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

Mamatha Thota[1] and Georgios Leontidis[2]

[1] School of Computer Science, University of Lincoln, Lincoln, UK
[2] Department of Computing Science, University of Aberdeen, Aberdeen, UK
mthota@lincoln.ac.uk, georgios.leontidis@abdn.ac.uk

Abstract. Retail food packaging contains information which informs choice and can be vital to consumer health, including product name, ingredients list, allergens, 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. 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 images 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 distributions of all pairs of source and target domains in a common feature space, along with the class boundaries. 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

1 Introduction & Related Work

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). 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. Labeling mistakes 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.

To the best of our knowledge this is the first study to consider a new multi-source deep learning domain adaptation approach for retail food packaging control, further supporting automation towards industry 4.0 and with high potential

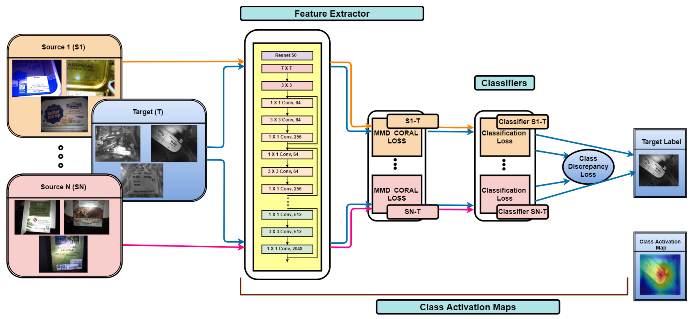

Fig. 1. Overview of our proposed Multi-Source Deep Domain Adaptation model, showing the Feature Extractor, Classifier and the Class Activation Maps.

to reduce errors and their related costs to the consumer and food business operators. Previous studies to consider deep learning (DL) techniques for OCV have primarily focused on one domain and/or using transfer learning (TL) to enhance the performance and generalization of the developed techniques [10], [11], [9]. In recent years, many single source domain adaptation methods have been proposed. Discrepancy-based approaches rely on aligning the distributions in order to minimize the divergence between the Maximum Mean Discrepancy (MMD) [8], Correlation Alignment (CORAL) [12]. The approach in [3] tries to minimize the feature distributions by integrating a gradient reversal layer. Rather than minimizing divergence, the method in [4] learns joint representations to classify the labeled source data, while reconstructing the target domain.

Learning from multiple different sources has originated from early theoretical analysis [1], [2], and has many practical applications. Initially many shallow models were proposed in order to tackle the multi-source domain adaptation problem [6] [7]. Deep Cocktail Network [13] 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 [15].

2 Methodology

As shown in Figure 1, our model comprises a feature extractor and a classification part. 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.

2.1 Multi-Source Domain Adaptation

The proposed approach comprises a feature extractor and source specific classification parts, aiming at minimizing the feature discrepancy, for learning domain-invariant representations; the class boundary discrepancy, for minimizing the mismatch among classifiers; and improving source data classification by reducing the classification loss, leading to better generalization on the target dataset.

Feature Discrepancy Loss: We reduce the feature discrepancy by minimizing both MMD and CORAL loss in order to align higher order statistics along with the first and second order statistics. MMD [5] defines the distance between the two distributions with their mean embeddings in the Reproducing Kernel Hilbert Space (1)

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

where ϕ(x) denotes the kernel associated with the feature map ϕ, N and M are the total number of items in the source and target respectively.

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 (2),

$$Loss_{CORAL} = \frac{1}{4d^2}||C_s - C_t||_F^2 \tag{2}$$

$||.||_F^2$ denotes the squared matrix Frobenius norm, C_s and C_t are the source and target covariance matrices. The total feature discrepancy loss is therefore given by the equation (3),

$$Loss_{FD} = Loss_{MMD} + Loss_{CORAL} \tag{3}$$

Class Discrepancy Loss: Classifiers are likely to misclassify 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 equation (4),

$$Loss_{CD} = \binom{N}{2}^{-1} \sum_{j=1}^{N-1}\sum_{i=j+1}^{N}[|E(X_i) - E(X_j)|] \tag{4}$$

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 cross-entropy loss as

$$Loss_{CL} = \frac{1}{N}\sum_{i=1}^{N} V(f(x_i^s), y_i^s) \tag{5}$$

where $V(.,.)$ 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 jointly minimizing these three losses,

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

Method	Loss Functions	S → T (Wi)	ACC	S → T (Ab)	ACC	S → T (Bo)	ACC	S → T (Bu)	ACC	S → T (Os)	ACC	S → T (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

*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.S → T: Source → Target .

our network can learn features that generalize and adapt well on the target dataset. The overall objective of our network can be formulated as:

$$Loss_{Total} = Loss_{CL} + \lambda Loss_{FD} + \gamma Loss_{CD} \tag{6}$$

where λ and γ are penalty parameters.

3 Experiments and Results

Our dataset as shown in Figure 2 includes 30,000 images from six UK locations (equal split), namely Abbeydale (Ab), Burton (Bu), Bourne (Bo), Listowel (Li),Windmill-Lane (Wi) and Ossett (Os), with two classes per location, i.e. OK (legible) vs NOT-OK (illegible). The combinations tested included all six locations and were conducted in the following manner: a) Transfer Learning, b) Single Source to Single Target, c) Combined Source to Single Target and d) Multi Source to Single Target (proposed method).

In the transfer learning setting, the labeled source and the unlabeled target images have been fed through the ResNet50 pre-trained on ImageNet which only uses the classification loss. In the Single-Source Domain Adaptation (SS-DA) setting, 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.

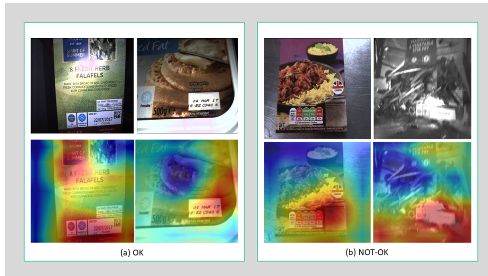

Fig. 2. (a) Acceptable Quality (OK) vs (b) Unacceptable Quality (NOT-OK) images. First row: raw images; second row: the images with class activation map [14].

Table 2. 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

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. We have combined and experimented using a) Two sources combined and b) Three sources combined.

In our proposed Multi-Source Domain Adaptation (MS-DA) 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 FD, CD and the classification losses using the techniques described in section 3. We performed various experiments per location as target domain with the results presented in Table 1 and Table 2. We categorized the experiments as a) Multi-Source with two datasets and b) Multi-Source with three datasets; our proposed approach significantly outperforms the baseline methods with an average classification accuracy improvement by more than 6%.

4 Conclusions

In this paper a multi-source domain 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. The results presented 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 other aspects of food packaging control, such as ingredients and allergen information.

References

1. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1-2), 151–175 (2010)
2. Crammer, K., Kearns, M., Wortman, J.: Learning from multiple sources. In: Advances in Neural Information Processing Systems. pp. 321–328 (2007)
3. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. arXiv preprint arXiv:1409.7495 (2014)
4. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W.: Deep reconstruction-classification networks for unsupervised domain adaptation. In: European Conference on Computer Vision. pp. 597–613. Springer (2016)
5. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. Journal of Machine Learning Research **13**(Mar), 723–773 (2012)
6. Jhuo, I.H., Liu, D., Lee, D., Chang, S.F.: Robust visual domain adaptation with low-rank reconstruction. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2168–2175. IEEE (2012)
7. Liu, H., Shao, M., Fu, Y.: Structure-preserved multi-source domain adaptation. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). pp. 1059–1064. IEEE (2016)
8. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. arXiv preprint arXiv:1502.02791 (2015)
9. Ribeiro, F.D.S., Caliva, F., Swainson, M., Gudmundsson, K., Leontidis, G., Kollias, S.: An adaptable deep learning system for optical character verification in retail food packaging. In: 2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). pp. 1–8. IEEE (2018)
10. Ribeiro, F.D.S., Gong, L., Calivá, F., Swainson, M., Gudmundsson, K., Yu, M., Leontidis, G., Ye, X., Kollias, S.: 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). pp. 2376–2380. IEEE (2018)
11. Suh, S., Lee, H., Lee, Y.O., Lukowicz, P., Hwang, J.: Robust shipping label recognition and validation for logistics by using deep neural networks. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 4509–4513. IEEE (2019)
12. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: European Conference on Computer Vision. pp. 443–450. Springer (2016)
13. Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L.: Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3964–3973 (2018)
14. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)
15. Zhu, Y., Zhuang, F., Wang, D.: Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5989–5996 (2019)