A Generative Framework for Zero-Shot Learning with Adversarial Domain Adaptation

Varun Khare*, Divyat Mahajan*, Homanga Bharadhwaj, Vinay Kumar Verma, Piyush Rai

Indian Institute of Technology Kanpur, India; Microsoft Research, India; University of Toronto, Canada

Introduction

- Zero-Shot Learning (ZSL) deals with the classification of novel classes at the test time
- Our framework addresses the problem of domain shift between the seen and unseen class distributions in Zero-Shot learning
- Domain shift can hinder the performance of Zero-Shot Learning models as most of the methods rely on the transfer of knowledge from the seen classes
- We minimise the domain shift by developing a generative model for ZSL and augmenting it with adversarial domain adaptation

Generative Framework

- We model the data distribution as a mixture of individual class conditional distributions:

$$\mathbf{x} \sim p(\mathbf{x}|\zeta_c)\ \forall c \in \mathcal{C}$$

- The class conditional distribution parameters (ζ_c) are modelled as functions of class attributes a_c by a fully connected neural network with parameters Θ:

$$\zeta_\mathbf{c} = f_\Theta(\mathbf{a_c})$$

- The parameters are learnt by Maximum Likelihood Estimation over the data from the seen classes S:

$$\underset{\mathbf{\Theta}}{\operatorname{argmax}}\ \underset{\mathbf{x},c\sim S}{\mathbb{E}}[log(p(\mathbf{x}|\zeta_\mathbf{c}))]$$

- Prediction for a data point x_+ among the unseen classes U at the test time:

$$\hat{y}_+ = \underset{\mathbf{c}\in\mathbf{U}}{\operatorname{argmax}}\ p(\mathbf{x}_+|\zeta_c)$$

- A simple choice for $p(\mathbf{x}|\zeta_c)$ can be Gaussian distribution which leads to the following setup:

$$\mathbf{x} \sim N\left(\mu_c, \mathbf{\Sigma_c}\right),\ \mu_\mathbf{c} = f_{\theta^\mu}(\mathbf{a_c}),\ \mathbf{\Sigma_c}^{-1} = diag(f_{\theta^\Sigma}(\mathbf{a_c}))$$

- The framework is trained by minimizing the following loss function:

$$\underset{\theta^\mu,\theta^\Sigma}{\operatorname{argmax}}\ \underset{(\mathbf{x},c)\sim S}{\mathbb{E}}\left[log(\mathbf{\Sigma_c}^{-1}) - (\mathbf{x}-\mu_\mathbf{c})^T\mathbf{\Sigma_c}^{-1}(\mathbf{x}-\mu_\mathbf{c})\right]$$

- End to end training of the generative framework helps in stable training w.r.t hyperparameters as compared to the previous generative approaches [1]
- Generative models can help in transductive learning as new samples from a particular class can be generated by conditional sampling from the above distribution
- Good prediction at the test time over unseen classes depends upon learning f_θ robustly from the seen classes at the training time

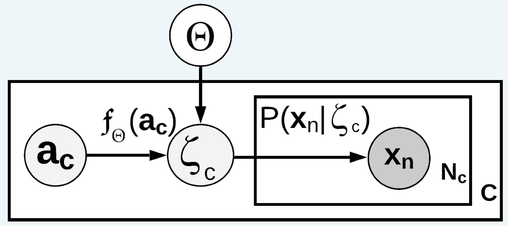


Figure: Plate notation of the generative framework

Adversarial Domain Adaptation

- Domain shift between the seen and the unseen classes would affect the estimation of unseen class distribution in the above approach
- We employ a Cycle GAN [2] based adversarial domain adaptation framework to bring the estimated unseen class distribution closer to the true unseen class distribution
- Wasserstein loss [3] was used for adversarial loss along with cyclic loss and classification loss
- Identity regularizer with $l1$ norm was added to the generator to ensure that the output domains for each generator remain unmodified if given as input
- The classification loss ($\mathcal{L}_{clf}$) of the target data (not generated by G) is added to the discriminator loss during the adversarial training
- Once the classification accuracy over the generated data becomes close or greater than the accuracy of pseudo-labels, corruption recovery is done by training the classifiers over both the transformed samples $G^T(y_{nc})$ and true data samples x_{nc}
- The framework is trained by minimizing the following loss function:

$$\mathcal{L} = \mathcal{L}^T_{adv} + \mathcal{L}^S_{adv} + \chi\mathcal{L}_{cyc} + \xi\mathcal{L}^T_{clf} + \xi\mathcal{L}^S_{clf}$$

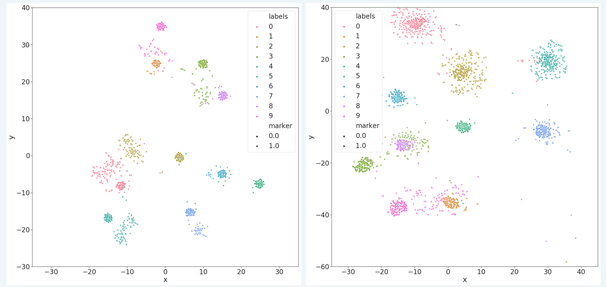
where $\mathcal{L}^{\{T,S\}}_{adv} = \{L_G + L_D\}^{\{T,S\}}$ with

$$L^T_G = \underset{c\sim p_c}{\mathbb{E}}[\beta\|G^T(x_{nc}) - x_{nc}\|_p - D^T_w \circ G^T(y_{nc})]$$

$$L^T_D = \underset{c\sim p_c}{\mathbb{E}}[D^T_w \circ G^T(y_{nc}) - D^T_w(x_{nc})]$$

$$\mathcal{L}_{cyc}(G^T, G^S) = \mathbb{E}_{c\sim p_c}[\|G^S \circ G^T(y_{nc}) - x_{nc}\|_p] + \mathbb{E}_{c\sim p_c}[\|G^T \circ G^S(x_{nc}) - y_{nc}\|_p]$$

$$L^T_{clf} = \mathbb{E}_{c\sim p_c}[L(C^T_{clf} \circ G^T(y_{nc}), Y^T)] + \mathbb{E}_{c\sim p_c}[L(C^T_{clf}(x_{nc}), \bar{Y}^U)]$$



(a) (b)

Figure: **(a)** shows the t-SNE plot for the output of the generative model as compared to the test data. Crosses represent the test data while dots represent the generated data. **(b)** shows the t-SNE plot after domain shift minimization with our model.

Proposed Approach

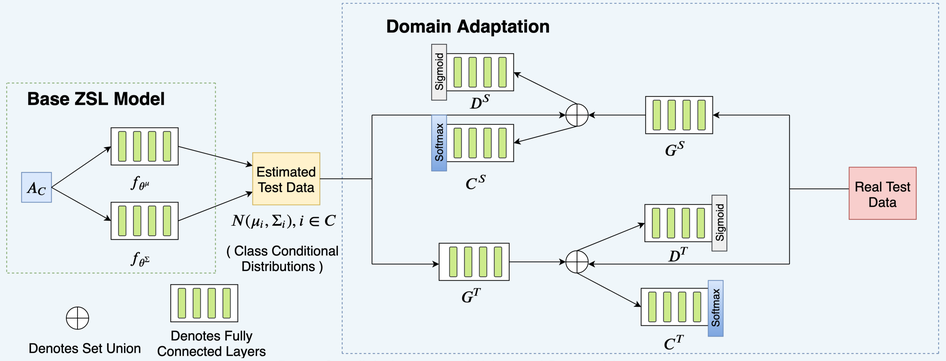


Figure: Overall Architecture with Adversarial Domain Adaptation

Results

	SUN	CUB	AWA2
Method	**PS**	**PS**	**PS**
CONSE	38.8	34.3	44.5
SSE	51.5	43.9	61.1
LATEM	55.3	49.3	55.8
DEVISE	56.5	52.0	59.7
SJE	53.7	53.9	61.9
ESZSL	54.5	53.9	58.6
SYNC	56.3	55.6	46.6
DEM	61.9	51.7	67.1
GFZSL	63.1	49.2	67.0
CVAE-ZSL	61.7	52.1	65.8
W/O ADA (Ours)	**63.3**	**70.9**	**70.4**

Table: Inductive Zero Shot Learning results with splits proposed in [4]

Method	SUN	CUB	AWA2
DSRL	56.8	48.7	72.8
ALE	55.7	54.5	70.7
GFZSL	64.2	50.5	**78.6**
With ADA (Ours)	**65.5**	**74.2**	**78.6**

Table: Transductive Zero-Shot Learning results with splits proposed in [4]

Ablation Results

- M1 Score: Class averaged top-1 accuracy of the predictions from the classifier C^T in the Cycle GAN
- M2 Score: 1 nearest neighbor classification accuracy using the gaussian distance between the transformed class conditionals $G^T(y_{nc})$ and the true features x_{nc}
- Std DA: Deep Classifier trained on the samples synthesized from the base ZSL model
- Vanilla ADA: Conventional GAN architecture replacing the Cycle GAN architecture in the proposed approach
- Cycle GAN w/o: Classifiers C^S and C^T removed from the proposed approach

	SUN		CUB		AWA2	
Variant	**M1**	**M2**	**M1**	**M2**	**M1**	**M2**
Std DA	64.8	NA	72.2	NA	71.3	NA
Vanilla ADA	64.9	47.1	71.5	57.8	77.3	56.1
CycleGAN w/o	NA	**57.2**	NA	**68.4**	NA	**75.8**
Ours	**65.5**	55.8	**74.2**	67.5	**78.6**	74.9

Table: Ablation study results for Zero-Shot Learning with splits proposed in [4]

References

[1] V. K. Verma and P. Rai, "A simple exponential family framework for zero-shot learning," in *ECML-PKDD*, pp. 792–808, Springer, 2017.

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks,"

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[4] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly," *IEEE transactions on pattern analysis and machine intelligence*, 2018.