

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

**DEEP NEURAL NETWORKS FOR
MULTI-SOURCE TRANSFER LEARNING**

by

Keqiuyin Li

A THESIS SUBMITTED
IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2022

Certificate of Original Authorship

I, Keqiuyin Li, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 17 Oct, 2022

ABSTRACT

Transfer learning is gaining incredible attention due to its ability to leverage previously acquired knowledge from source domain to assist in completing a task in a similar target domain. Many existing transfer learning methods deal with single source and single target transfer learning, but rarely consider the fact that information from a single source can be inadequate to a target domain and there can be multiple source domains. Few multi-source domain adaptations methods adapt all source and target data into a same latent feature space. However, domain shifts can be found among source domains and between each pair of source and target domains, thus, the model fitting all domains well may not exist. In addition, most transfer learning methods assume that the source and target domains share the same label space. But in practice, the source domain(s) sharing the same label space with the target domain may never be found. Third, data privacy and security are being magnificently conspicuous in real-world applications, which means the traditional transfer learning relying on data matching can trigger privacy concerns.

To solve the above-mentioned problems, this thesis develops a series of methods to tackle transfer learning with multiple source domains. Knowledge transfer with and without source data are explored under both homogeneous and heterogeneous label space settings.

To tackle knowledge transfer from multiple source domains, and measure contributions of source domains, multi-source contribution learning and dynamic classifier alignment methods are developed. In multi-source contribution learning method, the similarities and diversities of domains are learned simultaneously by extracting multi-view features. One view represents common features (similarities) among all domains. Other views represent different characteristics (diversities) in a target domain, in which each characteristic is expressed by features extracted in a source domain. Then multi-level distribution matching is employed to improve the transferability of latent features, aiming to reduce misclassification of boundary samples

by maximizing discrepancy between different classes and minimizing discrepancy between the same classes. Concurrently, when completing a target task by combining source predictions, instead of averaging source predictions or weighting sources using normalized similarities, the original weights learned by normalizing similarities between source and target domains are adjusted using pseudo target labels to increase the disparities of weight values, which is desired to improve the performance of the final target predictor if the predictions of sources exist significant difference.

In dynamic classifier alignment method, it aligns classifiers driven from multi-view features via a sample-wise automatic way. As proposed, both the importance of each view and the contribution of each source domain are investigated. To determine the important degrees of multiple views, an importance learning function is built by generating an auxiliary classifier. To learn the source combination parameters, a domain discriminator is developed to estimate the probability of a sample belonging to multiple source domains. Meanwhile, a self-training strategy is proposed to enhance the cross-domain ability of source classifiers with the assistance of pseudo target labels.

To learn similarity of source and target domains to define what to transfer, sample and source distillation method is proposed. It develops a two-step selective strategy to distill source samples and define the importance of source domains. To distill samples, the pseudo-labeled target domain is constructed to learn a series of category classifiers to identify transfer and inefficient source samples. To rank domains, a domain discriminator, which returns the degrees of a target sample belonging to the source domains, is developed based on selected transfer samples. Using the selected samples and ranked domains, transfer between the source and target domains is achieved by adapting multi-level distribution in a latent feature space. Furthermore, to explore more usable target information which is expected to enhance the cross-domain ability of source predictors, an enhancement mechanism is built by matching selected pseudo-labeled and unlabeled target samples. The degrees learned by the domain discriminator are finally employed to combine source predictors when predicting the target task.

To address transfer learning without the access to source data, generally auxiliary model training method is explored. The proposed method fits the source models to the target domain via fine-tuning under the supervision of pseudo target labels rather than matching data distributions. To collect high-quality initial pseudo target labels, both specific and generally auxiliary source models are pre-trained to improve the generality across domains of source models based on auxiliary learning, where source contributions are determined using an automatic way. Besides, the generally auxiliary model can take the benefit of sharing knowledge from multiple source domains without sharing data. Going further, it introduces a class balanced coefficient of each category based on the number of samples to reduce the misclassification caused by data imbalance.

To deal with soft information in transfer learning, fuzzy rule-based deep neural network is proposed to achieve multi-source data-free transfer learning. It takes advantage of a fuzzy system to handle data uncertainty in domain adaptation without source data. To learn source private models with high generality, which is important to collect low noisy pseudo target labels, auxiliary tasks are designed by jointly training source models from multiple domains which share source parameters and fuzzy rules while protecting source data. To transfer fuzzy rules and fit source private parameters to the target domain, self-supervised learning and anchor-based alignment are built to force target data to source feature spaces.

To handle transfer learning where source and target domains have unshared label space, partial and open-set transfer learning with generally auxiliary model training and fuzzy rules are explored under source-free setting. Universal transfer learning method is developed under multi-source-absent setting. In partial source-free transfer learning, a selection method is built to remove source samples from unshared categories, which is expected to eliminate the negative transfer resulting from the source outliers. In open-set transfer learning, a threshold generated from the predictions of the pre-trained source models is defined to identify the unknown target samples, aiming to eliminate the pseudo label noise caused by introducing unshared target samples.

In universal transfer learning, a unified learning model is proposed. The proposed method designs a module that can transfer knowledge from multi-source domains with both homogeneous and heterogeneous label spaces in universal scenario without accessing the source data. To classify known target classes, source anchors are generated to build data-matching between source and target domains via a contrastive method. In addition, class center consistency is adopted to distinguish source private samples when pseudo-labeling the target data to reduce label noise. To detect unknown classes, a clustering strategy which combines global and source local entropy assumptions is adopted to recognize the known and unknown target samples. By removing source private classes and target unknown samples, highly confident target samples are collected to self-supervise the adaptation of the pre-trained source model. At the same time, constraints enlarging the distance among target known classes and between the known and unknown samples are applied based on the pseudo-labels to enhance the performance of the proposed model.

Dedication

To myself and my family.

Acknowledgements

I would like to express my deepest gratitude to my principal supervisor Distinguished Professor Jie Lu, and my co-supervisors A/Professor Guangquan Zhang and Dr. Hua Zuo, who generously provided guidance and feedback during the past three years. Without their expertise and guidance, I could not have made it through my PhD degree.

Distinguished Professor Jie Lu played a decisive role in guiding me to develop research work as an academic researcher. She placed unconditional support in my research interests and had great patience to enlighten me to overcome the difficulties I encountered. Her invaluable suggestions benefited me in study and life. As an international student living alone in a foreign country, I felt honoured enough to have her as my supervisor, who loves students like her own children and respects our feelings and ideas. What I have learned from her will be great treasure in my life. A/Professor Guangquan Zhang gave me constructive advice on how to start research efficiently. His wisdom insights and expertise avoided me from losing my way in research. Without his critical comments, I would waste much time at the beginning to advance the first step of the research topic. Dr. Hua Zuo raised helpful points in our discussion which helped me completing my research. Her comments helped a lot to enrich the content and improve the quality of all my papers. I extremely appreciate to them for their help and support.

I am also grateful to the University of Technology Sydney and Australian Research Council (grand under FL190100149), who provided financial support for my research. Thanks to that, I had got the opportunity to experience this exciting journey as a PhD student in University of Technology Sydney, which absolutely expanded my knowledge and widened my view.

I had the pleasure of working with all members of Decision System and e-Service Intelligent Lab and Australian Artificial Intelligence Institute. They created positive study and work environment to keep physical and mental health for everyone. I would like to acknowledge honorary member Dr. Dianshuang Wu, Dr Feng Liu and Dr Yiliao Song, who helped me in both life and research. Dr. Yi Zhang provided valuable comments on reference management which benefited my research. Dr. Bin Wang helped me in building software environment. I would also like to thank Dr. Junyu Xuan, Dr. Qian Zhang, Dr. Anjin Liu, Dr. Zhen Fang, Dr. Ruxia Liang, Dr. Xiaoya Che, Tianyu Liu, Kun Wang and Mengjia Wu who shared their opinions and comments. Many thanks to Jim Howes who gave beneficial feedback in English to improve my writing. I also thank Camila Cremonese, Robyn Barden and Lily Qian, who helped in managing proofreading.

Thanks should also go to my parents Mr. Shixuan Li, Mrs. Meirong Guo and my younger brother Mr. Kesen Li. Their belief in me has kept my spirits and motivation high. Lastly, I'd like to mention other family members Jixiang and Ruyi, the Chinese rural dogs, and Kumar, the border collie, for all the entertainment and emotional support.

Keqiuyin Li
Sydney, Australia, 2022.

List of Publications

Journal Papers

- J-1. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Multi-source contribution learning for domain adaptation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, 2022, pp 5293 - 5307.[A*; Q1]
- J-2. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Dynamic classifier alignment for unsupervised multi-source domain adaptation,” *IEEE Transactions on Knowledge and Data Engineering*, DOI: 10.1109/TKDE.2022.3144423, 2022. [A*; Q1]
- J-3. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Multi-domain adaptation with sample and source distillation,” *IEEE Transactions on Cybernetics*, DOI 10.1109/TCYB.2023.3236008. [A; Q1]
- J-4. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Source-free multi-domain adaptation with fuzzy rule based deep neural networks,” *IEEE Transactions on Fuzzy System*. Under review. [A*; Q1]
- J-5. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Unified Learning for Source-Absent Universal Multi-Domain Adaptation,” *IEEE Transactions on Neural Networks and Learning Systems*. Under review. [A*; Q1]

Conference Papers

- C-1. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Multi-source domain adaptation with distribution fusion and relationship extraction,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. Virtual online: IEEE, July 19 - 24 2020, DOI: 10.1109/IJCNN48605.2020.9207556. [A]

- C-2. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Multi-source domain adaptation with fuzzy-rule based deep neural networks,” *in Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Virtual Online: IEEE, July 11 - 14 2021, DOI: 10.1109/FUZZ45933.2021.9494586. [A]
- C-3. **Keqiuyin Li**, Jie Lu, Hua Zuo, and Guangquan Zhang, “Source-free multi-domain adaptation with generally auxiliary model training,” *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. Padova, Italy, July 18 - 23 2022, DOI: 10.1109/IJCNN55064.2022.9892718. [A]

Contents

Certificate	ii
Abstract	iii
Dedication	vii
Acknowledgments	viii
List of Publications	x
List of Figures	xix
List of Tables	xxv
List of Algorithms	xxxiii
1 Introduction	1
1.1 Background	1
1.2 Research questions, objectives and expected outcomes	6
1.2.1 Research questions	6
1.2.2 Research objectives	8
1.2.3 Expected outcomes	10
1.2.4 Research significance	11
1.3 Thesis Organization	12
2 Literature Review	17
2.1 Discrepancy Measurement of Transfer Learning	17
2.1.1 Distribution Matching Based Methods	18

2.1.2	Parameter Matching Based Methods	20
2.2	Categories of Transfer Learning Based on Number of Domains	21
2.2.1	Single Source and Single Target Transfer Learning	21
2.2.2	Multi-Source and Single Target Transfer Learning	23
2.2.3	Single Source and Multi-Target Transfer Learning	25
2.3	Categories of Transfer Learning Based on Feature Space	26
2.3.1	Transfer Learning with Homogeneous Feature Space	26
2.3.2	Transfer Learning with Heterogeneous Feature Space	27
2.4	Categories of Transfer Learning Based on Label Space	28
2.4.1	Closed Set Transfer Learning	29
2.4.2	Open-Set Transfer Learning	30
2.4.3	Partial Transfer Learning	31
2.4.4	Universal Transfer Learning	32
2.5	Source-Free Transfer Learning	34
2.6	Methods for Transfer Learning	35
2.6.1	Weakly-Supervised Learning Methods	35
2.6.2	Unsupervised Learning Methods	37
2.7	Summary	40
3	Multi-Source Contribution Learning for Domain Adap-	
	tation	41
3.1	Introduction	41
3.2	Problem Setting and Notations	43
3.3	The Proposed Multi-Source Contribution Learning Method	44
3.3.1	Multi-view Feature Extracting	46

3.3.2	Multi-level Distribution Adapting	48
3.3.3	Predictions Learning Based on Weight Adjustment	52
3.3.4	Predictions Learning Based on Fuzzy Rules	55
3.4	Experiments	60
3.4.1	Datasets and Baselines	60
3.4.2	Parameter Setting and Effect of Different Similarity Metrics	62
3.4.3	Results and Analysis Based on Weight Adjustment	67
3.4.4	Results and Analysis Based on Fuzzy Rules	78
3.4.5	Visualization Analysis of Proposed Method	82
3.4.6	Ablation Study and Sample Complexity	84
3.5	Summary	89
4	Dynamic Classifier Alignment for Unsupervised Multi-Source Domain Adaptation	91
4.1	Introduction	91
4.2	Problem Setting and Notations	93
4.3	The Proposed Dynamic Classifier Alignment	95
4.3.1	Feature Extraction	95
4.3.2	Classifier Alignment	98
4.3.3	Pseudo Label Selection and Target Task Completion	100
4.4	Experiments	106
4.4.1	Baselines and Datasets	106
4.4.2	Parameter Setting	108
4.4.3	Results and Analysis	112

4.4.4	Comparison of Single-view Classifier and Multi-view Aligned Classifier	117
4.4.5	Comparison of Feature Alignment and Classifier Alignment	119
4.4.6	Influence of Pseudo Labels	124
4.4.7	Ablation Study	129
4.4.8	Data Visualization	131
4.5	Summary	135
5	Multi-Source Domain Adaptation with Sample and Source Distillation	136
5.1	Introduction	136
5.2	Problem setting and Notations	138
5.3	The Proposed Method: Sample and Source Distillation	140
5.3.1	Source Model Training and Pseudo Target Label Collection	140
5.3.2	The Two-step Selective Strategy	143
5.3.3	Domain Adaptation and Target Task Completion	145
5.4	Experiments	149
5.4.1	Datasets and Baselines	149
5.4.2	Parameter Setting	151
5.4.3	Results and Analysis	152
5.4.4	Influence of the Combination Rule	154
5.4.5	Sample Complexity Analysis	162
5.4.6	Ablation study	166
5.4.7	Application of Sample and Source Distillation in Existing Methods	168

5.4.8	Visualization Analysis	169
5.5	Summary	172
6	Multi-Source-Free Domain Adaptation with Generally Auxiliary Model Training	173
6.1	Introduction	173
6.2	Problem Setting and Notations	175
6.3	The Proposed Generally Auxiliary Model Method under Closed Set . .	176
6.3.1	Specific Source Model Pre-training	176
6.3.2	Generally Auxiliary Model Training	178
6.3.3	Pseudo Label Collecting and Target Task Predicting	179
6.4	The Proposed Generally Auxiliary Model Method under Closed Set with Sample and Source Distillation	183
6.5	The Proposed Generally Auxiliary Model Method under Partial Set . .	186
6.6	The Proposed Generally Auxiliary Model Method under Open-Set . .	188
6.7	Experiments	190
6.7.1	Datasets and Parameter Setting	190
6.7.2	Comparison and Analysis under Closed Set	190
6.7.3	Influence of General Model under Closed Set	192
6.7.4	Influence of Class Balanced Coefficient under Closed Set	197
6.7.5	Influence of Distance Measurements in Deep Clustering	197
6.7.6	Data Visualization under Closed Set	199
6.7.7	Comparison and Analysis under Closed Set with Sample and Source Distillation	199
6.7.8	Data Visualization under Closed Set with Sample and Source Distillation	206

6.7.9	Comparison and Analysis under Partial Set	206
6.7.10	Data Visualization under Partial Set	208
6.7.11	Comparison and Analysis under Open-Set	208
6.7.12	Data Visualization under Open-Set	211
6.8	Summary	213
7	Source-Free Multi-Domain Adaptation with Fuzzy Rule-Based Deep Neural Networks	214
7.1	Introduction	214
7.2	Problem Setting and Notations	217
7.3	The Proposed Fuzzy Source-Free Multi-Domain Adaptation Method	217
7.3.1	Source Private Model Training	218
7.3.2	Pseudo Target Label Collection	223
7.3.3	Model Adaptation and Target Task Prediction	226
7.4	Experiments	229
7.4.1	Datasets and Baselines	229
7.4.2	Results and Analysis	234
7.4.3	Generality Analysis of Source Private Model	235
7.4.4	Influence of Rule Numbers	239
7.4.5	Ablation Study	242
7.4.6	Trade-off Parameter Sensitivity Analysis	243
7.4.7	Validation under Heterogeneous Label Space Setting	244
7.4.8	Visualization Analysis	245
7.5	Summary	250
8	Unified Learning for Multi-Source-Free Universal Do-	

main Adaptation	251
8.1 Introduction	251
8.2 Problem Setting and Notations	255
8.3 The Proposed Unified Learning Model for Universal Multi-Source-Free Domain Adaptation	255
8.3.1 Unified Learning Model Training	256
8.3.2 Source Data Generation	258
8.3.3 Target Model Adaptation	260
8.4 Experiments	267
8.4.1 Results and Analysis	271
8.4.2 Ablation Study	280
8.4.3 Influence of Source Category Discriminator	281
8.4.4 Visualization Analysis	282
8.5 Summary	287
9 Conclusion and Future Research	288
9.1 Conclusion	288
9.2 Future Research	291

List of Figures

1.1	Thesis organisation.	13
3.1	The whole framework of the proposed method. Shared features are collected using pre-trained networks. Common features represent similarities among all source and target domains, while diverse features represent diversities contained in the target domain which can be expressed by different source domains. Target k means features of target collected using k th source features extraction networks.	45
3.2	An example of characteristics contained in source and target domains. Common means similar information among all domains, view means the information which is similar between a source and the target but different from other source domains. View k means information of the k th source domain.	47
3.3	An example of boundary samples.	50
3.4	Training loss of the proposed method. Taking task Amazon in Office-31 as an example.	63
3.5	Test accuracy of the proposed method without and with using threshold. Taking task Amazon in Office-31 as an example.	64
3.6	Number of same labels returned by source predictors. Taking target domain Amazon as example for Office-31, Pascal VOC 2012 for ImageCLEF-DA, Art for Office-Home and Caltech for Office-Caltech10.	65

3.7	Test accuracy of the proposed method with different threshold values. Taking task Amazon in Office-31 as an example, the accuracy is average result of three times experiments. The red line represents accuracy while the light red area signifies standard deviation.	66
3.8	Classification accuracy without and with adjusting weights. Figures (a) and (c) are results without adjustment, while figures (b) and (d) are results with adjustment.	77
3.9	Source weights without and with adjusting. Figures (a) and (c) are results without adjustment, while figures (b) and (d) are results with adjustment.	79
3.10	T-SNE visualization of target with different source domain, take task DSLR in Office-31 as example.	84
3.11	T-SNE visualization of target with different source domain, take task Art in Office-Home as example.	85
3.12	T-SNE visualization of features before and after adaptation, taking task DSLR in Office-31 as example for two-source domain adaptation.	86
3.13	T-SNE visualization of features before and after adaptation, taking task Art in Office-Home as example for three-source domain adaptation.	87
4.1	The whole framework of the proposed DCA. Shared features are extracted to learn a domain discriminator which defines the contributions of source domains when predicting target task. Specific multi-view features are extracted to adapt domain distributions and train classifiers. Auxiliary classifier in each source domain is generated to assist in aligning multi-view classifiers. Target samples are pseudo-labeled by the learned classifiers to supervise the further training.	96

4.2	The procedure of selecting pseudo labels. Solid arrows indicate training under the supervision of source labels and target pseudo labels. Dashed arrows indicate the prediction and selection of target samples. At the beginning, classifiers are supervised by source labels only. Unlabeled target samples are fed into the learned classifiers to be divided into easy-to-predict and hard-to-predict groups. Predicted target labels with high probability are selected to supervise further training of the classifiers.	102
4.3	Accuracy (%) on target domains A from Office-31 and C from Office-Home as the training progresses. Red line means training without adjusting learning rate when adding pseudo labels, blue line indicates that with adjusting learning rate.	111
4.4	Combination parameters of target domains A from Office-31 and A from Office-Home.	118
4.5	T-SNE visualization of multi-view features in target domain.	134
5.1	Example of inefficient source samples and domains.	137

5.2	The whole framework of the proposed method. Figure (a) indicates target label collection and source sample distillation. Given pre-trained source models comprised of shared backbone, specific feature extractors and predictors, pseudo target labels are collected first. The selected target labels are then employed to train the category classifier which distills inefficient source samples. Figure (b) indicates domain adaptation and dominant source domain selection. A domain discriminator is first built based on the distilled source domains to identify the dominant source domain and learn the relationships between source and target domains. Simultaneously, pre-trained source models are fine-tuned based on the distilled source domains by minimizing the discrepancy between the source and target distributions. Self-supervised training is also adopted based on the selected target labels to enhance the cross-domain ability of source models and guarantee the performance of transfer.	141
5.3	Quantities of the target samples belonging to the source domains. Red line indicates the threshold of the dominant source domain for each target task. Source order is described in section 5.4.1.	160
5.4	Accuracy (%) on dataset ImageCLEF-DA with different quantities of the transfer source samples.	166
5.5	Accuracy (%) on dataset OfficeHome with different quantities of the transfer source samples.	167
5.6	T-SNE visualization of target domain <i>A</i> from dataset Office-31.	170
5.7	T-SNE visualization of target domain <i>A</i> from OfficeHome.	171
6.1	The procedure of the proposed method.	177
6.2	T-SNE visualization of Office-31 under closed set.	200
6.3	T-SNE visualization of Office-Home under closed set.	201

6.4	T-SNE visualization with sample and source distillation.	207
6.5	T-SNE visualization under partial domain adaptation.	209
6.6	T-SNE visualization under open-set domain adaptation.	212
7.1	The procedure of the proposed method. The solid arrow means data-flow, the dashed arrow means loss computing. Figure (a) indicates source model training. Auxiliary tasks are constructed by sharing source parameters. Fuzzy C-means clustering is employed to learn the prototypes and memberships to build the fuzzy rules. Source anchors are extracted to describe the source class information without referring to the original data. Figure (b) demonstrates domain adaptation. By freezing source rules, self-supervised learning is employed to fine-tune feature extractors. Anchor-based alignment is built to match domains on the data-level by extracting invariant information. Deep clustering and a sample selection strategy are designed to predict pseudo labels with low noise which learn specific target information.	219
7.2	T-SNE visualization on target domain A from dataset Office-31. . . .	247
7.3	T-SNE visualization on target domain A from dataset Office-Home. . .	249
8.1	Universal domain adaptation with single and multiple source domains.	253
8.2	The procedure of the proposed method.	257
8.3	Settings of universal multi-source-free domain adaptation.	268
8.4	T-SNE visualization on target domain RealWorld from dataset OfficeHome under Set I.	284
8.5	T-SNE visualization on target domain RealWorld from dataset OfficeHome under Set II.	285

8.6 T-SNE visualization on target domain RealWorld from dataset OfficeHome under Set III.	286
--	-----

List of Tables

3.1	Accuracy (%) with different similarity metrics.	67
3.2	Comparison of classification accuracy (%) on dataset Office-31	68
3.3	Comparison of classification accuracy (%) on dataset ImageCLEF-DA	69
3.4	Comparison of classification accuracy (%) on dataset Office-Home . .	70
3.5	Comparison of classification accuracy (%) on dataset Office-Caltech10	71
3.6	Comparison of classification accuracy (%) on dataset Office-31 with different combination rules	73
3.7	Comparison of classification accuracy (%) on dataset ImageCLEF-DA with different combination rules	74
3.8	Comparison of classification accuracy (%) on dataset Office-Home with different combination rules	75
3.9	Comparison of classification accuracy (%) on dataset Office-Caltech10 with different combination rules	76
3.10	Comparison of classification accuracy (%) on dataset ImageCLEF-DA using fuzzy rules	80
3.11	Comparison of classification accuracy (%) on dataset Office-31 using fuzzy rules	81
3.12	Comparison of classification accuracy (%) on dataset ImageCLEF-DA without and with fuzzy rules.	82
3.13	Comparison of classification accuracy (%) on dataset Office-31 without and with fuzzy rules.	83

3.14	Accuracy (%) without different loss components	88
3.15	Accuracy (%) with different training sample size.	89
4.1	Notations and descriptions.	94
4.2	Accuracy (%) on dataset Office-31 with and without adjusting the learning rate when adding pseudo labels.	110
4.3	Accuracy (%) on dataset Office-Home with and without adjusting the learning rate when adding pseudo labels.	110
4.4	Training iterations on dataset Office-31 with and without adjusting the learning rate when adding pseudo labels.	111
4.5	Training iterations on dataset Office-Home with and without adjusting the learning rate when adding pseudo labels.	111
4.6	Accuracy (%) of the proposed and comparison methods on dataset Office-31.	113
4.7	Accuracy (%) of the proposed and comparison methods on dataset ImageCLEF-DA.	114
4.8	Accuracy (%) of the proposed and comparison methods on dataset Office-Caltech10.	115
4.9	Accuracy (%) of the proposed and comparison methods on dataset Office-Home.	116
4.10	Accuracy (%) of different alignment parameter calculation strategies on dataset Office-31	118
4.11	Accuracy (%) of classifier alignment and classifier trained on single view representations on dataset Office-31.	120
4.12	Accuracy (%) of classifier alignment and classifier with single view representations on dataset ImageCLEF-DA.	121

4.13 Accuracy (%) of classifier alignment and classifier with single view representations on dataset Office-Caltech10.	122
4.14 Accuracy (%) of classifier alignment and classifier with single view representations on dataset Office-Home.	123
4.15 Accuracy (%) of classifier alignment and feature concatenation on dataset Office-31.	125
4.16 Accuracy (%) of classifier alignment and feature concatenation on dataset ImageCLEF-DA.	125
4.17 Accuracy (%) of classifier alignment and feature concatenation on dataset Office-Caltech10.	126
4.18 Accuracy (%) of classifier alignment and feature concatenation on dataset Office-Home.	126
4.19 Accuracy (%) of adding pseudo labels at different epochs on dataset Office-31. Adding step is set as $q = 500$	127
4.20 Accuracy (%) of adding pseudo labels at different epochs on dataset Office-Home. Adding step is set as $q = 1000$	127
4.21 Accuracy (%) of adding pseudo labels using different steps on dataset Office-31. Start iteration is set as $E = 2000$	128
4.22 Accuracy (%) of adding pseudo labels using different steps on dataset Office-Home. Start iteration is set as $E = 2000$	128
4.23 Accuracy (%) of different constraints on dataset Office-31.	130
4.24 Accuracy (%) of different constraints on dataset OfficeHome.	130
4.25 MMD scores (10^{-3}) of different constraints on dataset Office-31.	132
4.26 MMD scores (10^{-3}) of different constraints on dataset Office-Home.	133
5.1 Notations and descriptions.	139

5.2	Accuracy (%) on dataset Office31 of the proposed and comparison methods	155
5.3	Accuracy (%) on dataset ImageCLEF-DA of the proposed and comparison methods	156
5.4	Accuracy (%) on dataset Office-Caltech10 of the proposed and comparison methods	157
5.5	Accuracy (%) on dataset OfficeHome of the proposed and comparison methods	158
5.6	Accuracy (%) on dataset DomainNet of the proposed and comparison methods	159
5.7	Accuracy (%) on dataset Office31 with different combination rules	162
5.8	Accuracy (%) on dataset ImageCLEF-DA with different combination rules	162
5.9	Accuracy (%) on dataset Office-Caltech10 with different combination rules	163
5.10	Accuracy (%) on dataset OfficeHome with different combination rules	163
5.11	Accuracy (%) on dataset DomainNet with different combination rules. Result with symbol * is the selected source domain.	164
5.12	Accuracy (%) on dataset ImageCLEF-DA with different quantities of the transfer source samples	165
5.13	Accuracy (%) on dataset OfficeHome with different quantities of the transfer source samples	165
5.14	Accuracy (%) of the ablation study on dataset ImageCLEF-DA	168
5.15	Accuracy (%) of the ablation study on dataset OfficeHome	169
5.16	Accuracy (%) on dataset OfficeHome of existing methods with transfer sample selection	170

6.1	Notations and descriptions.	175
6.2	Classification accuracy (%) of the proposed generally auxiliary model and the compared methods on Office-31.	193
6.3	Classification accuracy (%) of the proposed generally auxiliary model and the compared methods on Office-Home.	194
6.4	Source-only classification accuracy (%) of specific and general models on Office-31.	195
6.5	Source-only classification accuracy (%) of specific and general models on Office-Home.	195
6.6	Classification accuracy (%) of specific and general models on Office-31 after fitting.	196
6.7	Classification accuracy (%) of specific and general models on Office-Home after fitting.	196
6.8	Classification accuracy (%) of the general model with and without class balanced coefficient. “Bef” means source only model, “Aft” is adapted model.	197
6.9	Classification accuracy (%) of the specific source model 1 with and without class balanced coefficient. “Bef” means source only model, “Aft” is adapted model.	198
6.10	Classification accuracy (%) of the specific source model 2 with and without class balanced coefficient. “Bef” means source only model, “Aft” is adapted model.	198
6.11	Classification accuracy (%) of specif and general models on Office-Home with different distance measurements.	199
6.12	Accuracy (%) on dataset Office31 of the proposed and comparison methods under source-free domain adaptation.	204
6.13	Accuracy (%) on dataset Office-Caltech10 of the proposed and comparison methods under source-free domain adaptation.	205

6.14	Accuracy (%) on dataset Office31 of the source only model with and with out distillation.	206
6.15	Comparison of classification accuracy (%) on Office-Home under partial domain adaptation.	210
6.16	Comparison of classification accuracy (%) on Office-Home under open-set domain adaptation.	211
7.1	Symbol descriptions.	217
7.2	Classes, domains and samples in experiment datasets.	230
7.3	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset Office-31	236
7.4	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset ImageCLEF-DA	237
7.5	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset Office-Caltech10	237
7.6	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset Office-Home	238
7.7	Accuracy (%) on dataset Office-31 of source only models.	240
7.8	Accuracy (%) on dataset Office-Home of source only models.	241
7.9	Accuracy (%) on dataset Office-31 of ablation study.	241
7.10	Accuracy (%) on dataset Office-Home of ablation study.	242
7.11	Setting of ablation study.	243
7.12	Accuracy (%) on dataset Office-31 of ablation study.	243
7.13	Accuracy (%) on dataset Office-Home of ablation study.	244
7.14	Accuray (%) on dataset Office-31 with different values of parameter λ	244

7.15	Accuracy (%) on dataset Office-Home with different values of parameter λ .	245
7.16	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset Office-Home under partial domain adaptation.	246
7.17	Comparison (%) of the proposed fuzzy rule-based deep network and the baselines on dataset Office-Home under open-set domain adaptation.	246
8.1	Notations and descriptions.	255
8.2	Label set division under three settings. In source domain, the division is listed as source shared/private classes. In target domain, the division is listed as known/unknown classes.	269
8.3	HM (%) on dataset DomainNet of the ULMSFDA and baselines under Set I	273
8.4	HM (%) on dataset OfficeHome of the ULMSFDA and baselines under Set I	274
8.5	HM (%) on dataset Office31 of the ULMSFDA and baselines under Set I	275
8.6	Accuracy (%) on dataset OfficeHome of the ULMSFDA and baselines under Set I	276
8.7	Accuracy (%) on dataset Office31 of the ULMSFDA and baselines under Set I	277
8.8	HM (%) on dataset OfficeHome of the ULMSFDA and baselines under Set II	278
8.9	HM (%) on dataset Office31 of the ULMSFDA and baselines under Set II	278
8.10	HM (%) on dataset OfficeHome of the ULMSFDA and baselines under Set III	279

8.11 HM (%) on dataset Office31 of the ULMSFDA and baselines under Set III	279
8.12 Setting of ablation study	280
8.13 Ablation study (HM (%)) on dataset OfficeHome under Set I	281
8.14 Ablation study (HM (%)) on dataset OfficeHome under Set II	281
8.15 Ablation study (HM (%)) on dataset OfficeHome under Set III	282
8.16 HM (%) on dataset OfficeHome with and without source category discriminator	283

List of Algorithms

1	Weight adjustment-based multi-source domain adaptation	56
2	Fuzzy rule-based multi-source domain adaptation	59
3	Dynamic classifier alignment for multi-source domain adaptation . . .	105
4	Transfer sample selection	147
5	Target task prediction	148
6	Multi-source-free domain adaptation: Pre-training	182
7	Multi-source-free domain adaptation: Adapting	182
8	Multi-source-free domain adaptation with source and sample distillation.	185
9	Multi-source-free domain adaptation under partial set.	187
10	Multi-source-free domain adaptation under open-set.	189
11	SF-FDN: Source private model training.	228
12	SF-FDN: Target model adaptation.	228
13	ULMSFDA: Source model training.	265
14	ULMSFDA: Target model adaptation.	266