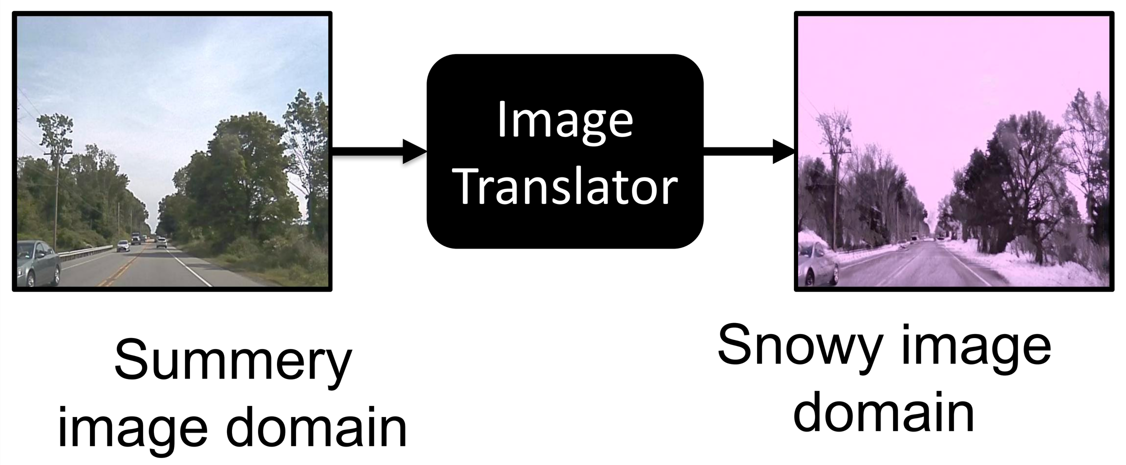
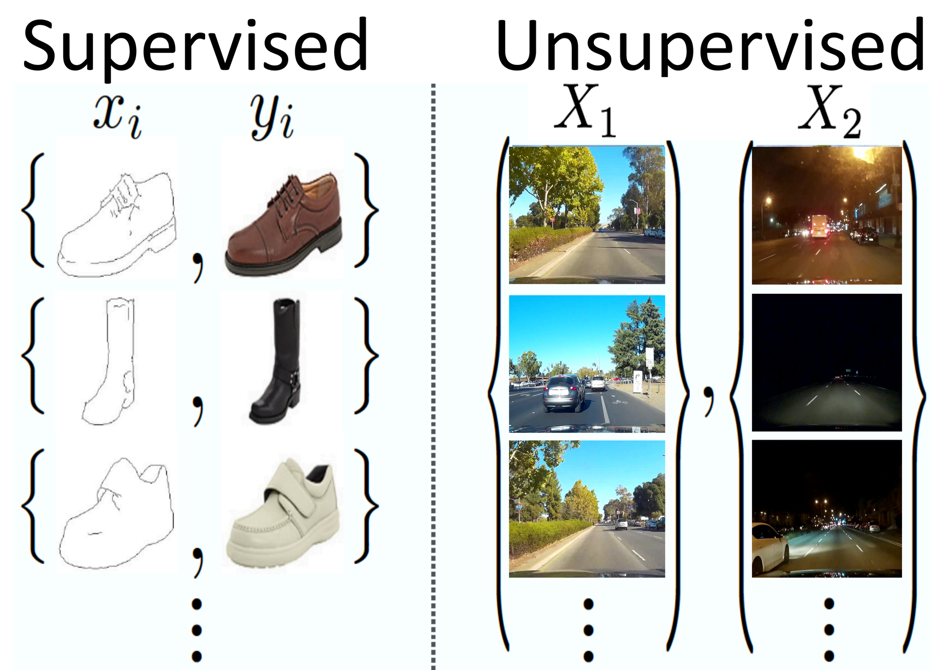
NIPS 2017
Monday December 04 -- Saturday December 09, 2017
Long Beach Convention Center, Long Beach

Unsupervised Image-to-Image Translation Networks

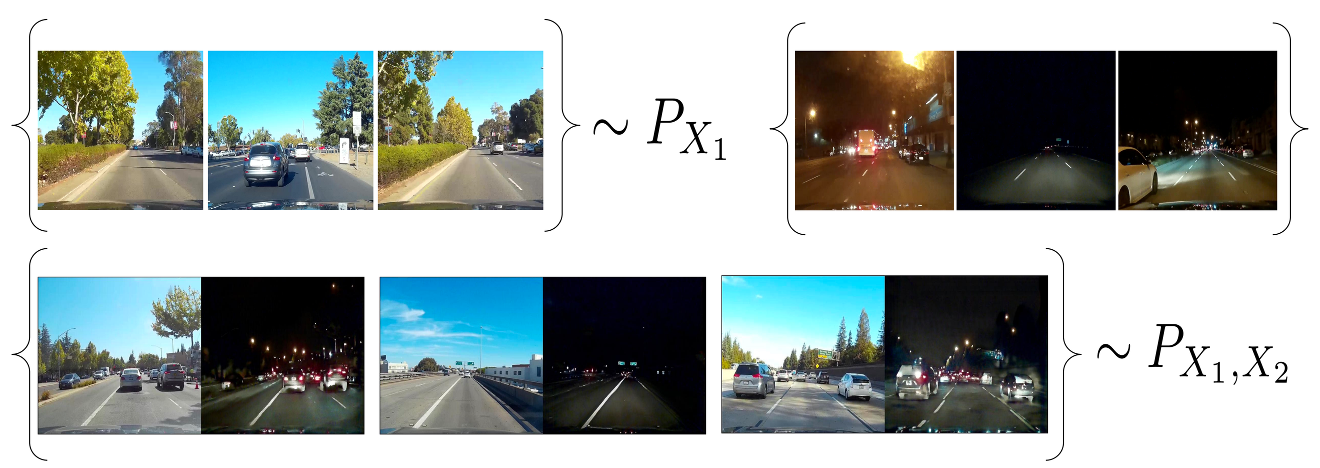
Ming-Yu Liu, Thomas Breuel, Jan Kautz
NVIDIA

Unsupervised Image-to-Image Translation



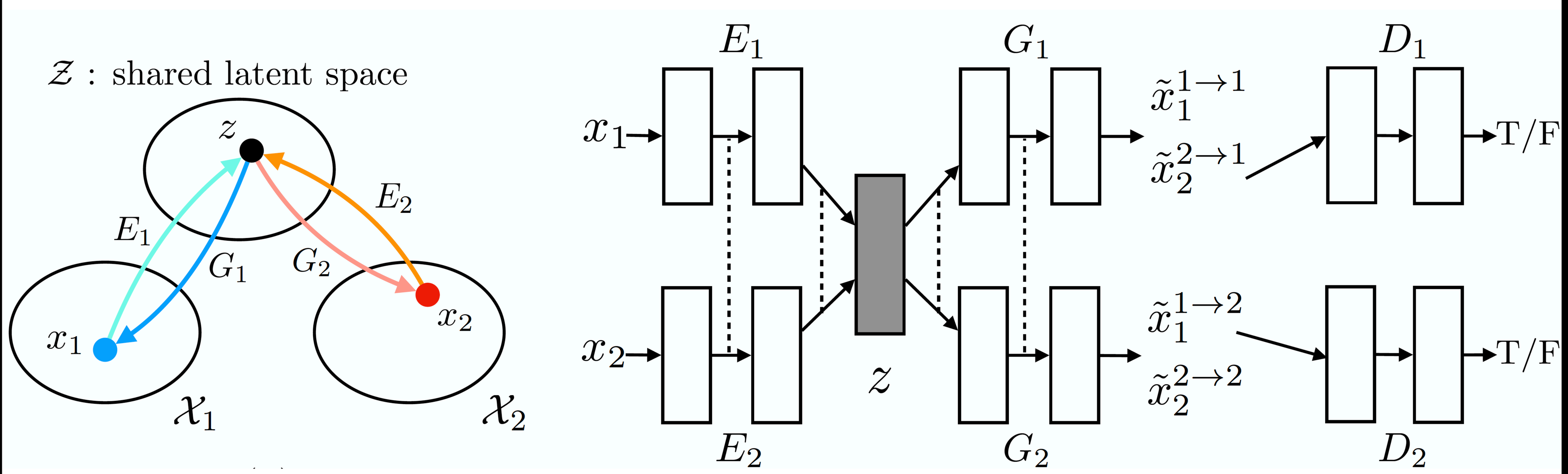


A Probabilistic Point of View



Estimate the joint distribution of images in different domains using samples from the marginal distributions.
- Ill-posed problem.
- Need additional assumptions.

Shared Latent Space Assumption



Learning Objective

$$\min_{E_1,E_2,G_1,G_2}\max_{D_1,D_2} \mathcal{L}_{\text{VAE}_1}(E_1,G_1)+\mathcal{L}_{\text{GAN}_1}(E_1,G_1,D_1)+\mathcal{L}_{\text{CC}_1}(E_1,G_1,E_2,G_2)$$
$$\mathcal{L}_{\text{VAE}_2}(E_2,G_2)+\mathcal{L}_{\text{GAN}_2}(E_2,G_2,D_2)+\mathcal{L}_{\text{CC}_2}(E_2,G_2,E_1,G_1).$$

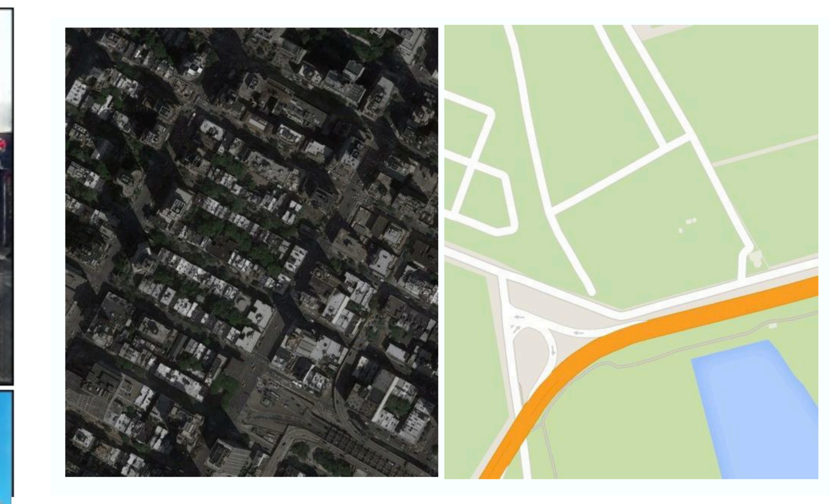
$$\mathcal{L}_{\text{GAN}_1}(E_1,G_1,D_1)=\lambda_0\mathbb{E}_{x_1\sim P_{\mathcal{X}_1}}[\log D_1(x_1)]+\lambda_0\mathbb{E}_{z_2\sim q_2(z_2|x_2)}[\log(1-D_1(G_1(z_2)))]$$
$$\mathcal{L}_{\text{GAN}_2}(E_2,G_2,D_2)=\lambda_0\mathbb{E}_{x_2\sim P_{\mathcal{X}_2}}[\log D_2(x_2)]+\lambda_0\mathbb{E}_{z_1\sim q_1(z_1|x_1)}[\log(1-D_2(G_2(z_1)))].$$

$$\mathcal{L}_{\text{VAE}_1}(E_1,G_1)=\lambda_1\text{KL}(q_1(z_1|x_1)||p_\eta(z))-\lambda_2\mathbb{E}_{z_1\sim q_1(z_1|x_1)}[\log p_{G_1}(x_1|z_1)]$$
$$\mathcal{L}_{\text{VAE}_2}(E_2,G_2)=\lambda_1\text{KL}(q_2(z_2|x_2)||p_\eta(z))-\lambda_2\mathbb{E}_{z_2\sim q_2(z_2|x_2)}[\log p_{G_2}(x_2|z_2)].$$

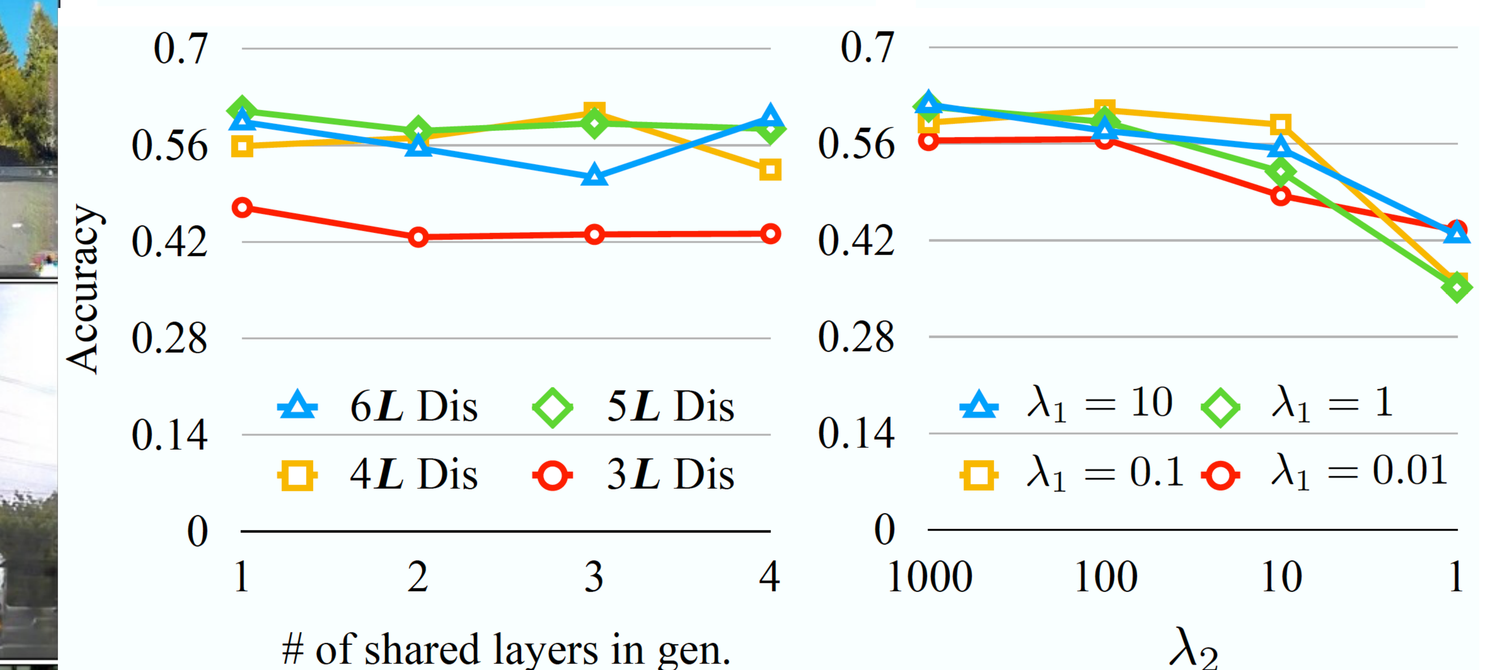
$$\mathcal{L}_{\text{CC}_1}(E_1,G_1,E_2,G_2)=\lambda_3\text{KL}(q_1(z_1|x_1)||p_\eta(z))+\lambda_3\text{KL}(q_2(z_2|x_1^{1\to2}))||p_\eta(z))-\lambda_4\mathbb{E}_{z_2\sim q_2(z_2|x_1^{1\to2})}[\log p_{G_1}(x_1|z_2)]$$
$$\mathcal{L}_{\text{CC}_2}(E_2,G_2,E_1,G_1)=\lambda_3\text{KL}(q_2(z_2|x_2)||p_\eta(z))+\lambda_3\text{KL}(q_1(z_1|x_2^{2\to1}))||p_\eta(z))-\lambda_4\mathbb{E}_{z_1\sim q_1(z_1|x_2^{2\to1})}[\log p_{G_2}(x_2|z_1)].$$

Visualization Results

Ablation & Hyper-parameter Study



Method	Accuracy
~~Weight Sharing~~	0.569±0.029
~~Cycle Consistenc~~	0.568±0.010
Proposed	0.600±0.015



Domain Adaptation

Method	DANN	DTN	CoGAN	UNIT
SVHN→MNIST	73.85	84.88		**90.53**
MNIST→USPS			95.65	**95.97**
USPS→MNIST			93.15	**93.58**

Conclusion

- A **Coupled GAN framework** for unsupervised image-to-image translation.
- Dealing with the ill-posed joint distribution learning problem via the shared latent space assumption and adversarial training.
- Shown translation results on various street scenes (640x480) and animal portraits (200x200).
- Good performance on digit domain adaptation

Implementation and More Results

https://github.com/mingyuliutw/UNIT

Reference

- [CoGAN] M-Y Liu, O. Tuzel "Coupled generative adversarial networks" NIPS'16
- [DTN] Y. Taigman, A. Polyak, L. Wolf "Unsupervised cross-domain image generation" ICLR'17
- [DANN] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky "Domain adversarial training for neural networks" JMLR16