Augmented Cyclic Adversarial Learning For Low Resource Domain Adaptation

Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, Richard Socher

salesforce RESEARCH

Overview

Motivation

- Training a model to perform a task typically requires a large amount of data from the domains in which the task will be applied
- Recent domain adaptation techniques, especially based on cyclic adversarial learning, deal with the challenge of adapting a model trained from a data-rich source domain to perform well in a **high-resource** unlabeled target domain.
- The conventional approach of enforcing cycle-consistency via reconstruction (CycleGAN [1]) is overly restrictive in cases where one or more domains have limited training data.

Proposal

- An augmented cyclic adversarial learning model (ACAL) that enforces the cycle-consistency constraint via an external task-specific model for **low-resource** domain adaptation
- It encourages the preservation of task-relevant content as opposed to exact reconstruction

Algorithm

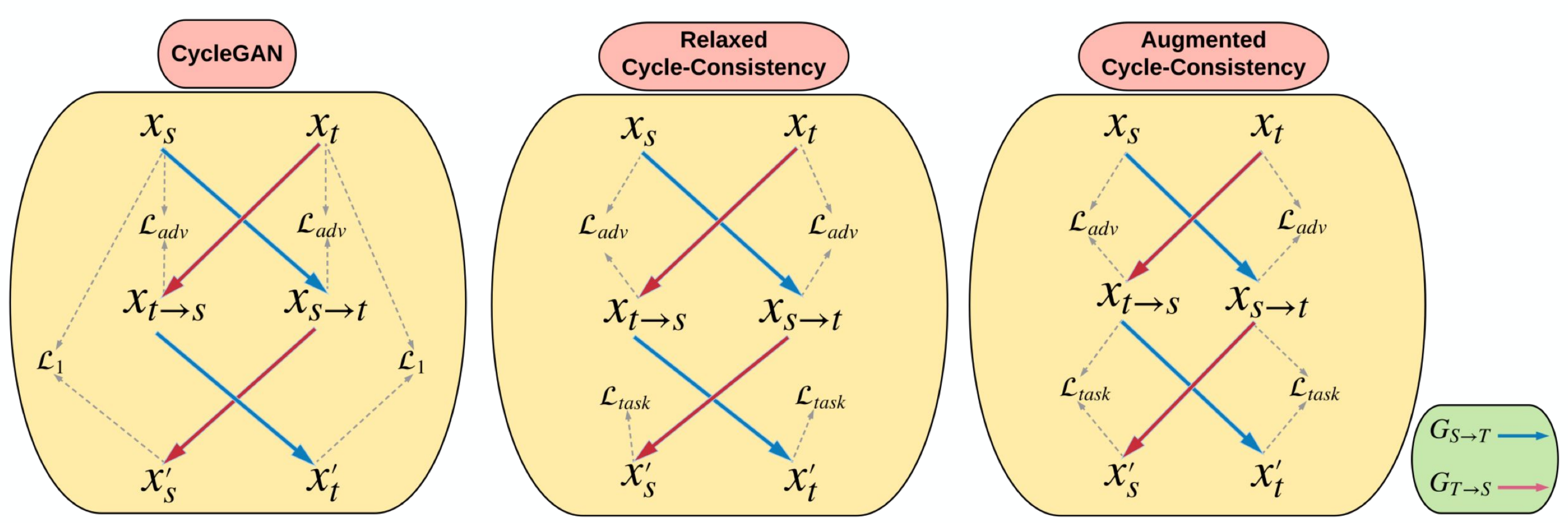
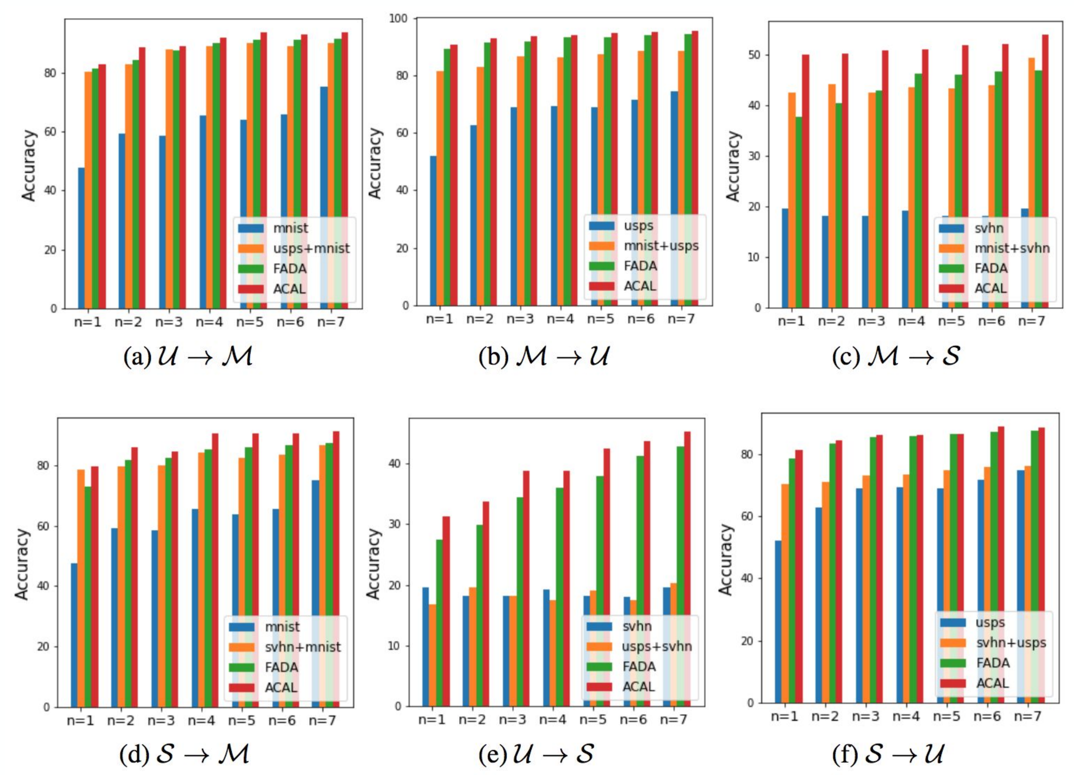


Illustration of proposed approach. In the diagrams X and L denote data and losses, respectively. The ultimate goal of our approach is to use the mapped Source → Target samples ($X_{S\to T}$) to augment the limited data of the target domain (X_T).

Algorithm 1 Augmented Cyclic Adversarial Learning (ACAL)

Input: source domain data $P_S(x,y)$, target domain data $P_T(x,y)$, *pretrained* source task model M_S
Output: target task model M_T
while *not converged* **do**
  Sample from (x_s, y_s) from P_S
  if y_t *in* P_T **then**
    %Supervised%
    Sample (x_t, y_t) from P_T
    Finetune source model M_S on (x_s, y_s) and $(G_{T\mapsto S}(x_t), y_t)$ samples (eq. 6)
    Train task model M_T on (x_t, y_t) and $(G_{S\mapsto T}(x_s), y_s)$ samples (eq. 7)
  else
    %Un-supervised%
    Sample x_t from P_T
    Finetune source model M_S on (x_s, y_s) samples (eq. 8)
    Train task model M_T $(G_{S\mapsto T}(x_s), y_s)$ and $(x_t, M_S(G_{T\mapsto S}(x_t)))$ samples (eq. 9)
  end
end

Visual Domain Adaptation (Supervised-Low Resource)



FADA model refers to Motiian et al. [2]. n = 1 means 1 labeled example per class. No unlabeled target sample is used.

Visual Domain Adaptation (Unsupervised-High Resource)

Model	Direction	Domain pairs $\mathcal{M}-\mathcal{U}$	$\mathcal{M}-\mathcal{MM}$	$\mathcal{M}-\mathcal{S}$	$\mathcal{S}-\mathcal{SD}$
Source-only	→	83.46	59.55	38.03	90.32
	←	71.14	98.36	71.11	88.17
DA (Häusser et al., 2017)	→	-	89.53	-	-
	←	-	-	**97.6**	91.86
VADA (Shu et al., 2018)	→	-	95.7	**73.3**	-
	←	-	-	94.5	94.9
Self-ensembling (MT+CT) (French et al., 2018)	→	88.14	-	33.87	-
	←	92.35	-	93.33	96.01
DupGAN (Hu et al., 2018)	→	96.01	-	62.65	-
	←	**98.75**	-	92.46	-
CyCADA (Hoffman et al., 2018)	→	95.6	57.21	14.56	81.19
	←	96.5	94.57	90.4	72.94
SBADA-GAN (Russo et al., 2018)	→	97.6	**99.4**	61.1	-
	←	95.0	-	76.1	-
ACAL (Ours)	→	**98.31**	97.29	60.85	**96.43**
	←	97.16	**99.26**	96.51	**97.98**
Target-only (completely supervised)	→	96.26	98.19	93.38	98.60
	←	99.49	99.49	99.49	93.38

MNIST (M), MNIST-M (MM), USPS (U) and SVHN (S), Synthetic Digits (SD)

Speech Domain Adaptation

Training Set	Domain Adaptation Model	Female (PER) Val	Test
$\mathcal{M}$	-	35.70	30.69
$\mathcal{F}$ (Baseline model)	-	24.51	23.22
$\mathcal{M}\to\mathcal{F}$	CycleGAN (Zhu et al., 2017)	32.95	30.07
	FHVAE (Hsu et al., 2017)	–	26.2
	MD-CycleGAN (Hosseini-Asl et al., 2018)	28.80	25.45
	ACAL (Ours)	**24.86**	**23.46**
$\mathcal{F}+(\mathcal{M}\to\mathcal{F})$	CycleGAN (Zhu et al., 2017)	28.32	28.43
	MD-CycleGAN (Hosseini-Asl et al., 2018)	21.15	19.08
	ACAL (Ours)	**20.32**	**19.02**
$\mathcal{F}+\mathcal{M}$	-	20.63	20.52
$\mathcal{F}+\mathcal{M}+(\mathcal{M}\to\mathcal{F})$	CycleGAN (Zhu et al., 2017)	21.03	22.81
	MD-CycleGAN (Hosseini-Asl et al., 2018)	20.26	19.60
	ACAL (Ours)	**20.02**	**18.44**

Speech domain adaptation results on TIMIT. We treat Male (M) and Female (F) voices for the source and target domains, respectively, based on the intrinsic imbalance of speaker genders in the dataset (about 7 : 3 male/female ratio). For the evaluation metric, lower is better.

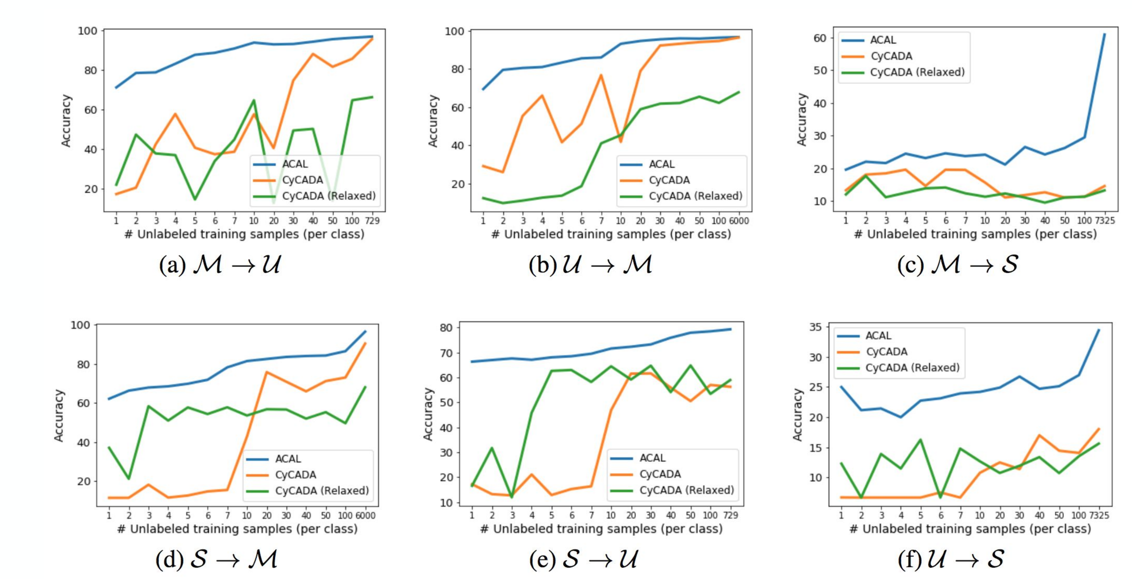
Ablation Study (One cycle vs. Two cycles)

SVHN (Source) -> MNIST-10 (Target)

Domain Adaptation Model	Test Accuracy (%)
No Adaptation (trained on SVHN)	71.11
Target Model (trained on MNIST-(10))	79.22±3.98
SVHN+MNIST-(10)	85.62±1.15
S→T	69.91±1.56
(S→T→S)-One Cycle	46.32±2.09
(T→S→T)-One Cycle	58.34±2.49
(S→T→S)-RCAL (Ours)	**72.51±1.71**
(T→S→T)-RCAL (Ours)	43.56±2.92
(S→T→S)-ACAL (Ours)	**79.40±0.73**
(T→S→T)-ACAL (Ours)	49.81±0.53
CycleGAN	45.54±1.05
RCAL (Ours)	88.62±1.77
ACAL (Ours)	**93.90±0.33**

This observation suggests that it would be beneficial to have cycles in both directions when applying the cycle-consistency constraint, since then both mappings can be learned via real examples

Ablation Study (Robustness)



Comparison of adaptation robustness between CyCADA (Hoffman et al., 2018), CyCADA with no l_1 reconstruction loss (Relaxed), and ACAL algorithms for variable number of unsupervised target samples. Note: No labeled sample is used.

Visual Domain Adaptation (Semi and Unsupervised – Low Resource)

# target samples per class	$\mathcal{S}\to\mathcal{M}$ 0%	10%	$\mathcal{M}\to\mathcal{U}$ 0%	10%	$\mathcal{U}\to\mathcal{M}$ 0%	10%	$\mathcal{S}\to\mathcal{U}$ 0%	10%
$n = 10$	81.43	77.63	93.86	94.01	93.22	94.89	71.54	75.98
$n = 50$	84.26	87.22	95.61	94.17	95.93	96.83	77.87	86.19
$n = 100$	86.49	91.75	96.31	96.01	96.43	96.92	78.42	89.03
$n =$ full train	96.51	99.41	96.91	95.71	96.74	98.45	79.23	93.17

Low-resource semi and unsupervised domain adaptation on MNIST (M), USPS (U) and SVHN (S) datasets. Note: n = 10 means 10 samples per class, and 10% denotes the percentage of target samples (per class) which have labels. 0% corresponds to low-resource unsupervised adaptation.

ASR Prediction

True	sil dh ah m aa r n ih ng sil d uw aa n dh ah s sil p ay dx er w eh sil g l ih s eh n sil d ih n dh ah s ah n sil
No adaptation	sil dh ah m aa r n ih ng sil d uw aa m ih s sil b ay er w ih sil b z l ih s ih n d ih n s ah n sil
CycleGAN	sil dh ih m aa r n ih ng sil d ih ah n dh ih s sil p ay ih w r eh sil dh l dh ih s ih n sil d ih n s ay n sil
ACAL	sil dh ah m aa r n ih ng sil d uw ah n dh ih s sil b ay dx y er w eh sil b l ih s ih n sil d ih n ih s ah n sil
True	sil iy v ih n ah s ih m sil p l v ah sil k ae sil b y ih l eh r iy sil k ah n sil t ey n sil t s ih m sil b l z sil
No Adaptation	sil iy dh ih n ah s ih m v l v ow sil k ae sil b y ih l eh r iy sil k eh n sil t ey n s ih m sil b l z sil
CycleGAN	sil iy ih m ah s eh m sil p l v dh aa sil k ey sil b y ih r ey ey sil k ih n sil t r ey n sil s ih m sil b ah l z sil
ACAL	sil iy v ih n ah s ih m sil p l v ow sil k ae sil b y ih l eh r iy sil k ih n sil t ay ey n s ih m sil b l z sil

ASR prediction improvement on low resource Female domain (TIMIT), when augmented with Male -> Female audios using domain adaptation

Contact
- Ehsan Hosseini-Asl, ehosseiniasl@salesforce.com
- Yingbo Zhou, yingbo.zhou@salesforce.com
- Caiming Xiong, cxiong@salesforce.com
- Richard Socher, rsocher@salesforce.com

Check out more of our work at https://einstein.ai

Paper:
https://openreview.net/forum?id=B1G9doA9F7

Selected References:

- [1] J.Y. Zhu, T. Park, P. Isola and A. A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. ICCV *2017*.
- [2] S. Motiian, Q. Jones, S. Iranmanesh, and G. Doretto. Few-shot adversarial domain adaptation. In Advances in Neural Information Processing Systems, 2017.
- [3] E. Hosseini-Asl, Y. Zhou, C. Xiong, and R. Socher. A multi-discriminator cyclegan for unsupervised non-parallel speech domain adaptation. In INTERSPEECH, 2018.
- [4] W.-N. Hsu, Y. Zhang, and J. R. Glass, Unsupervised learning of disentangled and interpretable representations from sequential data, in *NIPS*, 2017.