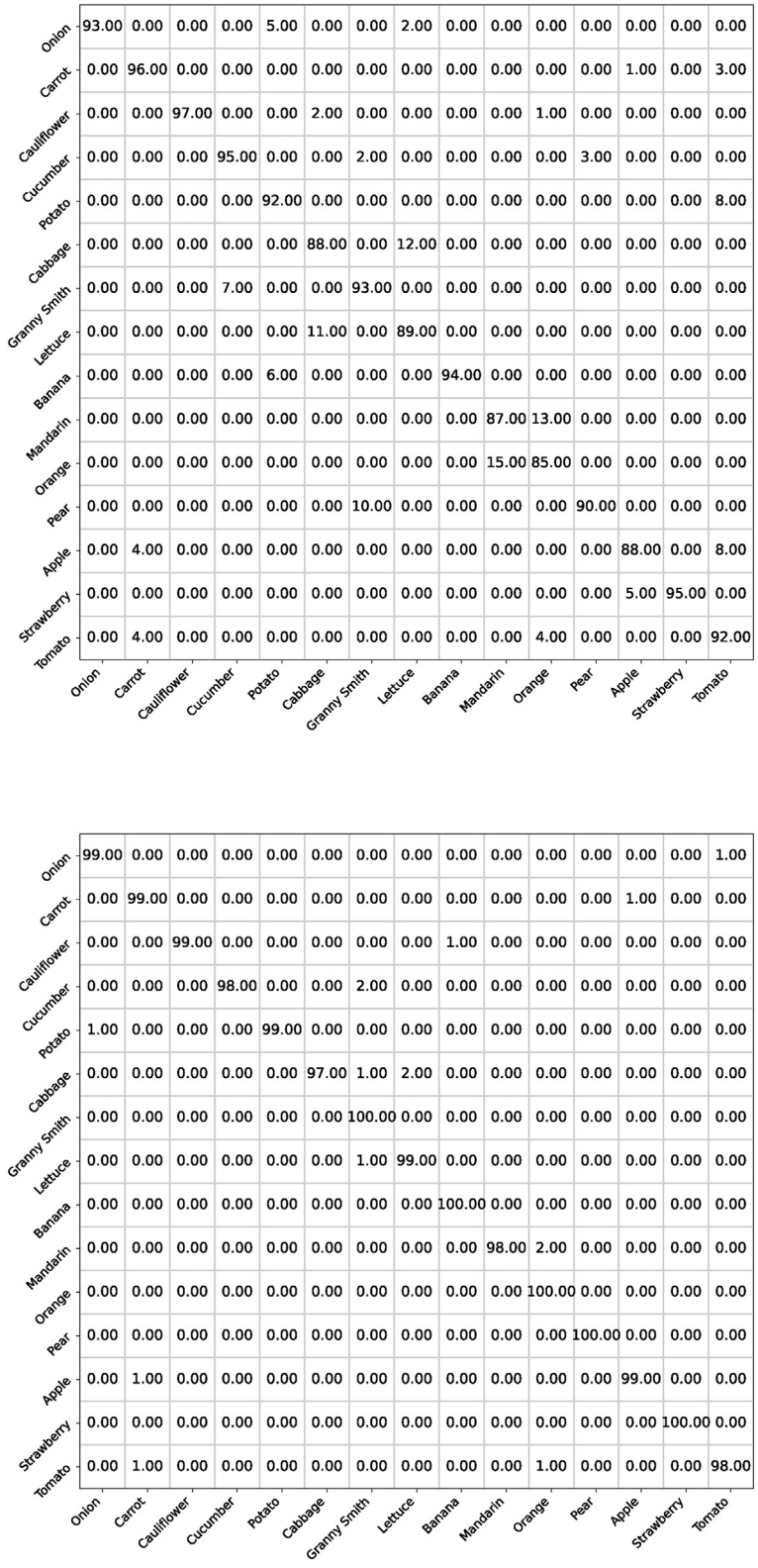

Fig. 11. A classification effectiveness comparison: [Upper] transfer learned ResNet50, and [Lower] adaptive classification margins with cluster embedding on transfer learned ResNet50 ($h = 1$).

adaptive classification margins is presented in Fig. 11. It can be observed that the proposed approach can achieve significantly classification effectiveness for classes with significant similar physical features.

Inference Analysis. An inference analysis is also performed to analyse the effectiveness of the proposed approach for a real-world supermarket environment. To achieve accurate inference times, we have used 20 images per class chosen randomly from

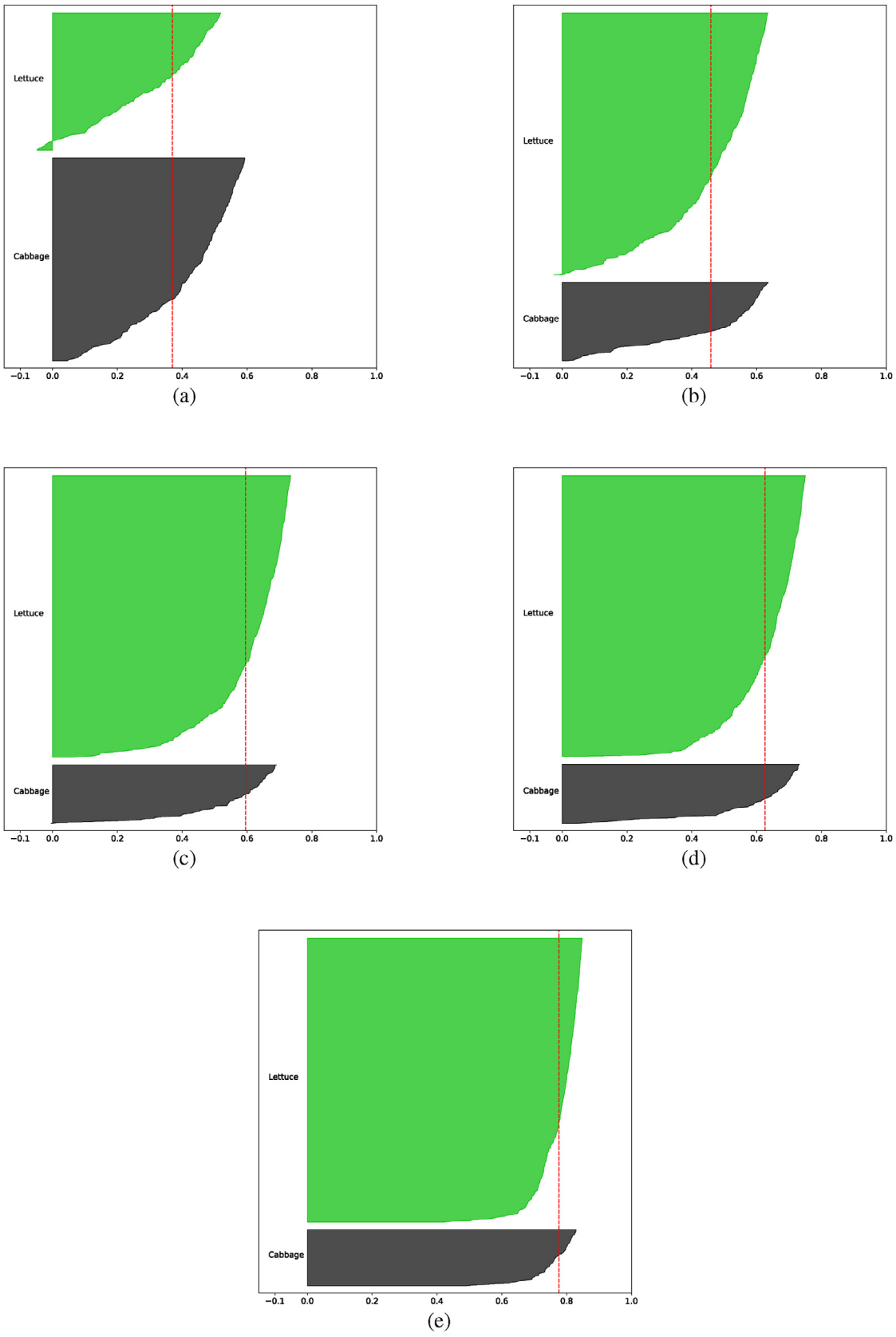

Fig. 12. Imbalanced dataset Silhouette score analysis: (a) $h = 8$, (b) $h = 4$, (c) $h = 2$, (d) $h = 1$ (pre-trained ResNet50), and (e) $h = 1$ (transfer learned ResNet50).

Table 3
Inference analysis of proposed approach for different values of adaptive angular margin.

Adaptive margin	Feature extraction (ms)	Feature embedding (ms)	Classification (ms)
$h = 1$ (TL*)	421.13	127.80	09.15
$h = 1$ (PT+)	415.23	121.50	09.12
$h = 2$	415.23	115.61	10.20
$h = 4$	415.23	98.60	13.26
$h = 8$	415.23	96.78	11.10

+ Pre-trained.
* Transfer learned.

the test dataset, where average time for each image classification is obtained by dividing the obtained value by 20. The inference times obtained from different values of adaptive angular margin for ResNet50 feature extraction, feature embedding (clustering) and classification (Softmax) is presented in Table 3. This inference analysis is performed on a 12 GB Tesla K80 (4992 cores) with 32 GB installed memory, however this inference time is prone to variation based on the execution environment and the underlying machine. Considering the obtained result, an important intuition can be established that feature embedding with large values of h penalises larger classification margin which makes the overall convergence harder and increases time complexity. Moreover, it can be observed that the proposed class distribution aware clustering has negligible time complexity as compared to the feature extraction part, and hence can be considered suitable for a real-world implementation.

6. Conclusion

The proposed class distribution aware adaptive classification margins approach with cluster-based embedding has been tested for cluster embedding and classification of seven and fifteen fruit and vegetables classes with significantly similar physical features. This intra-class variations and inter-class similarities limits the state-of-the-art DCNNs ability to estimate complex hyper-planes for classification. The intuition of the proposed approach is to embed the features from the ResNet50 to an imaginary feature space to enhance the intra-class compactness and inter-class margins. Extensive and detailed experiments have been performed for different adaptive classification margins. A vector projection based similarity is estimated between the class centroid and features vectors obtained by ResNet50 to achieve intra-class compactness. The adaptive angular margins are then used to enhance the classification margins between classes for fruit and vegetables. An imbalanced dataset distribution based clustering and classification effectiveness is also tested for the proposed approach. Significant positive results have been achieved for clustering and classification, where the proposed approach is able to achieve invariance w.r.t to the challenges described. The proposed approach adds a relatively negligible time complexity and can be considered suitable for real-world implementations. Considering the scalable property of the proposed approach, this technique can be also used for complex datasets with higher number of classes. Based on the experimental results it is concluded that a more sophisticated similarity measure for features embedding can be explored for further enhancement of the proposed approach specifically to deal with the imbalanced dataset distributions. Moreover, the capability of the deep features extractor (ResNet50) has a significant impact on the overall process. An enhancement of the features extractor to deal with complex and imbalanced datasets with large numbers of classes can also improve the clustering and classification effectiveness. The class distribution measure can be dynamically updated with the training iterations for complex datasets e.g. fruit and vegetables classification. Moreover, considering better statistical class distribution measures for example distribution dispersion and interquartile range can also improve the proposed approach.

CRediT authorship contribution statement

Khurram Hameed: Data curation, Methodology, Conceptualization, Formal analysis, Writing - original draft, Investigation, Software. **Douglas Chai:** Supervision, Validation, Writing - review & editing, Project administration. **Alexander Rassau:** Supervision, Validation, Writing - review & editing.

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.

References

[1] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Improved inception-residual convolutional neural network for object recognition, Neural Computing and Applications 32 (2020) 279–293, https://doi.org/10.1007/s00521-018-3627-6.
[2] M.B. Alvi, K. Hameed, M. Alvi, W. Javed, M. Afzal, Algorithmic state machine and data based modeling of superscalar processor of order 2, in: International Conference on Software Technology and Engineering (ICSTE), 2011, pp. 1–5, https://doi.org/10.1115/1.859797.paper73.
[3] J. Arbuthnot, An argument for divine providence, taken from the constant regularity observ'd in the births of both sexes, Philosophical Transactions of the Royal Society of London 27 (1710.) 325–336, https://doi.org/10.1098/rstl.1710.0011.
[4] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035. URL: https://dl.acm.org/doi/10.5555/1283383.1283494..
[5] J. Barnes, P. Hut, A hierarchical o (n log n) force-calculation algorithm, Nature 324 (1986) 446–449. URL: https://doi.org/10.1038/324446a0..
[6] G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, A study of the behavior of several methods for balancing machine learning training data, SIGKDD Explorations Newsletter 6 (2004) 20–29, https://doi.org/10.1145/1007730.1007735.
[7] R. Bobbit, J. Connell, N. Haas, C. Otto, S. Pankanti, J. Payne, Visual item verification for fraud prevention in retail self-checkout, in: IEEE Workshop on Applications of Computer Vision (WACV), 2011, pp. 585–590. URL: https://doi.org/10.1109/WACV.2011.5711557..
[8] R.M. Bolle, J.H. Connell, N. Haas, R. Mohan, G. Taubin, Veggievision: A produce recognition system, in: IEEE Workshop on Applications of Computer Vision (WACV), 1996, pp. 244–251. URL: https://doi.org/10.1109/ACV.1996.572062..
[9] H. Caesar, J. Uijlings, V. Ferrari, Joint calibration for semantic segmentation, in: Proceedings of the British Machine Vision Conference (BMVC), 2015, pp. 29.1–29.13. URL: https://doi.org/10.5244/C.29.29..
[10] C.L. Castro, A.P. Braga, Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data, IEEE Transactions on Neural Networks and Learning Systems 24 (2013) 888–899. URL: https://doi.org/10.1109/TNNLS.2013.2246188..
[11] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of Artificial Intelligence Research 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[12] H.C. Chi, M.A. Sarwar, Y.A. Daraghmi, K.W. Lin, T.U. Ik, Y.L. Li, Smart self-checkout carts based on deep learning for shopping activity recognition, in: Asia-Pacific Network Operations and Management Symposium (APNOMS), 2020, pp. 185–190, URL: https://doi.org/10.23919/APNOMS50412.2020.9237053..
[13] M. Cordea, B. Ionescu, C. Gadea, D. Ionescu, Dynface: A multi-label, dynamic-margin-softmax face recognition model, in: K. Arai, S. Kapoor (Eds.), Advances in Computer Vision, 2020, pp. 535–550. url:https://doi.org/10.1007/978-3-030-17795-9/_39..
[14] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, Li Fei-Fei, Imagenet: A large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.
[15] J. Deng, Y. Zhou, S. Zafeiriou, Marginal loss for deep face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 2006–2014, URL: https://doi.org/10.1109/CVPRW.2017.251..
[16] P. Diederik, J.B. Kingma, Adam: A method for stochastic optimization, in: International Conference on Learning Representations (ICLR), 2015. URL: http://arxiv.org/abs/1406.3269..

[17] Q. Dong, S. Gong, X. Zhu, Class rectification hard mining for imbalanced deep learning, in: IEEE International Conference on Computer Vision ICCV, 2017, pp. 1851–1860, https://doi.org/10.1109/ICCV.2017.205.
[18] C. Drummond, R.C. Holte, et al., C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling, in: Workshop on learning from imbalanced datasets II, 2003, pp. 1–8. URL: https://doi.org/10.1.1.68.6858..
[19] Y.C. Du, M. Muslikhin, T.H. Hsieh, M.S. Wang, Stereo vision-based object recognition and manipulation by regions with convolutional neural network, Electronics 9 (2020) 210, https://doi.org/10.3390/electronics9020210.
[20] S.R. Dubey, A.S. Jalal, Species and variety detection of fruits and vegetables from images, International Journal of Applied Pattern Recognition 1 (2013) 108–126, https://doi.org/10.1504/IJAPR.2013.052343.
[21] D. Eigen, R. Fergus, Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2650–2658, https://doi.org/10.1109/ICCV.2015.304.
[22] D. Erhan, A. Courville, Y. Bengio, P. Vincent, Why does unsupervised pre-training help deep learning?, in: International Conference on Artificial Intelligence and Statistics, 2010, pp 201–208, https://doi.org/10.5555/1756006.1756025.
[23] F. Femling, A. Olsson, F. Alonso-Fernandez, Fruit and vegetable identification using machine learning for retail applications, in: International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2018, pp. 9–15, https://doi.org/10.1109/SITIS.2018.00013.
[24] K. Fuchs, T. Grundmann, E. Fleisch, Towards identification of packaged products via computer vision: Convolutional neural networks for object detection and image classification in retail environments, in: International Conference on the Internet of Things, 2019, https://doi.org/10.1145/3365871.3365899.
[25] R. Ge, S.M. Kakade, R. Kidambi, P. Netrapalli, The step decay schedule: A near optimal, geometrically decaying learning rate procedure for least squares, in: Advances in Neural Information Processing Systems, 2019, pp. 1–12. URL: https://proceedings.neurips.cc/paper/2019/file/2f4059ce1227f021edc5d9c6f0f17dc1-Paper.pdf..
[26] I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, Maxout networks, in: International Conference on Machine Learning, 2013, pp. 1319–1327. url:https://doi.org/10.5555/3042817.3043084..
[27] R. Hadsell, S. Chopra, Y. LeCun, Dimensionality reduction by learning an invariant mapping, in: IEEE Conference on Computer Vision and Pattern Recognition CVPR, 2006, pp. 1735–1742, https://doi.org/10.1109/CVPR.2006.100.
[28] K. Hameed, D. Chai, A. Rassau, A comprehensive review of fruit and vegetable classification techniques, Image and Vision Computing 80 (2018) 24–44, https://doi.org/10.1016/j.imavis.2018.09.016.
[29] K. Hameed, D. Chai, A. Rassau, A progressive weighted average weight optimisation ensemble technique for fruit and vegetable classification, in: International Conference on Control, Automation, Robotics and Vision (ICARCV), 2020, pp. 1–6. URL: https://doi.org/10.1109/ICARCV50220.2020.9305474..
[30] K. Hameed, D. Chai, A. Rassau, A sample weight and adaboost cnn-based coarse to fine classification of fruit and vegetables at a supermarket self-checkout, Applied Sciences 10 (2020) 8667, https://doi.org/10.3390/app10238667.
[31] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Transactions on Knowledge and Data Engineering 21 (2009) 1263–1284. URL: https://doi.org/10.1109/TKDE.2008.239..
[32] H. He, Y. Ma, Imbalanced Learning: Foundations, Algorithms, and Applications, first ed., Wiley-IEEE Press, 2013. URL: https://doi.org/10.5555/2559492..
[33] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[34] C. Huang, C.C. Loy, X. Tang, Local similarity-aware deep feature embedding, in: International Conference on Neural Information Processing Systems, 2016, pp. 1270–1278. URL: https://dl.acm.org/doi/10.5555/3157096.3157238..
[35] P. Jeatrakul, K.W. Wong, C.C. Fung, Classification of imbalanced data by combining the complementary neural network and smote algorithm, in: International Conference on Neural Information Processing, 2010, pp. 152–159, https://doi.org/10.1007/978-3-642-17534-3-19.
[36] S.H. Khan, M. Hayat, M. Bennamoun, F.A. Sohel, R. Togneri, Cost-sensitive learning of deep feature representations from imbalanced data, IEEE Transactions on Neural Networks and Learning Systems 29 (2018) 3573–3587, https://doi.org/10.1109/TNNLS.2017.2732482.
[37] T.M. Khoshgoftaar, J. Van Hulse, A. Napolitano, Supervised neural network modeling: An empirical investigation into learning from imbalanced data with labeling errors, IEEE Transactions on Neural Networks 21 (2010) 813–830, https://doi.org/10.1109/TNN.2010.2042730.
[38] B. Krawczyk, M. Woźniak, G. Schaefer, Cost-sensitive decision tree ensembles for effective imbalanced classification, Applied Soft Computing 14 (2014) 554–562, https://doi.org/10.1016/j.asoc.2013.08.014.
[39] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Communications of the ACM 60 (2017) 84–90, https://doi.org/10.1145/3065386.
[40] G.C. Linderman, S. Steinerberger, Clustering with t-sne, provably, SIAM Journal on Mathematics of Data Science 1 (2019) 313–332, https://doi.org/10.1137/18m1216134.
[41] S. Liu, W. Deng, Very deep convolutional neural network based image classification using small training sample size, in: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730–734, https://doi.org/10.1109/ACPR.2015.7486599.
[42] W. Liu, R. Lin, Z. Liu, L. Liu, Z. Yu, B. Dai, L. Song, Learning towards minimum hyperspherical energy, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018, pp. 6225–6236, https://doi.org/10.5555/3327345.3327520.
[43] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: Deep hypersphere embedding for face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6738–6746, https://doi.org/10.1109/CVPR.2017.713.
[44] W. Liu, Y. Wen, Z. Yu, M. Yang, Large-margin softmax loss for convolutional neural networks, in: International Conference on Machine Learning (ICML), 2016, pp. 507–516, https://doi.org/10.5555/3045390.3045445.
[45] L.v.d. Maaten, G. Hinton, Visualizing data using t-sne, Journal of Machine Learning Research 9 (2008) 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html..
[46] T. Maciejewski, J. Stefanowski, Local neighbourhood extension of smote for mining imbalanced data, in: 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2011, pp. 104–111, https://doi.org/10.1109/CIDM.2011.5949434.
[47] M. Mostajabi, P. Yadollahpour, G. Shakhnarovich, Feedforward semantic segmentation with zoom-out features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3376–3385, https://doi.org/10.1109/CVPR.2015.7298959.
[48] W.W. Ng, G. Zeng, J. Zhang, D.S. Yeung, W. Pedrycz, Dual autoencoders features for imbalance classification problem, Pattern Recognition 60 (2016) 875–889, https://doi.org/10.1016/j.patcog.2016.06.013.
[49] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1717–1724, https://doi.org/10.1109/CVPR.2014.222.
[50] A. Patino-Saucedo, H. Rostro-Gonzalez, J. Conradt, Tropical fruits classification using an alexnet-type convolutional neural network and image augmentation, in: International Conference on Neural Information Processing, 2018, pp. 371–379, https://doi.org/10.1007/978-3-030-04212-7_32.
[51] J.B. Peter, R.S.S. Hancock, V.R. Mileva, Convolutional neural net face recognition works in non-human-like ways, Royal Society Open Science 7 (2020) 1–5. URL: https://doi.org/10.1098/rsos.200595..
[52] A. Rigner, Ai-based machine vision for retail self-checkout system, Master's Theses in Mathematical Sciences (2019), URL: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8985308&fileOId=8985340.
[53] A. Rocha, D.C. Hauagge, J. Wainer, S. Goldenstein, Automatic fruit and vegetable classification from images, Computers and Electronics in Agriculture 70 (2010) 96–104, https://doi.org/10.1016/j.compag.2009.09.002.
[54] S. Rota Bulo, G. Neuhold, P. Kontschieder, Loss max-pooling for semantic image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2126–2135, https://doi.org/10.1109/CVPR.2017.749.
[55] E.M. Rudd, M. Günther, T.E. Boult, Moon: A mixed objective optimization network for the recognition of facial attributes, in: European Conference on Computer Vision, 2016, pp. 19–35. URL: https://doi.org/10.1007/978-3-319-46454-1-2..
[56] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, International Journal of Computer Vision 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[57] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition CVPR, 2015, pp. 815–823. URL: https://doi.org/10.1109/CVPR.2015.7298682..
[58] W. Shen, X. Wang, Y. Wang, X. Bai, Z. Zhang, Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3982–3991, https://doi.org/10.1109/CVPR.2015.7299024.
[59] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations, ICLR, 2015, pp. 1–10. URL: http://arxiv.org/abs/1409.1556..
[60] R. Spezialetti, F. Stella, M. Marcon, L. Silva, S. Salti, L. Di Stefano, Learning to orient surfaces by self-supervised spherical cnns, in: Advances in Neural Information Processing Systems 33, 2020, URL: arxiv.org/abs/2011.03298..
[61] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.
[62] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708, https://doi.org/10.1109/CVPR.2014.220.
[63] P. Vuttipittayamongkol, E. Elyan, A. Petrovski, On the class overlap problem in imbalanced data classification, Knowledge-Based Systems 212 (2021), https://doi.org/10.1016/j.knosys.2020.106631 106631.
[64] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, R. Fergus, Regularization of neural networks using dropconnect, in: International Conference on Machine Learning, 2013, pp. 1058–1066. URL: https://doi.org/10.5555/3042817.3043055..

[65] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: Large margin cosine loss for deep face recognition, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5265–5274, https://doi.org/10.1109/CVPR.2018.00552.
[66] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, P.J. Kennedy, Training deep neural networks on imbalanced data sets, in: 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 4368–4374, https://doi.org/10.1109/IJCNN.2016.7727770.
[67] T. Wang, G. Zeng, W.W.Y. Ng, J. Li, Dual denoising autoencoder features for imbalance classification problems, in: IEEE International Conference on Internet of Things iThings and IEEE Green Computing and Communications GreenCom and IEEE Cyber, Physical and Social Computing CPSCom and IEEE Smart Data SmartData, 2017, pp. 312–317, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.52.
[68] Y.X. Wang, D. Ramanan, M. Hebert, Learning to model the tail, Advances in Neural Information Processing Systems 30 (2017) 7029–7039, https://doi.org/10.5555/3295222.3295446.
[69] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: European Conference on Computer Vision (ECCV), 2016, pp. 499–515. URL: https://doi.org/10.1007/978-3-319-46478-7_31..
[70] B. Wu, W. Tseng, Y. Chen, S. Yao, P. Chang, An intelligent self-checkout system for smart retail, in: International Conference on System Science and Engineering (ICSSE), 2016, pp. 1–4, https://doi.org/10.1109/ICSSE.2016.7551621.
[71] R. Yuda, C. Aroef, Z. Rustam, H. Alatas, Gender classification based on face recognition using convolutional neural networks (cnns), Journal of Physics: Conference Series (2020), https://doi.org/10.1088/1742-6596/1490/1/012042 012042.
[72] M.D. Zeiler, R. Fergus, Stochastic pooling for regularization of deep convolutional neural networks, in: 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2–4, 2013, Conference Track Proceedings, 2013. URL: http://arxiv.org/abs/1301.3557..
[73] X. Zhang, Z. Fang, Y. Wen, Z. Li, Y. Qiao, Range loss for deep face recognition with long-tailed training data, in: IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5419–5428, https://doi.org/10.1109/ICCV.2017.578.
[74] Z.H. Zhou, X.Y. Liu, Training cost-sensitive neural networks with methods addressing the class imbalance problem, IEEE Transactions on Knowledge and Data Engineering 18 (2005) 63–77. URL: https://doi.org/10.1109/TKDE.2006.17..
[75] G. Zong, Q. Li, P. Zhang, G. Zhang, Refined cnns for face recognition applications on embedded devices, 2020, pp. 307–312. URL: https://doi.org/10.1145/3383972.3384025..

Khurram Hameed received his BSc and MSc in Computer Systems Engineering at The Islamia University of Bahawalpur and University of Science and Technology Taxila, Pakistan, respectively. He is currently a PhD scholar at School of Engineering, Edith Cowan University, Australia. He is also recipient of HEC-ECU PhD scholarship (2017). He has been working as Assistant Professor at The Islamia University of Bahawalpur Pakistan for a decade. His research interests are computer vision, pattern recognition, image processing, parallel and distributed computing, artificial intelligence and machine learning.

Douglas Chai completed the BE(Hons) and PhD degrees in Electrical and Electronic Engineering from the University of Western Australia, Australia, in 1994 and 1999, respectively. He is currently the Associate Dean (Academic) with the School of Engineering at Edith Cowan University, Australia. His research interests include image analysis, pattern recognition, computer vision, barcode technology and document imaging. He has published over 80 technical papers, and received over 3,900 citations according to Google Scholar. He was an associate editor of the Australian Journal of Electrical and Electronics Engineering (AJEEE) in 2014–2017. Associate Professor Chai is a senior member of the Institute of Electrical and Electronics Engineers (IEEE). He has served in various IEEE committees for over 19 years, including chairmanship of IEEE Western Australia Section (in 2003–2004, 2007–2008) and the IEEE Signal Processing Western Australia Chapter (in 2005–2006, 2008, 2011–2013, 2018–2019).

Alexander Rassau received a Bachelor of Science (Cybernetics and Control Engineering) and a PhD in microelectronics from the University of Reading in the United Kingdom in 1997 and 2000 respectively. He is currently the Associate Dean Teaching and Learning for the School of Engineering at Edith Cowan University. He has played an active role in the very rapid expansion of the School over the past decade, overseeing the introduction of a wide range of new and innovative engineering programs and courses. Since 1998, he has also been actively involved as a researcher and educator in the areas of embedded systems, intelligent control, machine learning, automation and robotics.