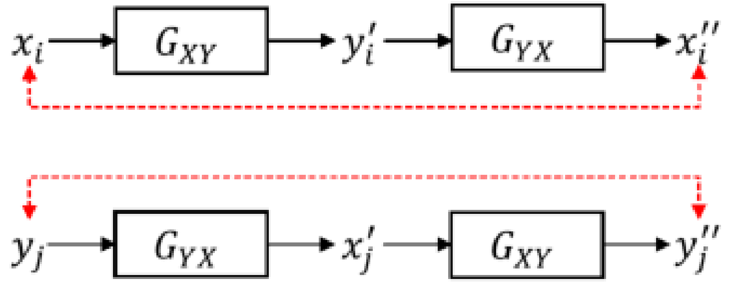
Unsupervised Image-to-image translation

段永杰

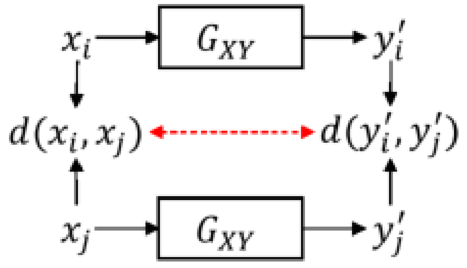
2020/7/30

无监督图像转换

- 面临的问题
  - 缺少成对图像信息（无监督）
  - 数学模型为根据两个边缘分布估计联合分布，是多解问题（ill-posed）
  - 无监督信号带来的内容及结构约束缺失
- 思路
  - CycleGAN —— 双向生成约束
  - DistanceGAN —— 通过归一化信号去除域的影响
  - UNIT —— 投影至同一隐空间进行约束
  - MUNIT —— 解决无监督下“一对多”生成问题
  - GcGAN —— 解决无监督下学习几何结构特征问题
  - TransGaGa —— 解决无监督下大差异域之间的转换问题



cyclic reconstruction
for cycle consistency



preserving $d(\cdot)$
for distance consistency

Unsupervised image-to-image translation networks

Ming-Yu Liu, Thomas Breuel, Jan Kautz

NVIDIA

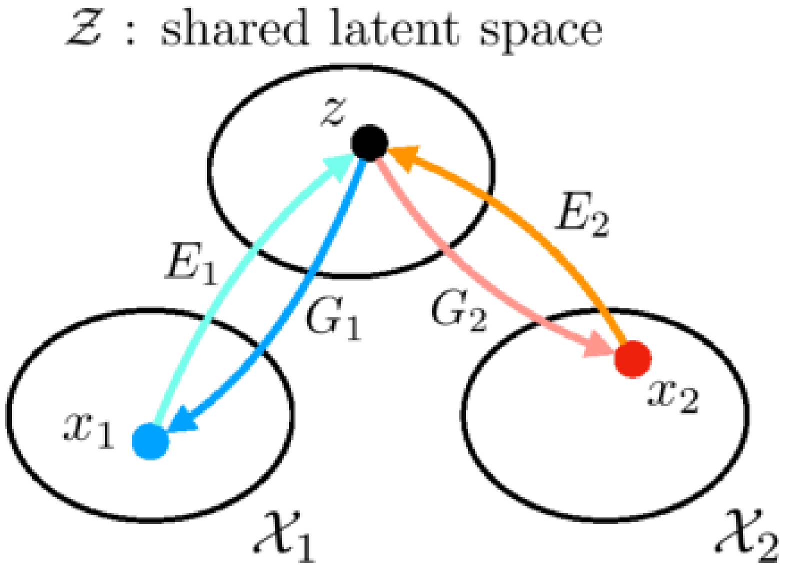
NIPS 2017

方法

- 要解决的问题
  - 缺少成对的训练数据

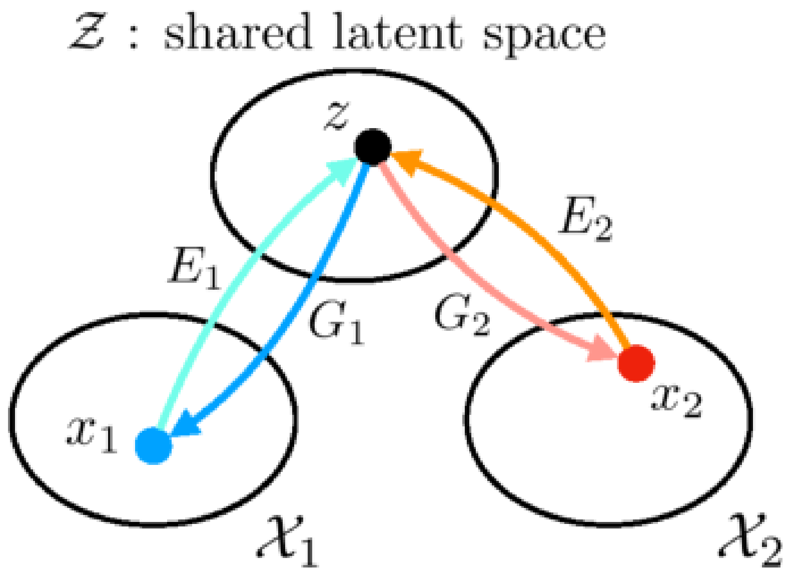
- 思路
  - 假设存在共享隐空间Z
  - 采取cycle-consistency约束
  - 利用变分自编码器（VAE）引入条件约束

- 符号
  - X_1, E_1, G_1, D_1
  - X_2, E_2, G_2, D_2

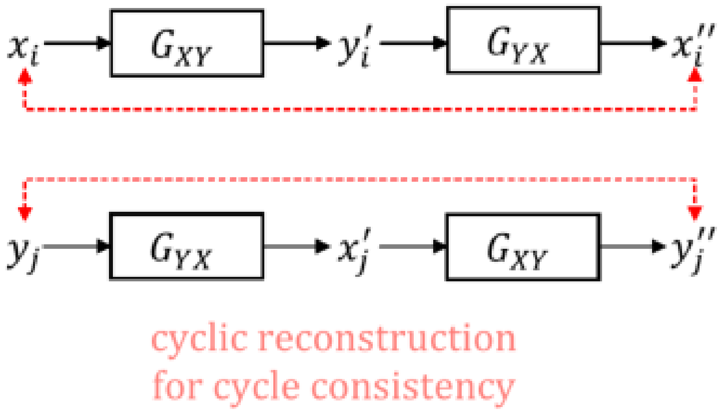


方法

- 对于成对数据(x_1, x_2)
  - $z = E_1^*(x_1) = E_2^*(x_2)$
  - $x_1 = G_1^*(z)$， $x_2 = G_2^*(z)$
  - $x_2 = F_{1\to2}^*(x_1) = G_2^*(E_1^*(x_1))$
  - $x_1 = F_{2\to1}^*(x_2) = G_1^*(E_2^*(x_2))$



- 类似cycle-consistency loss，但 cycle loss 中缺少同一隐空间的约束
  - $x_1 = F_{2\to1}^*(F_{1\to2}^*(x_2))$
  - $x_2 = F_{1\to2}^*(F_{2\to1}^*(x_1))$



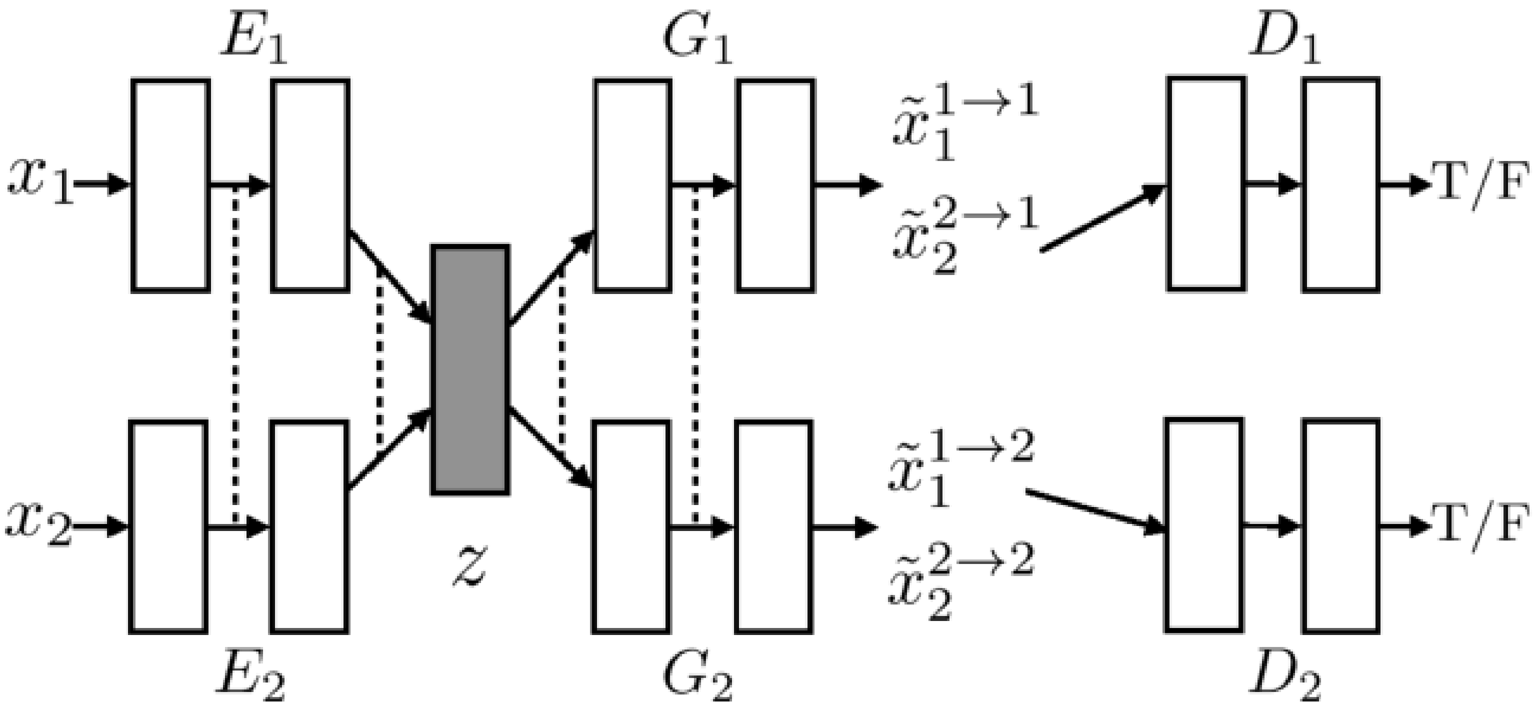
方法

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

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

- 推广至非成对数据
  - 两路分支同时训练，部分权值共享
  - 隐变量 z 服从变分自编码器中的假设：标准方差的正态先验

1. VAE中对噪声的鲁棒性
2. 同一隐空间的约束



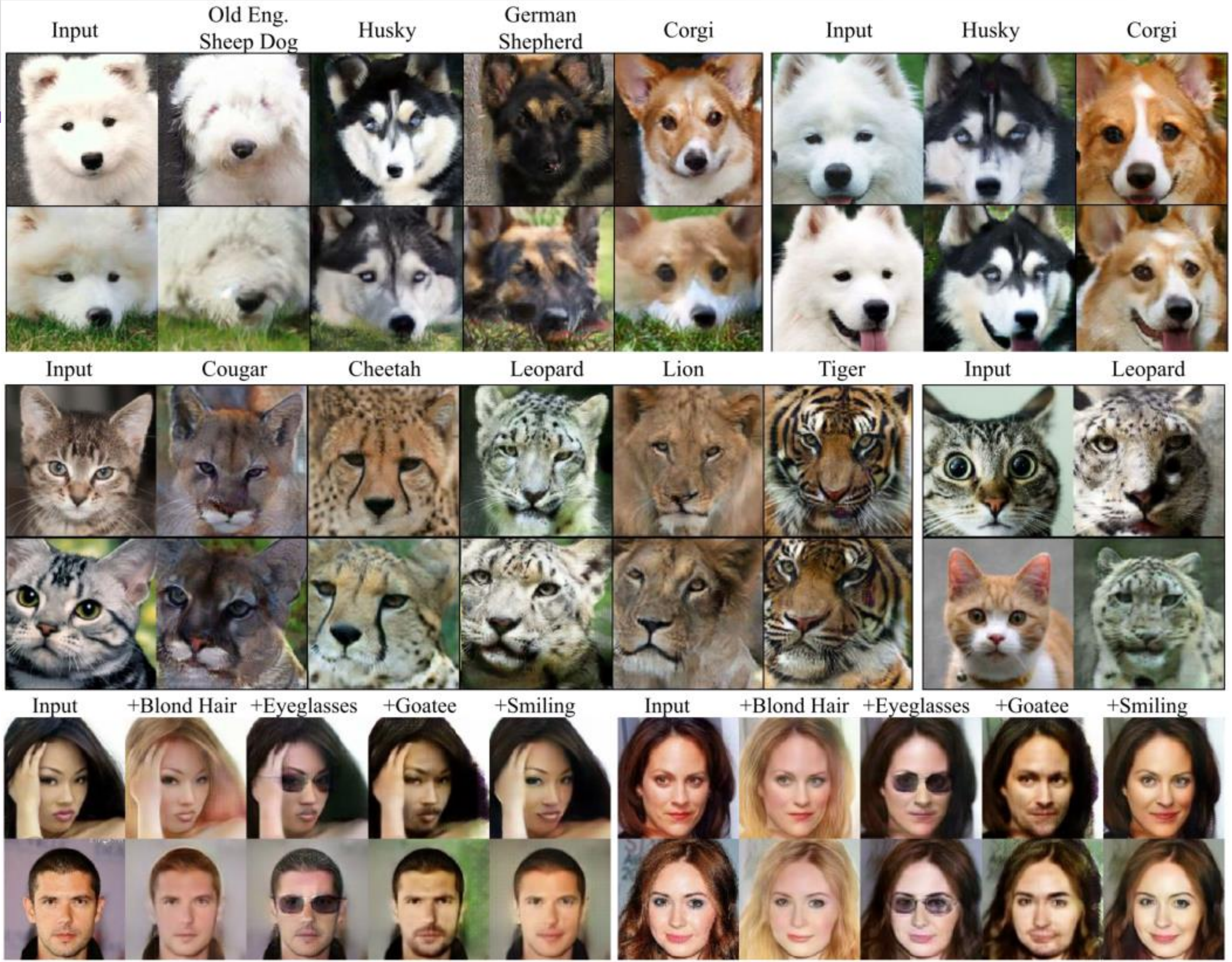
$$\min_{E_1,E_2,G_1,G_2} \max_{D_1,D_2} \mathcal{L}_{\text{VAE}_1}(E_1,G_1) + \mathcal{L}_{\text{GAN}_1}(E_1,G_1,D_1) + \mathcal{L}_{\text{CC}_1}(E_1,G_1,E_2,G_2)$$
$$\mathcal{L}_{\text{VAE}_2}(E_2,G_2) + \mathcal{L}_{\text{GAN}_2}(E_2,G_2,D_2) + \mathcal{L}_{\text{CC}_2}(E_2,G_2,E_1,G_1).$$

结果

- 街景
  - 白天/黑夜
  - 晴天/雨天
  - 夏天/冬天
  - 现实/模拟

结果

- 动物品种
- 人脸配饰


Input
Old Eng. Sheep Dog
Husky
German Shepherd
Corgi
Input
Husky
Corgi
Input
Cougar
Cheetah
Leopard
Lion
Tiger
Input
Leopard
Input
+Blond Hair
+Eyeglasses
+Goatee
+Smiling
Input
+Blond Hair
+Eyeglasses
+Goatee
+Smiling

结果

- 域适应（Domain adaption）
  - Source domain：训练，有标注
  - Target domain：测试，无标注

SVHN

MNIST

USPS

1. SVHN: Street View House Number
2. USPS: Normalized handwritten digits scanned from envelopes by the U.S. Postal Service

结果

- 方案
  - 训练域之间的转换
  - 使用source domain中判别器（D）的特征训练分类任务

Table 2: Unsupervised domain adaptation performance. The reported numbers are classification accuracies.

Method	SA [4]	DANN [5]	DTN [26]	CoGAN	UNIT (proposed)
SVHN→ MNIST	0.5932	0.7385	0.8488	-	**0.9053**
MNIST→ USPS	-	-	-	0.9565	**0.9597**
USPS→ MNIST	-	-	-	0.9315	**0.9358**

Multimodal unsupervised image-to-image translation

Xun Huang[1] , Ming-Yu Liu[2], Serge Belongie[1], Jan Kautz[2]

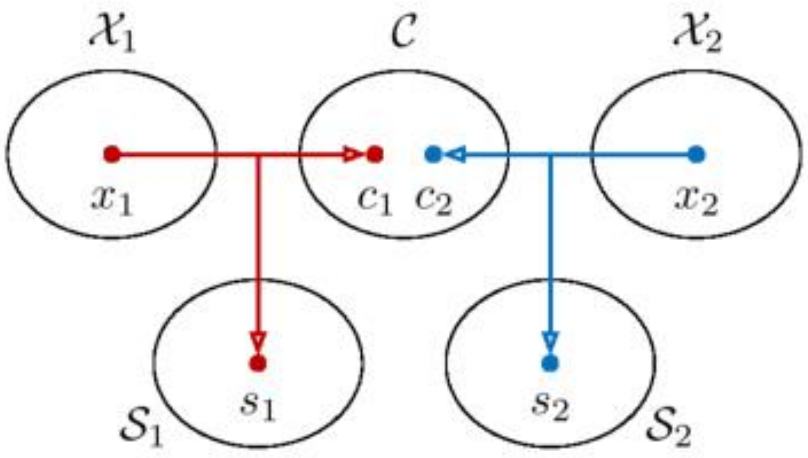
1. Cornell University
2. NVIDIA

方法

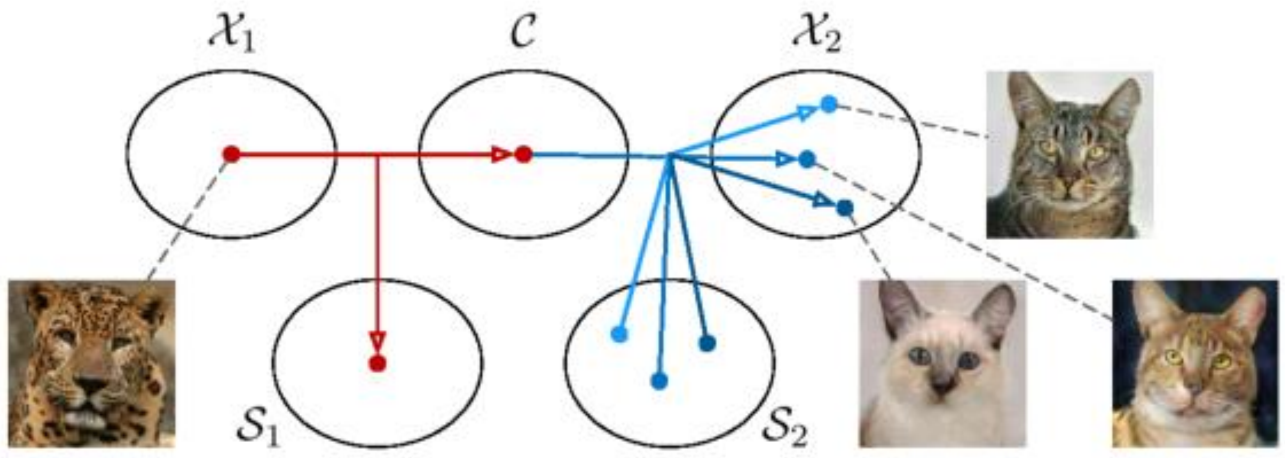
- 要解决的问题
  - 一对多：根据原始图像生成不同风格的目标图像（区别于多域图像转换问题）

例如：夏天 -> 冬天，但同时会受到光照、时间、天气等影响

- 思路
  - 将隐空间进一步划分为内容（content code）空间和风格（style code）空间
  - 内容空间共享，风格空间独立



(a) Auto-encoding

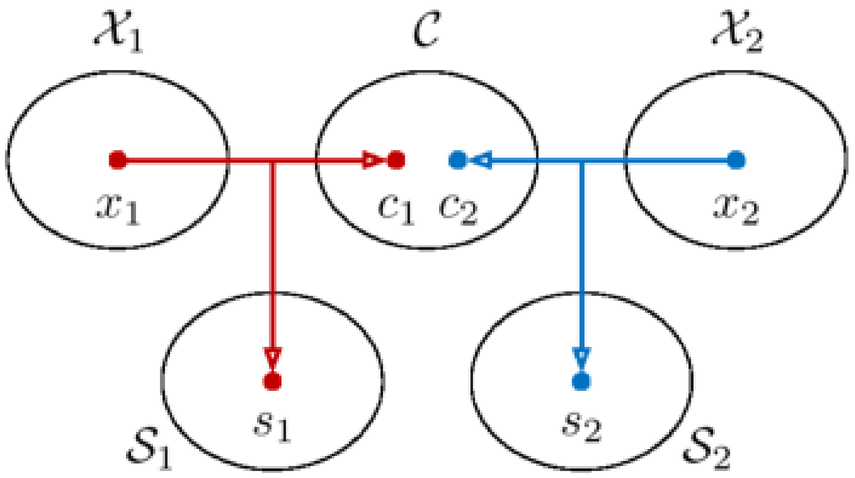


(b) Translation

方法

- 对于成对数据(x_1, x_2)
  - $c = E_1^{c*}(x_1) = E_2^{c*}(x_2)$
  - $x_1 = G_1^*(c, s_1)$
  - $x_2 = G_2^*(c, s_2)$

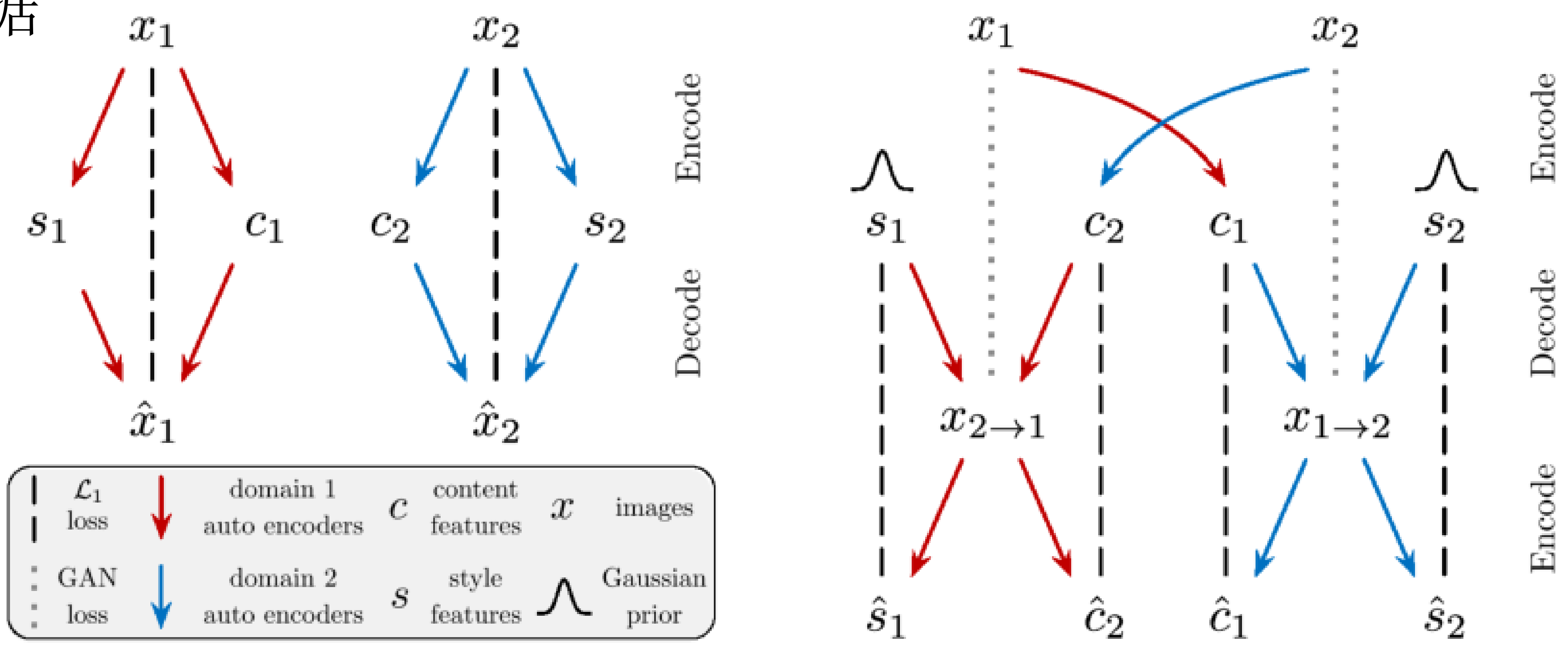
  - $x_2 = G_2^*\left(E_1^{c*}(x_1), E_2^{s*}(x_2)\right)$
  - $x_1 = G_1^*(E_2^{c*}(x_2), E_1^{s*}(x_1))$



- 隐空间仅有部分（content part）共享

方法

- 推广至非成对数据



(a) Within-domain reconstruction

(b) Cross-domain translation

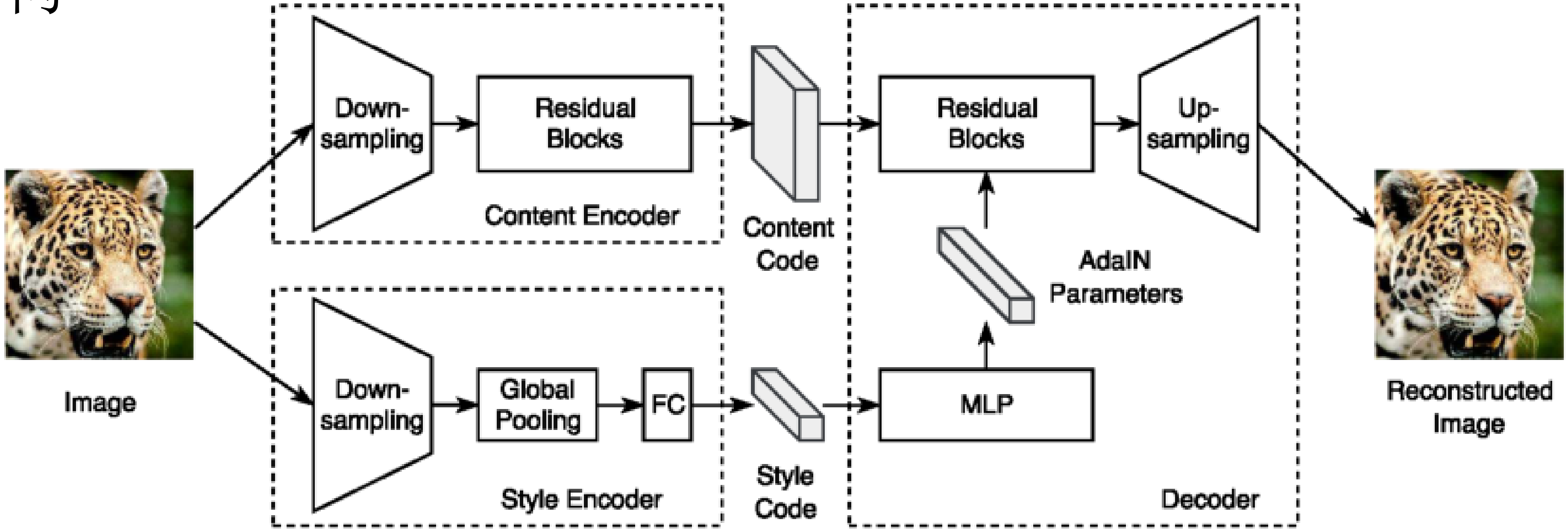
1. 图像重建
2. 内容编码重建
3. 风格编码重建

隐式地体现了 cycle-consistency 约束

$$\min_{E_1,E_2,G_1,G_2}\ \max_{D_1,D_2}\ \mathcal{L}(E_1,E_2,G_1,G_2,D_1,D_2) = \mathcal{L}^{x_1}_{\text{GAN}} + \mathcal{L}^{x_2}_{\text{GAN}} +$$
$$\lambda_x(\mathcal{L}^{x_1}_{\text{recon}} + \mathcal{L}^{x_2}_{\text{recon}}) + \lambda_c(\mathcal{L}^{c_1}_{\text{recon}} + \mathcal{L}^{c_2}_{\text{recon}}) + \lambda_s(\mathcal{L}^{s_1}_{\text{recon}} + \mathcal{L}^{s_2}_{\text{recon}})$$

方法

- 网络结构



$$\mathrm{AdaIN}(z,\gamma,\beta)=\gamma\left(\frac{z-\mu(z)}{\sigma(z)}\right)+\beta$$

方法

- 域不变感知损失函数（domain-invariant perceptual loss）
  - 一般的感知损失用于监督学习（成对数据）
  - 通过使用IN（instance normalization）归一化VGG特征，去除与域相关的特征，与原始输入图像的归一化特征进行距离计算

结果

- 轮廓

(a) edges ↔ shoes　　(b) edges ↔ handbags

结果

- 街景
  - 现实->游戏
  - 游戏->现实
  - 夏天->冬天
  - 冬天->夏天

结果

- 控制转换方向

(a) edges → shoes

(b) big cats → house cats

Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping

Huan Fu* [1], Mingming Gong* [2,3], Chaohui Wang[4], Kayhan Batmanghelich[2], Kun Zhang[3], Dacheng Tao[1]

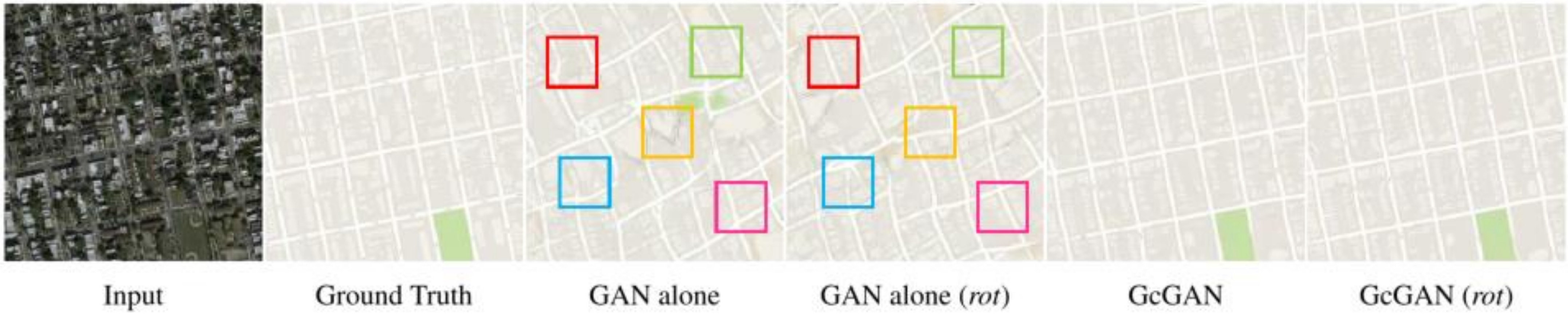
1. UBTECH Sydney AI Centre, School of Computer Science, FEIT, University of Sydney, Darlington, NSW 2008, Australia
2. Department of Biomedical Informatics, University of Pittsburgh
3. Department of Philosophy, Carnegie Mellon University
4. Universit´ e Paris-Est, LIGM (UMR 8049), CNRS, ENPC, ESIEE Paris, UPEM, Marne-la-Vall´ ee, France

CVPR 2019

方法

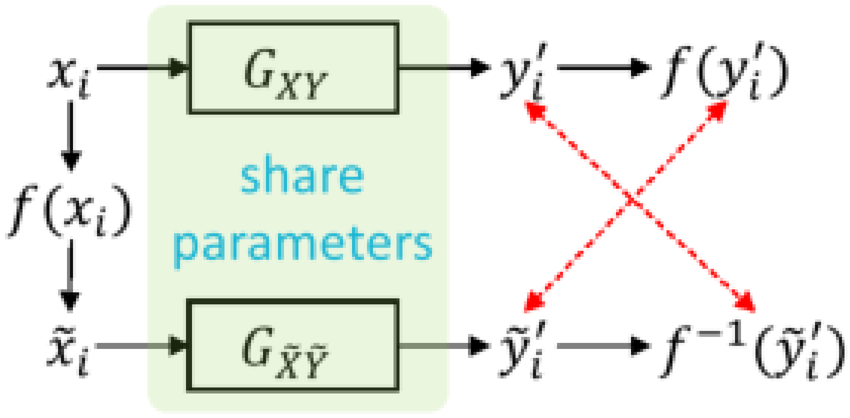
- 要解决的问题
  - 无监督学习时，生成图像时学习图像的几何（语义）特征

- 思路
  - 人为添加几何变换（无监督学习的思路）
  - 保证几何变换与生成变换的可交换性



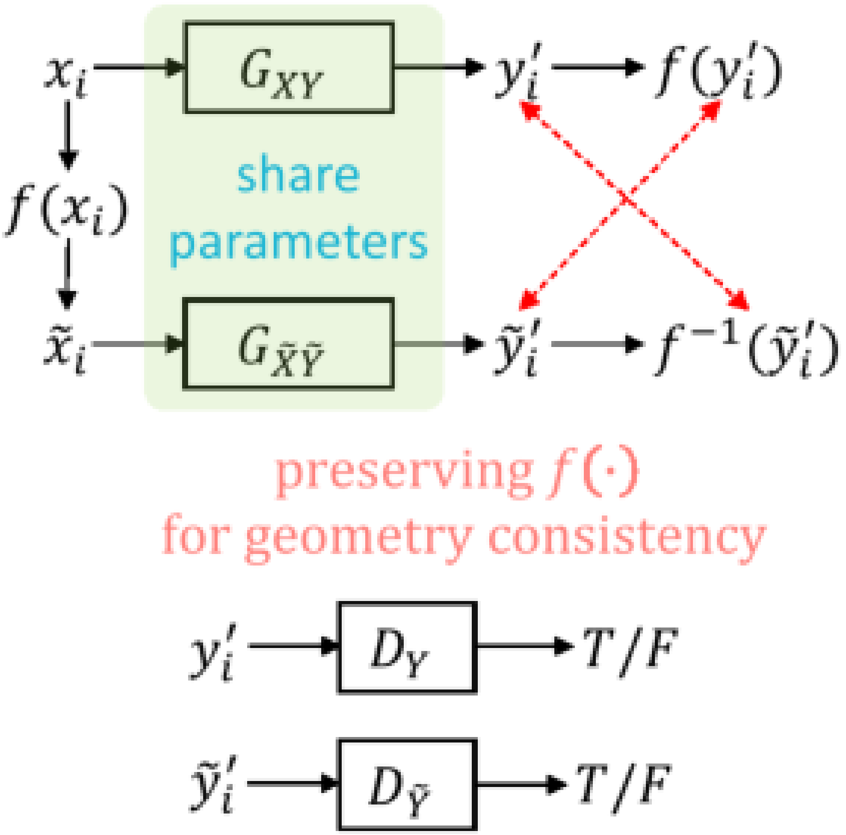
方法

- 学习几何结构特征
  - $f(x)$ —— 可逆变换：翻转、旋转
  - $X, Y, \tilde{X}, \tilde{Y}$
- 对于输入x_i
  - $y_i' = G_{XY}(x_i)$
  - $\tilde{y}_i' = G_{XY}(\tilde{x}_i) = G_{XY}\big(f(x_i)\big)$
  - $\tilde{y}_i' = f(y_i') = f\big(G_{XY}(x_i)\big) = G_{XY}\big(f(x_i)\big)$
  - $y_i' = f^{-1}(\tilde{y}_i') = f^{-1}\Big(G_{XY}\big(f(x_i)\big)\Big)$



方法

- 学习几何结构特征
  - G_{XY} 与 $G_{\tilde{X}\tilde{Y}}$ 共享网络结构及参数

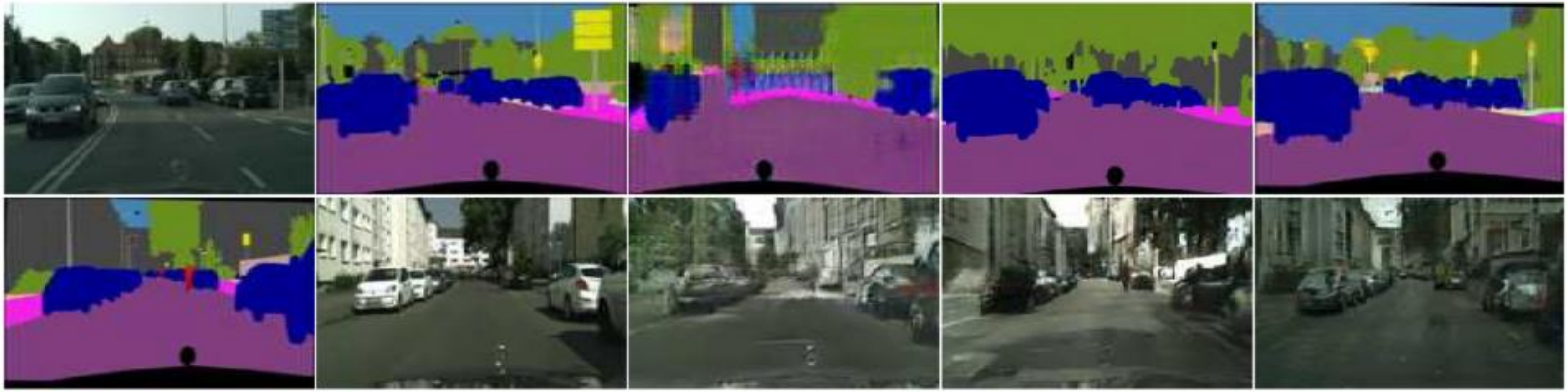


$$
\begin{aligned}
\mathcal{L}_{GcGAN} = & \ \mathcal{L}_{gan}(G_{XY}, D_Y, X, Y) \\
& + \mathcal{L}_{gan}(G_{\tilde{X}\tilde{Y}}, D_{\tilde{Y}}, X, Y) \\
& + \lambda \mathcal{L}_{geo}(G_{XY}, G_{\tilde{X}\tilde{Y}}, X, Y)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_{geo} = & \ \mathbb{E}_{x \sim P_X}[\|G_{XY}(x) - f^{-1}(G_{\tilde{X}\tilde{Y}}(f(x)))\|_1] \\
& + \mathbb{E}_{x \sim P_X}[\|G_{\tilde{X}\tilde{Y}}(f(x)) - f(G_{XY}(x))\|_1].
\end{aligned}
$$

结果

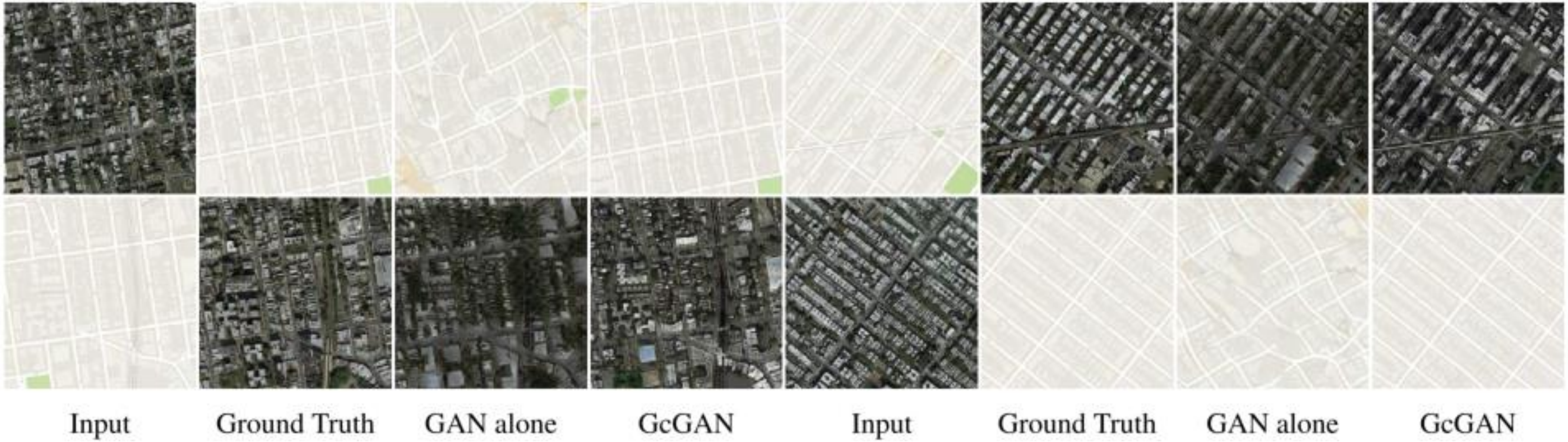
- 街景
  - Parsing -> image：对生成的图像进行场景解析（分割），分析分割结果
  - Image -> parsing：直接比较生成的分割图的性能



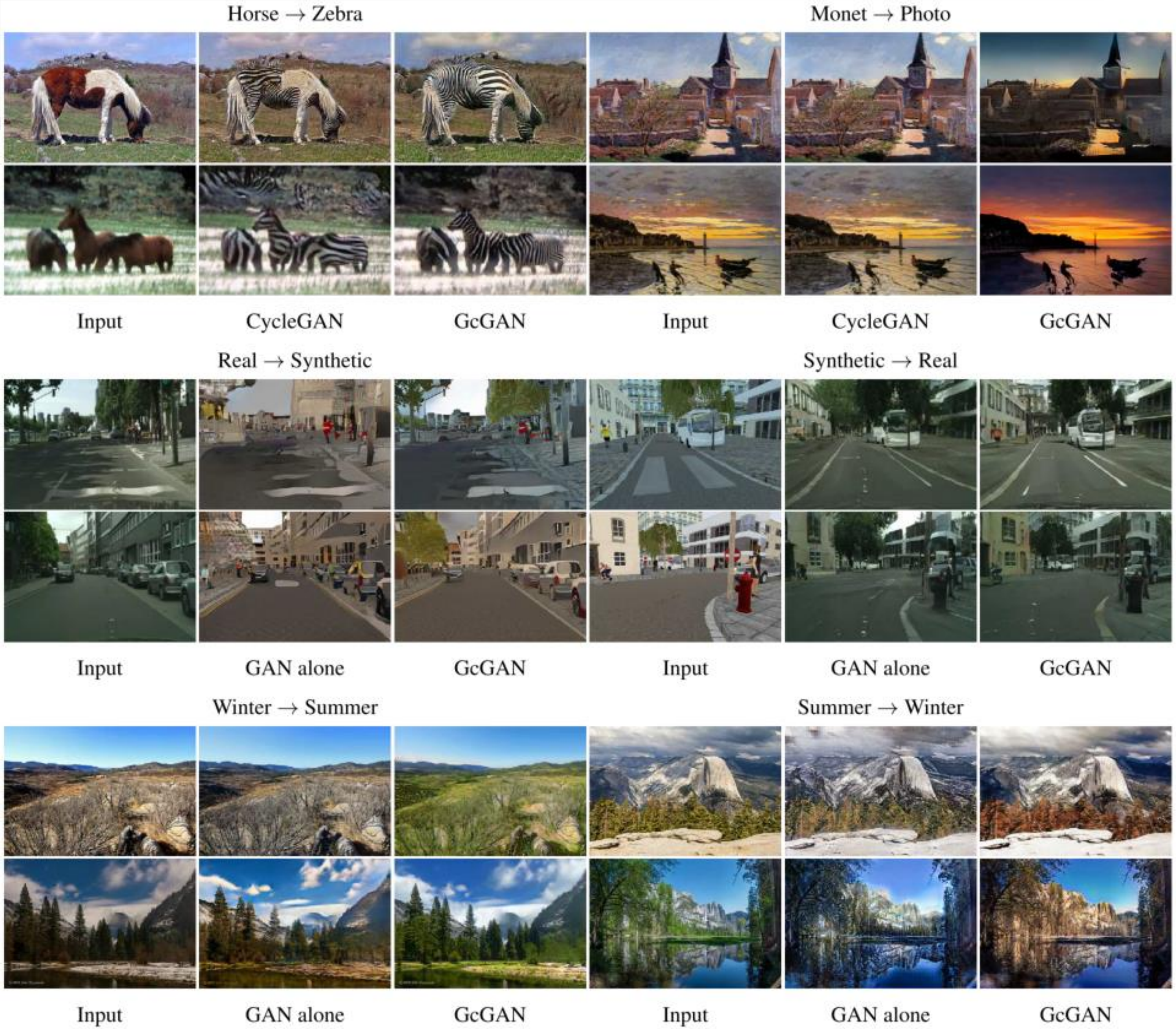
Input　Ground Truth　GAN alone　CycleGAN　GcGAN

结果

- 谷歌地图与航拍
  - 航拍 -> 地图：RMSE + pixel accuracy，RGB值的差异在设置阈值内
  - 地图 -> 航拍：仅定性分析



结果



Horse → Zebra
Monet → Photo
Input
CycleGAN
GcGAN
Input
CycleGAN
GcGAN
Real → Synthetic
Synthetic → Real
Input
GAN alone
GcGAN
Input
GAN alone
GcGAN
Winter → Summer
Summer → Winter
Input
GAN alone
GcGAN
Input
GAN alone
GcGAN

TransGaGa: Geometry-Aware Unsupervised Image-to-Image Translation

Wayne Wu[1], Kaidi Cao[2], Cheng Li[1], Chen Qian[1], Chen Change Loy[3]

1. SenseTime Research
2. Stanford University
3. Nanyang Technological University

CVPR 2019

方法

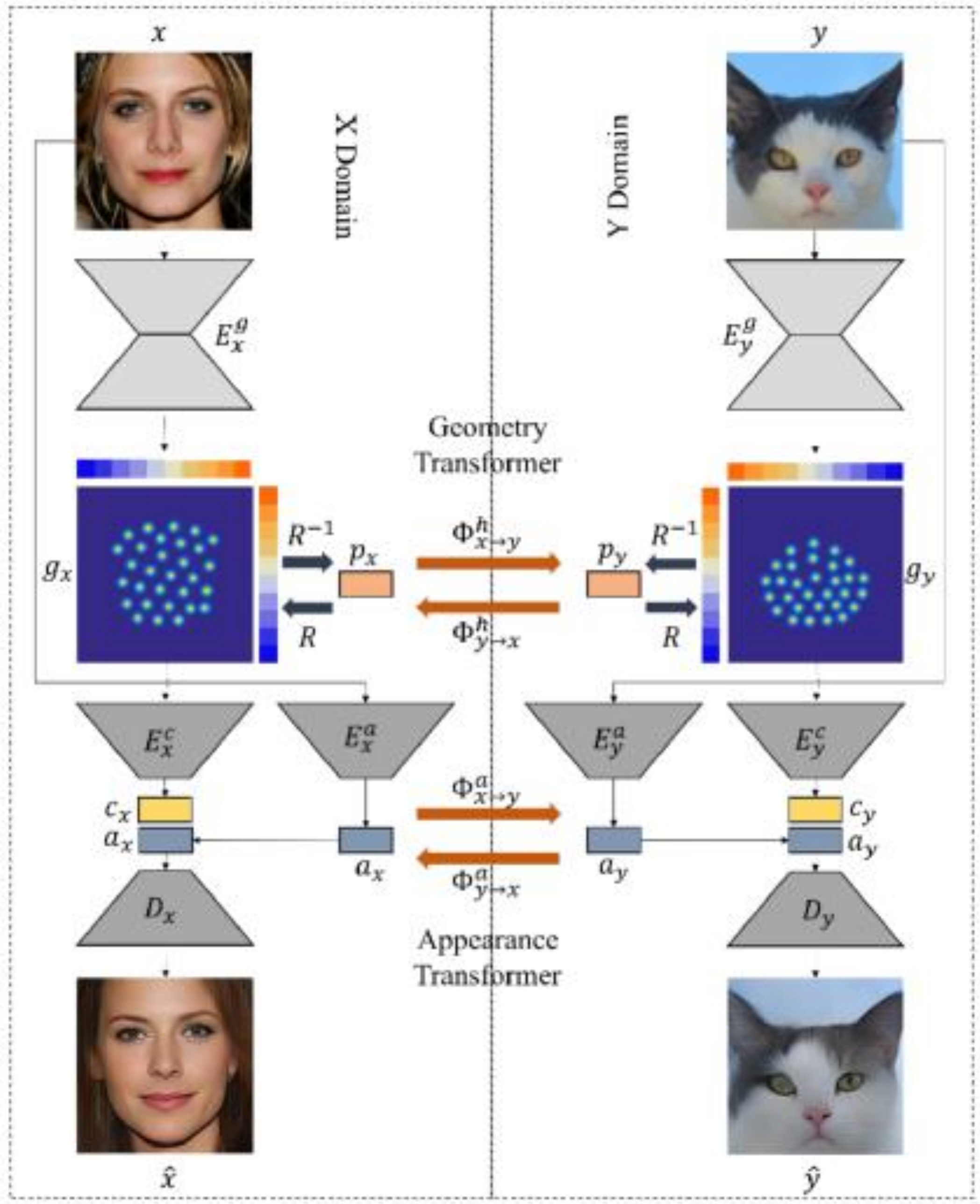
- 在两个具有较大差异的域之间进行转换
  - 例如：人脸 -> 动物，马 -> 长颈鹿
  - 局部纹理特征难以在原始图与生成图之间得到保留

- 思路
  - 间接：对隐空间进行分解（几何点集+表征空间）

方法

- 隐空间分解
  - 几何点集
    - E^g
    - g —— 热图（30个独立的通道）
    - c —— 几何编码
  - 表征空间
    - E^a
    - a —— 表征编码
- Conditional VAE

$$\mathcal{L}_{\text{CVAE}}(\pi,\theta,\phi,\omega) = -KL(q_\phi(c|x,g)||p(a|x)) + \|x - D(E^c(E^g(x)), E^a(x))\|,$$



方法

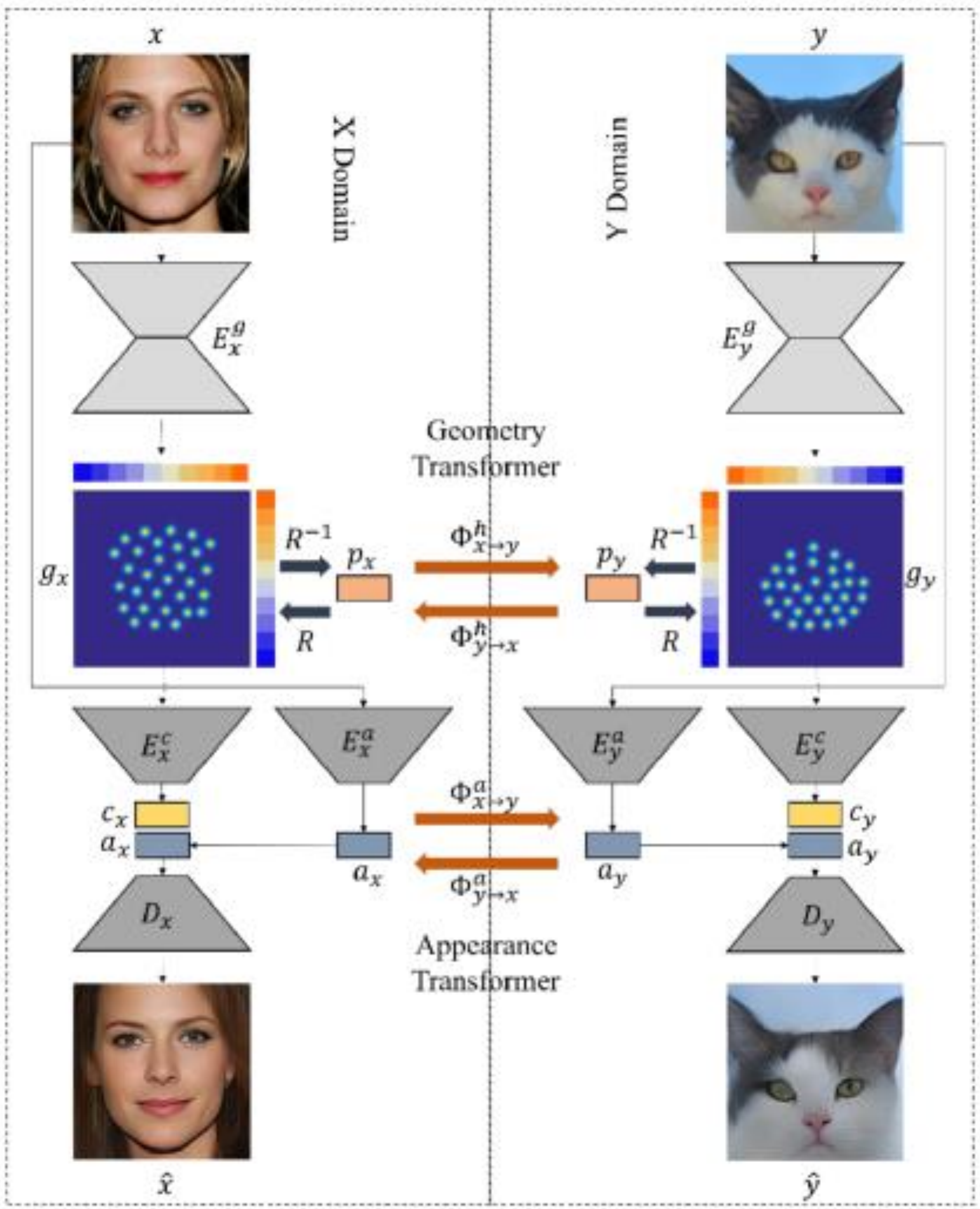
- 仅建模几何信息（热图）
  - 没有监督信息？[1]

$$\mathcal{L}\text{prior} = \sum_{i\neq j} \exp\left(-\frac{\|g^i - g^j\|^2}{2\sigma^2}\right) + \mathrm{Var}(g)$$

分离　　　　聚焦

- 几何变换
  - 计算热图一阶矩
  - 利用PCA去除噪声影响（embedding）
- 表征变换
  - 计算输入与生成图像的Gram矩阵

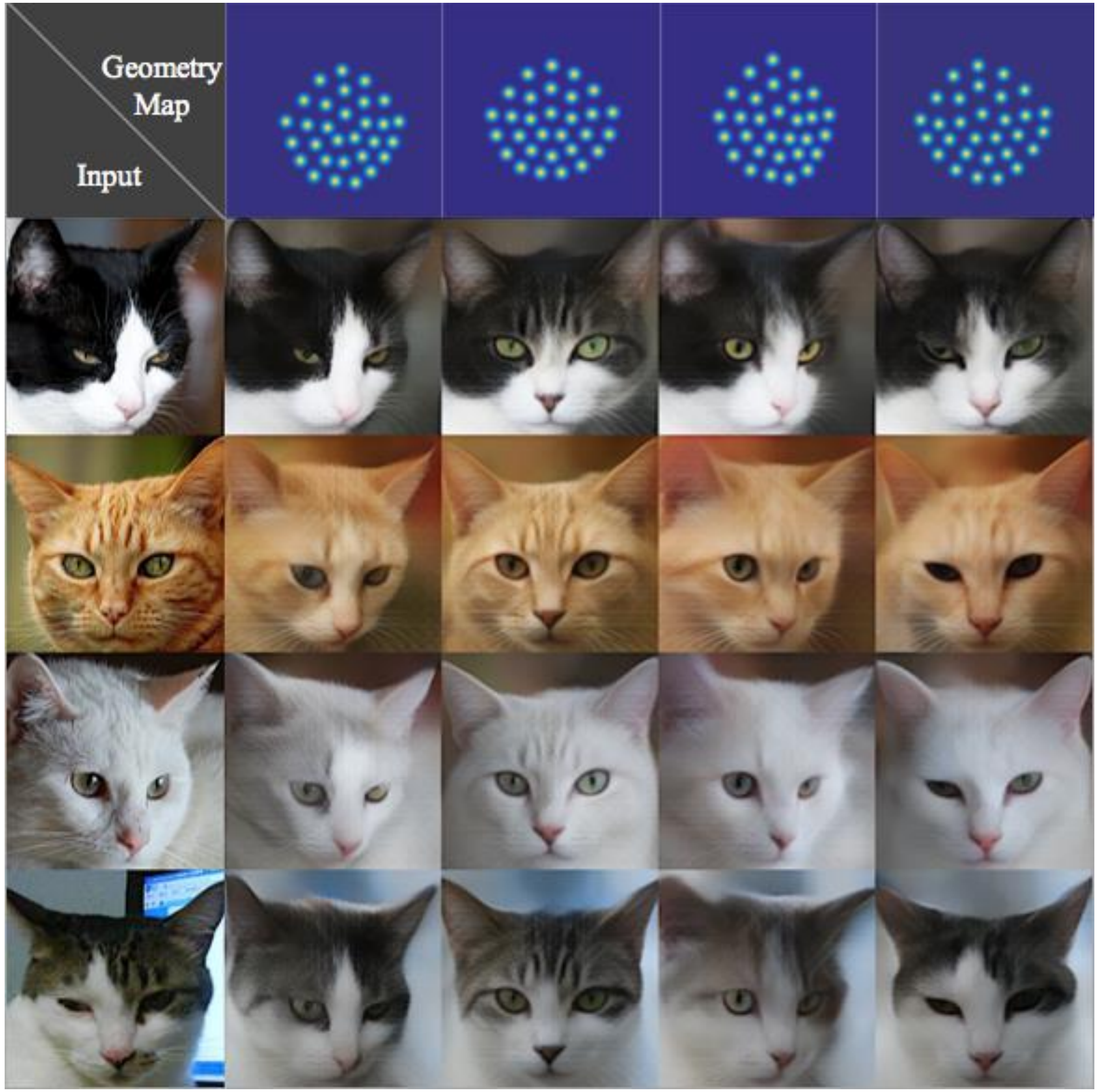
$$\mathcal{L}^a_{\text{con}} = \|\zeta(x) - \zeta(D_y(\Phi^g_{x\to y}\cdot E^g_x(x), \Phi^a_{x\to y}\cdot E^a_x(x)))\|$$



1. Zhang, Yuting, et al. "Unsupervised discovery of object landmarks as structural representations." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

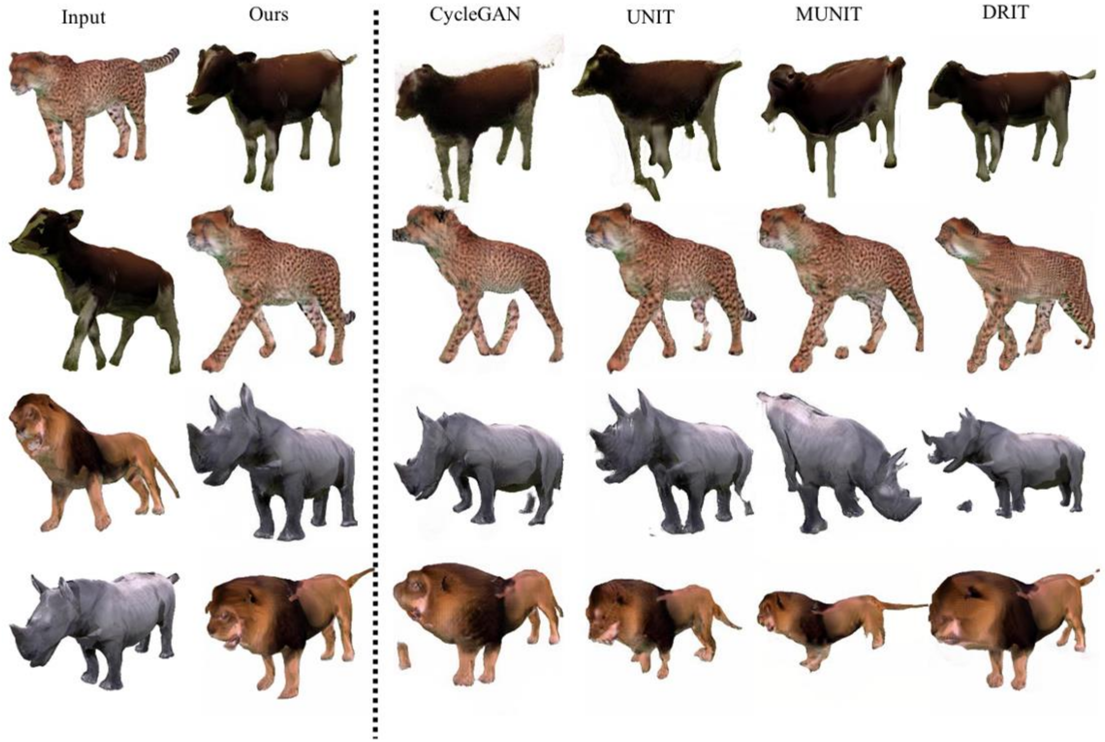
方法

- 多模式输出
  - 几何点集与表征特征结合
  - 采样

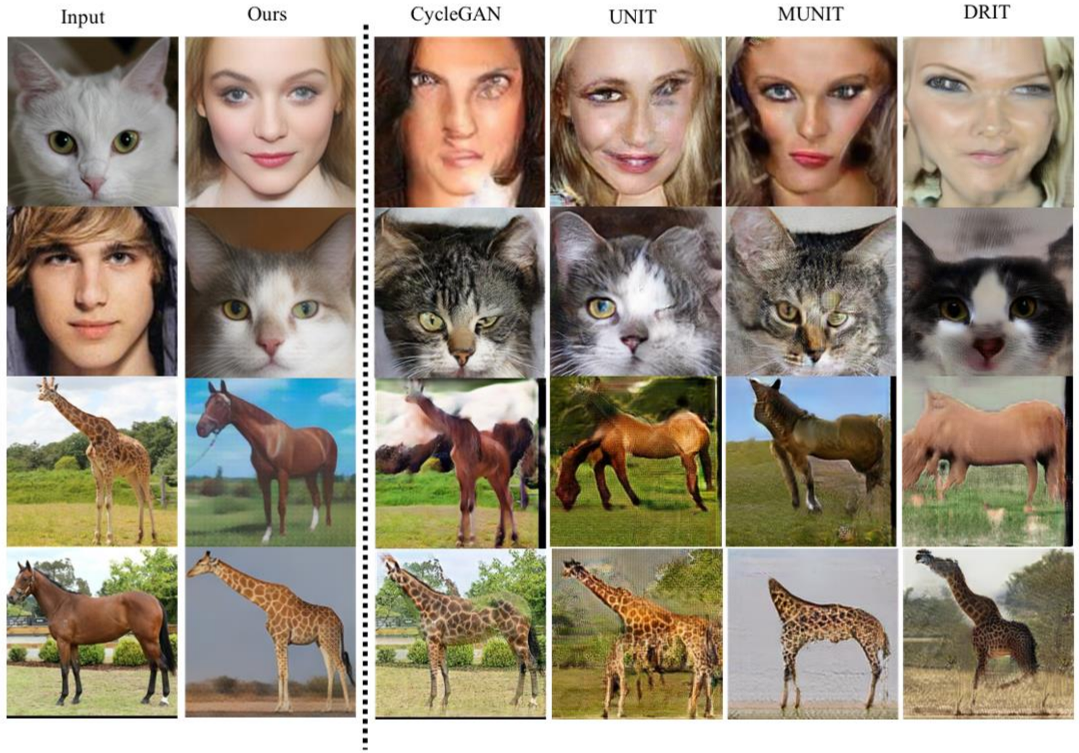


结果

- 对几何信息的保持



(a) cow↔cheetah and lion↔rhino



(b) cat↔human face and giraffe↔horse

结果

- 定性评价
  - 该算法结果是否比其他算法的结果更好？A/B/Not Sure

Method	**horse → giraffe** % Testers labeled *better*	**human → cat face** % Testers labeled *better*
CycleGAN [52]	11.9%	25.7%
UNIT [27]	16.5%	23.3%
MUNIT [14]	19.2%	31.7%
DRIT [23]	23.6%	34.4%
Ours	**50.0%**	**50.0%**

(b) Score of "geometry-consistency".

Method	**horse → giraffe** % Testers labeled *better*	**human → cat face** % Testers labeled *better*
CycleGAN [52]	15.0%	15.4%
UNIT [27]	19.3%	18.9%
MUNIT [14]	20.4%	17.8%
DRIT [23]	16.1%	23.4%
Ours	**50.0%**	**50.0%**

(a) Score of "realism".

结果

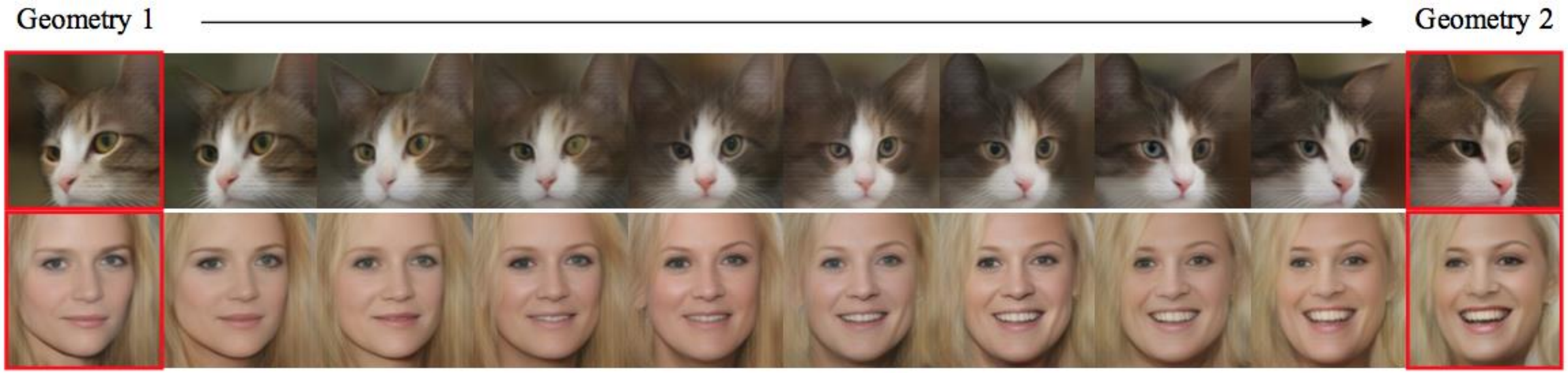
- Exemplar-guided generation
  - 保留结构 + 替换表征

(a) Cats to Human Faces

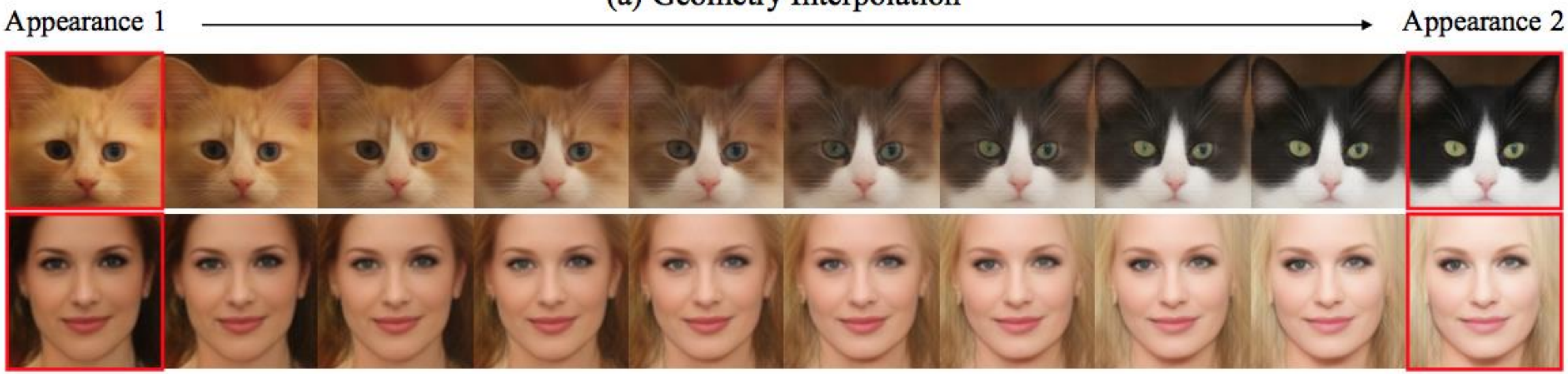
(b) Dogs to Cats

结果

- 隐空间插值



(a) Geometry Interpolation



(a) Appearance Interpolation

- 参考文献

  - Liu, Ming-Yu, Thomas Breuel, and Jan Kautz. "Unsupervised image-to-image translation networks." Advances in neural information processing systems. 2017.
  - Huang, Xun, et al. "Multimodal unsupervised image-to-image translation." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
  - Fu, Huan, et al. "Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
  - Wu, Wayne, et al. "Transgaga: Geometry-aware unsupervised image-to-image translation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.