

Multi-Source Domain Adaptation for Quality Control in Retail Food Packaging

Mamatha Thota^a, Stefanos Kollias^a, Mark Swainson^b, Georgios Leontidis^{c,*}

^a*School of Computer Science, University of Lincoln, Brayford Pool Campus, LN6 7TS, Lincoln, UK*

^b*National Centre for Food Manufacturing, University of Lincoln, PE12 7PT, Holbeach, UK*

^c*Department of Computing Science, University of Aberdeen, AB24 3UE, Aberdeen, UK*

Abstract

Retail food packaging contains information which informs choice and can be vital to consumer health, including product name, ingredients list, nutritional information, allergens, preparation guidelines, pack weight, storage and shelf life information (use-by / best before dates). The presence and accuracy of such information is critical to ensure a detailed understanding of the product and to reduce the potential for health risks. Consequently, erroneous or illegible labeling has the potential to be highly detrimental to consumers and many other stakeholders in the supply chain. In practice, due to the high volume of food packages that go through the supply chain, mistakes do occur therefore good quality of images are needed to verify the correctness of the information. In this paper, a multi-source deep learning-based domain adaptation system is proposed and tested to identify and verify the presence and legibility of use-by date information from food packaging photos taken as part of the validation process as the products pass along the food production line. This was achieved by improving the generalization of the techniques via incorporating new loss functions and making use of multi-source datasets in order to extract domain-invariant representations for all domains and aligning distribution of all pairs of source and target domains in a common feature space, along with the class boundaries. The proposed system performed very well in the conducted experiments, for automating the verification process and reducing labeling errors that could otherwise threaten public health and contravene legal requirements for food packaging information and accuracy. Comprehensive experiments on our food packaging datasets demonstrate that the proposed multi-source deep domain adaptation method significantly improves the classification accuracy and therefore has great potential for application and beneficial impact in food manufacturing control systems.

Keywords: deep learning, convolutional neural networks, multi-source domain adaptation, optical character verification, retail food packaging

*Corresponding author. Tel.: +44(0)7951053731

Email address: georgios.leontidis@abdn.ac.uk (Georgios Leontidis)

1. Introduction

Europe's food and drink sector employs 4.57 million people and has a turnover of €1.1 trillion, making it the largest manufacturing industry in the EU (Source: Data & Trends. EU Food & Drink Industry 2018. FoodDrink Europe 2018). The priority of all nations is to protect and feed their citizens. To assure public health, food safety is a legal requirement across the food supply chain. As part of this control approach all pre-packaged food products are required to display mandatory information on the food pack label. This information serves to ensure that the consumer can make a clear and informed choice as to the nature of their food purchases and are warned of any particular issues which could affect their health, e.g. Product shelf life information, the presence of allergens and also other warnings such as the potential for presence of physical or microbiological hazards associated with the particular food type. Labeling mistakes or legibility issues can therefore create major food safety problems including: Food poisoning, caused by the consumption of a product that has exceeded its actual use-by date (due to the date on the packaging either being incorrect or illegible); the triggering of an allergic reaction in a consumer who is susceptible to a particular food allergen, but due to a label fault was not aware that this allergen was present in their chosen food.

Food product traceability is a legal requirement in the EU, typically achieved by the presence of production date, time and process line code information on the food packaging. The presence of this information ensures that in the event of an emergency it is possible to identify and remove the affected food products in the supply chain. As a result, any fault in the accuracy or legibility of the pack traceability information will result in the supply chain stakeholders having to recall far more product than necessary due to the fact that the specific batch of product actually affected cannot be individually identified [1]. Such circumstances are financially expensive to the food manufacturers both in monetary terms and reputation. The environmental impact of such events also cannot be overlooked in terms of the impact of food waste on sector carbon footprints and ultimately in exacerbating climate change.

There are various reasons that could result in the label faults during the packaging process. They are mainly comprised of but are not limited to human error and equipment faults. For example they can occur due to the result of a faulty equipment, such as a clogged printer nozzle that could fail to produce readable labels, or due to the breakdown of a label printer with the production line still running, or it could also be due to simple human error where manual inspection cannot keep up with production line speeds. Preventing incorrect labels enable operators to take immediate action, such as rejecting a mislabeled product from the line, or stopping the assembly line altogether so that the problem can be corrected immediately.

The most common approach to overcome these labeling risks in the food supply chain is to read and verify the use by dates on the pre-packaged products during the manufacturing process. These Quality Assurance (QA) checks are typically conducted by an operator, but as such practices are very laborious and repetitive in nature, it places the operator in an error-prone working environment. Another less popular approach aiming at overcoming these labeling risks in the food supply chain is to use Optical Character Verification (OCV) approaches, whereby a supervisory system has the correct date code format, which gets transferred

both to the printer and the vision system. The vision system then verifies the date, heavily relying on the consistency in the date format, packaging and camera view angle, which is often hard to achieve in the food and drink manufacturing environment. OCV systems also require accurately labeled data to be utilized for training; but labeling a large number of target samples is overly laborious and a very cost-ineffective process, hence the need for a more robust solution. Previous studies to consider deep learning (DL) techniques for OCV have primarily focused on one domain and/or using transfer learning to enhance the performance and generalization of the developed techniques [2, 3, 4].

1.1. State-of-the-art

Over the last few years, DL [5] has been successfully applied to numerous applications and domains due to the availability of large amounts of labeled data, such as computer vision and image processing [6], signal processing [7], time series analysis and forecasting [8], and autonomous driving [9]. As most of the applications of DL techniques, such as the aforementioned ones, refer to supervised learning, labeling large number of datasets consume a lot of time and is very cost ineffective. In addition, when deploying a trained model to real-life applications the assumption is that both the source (training set) and the target one (application-specific) are drawn from the same distributions. When this assumption is violated, the DL model trained on the source domain will not generalize well on the target domain due to the distribution differences between the source and the target domains known as domain shift.

Learning a discriminative model in the presence of domain shift between source and target datasets is known as *Domain Adaptation*. Typically, domain adaptation methods are used to model labeled data from one single source domain to another called target domain. In real life applications, data from multiple sources and domains do exist, which could be leveraged to develop more robust and generalizable models based on novel approaches, such as introducing new loss functions. The information extracted from these multiple sources will be able to better fit the target distribution data that might not match any of the available source data – so it is more valuable in the performance improvement and is therefore receiving considerable attention in the real world applications like those described in this paper.

One simple approach for predicting the labels of the target domain data is to combine the training samples from all source domains and build a single model based on the pooled training samples. Due to the data expansion, the methods might improve the performance slightly, however, this simple approach will not work well in our application as there are significant conditional probability differences across domains in the food image data. Another approach is to extract multiple domain-invariant representations for each source and target domain pairs and map each of them into specific feature spaces in order to match their distributions, by training multiple models. However, this would take a lot of time as it involves training multiple models, therefore it is necessary to find a better way to make full use of multiple source domains. So here we propose an approach to overcome the above mentioned issues and present ablation studies using all six available datasets. To the best of our knowledge this is the first study to consider a new multi-source deep learning domain adaptation in the retail food packaging control, showing the great potential of advancing related

quality assurance systems in the food supply chain, further supporting automation towards industry 4.0 and with high potential to reduce errors and their related costs to the consumer and food business operators. Therefore, this paper claims two main contributions:

- A new multi-source deep learning-based domain adaptation approach is proposed that makes use of all available datasets in a multi-source combined manner that can improve the overall performance extensively via incorporating more than one loss functions.
- Demonstrate how a quality control system based on machine learning methodologies can help to improve the quality control processes applied to the retail food packaging sector. The importance and potential impact of such a system is exacerbated by the number of errors found in retail food packaging labels that contribute to the overall food waste, the loss of billions of pounds to the sector and most importantly jeopardising public health and food safety, which is a legal requirement across the food supply chain

The rest of the paper is laid out as follows: Section 2 presents the related work in single source and multi source domain adaptation. Section 3 describes the dataset and our proposed approach, focusing on multi-source domain adaptation techniques, Section 4 presents the experimental results obtained after applying our model to the food packaging data and Section 5 presents the commercial impact associated with our method and Section 6 concludes the paper.

2. Related Work

In recent years, many single source domain adaptation methods have been proposed. Discrepancy-, Adversarial- and Reconstruction- based approaches are the three primary domain adaptation approaches currently being applied to address the distribution shift [10]. Discrepancy-based approaches rely on aligning the distributions in order to minimize the divergence between them. The most commonly used discrepancy-based methods are Maximum Mean Discrepancy (MMD) [11], Correlation Alignment (CORAL) [12] and Kullback–Leibler divergence [13]. Adversarial-based approaches minimize the distance between the source and the target distributions using domain confusion, an adversarial method used in Generative Adversarial Networks. The approach proposed in [14] tries to minimize the feature distributions by integrating a gradient reversal layer, whereas in [15] the aim is to minimize the distance between source and target samples in an adversarial manner using Wasserstein distance. Another class of approaches known as Reconstruction-based approaches create a shared representation between the source and the target domains whilst preserving the individual characteristics of each domain. Rather than minimizing the divergence, the method in [16] learns joint representations that classify the labeled source data and at the same time reconstruct the target domain.

In contrast to the single source domain adaptation techniques, multi-source domain adaptation techniques assume the data is available from multiple source domains. In domain adaptation, multi-source domains are even more critical than single source domain as they need to handle the domain alignment between source and target domains, along with alignments between multiple available sources. The previous single source domain

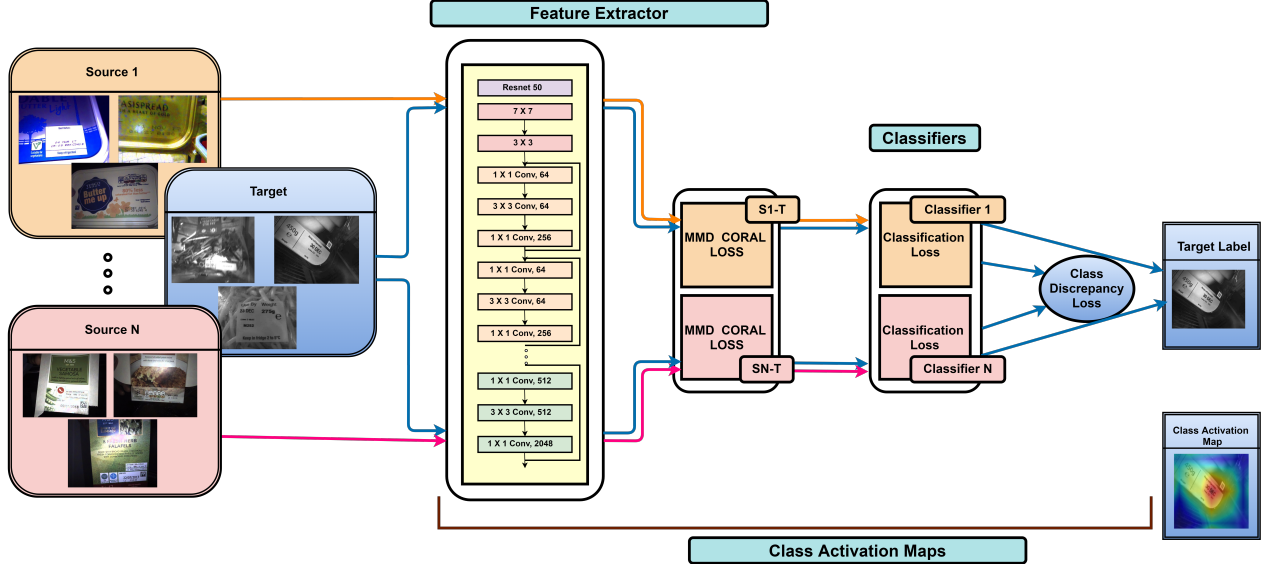


Figure 1: Overview of our proposed Multi-Source Deep Domain Adaptation model, showing the Feature Extractor, Classifier and the Class Activation Maps.

adaptation methods, as the name suggests, only considered a single source and a single target, however in real world applications, there are multiple source domains available to extract knowledge from.

Learning from multiple different sources has originated from early theoretical analysis [17],[18], and has many practical applications. Initially many shallow models were proposed in order to tackle the multi-source domain adaptation problem [19] [20]. The work in [18] established a general bound on the expected loss of the model by minimizing the empirical loss on the nearest k sources. Deep Cocktail Network [21] proposed a multi-way adversarial learning to minimize the discrepancy between the target and each of the multiple source domains. The work most related to ours has been the one in [22], where only MMD loss was used to minimize the feature discrepancy. Our proposed approach aims at minimizing the feature discrepancy through implementing a new loss function that includes both MMD and CORAL losses in for improved generalization. Additionally, the Class Activation Mapping (CAM) component of our model [23] adds an extra step to the algorithm that provides a visualization of which areas in the image contributed the most to the decision-making process, enabling the trust of the end-users.

3. Methodology

3.1. Problem Statement

Multi-source domain adaptation with N different source distributions is denoted as $\{p_{s_j}(x, y)\}_{j=1}^N$, and the labeled source domain image data $\{(X_{s_j}, Y_{s_j})\}_{j=1}^N$ are drawn from these distributions, where $X_{s_j} = \{x_i^{s_j}\}_{i=1}^{c_j}$ represents image data sampled with c number of images from source domain j and $Y_{s_j} = \{y_i^{s_j}\}_{i=1}^{c_j}$ is the corresponding ground truth labels. We also have the target distribution $p_t(x, y)$, from which the target domain image data $X_t = \{x_i^t\}_{i=1}^c$ with a total number of c images sampled without any label observation Y_t .

The end goal is to predict the labels of the unlabeled data in the target domain using the labeled multi-source domain data.

3.2. Dataset Description

Our dataset includes data from six domains Abbeydale (Ab), Burton (Bu), Bourne (Bo), Listowel (Li), Windmill-Lane (Wi) and Ossett (Os), with two classes in each location, i.e. OK (legible) vs NOT-OK (illegible). Each domain contains images from within the food packaging supply chain but with varying background, colour, context, etc. To make sure any proposed implementation is not affected by imbalanced classes or under-/over- represented domains, each domain consists of 5,000 images, balanced equally with 2,500 images for each class. The best before date on the OK category images is readable, whereas the one on the NOT-OK category is unreadable, possibly due to incomplete or missing dates, packaging which had been folded at crucial points, digits fading over time, human-made date obstructions, heavy distortion, lighting, blur, etc. Initially, three people manually annotated this dataset, with two more annotators further sampling and verifying the manual annotations for quality control, hence keeping those 30,000 images that both annotators were in full agreement. The main challenges of such real-life datasets are the unavailability of labeled data and high variability across the datasets, such as heavy distortion, varying background, illumination/blur, date format, angle and orientation of the label, etc. In this paper various techniques are explored and a new approach is proposed for adapting knowledge learnt by one dataset to another. Examples of the images used in this study can be seen in Figure 2.

3.3. Overview of our Architecture

In this section our proposed multi-source DL-based domain adaptation approach is introduced that aims at improving the binary classification of the food packaging dataset across all locations. We use labeled source data from multiple locations and unlabeled target data from a single location. As shown in Figure 1, our model comprises of a feature extractor and a classification part, where additional losses are introduced to enhance the learning process and the overall domain adaptation performance. The feature extractor part learns useful representations for all domains, whereas its sub-network learns features specific to each source-target domain pairs. The classification part of the model learns domain-specific attributes for each target image and provides N categorization results. The classification part of the model aligns with the domain-specific classifiers', as the class boundaries are highly likely to be misclassified, because they are learned from different classifiers. The CAM part of the model helps with the visualization of the CNN's interpretation of the model when making predictions.

3.4. Multi-Source Domain Adaptation

Domain Adaptation is challenging due to the domain shift between the source and the target datasets. Multi-source domain adaptation is even more challenging as the aim is to deal with domain shift between

multiple source datasets. The proposed approach is composed of a feature extractor and source specific classification parts and aims at minimizing the feature discrepancy, for learning domain-invariant representations, the class boundary discrepancy, for minimizing the mismatch among all classifiers, and finally improving source data classification by reducing the classification loss, leading to improved generalization on the target dataset, as will be demonstrated in the ablation studies below.

Feature Discrepancy Loss: We reduce the feature discrepancy by minimizing both MMD and CORAL loss. The proposed deep domain adaptation architecture jointly adapts features using two popular feature adaptation metrics that combine MMD with CORAL in order to align higher order statistics along with the first and second order statistics.

MMD defines the distance between the two distributions with their mean embeddings in the Reproducing Kernel Hilbert Space (RKHS). MMD is a two sample kernel test to determine whether to accept or reject the null hypothesis $p = q$ [24], where p and q are source and target domain probability distributions. Let H be the RKHS with a characteristic kernel k , the squared distance formulation of MMD is shown in equation (1)

$$d_k^2(p, q) = \|E_p[\phi(x^s)] - E_q[\phi(x^t)]\|_H^2 \quad (1)$$

The distance between the distributions with their mean embeddings is given in equation (2)

$$Loss_{MMD} = \left\| \frac{1}{N} \sum_{i=1}^N \phi(x_i^s) - \frac{1}{M} \sum_{j=1}^M \phi(x_j^t) \right\|^2 \quad (2)$$

$$\begin{aligned} = & \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N \phi(x_i^s)^T \phi(x_{i'}^s) - \frac{2}{NM} \sum_{i=1}^N \sum_{j=1}^M \phi(x_i^s)^T \phi(x_j^t) \\ & + \frac{1}{M^2} \sum_{j=1}^M \sum_{j'=1}^M \phi(x_j^t)^T \phi(x_{j'}^t) \end{aligned} \quad (3)$$

where N and M are the total number of items in the source and target respectively.

The kernel trick can be applied as each term in the equation (3) involves inner products between ϕ vectors in order to estimate the squared distance between the empirical kernel mean embeddings as follows:

$$\begin{aligned} Loss_{MMD} = & \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N k(x_i^s, x_{i'}^s) - \frac{2}{NM} \sum_{i=1}^N \sum_{j=1}^M k(x_i^s, x_j^t) \\ & + \frac{1}{M^2} \sum_{j=1}^M \sum_{j'=1}^M k(x_j^t, x_{j'}^t) \end{aligned} \quad (4)$$

CORAL loss [12] is also used to minimize the discrepancy between source and target data by reducing the distance between the source and target feature representations. We define the CORAL loss as the distance between the second-order statistics (covariances) of the source and target features:

$$Loss_{CORAL} = \frac{1}{4d^2} \|C_s - C_t\|_F^2 \quad (5)$$

$\|\cdot\|_F^2$ denotes the squared matrix Frobenius norm, C_s is the source covariance matrix and C_t is the target covariance matrix.

The covariance matrices of the source and target data are given by:

$$C_s = \frac{1}{N-1} (D_s^T D_s - \frac{1}{N} (1^T D_s)^T (1^T D_s)) \quad (6)$$

$$C_t = \frac{1}{M-1} (D_t^T D_t - \frac{1}{M} (1^T D_t)^T (1^T D_t)) \quad (7)$$

1 is a column vector with all elements equal to 1, N and M are the total number of items in the source and target respectively.

The total feature discrepancy loss is therefore given by the equation (8),

$$Loss_{FD} = Loss_{MMD} + Loss_{CORAL} \quad (8)$$

Class Discrepancy Loss: Classifiers are likely to mis-classify the target samples near the class boundary as they are trained using different source domains, each having different target prediction. Therefore we aim at minimizing the discrepancy among all classifiers by making their probabilistic outputs similar. The class discrepancy is calculated by the equation (9),

$$Loss_{CD} = \binom{N}{2}^{-1} \sum_{j=1}^{N-1} \sum_{i=j+1}^N [|E(X_i) - E(X_j)|] \quad (9)$$

where N is total number of classifiers.

Classification Loss: The network reduces the discrepancy among classifiers by minimizing the classification loss. We train the network with labeled source data and calculate the empirical loss through minimizing the cross-entropy loss as follows

$$Loss_{CL} = \frac{1}{N} \sum_{i=1}^N V(f(x_i^s), y_i^s) \quad (10)$$

where $V(\cdot, \cdot)$ is the cross-entropy loss function and $f(x_i^s)$ is the conditional probability that the CNN assigns to label y_i^s .

Our total loss is made up of classification loss (CL), feature discrepancy loss (FD) and class discrepancy loss (CD). By minimizing each of these losses the network can classify the source domain data more accurately and reduce the dataset bias and the discrepancy among classifiers. Jointly minimizing these three losses, the network can learn features that generalize and adapt well on the target dataset. The overall objective of our network can be formulated as:

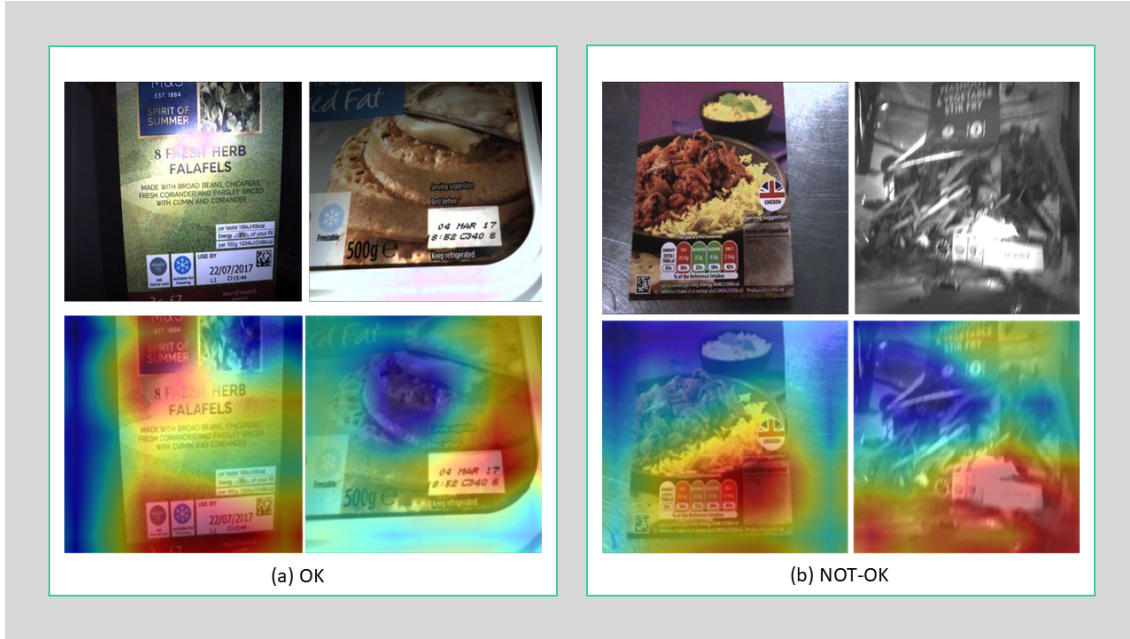


Figure 2: (a) Acceptable Quality (OK) vs (b) Unacceptable Quality (NOT-OK) images. First row: raw images; second row: images with class activation map

$$Loss_{Total} = Loss_{CL} + \lambda Loss_{FD} + \gamma Loss_{CD} \quad (11)$$

where λ and γ are penalty parameters.

3.5. Class Activation Maps

Deep neural networks are often considered to be black boxes that offer no straightforward way of understanding what a network has learned or which part of an input to the network was responsible for the prediction of the network. When such models make predictions, there is often no explanation to justify them. Class activation maps are an efficient way to visualize the importance the model places on various regions in an image while training, offering insights that are crucial for the model’s usability and explainability [23]. We have incorporated CAM to our approach in order to visualize which areas of the food packaging images contribute the most to the decision-making process of the algorithm.

CAM provides some insight into the process of CNN interpretability and explainability by overlaying a heat map over the original image to demonstrate where the model is paying more attention for its decision-making process. It visually demonstrates how the algorithm comes up with its prediction by highlighting the pixels of the image that trigger the model to associate the image with a particular class. CAM help us understand which regions of an input image influence the convolutional neural network’s output prediction. Such an information can be used for examining the bias of the algorithm and the lack of generalization capabilities, allowing to take steps to enhance the robustness of the model and potentially increase its accuracy.

Table 1: Classification accuracy (%) and loss functions used for each method, per target for source - target combinations*

Method	Loss Functions	Source → Target(Wi)	ACC	Source → Target(Ab)	ACC	Source → Target(Bo)	ACC	Source → Target(Bu)	ACC	Source → Target(Os)	ACC	Source → Target(Li)	ACC
TL	CL	Ab	77.9	Bu	79.4	Ab	78.2	Ab	79.2	Ab	80.1	Ab	77.6
SS-DA	CL, FD	Ab	83.2	Bu	84.6	Ab	82.6	Ab	83.5	Ab	84.7	Ab	86.2
SC-DA2	CL, FD	Ab,Bo	84.5	Bu,Bo	82.3	Ab,Bu	84.8	Ab,Bo	84.1	Ab,Bo	85.2	Ab,Bo	84.4
		Ab,Bu	85.6	Bu,Li	86.6	Ab,Li	83.5	Ab,Li	84.7	Ab,Bu	86.1	Ab,Bu	87.4
		Ab,Li	82.1	Bu,Os	83.8	Ab,Os	86.3	Ab,Os	88.3	Ab,Li	85.3	Ab,Os	87.6
		Ab,Os	80.9	Bu,Wi	87.9	Ab,Wi	84.2	Ab,Wi	82.7	Ab,Wi	84.6	Ab,Wi	88.3
SC-DA3	CL, FD	Ab,Bu,Li	83.6	Bu,Bo,Wi	83.7	Ab,Bu,Li	86.2	Ab,Bo,Wi	84.9	Ab,Bo,Wi	86.2	Ab,Bo,Wi	85.2
		Ab,Bu,Os	86.1	Bu,Li,Os	87.5	Ab,Bu,Os	84.9	Ab,Li,Os	85.4	Ab,Bu,Li	85.4	Ab,Bu,Os	88.6
		Ab,Bu,Bo	86.3	Bu,Li,Bo	88.2	Ab,Bu,Wi	85.2	Ab,Li,Bo	85.1	Ab,Bu,Bo	86.6	Ab,Bu,Bo	88.1
		Ab,Li,Os	84.1	Bu,Li,Wi	87.2	Ab,Li,Os	86.9	Ab,Li,Wi	84.9	Ab,Bu,Wi	87.2	Ab,Bu,Wi	88.7
		Ab,Li,Bo	84.6	Bu,Os,Bo	85.6	Ab,Li,Wi	84.1	Ab,Os,Bo	88.6	Ab,Li,Bo	85.6	Ab,Os,Bo	82.1
		Ab,Os,Bo	83.7	Bu,Os,Wi	88.3	Ab,Os,Wi	86.8	Ab,Os,Wi	87.9	Ab,Li,Wi	87.3	Ab,Os,Wi	90.2
MS-DA2	CL, FD, CD	Ab,Bo	90.3	Bu,Bo	89.1	Ab,Bu	90.2	Ab,Bo	90.6	Ab,Bo	88.7	Ab,Bo	89.6
		Ab,Bu	89.5	Bu,Li	88.7	Ab,Li	91.1	Ab,Li	88.7	Ab,Bu	90.9	Ab,Bu	92.1
		Ab,Li	91.6	Bu,Os	90.6	Ab,Os	90.3	Ab,Os	92.9	Ab,Li	89.9	Ab,Os	91.3
		Ab,Os	92.8	Bu,Wi	90.2	Ab,Wi	91.2	Ab,Wi	90.2	Ab,Wi	91.9	Ab,Wi	90.5
MS-DA3	CL, FD, CD	Ab,Bu,Li	92.8	Bu,Bo,Wi	91.3	Ab,Bu,Li	94.2	Ab,Bo,Wi	91.4	Ab,Bo,Wi	93.6	Ab,Bo,Wi	91.2
		Ab,Bu,Os	93.2	Bu,Li,Os	91.5	Ab,Bu,Os	92.6	Ab,Li,Os	92.6	Ab,Bu,Li	91.5	Ab,Bu,Os	92.3
		Ab,Bu,Bo	92.7	Bu,Li,Bo	91.9	Ab,Bu,Wi	92.1	Ab,Li,Bo	92.9	Ab,Bu,Bo	91.8	Ab,Bu,Bo	92.2
		Ab,Li,Os	93.1	Bu,Li,Wi	92.1	Ab,Li,Os	92.3	Ab,Li,Wi	91.3	Ab,Bu,Wi	92.7	Ab,Bu,Wi	92.6
		Ab,Li,Bo	92.5	Bu,Os,Bo	92.3	Ab,Li,Wi	92.9	Ab,Os,Bo	93.4	Ab,Li,Bo	93.5	Ab,Os,Bo	92.1
		Ab,Os,Bo	93.2	Bu,Os,Wi	92.6	Ab,Os,Wi	92.4	Ab,Os,Wi	93.2	Ab,Li,Wi	92.6	Ab,Os,Wi	93.7

*TL:Transfer Learning, SS-DA:Single source Domain Adaptation, SC-DA2:Source Combined Domain Adaptation with 2 sources combined into a single source, SC-DA3:Source Combined Domain Adaptation with 3 sources combined into a single source, MS-DA2:Multi Source Domain Adaptation with 2 sources combined into a single source, MS-DA3:Multi Source Domain Adaptation with 3 sources combined into a single source. Wi:Windmill-Lane, Ab:Abbeyleale, Bo:Bourne, Bu:Burton, Os:Ossett, Li>Listowel. Loss functions - CL:Classification Loss, FD:Feature Discrepancy Loss, CD:Class Discrepancy Loss.

4. Experiments and Results

As explained earlier, some of the benefits of applying the proposed approach to this application is the elimination of monotonous and inconsistent manual labor and reducing the human errors along with increasing speed and productivity. Firstly, we conducted experiments using labeled single source dataset and unlabeled single target dataset for all the six locations. The goal of this experiment has been to establish a baseline for images that would be classified as readable and acceptable according to human standards. Further experiments were conducted using our proposed multi-source domain adaptation approach this time, i.e. adapting two labeled source datasets with a single unlabeled target domain and three labeled source datasets with a single unlabeled target domain. We compared the obtained results with the baseline Transfer Learning and single source adaptation experiment conducted initially. Additional experiments were carried out by combining the two/three source datasets into a single source dataset. Our overall aim has been to improve the image classification accuracy in the provided food packaging datasets, hence allowing for enhanced quality control in the food supply chain. The combinations tested included all six locations available and were conducted in the following manner:

Table 2: Average classification accuracy (%) per target for each target dataset

Method	Loss Functions	Avg-ACC(Wi)	Avg-ACC(Ab)	Avg-ACC(Bo)	Avg-ACC(Bu)	Avg-ACC(Os)	Avg-ACC(Li)
TL	CL	77.90	79.42	78.25	79.26	80.10	77.67
SS-DA	CL, FD	83.20	84.60	82.60	83.50	84.70	86.20
SC-DA2	CL, FD	83.27	85.15	84.70	84.95	85.30	86.92
SC-DA3	CL, FD	84.73	86.75	85.68	86.14	86.38	87.15
MS-DA2	CL, FD, CD	91.05	89.65	90.70	90.60	90.35	90.87
MS-DA3	CL, FD, CD	92.91	91.95	92.75	92.46	92.61	92.35

- Transfer Learning (Vanilla Resnet-50)
- Single Source to Single Target
- Combined Source to Single Target
- Multi Source to Single Target

We have trained our model on labeled data from a source domain to achieve better performance on data from a target domain, with access to only unlabeled data in the target domain. We have conducted Transfer Learning on the food packaging datasets by adapting the pre-trained CNN weights of imagenet [25] on Alexnet [26], VGG16 [27], Resnet50 and Resnet101 [28]. Each food packaging image was fed through the previously trained network on the ImageNet dataset. The resultant accuracy from the experiments conducted as described above showed that Resnet-50 had a better accuracy compared to other methods, hence we used ResNet-50 [28] pretrained on ImageNet [25] as our backbone/baseline network.

We have therefore replaced the last fully connected (FC) layer with the task specific FC layer and fine-tuned all the convolutional and pooling layers and trained the classifier layer via back propagation. Adam optimizer with a learning rate of 0.001 was used [29]. It should be pointed out that the focus of the paper is not to propose a new architecture; the baseline ResNet model is a common practice when developing domain adaptation techniques so that different methods can be compared without the results being affected by the backbone CNN model, hence the decision to use the Resnet-50 model as our baseline model to demonstrate how our proposed multi-source domain adaptation approach can be used in a real-life problem.

4.1. Baseline Transfer Learning Process (Vanilla Resnet-50)

The labeled source and the unlabeled target images have been fed through the ResNet50 pre-trained on ImageNet which only uses the classification loss. We performed various experiments with each location as the target domain for multiple source-target pairs and presented the results in Table 1. Table 2 shows the average accuracy per target by each method, which can be visualized in Figure 3, and finally Table 3 shows the overall accuracy for all targets per method.

4.2. Single-Source Domain Adaptation

The labeled source and the unlabeled target images have been fed through the model where the discrepancy between the pair of datasets was minimized by jointly reducing the feature discrepancy and the classification losses. We performed various experiments with each location as the target domain for multiple source-target pairs and presented the results in Table 1. Table 2 shows the average accuracy per target by each method which can be visualized in Figure 3, and finally Table 3 shows the overall accuracy for all targets per method.

4.3. Combined Source to Single Target

In the source combined setting, all the source domains are combined into a single domain, and the experiments are conducted in a traditional single domain adaptation manner. The labeled source and the unlabeled target images have been fed through the model where the discrepancy between the pair of datasets was minimized by jointly reducing the feature discrepancy and the classification losses. We have combined and experimented using

- Two sources combined
- Three sources combined

We performed various experiments with each location as the target domain for multiple source-target pairs. In the first set of experiments, two datasets were combined into a single source dataset, and in the next set of experiments, three source datasets were combined into a single source dataset. The results of this method can be seen in Table 1 for multiple source-target pairs. Table 2 shows the average accuracy per target by each method which can be visualized in Figure 3, and finally Table 3 shows the overall accuracy for all targets per method.

4.4. Multi-Source Domain Adaptation (Proposed Method)

The labeled sources and the unlabeled target images are fed through the model where the discrepancy between the pair of datasets was minimized by jointly reducing the feature discrepancy, class discrepancy and the classification losses using the techniques described in section 3. We performed various experiments per location as target domain and presented the results from Table 1 through Table 3. We categorized the experiments as follows:

- Multi-Source with two datasets
- Multi-Source with three datasets

Initially, we performed the experiments using two source datasets as input domains and in the second set of experiments, we used three sources as the input domains. The results of this method can be seen in Table 1 for multiple source-target pairs. Table 2 shows the average accuracy per target by each method which can be visualized in Figure 3, and Table 3 shows the overall accuracy for all targets per method.

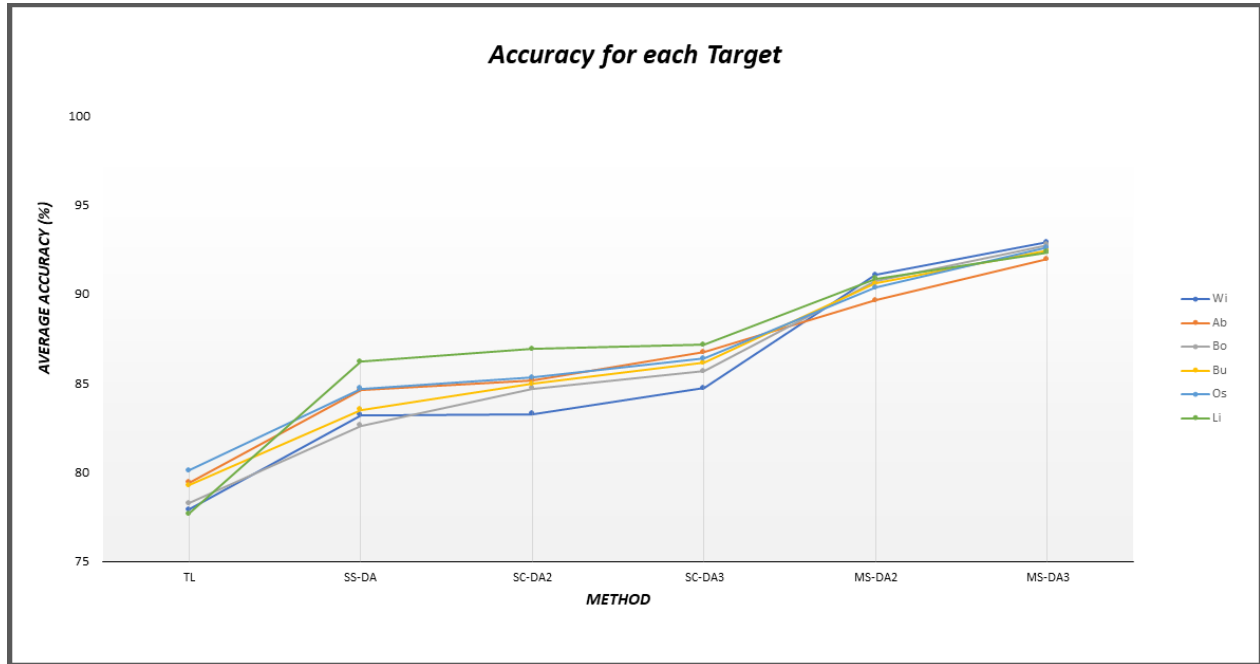


Figure 3: Graph showing analysis of Average Accuracy per method for each target dataset.

Table 3: Comparison of average classification accuracy across all methods

Method	Loss Functions	Avg-ACC
TL	CL	78.77
SS-DA	CL, FD	84.14
SC-DA2	CL, FD	85.05
SC-DA3	CL, FD	86.13
MS-DA2	CL, FD, CD	90.53
MS-DA3	CL, FD, CD	92.50

From the experimental results as per Table 3, we can observe that our proposed multi-source domain adaptation approach significantly outperforms two of the above mentioned baseline methods on most of the tasks with an average classification accuracy improvement by more than 6%. We can also note that learning domain-invariant features for each source-target pairs along with exploiting the domain-specific class boundary information significantly increased the average classification accuracy in the target dataset.

The overall average accuracy for the Transfer Learning is 78.77%, single-source approach is 84.14%, two and three source combined approach is 85.05% and 86.13% respectively and finally our multi-source approach with two and three sources is 90.53 % and 92.50% respectively.

5. Discussion on Commercial Impact

Enhancing image quality control systems in the food supply chain will significantly improve the assurance of food safety, quality and legality. There are major commercial benefits associated with these positive impacts. Food manufacturers and retailers will save money associated with product trade withdrawals and recalls linked to incorrect or illegible critical product information on food packaging. With the average cost of a product recall in the UK being £38,075 [30] the financial impact of such an issue can be sufficient to render smaller businesses insolvent. This is high risk in a sector comprising 96% (7,030) micro, small and medium sized enterprises [31]. Economic damage is exacerbated by serious operational disruption and reputational harm caused by product recall, which adds significantly to the direct financial costs incurred.

In addition, increased control will ensure that food manufacturers are far less susceptible to potential legal costs associated with print image quality faults. Enforcement bodies, such as the UK Food Standards Agency, via Local Authority Trading Standards and Environmental Health Departments, have the capacity to issue improvement notices and prosecute food businesses for failure to control food safety and quality. For example, Food Information Regulation 1169/2011 is specific to the presence and accuracy of product information, mandating the provision of clear consumer information and the labelling of both prepacked and non-prepacked food [32]. Further to statutory routes, with an increasingly litigious society there is also the potential for private prosecution relating to food safety offences if a lack of due diligence can be seen to have contributed to the fault / failure.

Reflecting that food and drink is the UKs largest manufacturing sector - contributing an annual GVA of £28.8 billion [31] and with proportionate global reputation and impact – there are significant commercial benefits to be realised from label and date code verification technologies. The major benefits of image control systems such as APRIL Eye [33] ripple through the supply chain, from technology developers, manufacturers and suppliers through to end consumers.

6. Conclusion and future Work

In this paper a multi-source adaptation methodology is proposed that attempts to adapt and generalize information from one dataset to another for automating the verification of the use-by dates on food packaging datasets. We implemented and applied the proposed DL-based domain adaptation technique with multiple source domains that simultaneously align covariances along with maximum mean discrepancy for each pair of source and target domains by learning multiple domain-invariant representations and outputs of classifiers from multiple sources. We have incorporated a domain-specific class boundary alongside extracting common domain invariant representations. Within this framework, we have also integrated class activation maps to visualize the importance the model places on various regions in an image while training, offering insights that are crucial for the model’s usability and explainability. Single source and Source combined methods apply feature discrepancy and the classification losses, whereas the Multi-source domain adaptation method utilizes

feature discrepancy, class discrepancy and the classification losses together to reduce the discrepancy between the datasets. Extensive experiments conducted on our food packaging datasets demonstrate the effectiveness of the proposed framework; results shown in Table 3 illustrate that the performance of the food packaging classification improved significantly with our proposed approach compared to both the baseline approach (transfer learning) and the common approach of single source/source-combined. The proposed approach can also be applied to wider aspects of food packaging control, such as the verification of the allergen labeling, barcode, nutritional information and many more. Our future work will extend this study to a much larger dataset, consisting of about half a million food packaging images, and we are also aiming at improving the domain adaptation approach by incorporating adversarial components.

Acknowledgment

We would like to thank Olympus Automation Ltd. Company for providing us with the data used in this study.

References

- [1] Simon Pearson, David May, Georgios Leontidis, Mark Swainson, Steve Brewer, Luc Bidaut, Jeremy G Frey, Gerard Parr, Roger Maull, and Andrea Zisman. Are distributed ledger technologies the panacea for food traceability? *Global Food Security*, 20:145–149, 2019.
- [2] Fabio De Sousa Ribeiro, Liyun Gong, Francesco Calivá, Mark Swainson, Kjartan Gudmundsson, Miao Yu, Georgios Leontidis, Xujiang Ye, and Stefanos Kollias. An end-to-end deep neural architecture for optical character verification and recognition in retail food packaging. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2376–2380. IEEE, 2018.
- [3] Sungho Suh, Haebom Lee, Yong Oh Lee, Paul Lukowicz, and Jongwoon Hwang. Robust shipping label recognition and validation for logistics by using deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4509–4513. IEEE, 2019.
- [4] Fabio De Sousa Ribeiro, Francesco Caliva, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis, and Stefanos Kollias. An adaptable deep learning system for optical character verification in retail food packaging. In *2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–8. IEEE, 2018.
- [5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [6] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis, and Stefanos Kollias. Adaptation and contextualization of deep neural network models. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2017.

- [7] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazière, Paolo Vinai, Georgios Leontidis, and Stefanos Kollias. A deep learning approach to anomaly detection in nuclear reactors. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.
- [8] George Onoufriou, Ronald Bickerton, Simon Pearson, and Georgios Leontidis. Nemesyst: A hybrid parallelism deep learning-based framework applied for internet of things enabled food retailing refrigeration systems. *Computers in Industry*, 113:103133, 2019.
- [9] Swarnendu Ghosh, Anisha Pal, Shourya Jaiswal, KC Santosh, Nibaran Das, and Mita Nasipuri. Segfast-v2: Semantic image segmentation with less parameters in deep learning for autonomous driving. *International Journal of Machine Learning and Cybernetics*, 10(11):3145–3154, 2019.
- [10] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.
- [11] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*, 2015.
- [12] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision*, pages 443–450. Springer, 2016.
- [13] Fuzhen Zhuang, Xiaohu Cheng, Ping Luo, Sinno Jialin Pan, and Qing He. Supervised representation learning: Transfer learning with deep autoencoders. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
- [14] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.
- [15] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. *arXiv preprint arXiv:1707.01217*, 2017.
- [16] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 597–613. Springer, 2016.
- [17] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010.
- [18] Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. In *Advances in Neural Information Processing Systems*, pages 321–328, 2007.
- [19] I-Hong Jhuo, Dong Liu, DT Lee, and Shih-Fu Chang. Robust visual domain adaptation with low-rank reconstruction. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2168–2175. IEEE, 2012.

- [20] Hongfu Liu, Ming Shao, and Yun Fu. Structure-preserved multi-source domain adaptation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 1059–1064. IEEE, 2016.
- [21] Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3964–3973, 2018.
- [22] Yongchun Zhu, Fuzhen Zhuang, and Deqing Wang. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5989–5996, 2019.
- [23] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.
- [24] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.
- [25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
- [29] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [30] James Ridler. Cost-cutting sparks rise in food and drink recalls, January 2018. <https://www.foodmanufacture.co.uk/Article/2017/12/22/Food-and-drink-recalls-rise-sparked-by-cost-cutting/>.
- [31] Department for Environment Food Rural Affairs. National statistics, food statistics in your pocket: Food chain, 2018. <https://www.gov.uk/government/publications/food-statistics-pocketbook/food-statistics-in-your-pocket-food-chain>.
- [32] Mark Swainson. *Swainson’s Handbook of Technical and Quality Management for the Food Manufacturing Sector*. Woodhead Publishing, 2018.

[33] Olympus Automation Ltd. (OAL). April™ eye label and date code verification, 2020. <https://connected.oalgroup.com/april-eye>.