Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation

M. Ghifary[1,2], W. Bastiaan Kleijn[2], M. Zhang[2], D. Balduzzi[2], W. Li[3]

mghifary@gmail.com, {bastiaan.kleijn,mengjie.zhang}@ecs.vuw.ac.nz,
david.balduzzi@vuw.ac.nz, liwen@vision.ee.ethz.ch

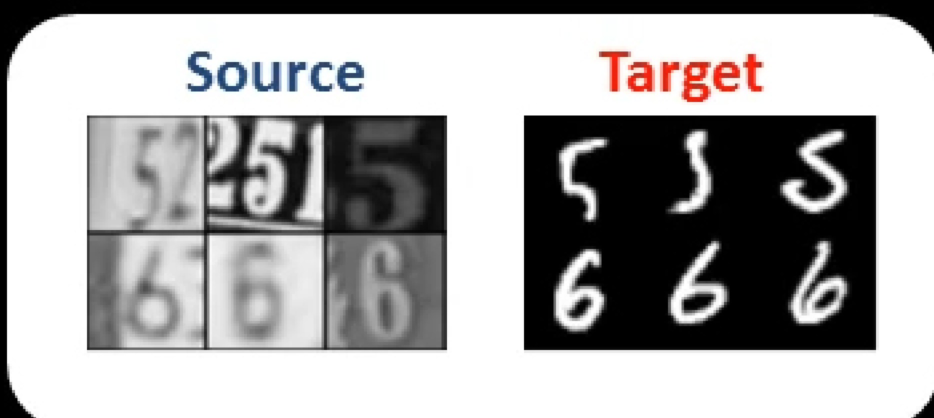
1 weta DIGITAL

2 TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI VICTORIA UNIVERSITY OF WELLINGTON

3 ETH Zürich

Unsupervised Domain Adaptation

- Dataset bias [TOR'2011]
- Domain adaptation (DA): solve dataset bias between source and target domains
- *Unsupervised* domain adaptation (uDA): DA without labeled *target* data
- Deep learning has played an important role [DON'14, GAN'15], but still awaiting more effective approaches to come



[TOR'11] A. Torralba and A. Efros. Unbiased look at dataset bias. In *CVPR*, 2011.
[DON'14] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell. DeCAF: a deep convolutional activation feature for generic visual recognition. In *ICML*, 2014.
[GAN'15] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.

Main Contribution

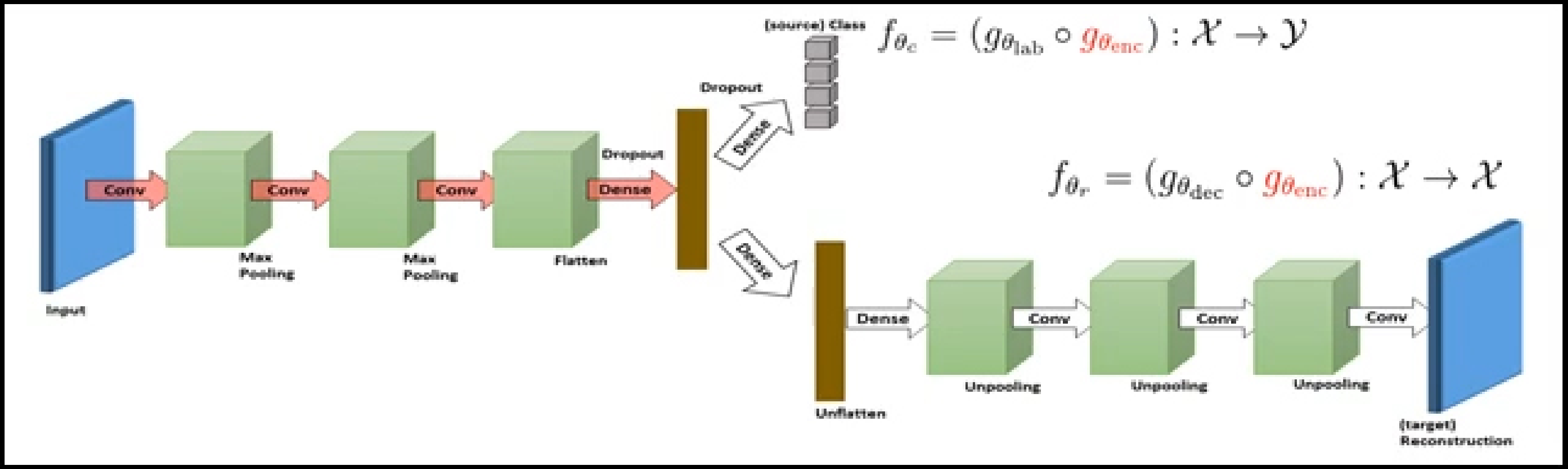
- Deep multi-task learning method for uDA: Deep Reconstruction-Classification Networks (DRCN)
- Motivated from semi-supervised learning hypothesis
  - *When labels are lacking, modelling both the labels, Q(y|x), and the structure of the data, Q(x), would produce a better labelling function*
- In uDA, we could model Q(x) from the target domain , but not Q(y|x) since no labelled samples -- need to borrow labels from somewhere else (source domain)
- Our proposed method jointly learns two functions with a shared encoding representation:

1. (*Source*) labelling function $f_{\theta_c} = (g_{\theta_{\mathrm{lab}}} \circ g_{\theta_{\mathrm{enc}}}) : \mathcal{X} \rightarrow \mathcal{Y}$

2. (*Target*) reconstruction function $f_{\theta_r} = (g_{\theta_{\mathrm{dec}}} \circ g_{\theta_{\mathrm{enc}}}) : \mathcal{X} \rightarrow \mathcal{X}$

Deep Reconstruction-Classification Networks (DRCN)

- Architecture



- Algorithm
  - Define two empirical losses:

$$\mathcal{L}^s(\theta_c) = \frac{1}{n_s}\sum_{i=1}^{n_s} \ell_c(f_{\theta_c}(x_i^s), y_i^s)$$

$$\mathcal{L}^t(\theta_r) = \frac{1}{n_t}\sum_{i=1}^{n_t} \ell_r(f_{\theta_r}(x_i^t), x_i^t)$$

  - Solve

$$\{\hat{\theta}_c, \hat{\theta}_r\} := \arg\min_{\theta_c,\theta_r} \lambda\mathcal{L}^s(\theta_c) + (1-\lambda)\mathcal{L}^t(\theta_r)$$

  - Optimized by stochastic gradient descent

Results (1)

- Cross-domain object classification accuracy
  - Benchmarks: MNIST (MN), USPS (US), SVHN (SV), STL (ST), CIFAR (CI)
  - ~8% improvement on SVHN -> MNIST over the prior state-of-the-art

Methods	MN → US	US → MN	SV → MN	MN → SV	ST → CI	CI → ST
ConvNet$_{src}$	85.55 ± 0.12	65.77 ± 0.06	62.33 ± 0.09	25.95 ± 0.04	54.17 ± 0.21	63.61 ± 0.17
SDA$_{sh}$ [32]	43.14 ± 0.16	37.30 ± 0.12	55.15 ± 0.08	8.23 ± 0.11	35.82 ± 0.07	42.27 ± 0.12
SA [27]	85.89 ± 0.13	51.54 ± 0.06	63.17 ± 0.07	28.52 ± 0.10	54.04 ± 0.19	62.88 ± 0.15
SCAE [44]	85.78 ± 0.08	63.11 ± 0.04	60.02 ± 0.16	27.12 ± 0.08	54.25 ± 0.13	62.18 ± 0.04
SCAE$_t$ [44]	86.24 ± 0.11	65.37 ± 0.03	65.57 ± 0.09	27.57 ± 0.13	54.68 ± 0.08	61.94 ± 0.06
ReverseGrad [18]	91.11 ± 0.07	**74.01 ± 0.05**	73.91 ± 0.07	35.67 ± 0.04	56.91 ± 0.05	66.12 ± 0.08
DRCN	**91.80 ± 0.09**	73.67 ± 0.04	**81.97 ± 0.16**	**40.05 ± 0.07**	**58.86 ± 0.07**	**66.37 ± 0.10**
ConvNet$_{tgt}$	96.12 ± 0.07	98.67 ± 0.04	98.67 ± 0.04	91.52 ± 0.05	78.81 ± 0.11	66.50 ± 0.07

- Comparison with other DRCN versions: DRCN$_s$ and DRCN$_{st}$
  - DRCN$_s$ and DRCN$_{st}$ underperform the original DRCN – unlabeled *source* data do not help
  - Partially explained by our theoretical analysis in connection to SSL

Methods	MN → US	US → MN	SV → MN	MN → SV	ST → CI	CI → ST
DRCN$_s$	89.92 ± 0.12	65.96 ± 0.07	73.66 ± 0.04	34.29 ± 0.09	55.12 ± 0.12	63.02 ± 0.06
DRCN$_{st}$	91.15 ± 0.05	68.64 ± 0.05	75.88 ± 0.09	37.77 ± 0.06	55.26 ± 0.06	64.55 ± 0.13
DRCN	**91.80 ± 0.09**	**73.67 ± 0.04**	**81.97 ± 0.16**	**40.05 ± 0.07**	**58.86 ± 0.07**	**66.37 ± 0.10**

Results (2)

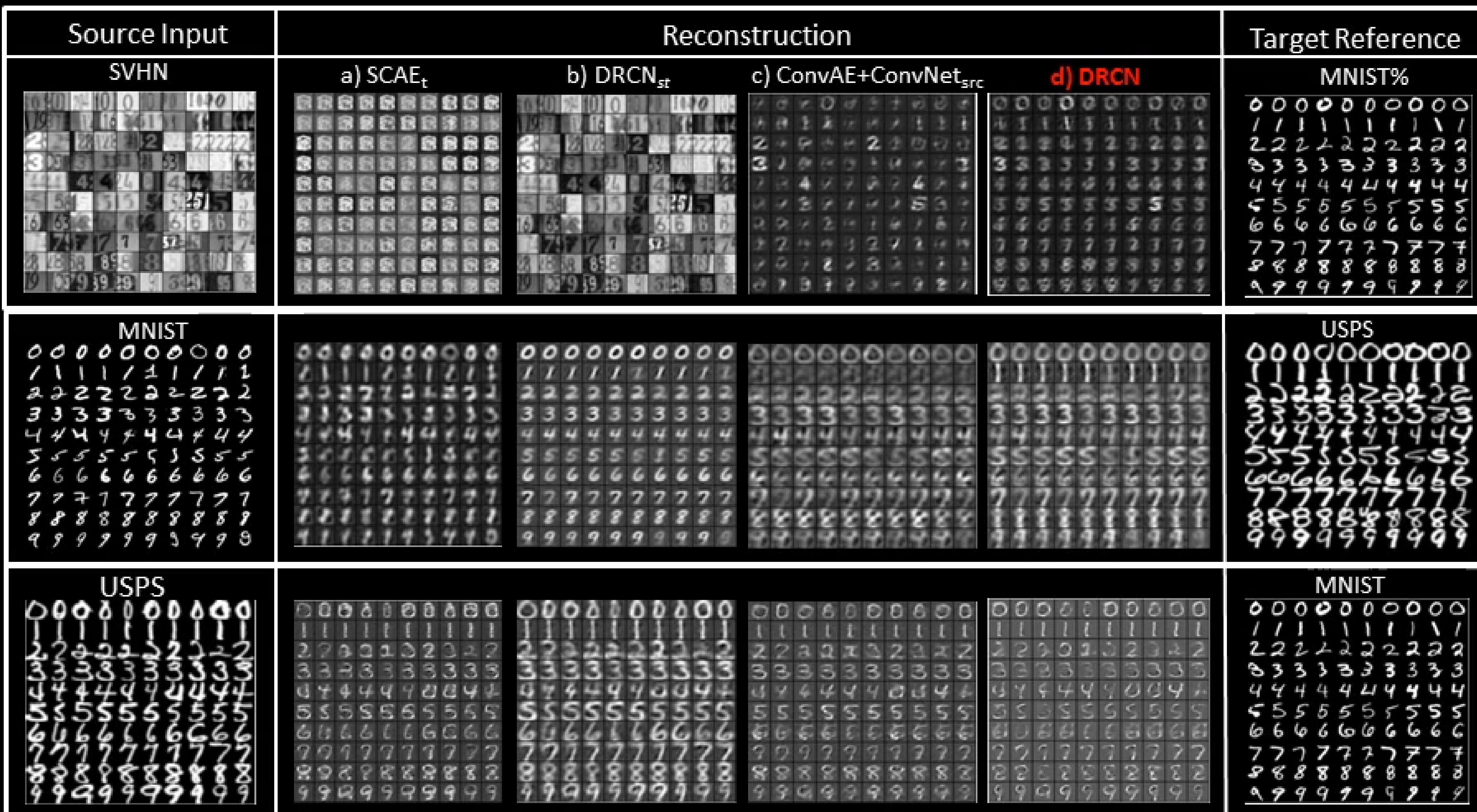
- The Office dataset [SAE'10] experiments
  - Use AlexNet
  - Competitive performance
  - Best when the target is Amazon, which has the most number of unlabelled data among others

Method	A → W	W → A	A → D	D → A	W → D	D → W
DDC [41]	61.8 ± 0.4	52.2 ± 0.4	64.4 ± 0.3	52.1 ± 0.8	98.5 ± 0.4	95.0 ± 0.5
DAN [40]	68.5 ± 0.4	53.1 ± 0.3	67.0 ± 0.4	54.0 ± 0.4	99.0 ± 0.2	96.0 ± 0.3
ReverseGrad [18]	**72.6** ± 0.3	52.7 ± 0.2	**67.1** ± 0.3	54.5 ± 0.4	**99.2** ± 0.3	**96.4** ± 0.1
DRCN	68.7 ± 0.3	**54.9** ± 0.5	66.8 ± 0.5	**56.0** ± 0.5	99.0 ± 0.2	**96.4** ± 0.3

[SAE'10] K. Saenko, B. Kulis, M. Friz, and T. Darrell. Adapting visual category models to new domains. In *ECCV*, 2010.

Results (3)

- Insights from data reconstruction $f_{\hat{\theta}_r}(x^t)$
  - Reconstructed source images from DRCN resemble the appearance of the target images



Summary

1. Propose a deep learning approach for unsupervised domain adaptation: Deep Reconstruction Classification Networks (DRCN) that jointly learns a shared representation for two tasks: i) ***source* output classification**, ii) ***target* input reconstruction**
2. DRCN achieves state-of-the-art performance over a range of cross-domain object classification benchmarks
3. Reconstructed source images from DRCN resemble the appearance of target images
4. DRCN's learning objective is closely related to a semi-supervised learning framework, which leads to the soundness of the approach