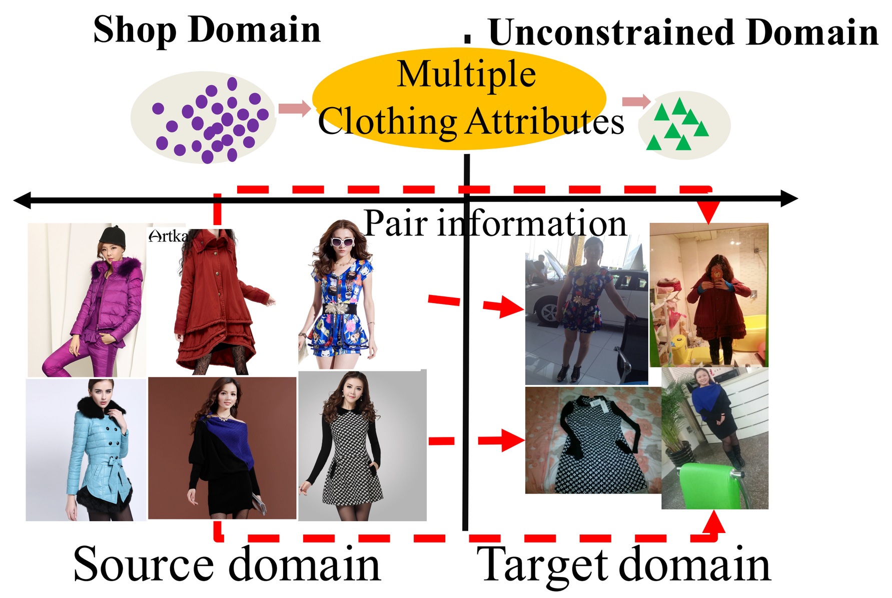
Multi-Task Curriculum Transfer Deep Learning of Clothing Attributes

Queen Mary University of London

Qi Dong q.dong@qmul.ac.uk
Shaogang Gong s.gong@qmul.ac.uk
Xiatian Zhu xiatian.zhu@qmul.ac.uk

WACV 2017
IEEE 2017 Winter conference on Applications of Computer Vision

1 Introduction



➢ **Problem**

Domain transfer learning for recognising fine-grained multi-label clothing attributes in the street (wild) given limited training data.

➢ **Limitation of Existing Methods**

- Hand crafted features.
- Single task deep learning for multi-label recognition.
- Lack of end-to-end cross domain transfer learning.

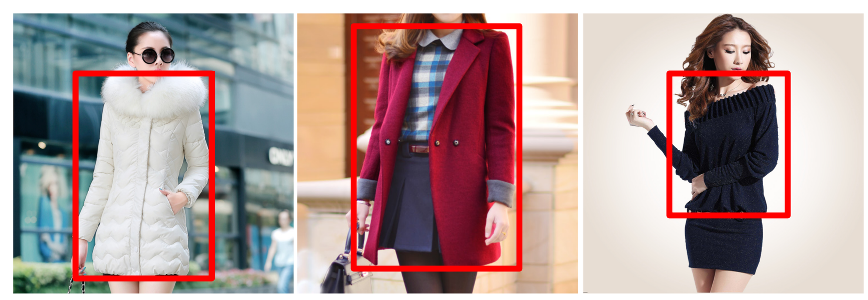
➢ **Contributions**

- Novel Multi-Task Curriculum Transfer (MTCT) deep learning strategy.
- Effective Multi-Task Network (MTN) for learning from sparse target data.

2 Overview of method

➢ **Clothing detection**

Faster R-CNN[4] for clothing detection



➢ **Stage1: Shop domain (clean)**

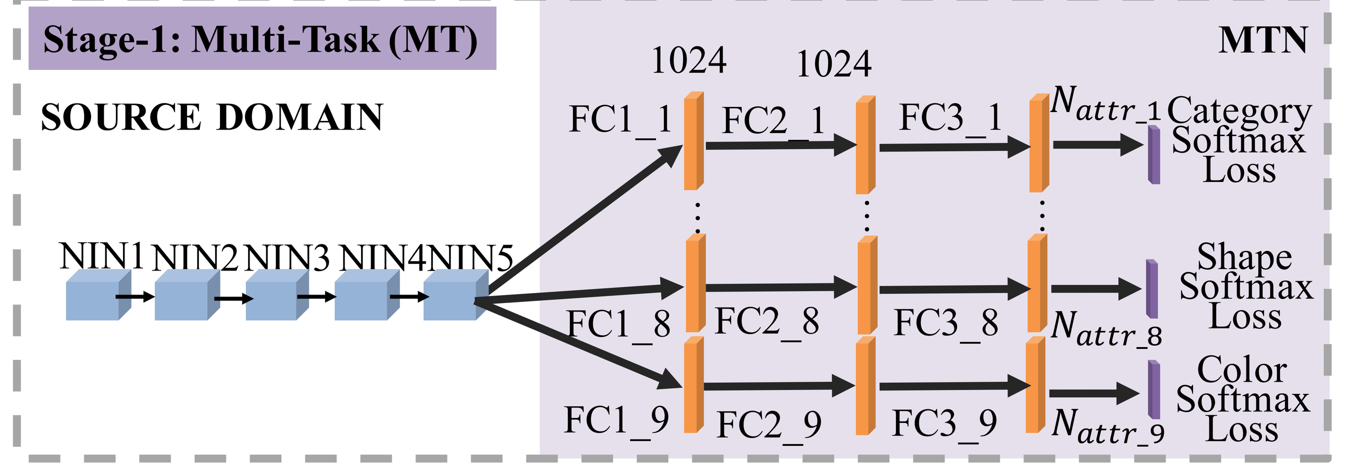
Pretrain MTN on ImageNet and train on shop images.

➢ **Stage 2: Street domain (wild)**

Initialize 3MTN by shop domain images trained model and then fine-tune FC layers using cross-domain triplet information for transfer learning

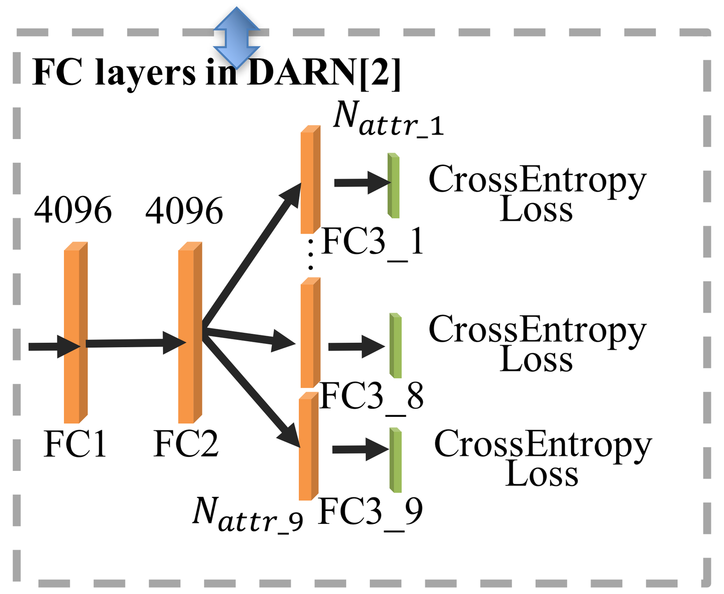
3 Multi-task deep learning

SHOP DOMAIN



➢ **MTN**

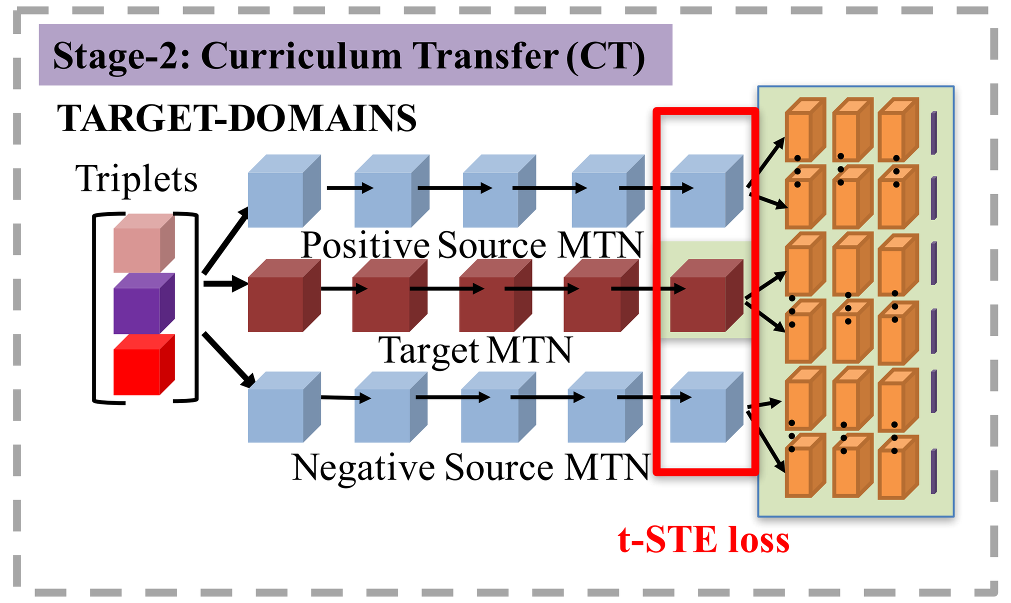
A three-layer branch for learning specifics of each attribute category, with shared learning of generic features in conv layers.



	Layer Name	Parameters
NIN1	Conv1	7x7-96-2
	Conv1_1	1x1-96-1
	Conv1_2	1x1-96-1
	Maxpooling-3x3	
NIN2	Conv2	5x5-256-2
	Conv2_1	1x1-256-1
	Conv2_2	1x1-256-1
	Maxpooling-3x3	
NIN3	Conv3	3x3-512-1
	Conv3_1	1x1-512-1
	Conv3_2	1x1-512-1
NIN4	Conv4	3x3-1024-1
	Conv4_1	1x1-1024-1
	Conv4_2	1x1-512-1
	Conv4_3	1x1-384-1
NIN5	Conv5	3x3-512-2
	Conv5_1	1x1-512-1
	Conv5_2	1x1-512-1
	Maxpooling-3x3	
FC-1024 for any branch		
FC-1024 for any branch		
FC- N_{attr_i} for any branch		
Softmax for any branch		

4 Curriculum transfer learning

STREET DOMAIN



➢ **Cross-domain loss function**

$$l_{\text{t-STE}} = \sum_{\{\boldsymbol{I}_t, \boldsymbol{I}_{ps}, \boldsymbol{I}_{ns}\} \in T} \log \frac{(1 + \frac{\|f_t(\boldsymbol{I}_t) - f_s(\boldsymbol{I}_{ps})\|^2}{\alpha})^{\beta}}{(1 + \frac{\|f_t(\boldsymbol{I}_t) - f_s(\boldsymbol{I}_{ps})\|^2}{\alpha})^{\beta} + (1 + \frac{\|f_t(\boldsymbol{I}_t) - f_s(\boldsymbol{I}_{ns})\|^2}{\alpha})^{\beta}}$$

[1]Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan. Deep domain adaptation for describing people based on fine-grained clothing attributes. CVPR2015.
[2]Huang, R. S. Feris, Q. Chen, and S. Yan. Cross-domain image retrieval with a dual attribute-aware ranking network. ICCV2015.
[3]Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. CVPR2016.
[4] S.Ren, et.al.Fasterr-cnn:Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems, pages 91–99, 2015.

5 Experiments

➢ **Comparison to the State-of-The-Arts**

Methods	Category	Button	Colour	Length	Pattern	Shape	Collar	Slv-Len	Slv-shp	mAP^{cls}	mP^{ins}	mR^{ins}
DDAN[1]	12.56	24.13	20.72	35.91	61.67	47.14	31.17	80.63	73.96	43.10	45.41	52.20
DARN[2]	52.55	37.48	58.24	51.49	67.53	47.70	47.77	82.04	73.72	57.61	57.79	67.29
FashionNet[3]	55.85	39.52	60.33	53.08	68.65	49.79	51.27	83.79	75.34	59.84	59.97	69.74
MTCT	**65.96**	**43.57**	**66.86**	**58.27**	**70.55**	**51.40**	**58.79**	**86.05**	**77.54**	**64.35**	**64.97**	**75.66**

➢ **MTN and Transfer learning**

Method	mAP^{cls}	mP^{ins}	mR^{ins}
JAN(No Adpt)	50.46	50.39	58.40
MTN(No Adpt)	51.38	51.82	60.00
MTN(UD)	58.76	60.16	70.00
MTN(FTT)	61.82	62.53	72.76
MTCT	**64.35**	**64.97**	**75.66**

*No Adpt: without adaptation
*UD: United domains *FTT: Finetuning on the target domain

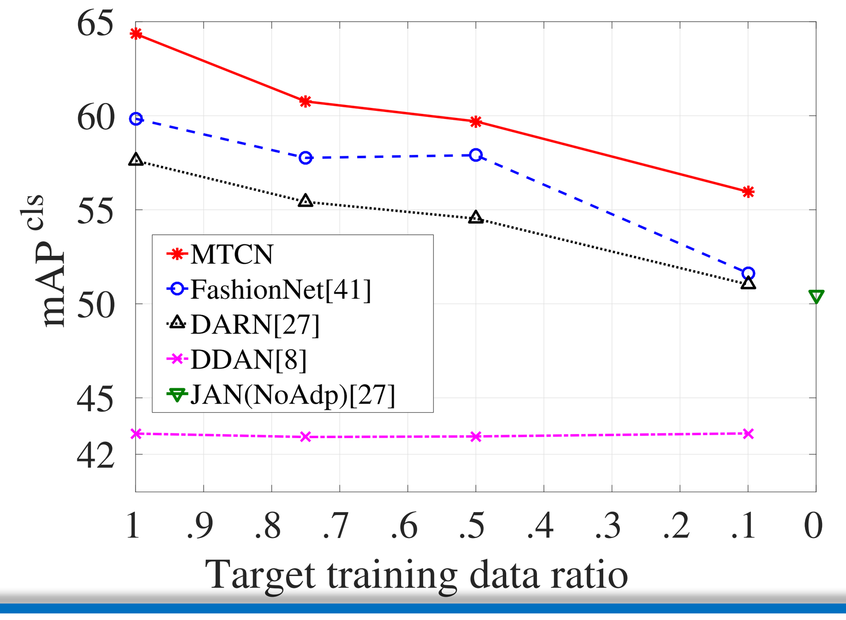
➢ **End-to-End vs. Curriculum Transfer learning**

Method	mAP^{cls}	mP^{ins}	mR^{ins}
End-to-End	62.30	63.00	73.37
Curriculum	**64.35**	**64.97**	**75.66**

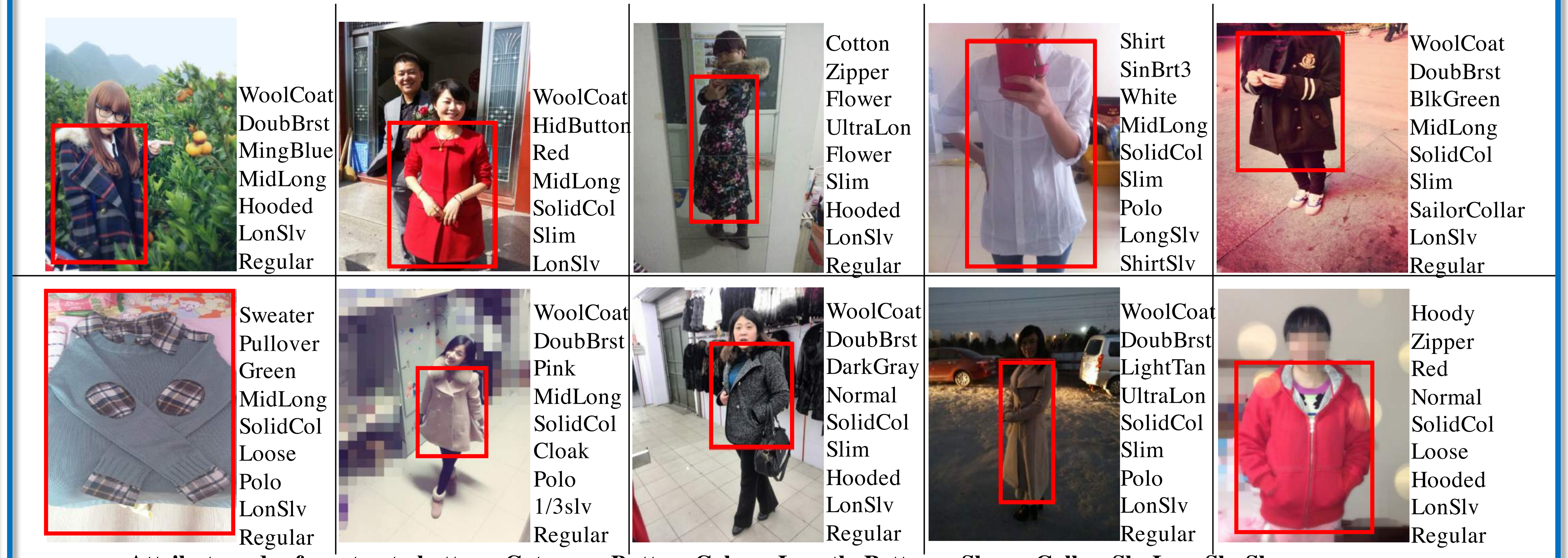
➢ **Different cross-domain loss functions**

Method	mAP^{cls}	mP^{ins}	mR^{ins}
TripletRanking	62.60	63.45	73.83
t-STE	**64.35**	**64.97**	**75.66**

➢ **Model Robustness vs. target data size**



6 A qualitative evaluation of MTCT



Attribute order from top to bottom: Category, Button, Colour, Length, Pattern, Shape, Collar, Slv-Len, Slv-Shape