Image-to-Image Translation

UNIT

NeurIPS

2017

Unsupervised Image-to-Image Translation Networks

Ming-Yu Liu, Thomas Breuel, Jan Kautz
NVIDIA
{mingyul,tbreuel,jkautz}@nvidia.com

Abstract

Unsupervised image-to-image translation aims at learning a joint distribution of images in different domains by using images from the marginal distributions in individual domains. Since there exists an infinite set of joint distributions that can arrive the given marginal distributions, one could infer nothing about the joint distribution from the marginal distributions without additional assumptions. To address the problem, we make a shared-latent space assumption and propose an unsupervised image-to-image translation framework based on Coupled GANs. We compare the proposed framework with competing approaches and present high quality image translation results on various challenging unsupervised image translation tasks, including street scene image translation, animal image translation, and face image translation. We also apply the proposed framework to domain adaptation and achieve state-of-the-art performance on benchmark datasets. Code and additional results are available in https://github.com/mingyuliutw/unit.

1 Introduction

Many computer visions problems can be posed as an image-to-image translation problem, mapping an image in one domain to a corresponding image in another domain. For example, super-resolution can be considered as a problem of mapping a low-resolution image to a corresponding high-resolution image; colorization can be considered as a problem of mapping a gray-scale image to a corresponding color image. The problem can be studied in supervised and unsupervised learning settings. In the supervised setting, paired of corresponding images in different domains are available [8, 15]. In the

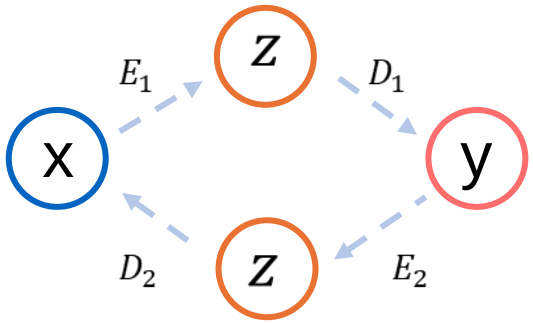
Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

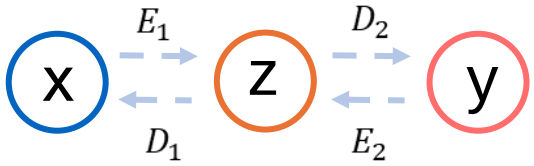
UNIT

Comparison of Approaches

CycleGAN



UNIT



Shared Latent Space

Image-to-Image Translation

UNIT

Background & Goal

Previous research limitation

- Inferring a joint distribution from separate individual distributions is challenging.

Goal

- To achieve high-quality image translation results based on a shared latent space assumption

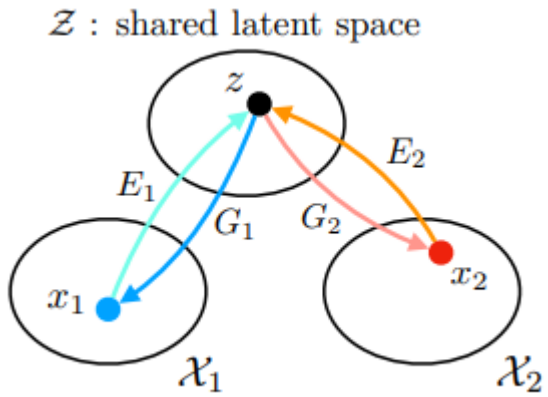


Fig 1. The shared latent space assumption

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Network Architecture

VAE + GAN

- The encoder maps the input image x to the latent space z, and the decoder reconstructs the image from z.
- VAE : enforces the latent space to follow a normal distribution.

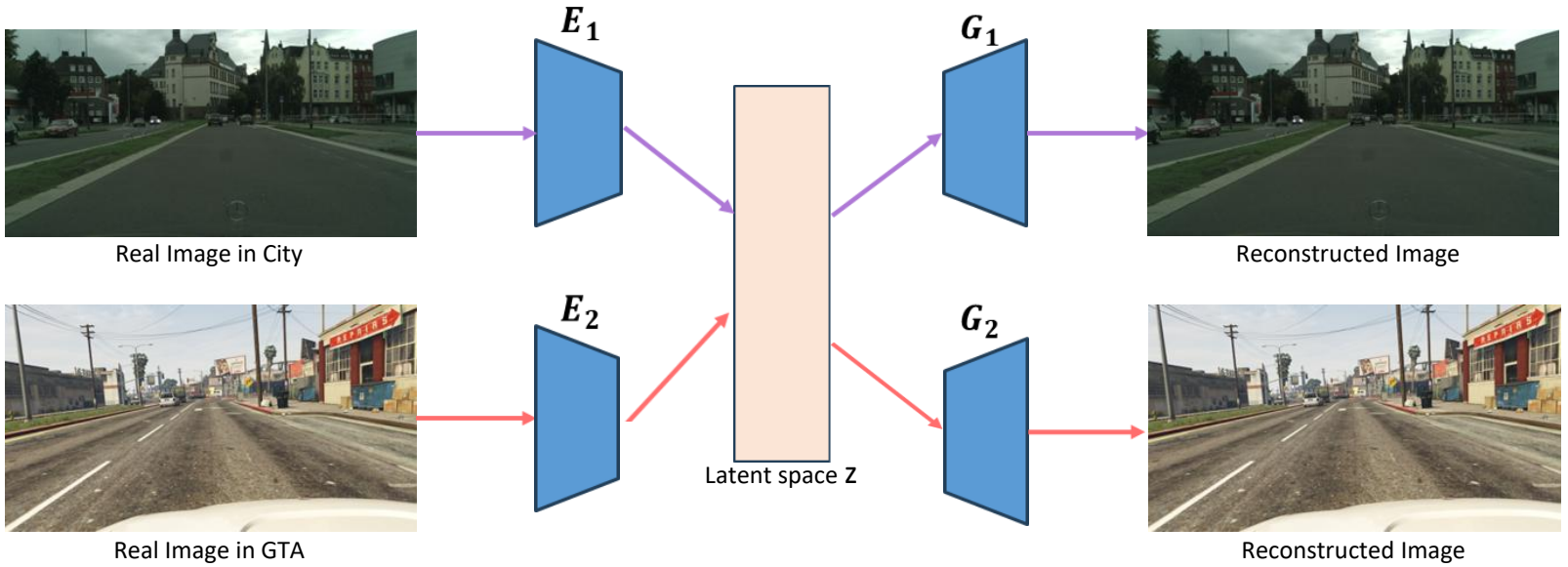


Fig 2. Overview of VAE in UNIT

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Network Architecture

VAE + GAN

- The encoder maps the input image x to the latent space z, and the decoder reconstructs the image from z.
- VAE : enforces the latent space to follow a normal distribution.
- GAN : ensures that the generated image appears as a real image in the corresponding domain.

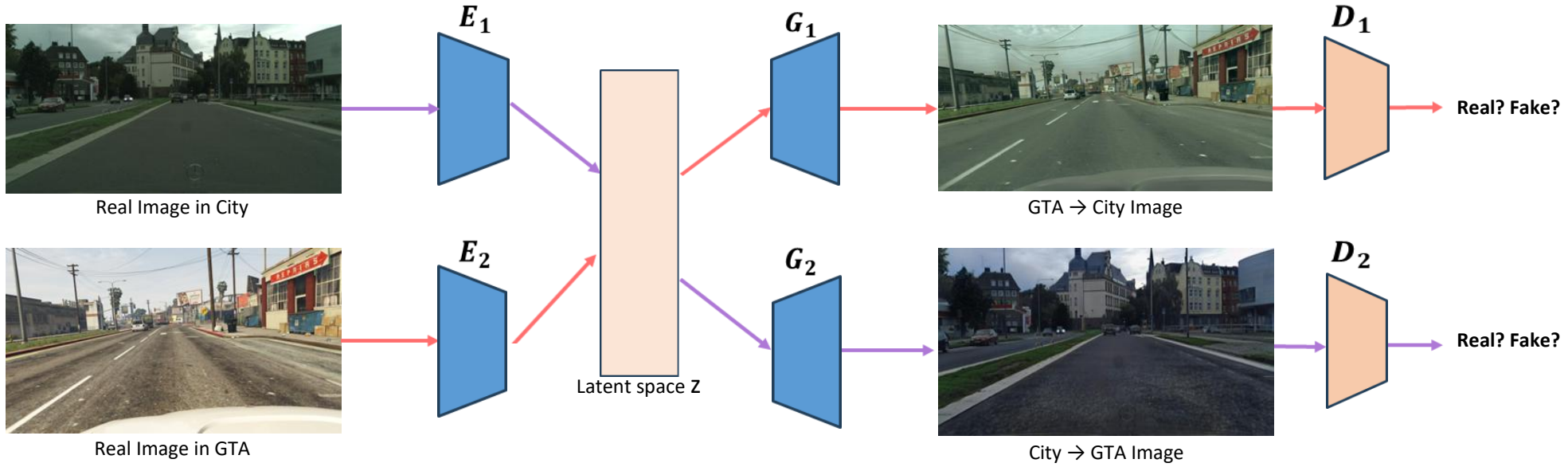


Fig 3. Overview of GAN in UNIT

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Network Architecture

Weight-sharing

- The last few layers of the two encoders and the first few layers of the two generators share weights

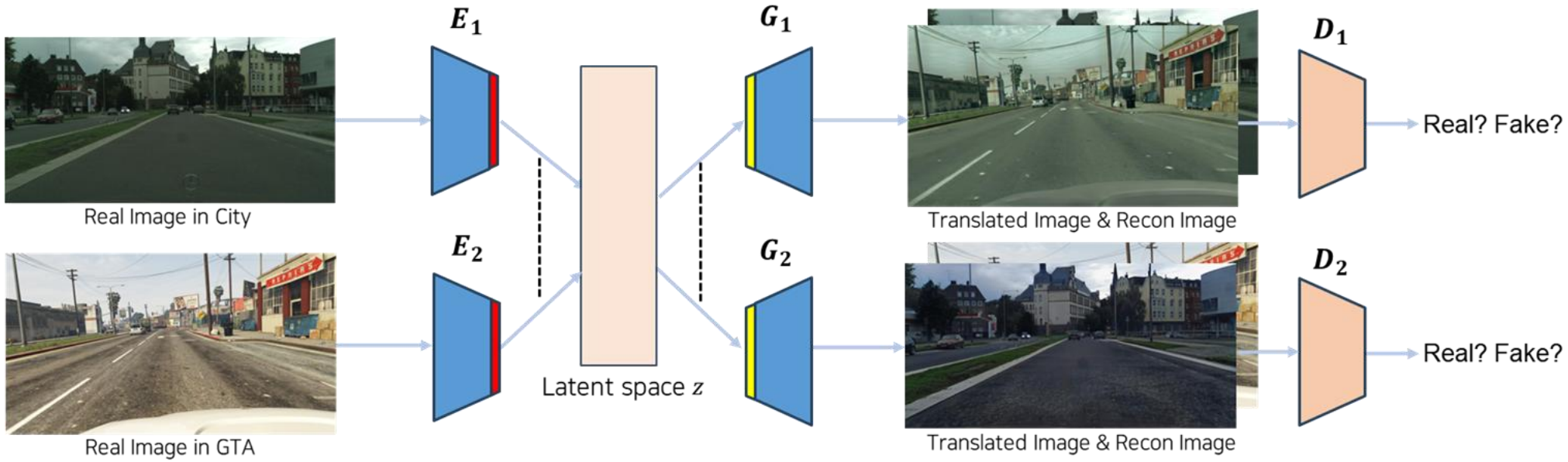


Fig 4. Weight sharing between several layers

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Loss Functions

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

prior distribution is a zero mean Gaussian $p_\eta(z) = \mathcal{N}(z|0, I)$

Eq 1. VAE loss

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

Eq 2. GAN loss

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

Eq 3. VAE-based cycle consistency loss

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Experiment Results

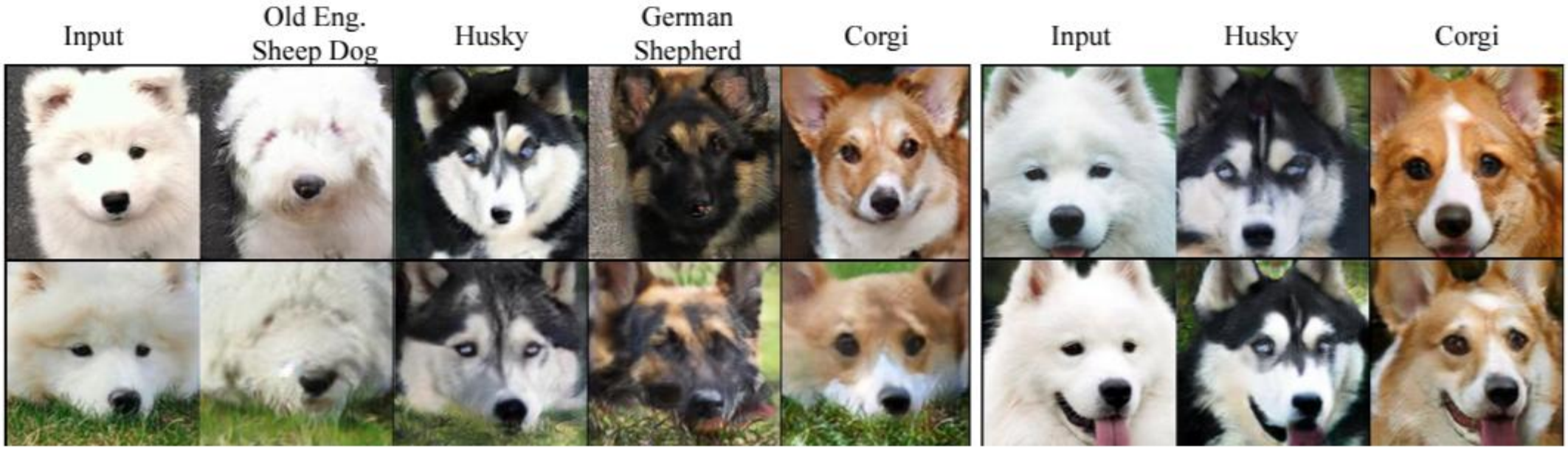


Fig 5. Example results of dog breed image translation

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.

Image-to-Image Translation

UNIT

Implications & Limitations

Implications

- A framework for unsupervised image-to-image translation without paired training images.
- Aligning different domains by mapping them to a shared latent space.

Limitations

- The current framework is unimodal, allowing only one-to-one mapping.
- The strict constraint of a shared latent space makes translating between highly different images increasingly challe nging.