

Pix2Pix

摘要

1. 研究条件GAN网络在图像翻译任务中的通用解决方案
2. 网络不仅学习到从输入图像到输出图像的映射，还学习用于训练该映射的损失函数
3. 证明了这种方法可以有效应用在图像合成、图像上色等多种图像翻译任务上
4. 使用作者发布的pix2pix软件，大量用户已经成功进行了自己的实验，进一步证明了方法的泛化性
5. 表明可以在不手工设计损失函数的情况下，也能获得理想的结果

研究背景

数字图像任务：

- 计算机视觉

模仿人眼和大脑对视觉信息的处理和理解。

图像分类，目标检测，人脸识别

- 计算机图像学

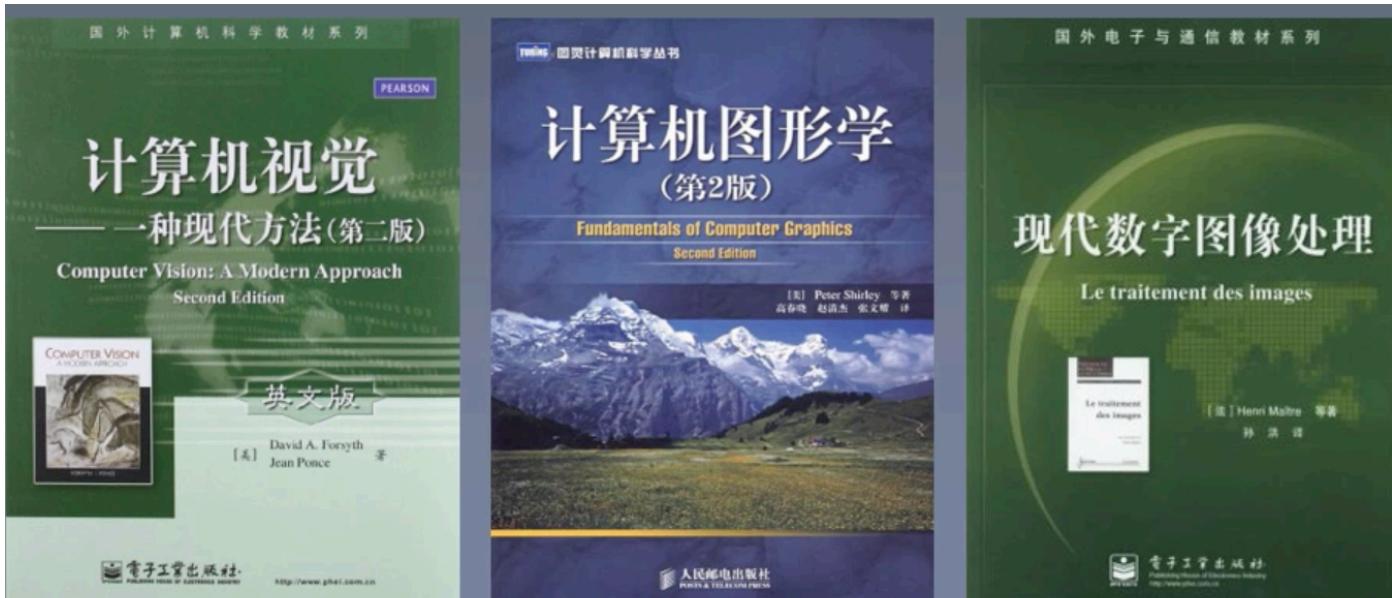
在数字空间中模拟物理世界的视觉感知

动画制作，3D建模，虚拟现实

- 数字图像处理

依据先验知识，对图像的展现形式进行转换

图像增强，图像修复，相机ISP



	Input	Output	Image	Knowledge
Image			Digital Image Processing	Computer Vision
Knowledge			Computer Graphics	Artificial Intelligence

图像翻译

图像与图像之间以不同形式的转换。根据source domain的图像生成target domain中的对应图像，约束生成的图像和source图像的分布在某个维度上尽量一致

- 图像修复
- 视频插帧
- 图像编辑
- 风格迁移
- 超分辨率

图像质量评价 (image Quality Assessment, IQA)

有参考图像损失

- 像素损失。 MSE -> PSNR (加了log)
- 结构性损失 SSIM
- 色彩损失
- 锐度损失. GMSD
- 感知损失.



数据集

CMP Facade Database

Paris StreetView Dataset

Cityscapes Dataset

CMP Facade Database

- 由捷克理工大学的机器感知中心（CMP）发布，包含606张建筑正面的校正图像，来自世界各地的不同城市，包含12类语义分割标注

<http://cmp.felk.cvut.cz/~tylecr1/facade/>

Paris StreetView Dataset

- 由牛津大学发布，包含6412张从Flickr上下载的包含巴黎标志性建筑的街景图像

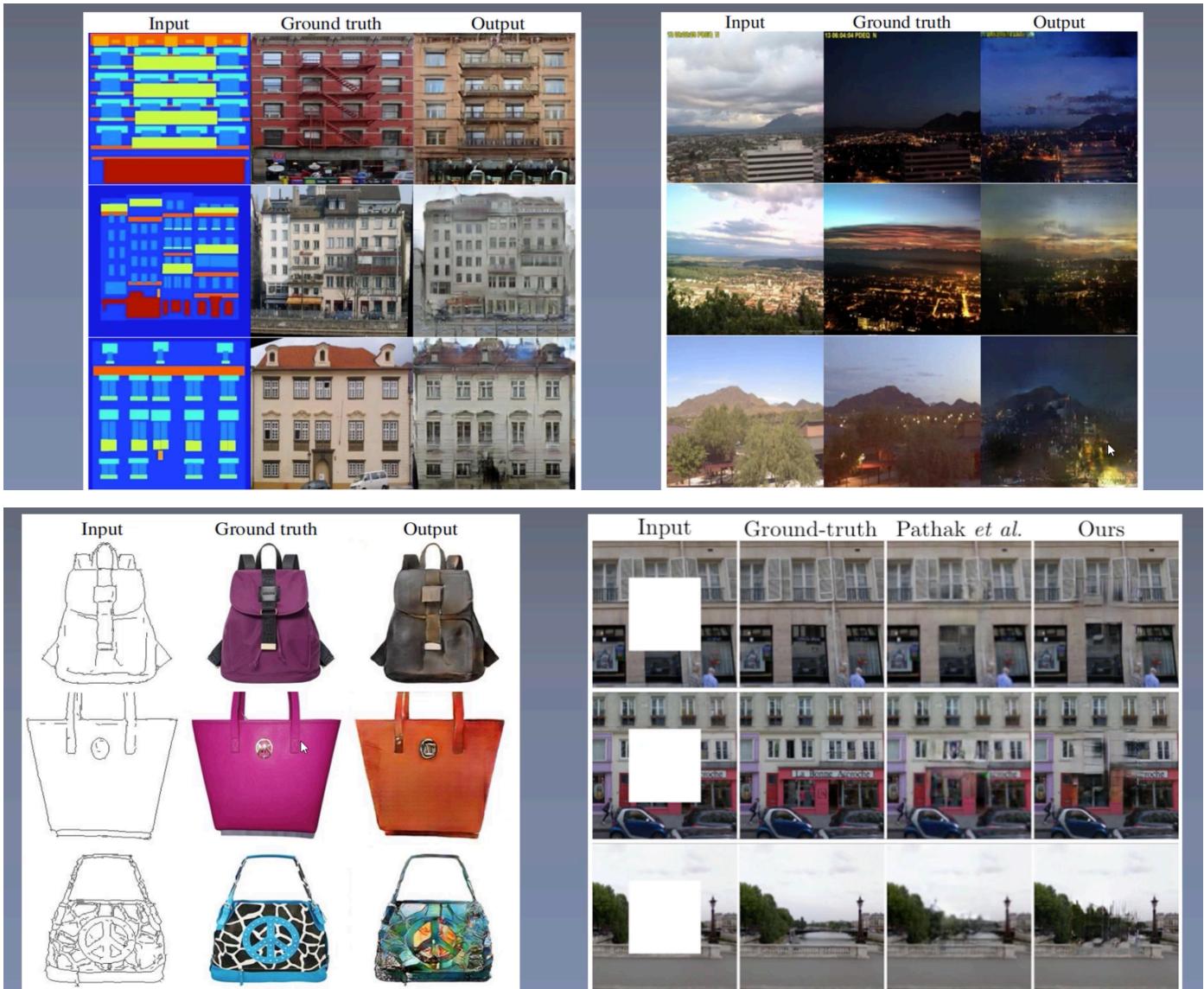
<http://www.robots.ox.ac.uk/~vgg/data/parisbuildings/>

Cityscapes Dataset

- 由德国三个研究机构联合发布的城市景观数据集，拥有5000张带语义理解标注的城市街景图像

<https://www.cityscapes-dataset.com/>

研究成果



研究意义

给出了图像翻译任务的通用框架，对于不同类型的任务，不需要一个特定的算法和损失函数

使GAN从此统治了各类图像翻译任务，促使各类图像翻译任务的效果逐渐接近了实用级别