AlignFlow: Cycle Consistent Learning from Multiple Domains via Normalizing Flows [AAAI 2020]

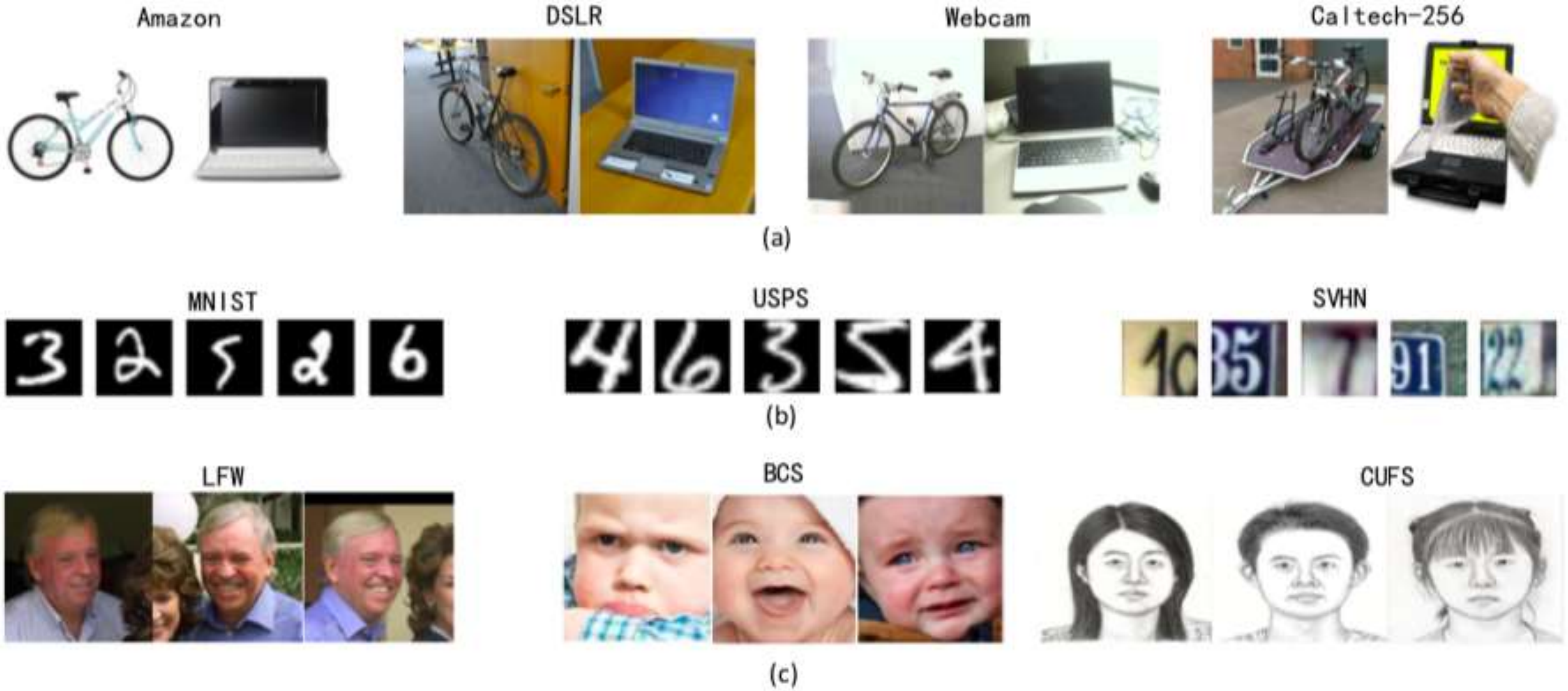
Presentation for CS 594 with Prof. Sathya

Dec 3, 2020

Presented By: Krishna Garg

Motivation

- Learn shared structure from different domains

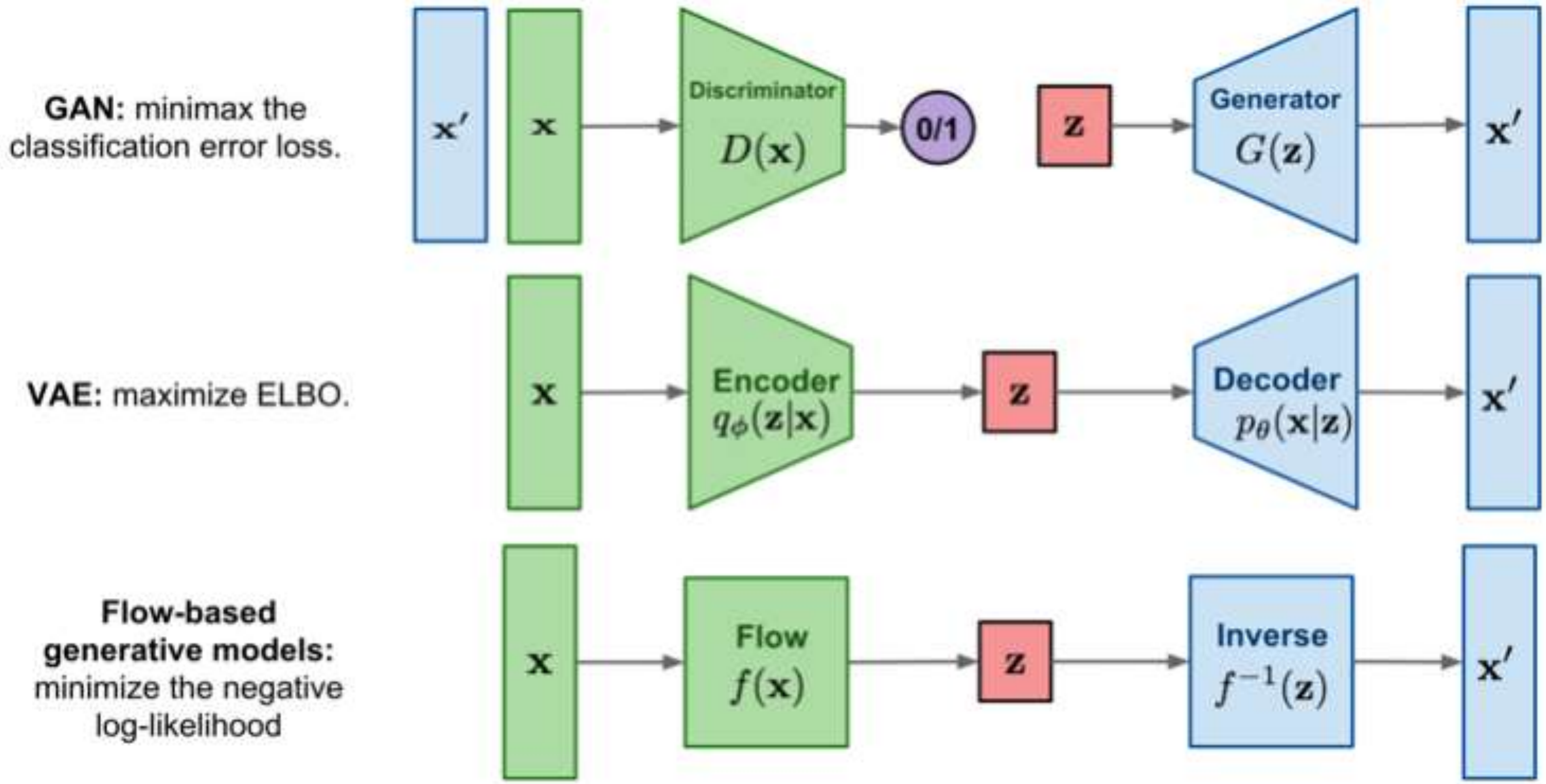


Source: Deep Visual Domain Adaptation: A Survey

The high-level idea

- Latent variable generative modelling to discover the shared structure
- Model the data from each domain via “invertible generative model” (“Normalizing flows method”)
- Shared latent space across all domains

Generative Modelling at a glance



(+) Likelihood-free
(+) Sampling-based adversarial loss
(-): Mode collapse
(-): Slow and unstable

(+) Statistically efficient in certain conditions
(-) Intractable/ ill-defined likelihood often

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

(+) Evaluate exact likelihoods

AlignFlow, FlowGAN : BEST OF ALL WORLDS $\min_\theta \max_\phi V(G_\theta, D_\phi) - \lambda \mathbb{E}_{\mathbf{x}\sim P_{\text{data}}}[\log p_\theta(\mathbf{x})]$

Source: https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.html#types-of-generative-models

Flow-based Generative Modeling (FlowGAN)

- Change-of-variables formula

Density of observed variable

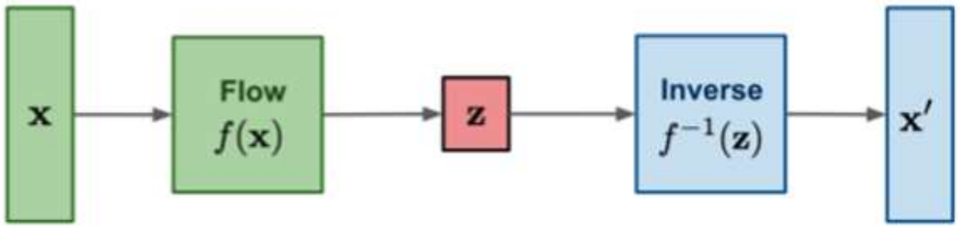
$$p_\theta(\mathbf{x}) = p(\mathbf{z}) \left| \det \frac{\partial f_\theta(\mathbf{x})}{\partial \mathbf{x}} \right|$$

Prior over latent variable

Jacobian
[Requirement: Should be easy to compute]

Requirements:

- G_θ invertible. Let $f_\theta = G_\theta^{-1}$
- $z = f_\theta(x)$

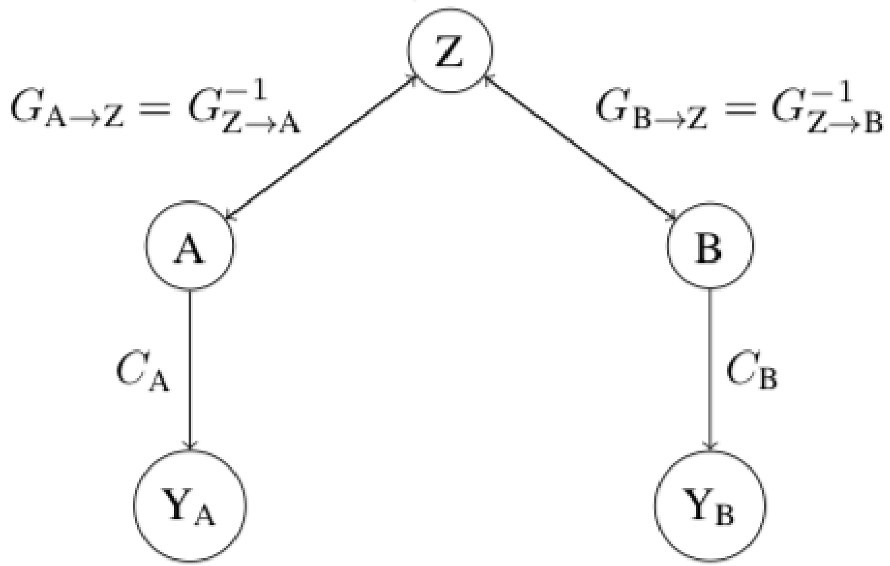


Problem Setup

- Given: Unpaired datasets D_A, D_B with unknown densities p_A^*, p_B^*
- Goals: Estimate
  - Marginal likelihoods p_A, p_B that approximate p_A^*, p_B^*
  - Conditional distributions $p_{A|B}, p_{B|A}$

Representation

- Bayesian Network: $A \leftarrow Z \rightarrow B$
- Define Generator Mappings: $G_{Z\rightarrow A}, G_{Z\rightarrow B}$ s.t. their inverse exists



- Domain Translation
  - $G_{A\rightarrow B} = G_{Z\rightarrow B} \circ G_{A\rightarrow Z}$
  - $G_{B\rightarrow A} = G_{Z\rightarrow A} \circ G_{B\rightarrow Z}$
- Proving $G_{B\rightarrow A} = G^{-1}_{A\rightarrow B}$

$$\begin{aligned} G^{-1}_{A\rightarrow B} &= (G_{Z\rightarrow B} \circ G_{A\rightarrow Z})^{-1} = G^{-1}_{A\rightarrow Z} \circ G^{-1}_{Z\rightarrow B} \\ &= G_{Z\rightarrow A} \circ G_{B\rightarrow Z} = G_{B\rightarrow A}. \end{aligned}$$

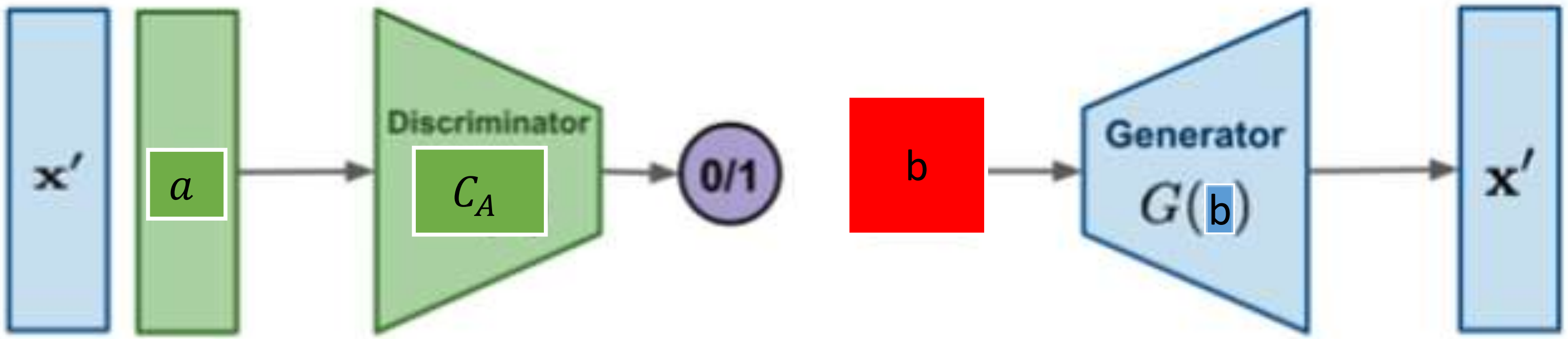
Learning Algorithms & Objectives

- Both flows $G_{Z \to A}, G_{Z \to B}$ can be trained independently via MLE, ADV, HYBRID
  - ADV $\min_{\theta} \max_{\phi} V(G_\theta, D_\phi)$ MLE $\max_{\theta \in \mathcal{M}} \mathbb{E}_{\mathbf{x} \sim P_{\text{data}}} [\log p_\theta(\mathbf{x})]$
  - HYBRID $\min_{\theta} \max_{\phi} V(G_\theta, D_\phi) - \lambda \mathbb{E}_{\mathbf{x} \sim P_{\text{data}}} [\log p_\theta(\mathbf{x})]$

- But how can we exploit data from the two domains??

Learning Algorithms & Objectives

- Both flows $G_{Z\to A}, G_{Z\to B}$ can be trained independently via MLE, ADV, HYBRID
- But how can we exploit data from the two domains??
  - Directly perform ADV training of $G_{B\to A}$



$$\mathcal{L}_{\mathrm{GAN}}(C_{\mathrm{A}}, G_{\mathrm{B}\to\mathrm{A}}) = \mathbb{E}_{a\sim p^*_{\mathrm{A}}}[\log C_{\mathrm{A}}(a)] + \mathbb{E}_{b\sim p^*_{\mathrm{B}}}[\log(1 - C_{\mathrm{A}}(G_{\mathrm{B}\to\mathrm{A}}(b)))].$$

Learning Algorithms & Objectives

- Both flows $G_{Z\to A}, G_{Z\to B}$ can be trained independently via MLE, ADV, HYBRID

- But how can we exploit data from the two domains??
  - Directly perform ADV training of $G_{B\to A}, G_{A\to B}$ but keep MLE separate
  - Hybrid Learning Objective
  - Combine the two generators

$$
\begin{aligned}
&\mathcal{L}_{\mathrm{AlignFlow}}(\boxed{G_{\mathrm{B}\to\mathrm{A}}}, C_{\mathrm{A}}, C_{\mathrm{B}}; \lambda_{\mathrm{A}}, \lambda_{\mathrm{B}}) \\
&= \mathcal{L}_{\mathrm{GAN}}(C_{\mathrm{A}}, G_{\mathrm{B}\to\mathrm{A}}) + \mathcal{L}_{\mathrm{GAN}}(C_{\mathrm{B}}, G_{\mathrm{A}\to\mathrm{B}}) \\
&- \cancelto{0}{\lambda_{\mathrm{A}}\mathcal{L}_{\mathrm{MLE}}(G_{\mathrm{Z}\to\mathrm{A}})} - \cancelto{0}{\lambda_{\mathrm{B}}\mathcal{L}_{\mathrm{MLE}}(G_{\mathrm{Z}\to\mathrm{B}})}
\end{aligned}
$$

$\lambda_A = \lambda_B = 0$
Pure ADV training

$\lambda_A = \lambda_B \to \infty$
Pure MLE training

Recall Problem Setup -> Inference Stage

- Given: Unpaired datasets D_A, D_B with unknown densities p_A^*, p_B^*
- Goals: Estimate
  - Marginal likelihoods p_A, p_B that approximate p_A^*, p_B^*
    - Assume shared latent space $z \sim p(z), \lambda_A \neq 0, \lambda_B \neq 0$
    - Paired Dataset becomes $\left(G_{Z\to A}(z), G_{Z\to B}(z)\right)$
  - Conditional distributions $p_{A|B}, p_{B|A}$
    - Given by $G_{B\to A}, G_{A\to B}$

Theoretical Analysis – Optimal Generators

- Def. An invertible mapping $G_{Y\to X}$ is marginally consistent(M.C.) w.r.t. (p_X, p_Y):

$$p_X(x) = \begin{cases} p_Y(y)\left|\det \frac{\partial G_{Y\to X}}{\partial X}^{-1}\right|_{X=x}, & if\ x = G_{Y\to X}(y) \\ 0, & otherwise. \end{cases}$$

- Lemma. If there exist mappings $G^*_{Z\to A}, G^*_{Z\to B}$ that are M.C. w.r.t (p^*_A, p_Z) & (p^*_B, p_Z), then $G^*_{B\to A}$ is M.C. w.r.t. (p^*_A, p^*_B) [From def. & Change of Variables Theorem]
- Theorem. Given critics C_A, C_B, $G^*_{B\to A}$ globally minimizes the AlignFlow objective

Theoretical Analysis – Optimal Generators (Proof)

- Theorem. Given critics C_A, C_B, $G^*_{B\to A}$ globally minimizes the AlignFlow objective
- Proof.
  1. $L_{MLE}(\mathrm{G}_{\mathrm{Z}\to\mathrm{X}})$ minimized at marginally consistent mapping $G_{Z\to X^*}$ (X=A,B)
  2. $L_{GAN}(C_A, G_{B\to A})$ minimized when $p_A = p^*_A$ and critic is Bayes optimal [Goodfellow et al., Theorem 1]. $\Longrightarrow L_{GAN}(C_A, G_{B\to A})$ minimized by marginally consistent mapping $G^*_{B\to A}$ w.r.t. (p^*_A, p^*_B). Similar for $L_{GAN}(C_B, G_{A\to B})$.
  3. $G^*_{B\to A}$ = $G^*_{Z\to A} \circ G^*_{Z\to B}$ [by design]

  Thus, $G^*_{B\to A}$ globally minimizes all the terms in the AlignFlow learning objective.

Theoretical Analysis – Optimal Critics

Theorem 2. *Let p^*_{A} and p^*_{B} denote the true data densities for domains* A *and* B *respectively. Let C^*_{A} and C^*_{B} denote the optimal critics for the AlignFlow objective with the cross-entropy GAN loss for any fixed choice of the invertible mapping $G_{\mathrm{A}\to\mathrm{B}}$. Letting $b = G_{\mathrm{A}\to\mathrm{B}}(a)$ for any $a \in \mathrm{A}$, we have:*

$$C^*_{\mathrm{A}}(a) = \frac{C^*_{\mathrm{B}}(b) p^*_{\mathrm{A}}(a)}{p^*_{\mathrm{A}}(a) + p^*_{\mathrm{B}}(b)(1 - C^*_{\mathrm{B}}(b)) \left| \det \frac{\partial G^{-1}_{\mathrm{B}\to\mathrm{A}}}{\partial \mathrm{A}} \right|_{\mathrm{A}=a}}.$$

Theoretical Analysis – Optimal Critics (Proof)

$$C^*_{\mathrm{A}}(a) = \frac{p^*_{\mathrm{A}}(a)}{p^*_{\mathrm{A}}(a) + p_{\mathrm{A}}(a)}$$

[Proposition 1, Goodfellow et al.]

$$p_{\mathrm{A}}(a) = p_{\mathrm{B}}(b) \left| \det \frac{\partial G^{-1}_{\mathrm{B}\to\mathrm{A}}}{\partial \mathrm{A}} \right|_{\mathrm{A}=a}$$

[Using change of variables theorem]

where $b = G_{\mathrm{A}\to\mathrm{B}}(a)$.

[Condition for optimal critic]

$$C^*_{\mathrm{A}}(a) = \frac{p^*_{\mathrm{A}}(a)}{p^*_{\mathrm{A}}(a) + p_{\mathrm{B}}(b) \left| \det \frac{\partial G^{-1}_{\mathrm{B}\to\mathrm{A}}}{\partial \mathrm{A}} \right|_{\mathrm{A}=a}}$$

$$C^*_{\mathrm{B}}(b) = \frac{p^*_{\mathrm{B}}(b)}{p^*_{\mathrm{B}}(b) + p_{\mathrm{B}}(b)}.$$

[Proposition 1, Goodfellow et al.]

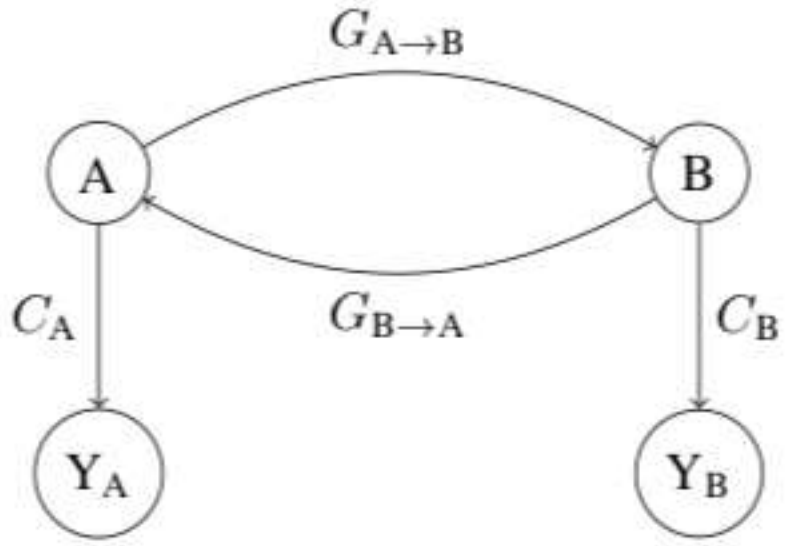
$$p_{\mathrm{B}}(b) = p^*_{\mathrm{B}}(b) \left(\frac{1}{C^*_{\mathrm{B}}(b)} - 1 \right)$$

[Rearranging terms]

$$C^*_{\mathrm{A}}(a) = \frac{C^*_{\mathrm{B}}(b) p^*_{\mathrm{A}}(a)}{p^*_{\mathrm{A}}(a) + p^*_{\mathrm{B}}(b)(1 - C^*_{\mathrm{B}}(b)) \left| \det \frac{\partial G^{-1}_{\mathrm{B}\to\mathrm{A}}}{\partial \mathrm{A}} \right|_{\mathrm{A}=a}}$$

Theoretical Analysis - Exact Cycle Consistency

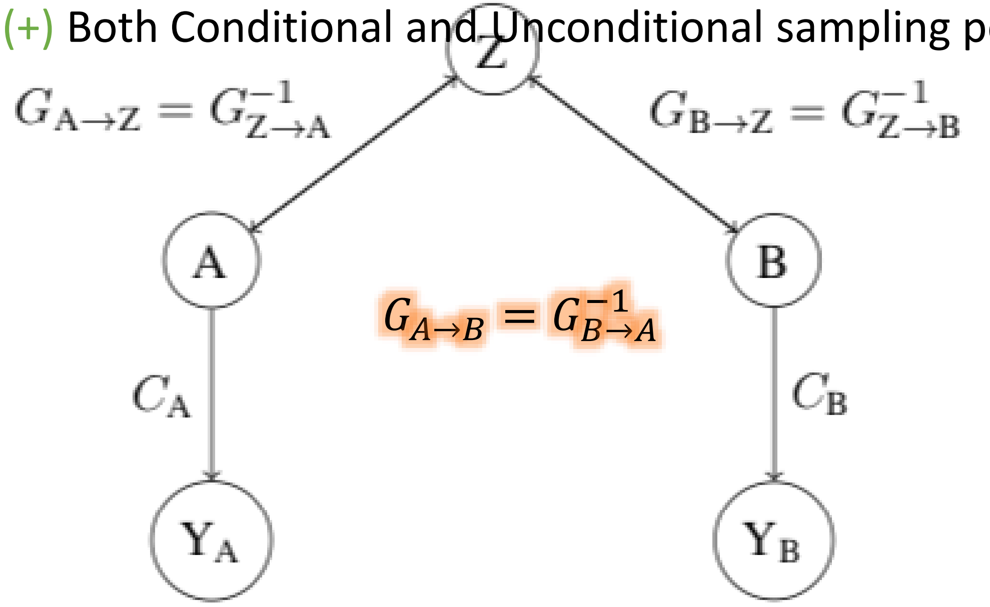
(-) Two independent domain mappings
(-) Only ADV training
(-) Additional Cycle Loss required
(-) Only Conditional sampling possible



(a) CycleGAN

$$\mathcal{L}_{\text{Cycle}}(G_{B\to A}, G_{A\to B}) = E_{a\sim p^*_A}[\|G_{B\to A}(G_{A\to B}(a)) - a\|_1].$$

(+) Single invertible mapping
(+) ADV / MLE / HYBRID
(+) Additional Cycle Loss not required
(+) Both Conditional and Unconditional sampling possible



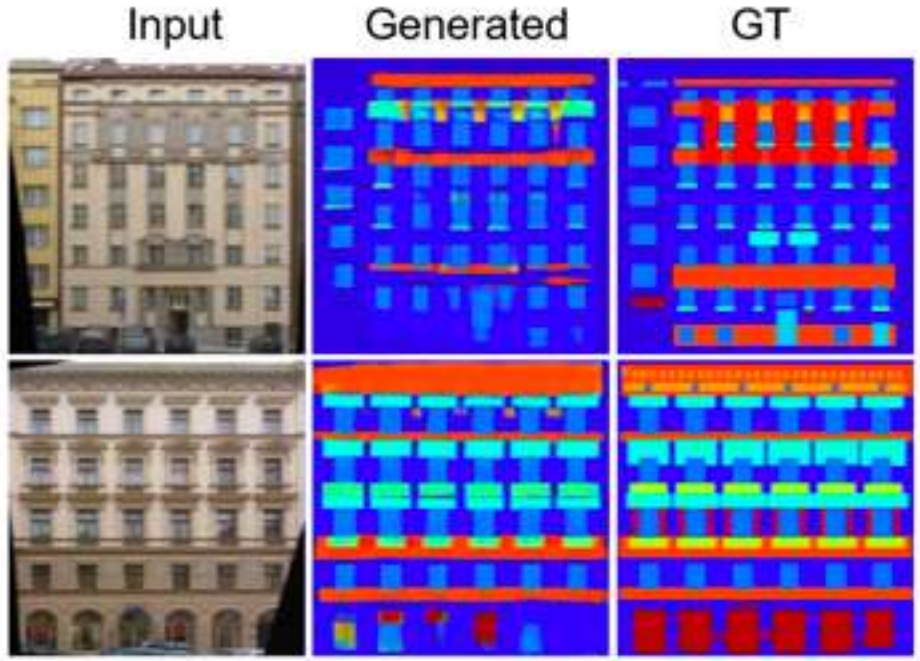
(b) AlignFlow

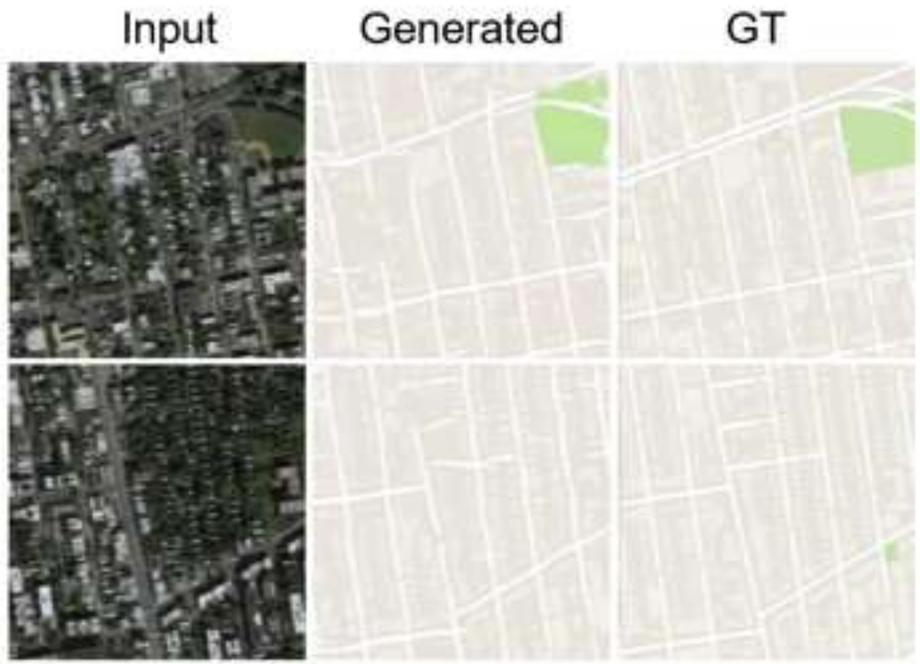
$$\mathcal{L}_{\text{Cycle}}(G_{B\to A}, G_{A\to B}) = 0$$
$$\mathcal{L}_{\text{Cycle}}(G_{A\to B}, G_{B\to A}) = 0$$

Empirical Results

- Unsupervised Image-To-Image Translation
  - Unsupervised during training
  - Supervised during validation

Datasets: Facades, Maps, CityScape





Empirical Results

Dataset	Model	MSE ($A \rightarrow B$)	MSE ($B \rightarrow A$)
Facades	CycleGAN	0.7129	0.3286
	AF (ADV only)	0.6727	0.2679
	AF (Hybrid)	**0.5801**	**0.2512**
	AF (MLE only)	0.9014	0.5960
Maps	CycleGAN	0.0245	0.0953
	AF (ADV only)	0.0385	0.1123
	AF (Hybrid)	**0.0209**	**0.0897**
	AF (MLE only)	0.0452	0.1746
CityScapes	CycleGAN	0.1252	**0.1200**
	AF (ADV only)	0.2569	0.2196
	AF (Hybrid)	**0.1130**	0.1462
	AF (MLE only)	0.2526	0.2272

Empirical Results

- Unsupervised Domain Adaptation
  - Source Domain: Access to both inputs and labels
  - Target Domain: Access to only inputs
  - Goal: Learn a classifier for target domain

(a) MNIST→USPS

Empirical Results

Model	MNIST $\rightarrow$ USPS	USPS $\rightarrow$ MNIST	SVHN $\rightarrow$ MNIST
source only	82.2 ± 0.8	69.6 ± 3.8	67.1 ± 0.6
ADDA (Tzeng et al. 2017)	89.4 ± 0.2	90.1 ± 0.8	76.0 ± 1.8
CyCADA (Hoffman et al. 2017)	95.6 ± 0.2	96.5 ± 0.1	90.4 ± 0.4
UNIT (Liu, Breuel, and Kautz 2017)	95.97	93.58	90.53
AlignFlow	$\mathbf{96.2} \pm \mathbf{0.2}$	$\mathbf{96.7} \pm \mathbf{0.1}$	$\mathbf{91.0} \pm \mathbf{0.3}$
target only	96.3 ± 0.1	99.2 ± 0.1	99.2 ± 0.1

Qualitative Results – Cross-domain translation

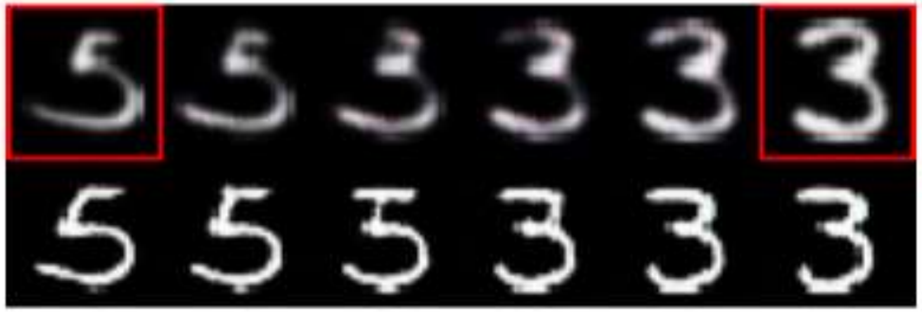
Failure cases for CyCADA model using CycleGAN (SVHN->MNIST)

(a) (b)

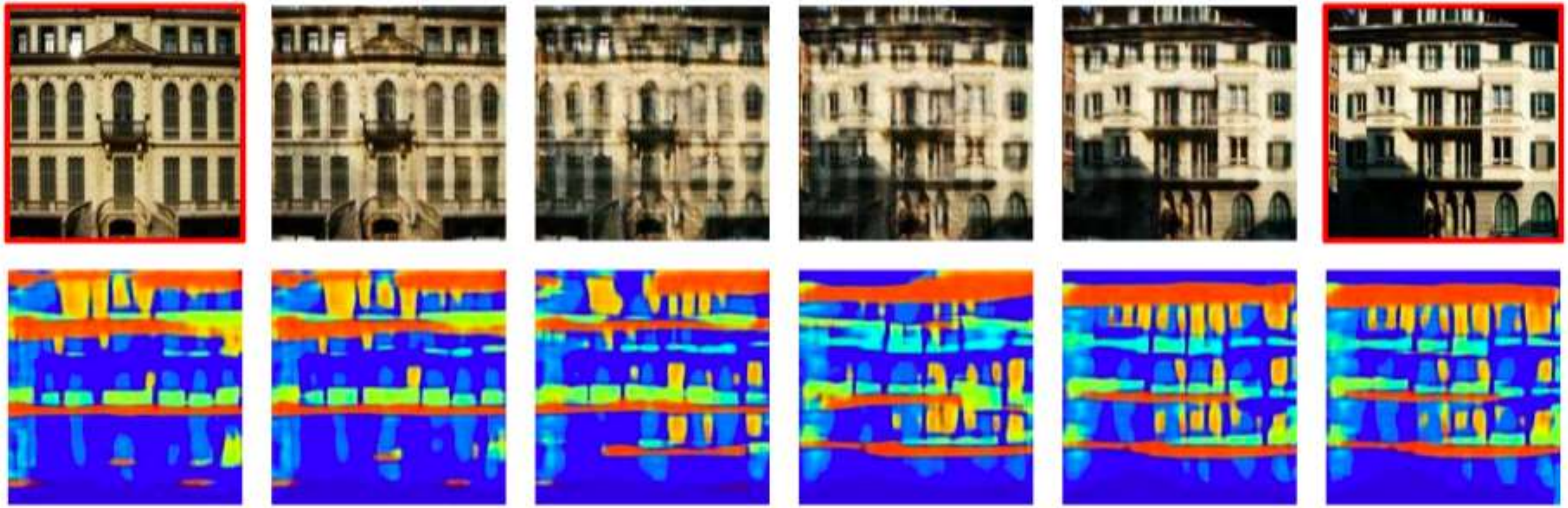
(c) (d)

Qualitative Results – Latent Space Interpolations

(a) MNIST→USPS



(b) USPS→MNIST



Summary

- A promising combination of FlowGAN and CycleGAN
- Theoretical Guarantees
  - Exact cycle consistency using shared latent space
- Empirical Results surpass the existing baselines on 2 different tasks
  - Code released
- Qualitative Results (Latent Space Interpolations)
- Interesting Future direction
  - Translation across more than two domains

Thank You

References

- https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.html#types-of-generative-models
- Wang, Mei, and Weihong Deng. "Deep visual domain adaptation: A survey." *Neurocomputing* 312 (2018): 135-153.
- Grover, Aditya, Manik Dhar, and Stefano Ermon. "Flow-gan: Combining maximum likelihood and adversarial learning in generative models." *arXiv preprint arXiv:1705.08868* (2017).