

# 工作汇报

Shilong Zhang

9.10至 9.23

- 1.因为最后张老师决定做动作迁移，读了几篇与此相关的文章。
- 2.因为要用GAN，补了的GAN相关的知识。
- 3.学习了一些机器学习理论。

## 1 Paper Reading

### 1.1 Everybody Dance Now

#### 1.1.1 Abstreact

Transfer motion between human subjects in dierent videos.

#### 1.1.2 Method overview

pose detection  $\rightarrow$  global pose normalization  $\rightarrow$  mapping from normalized pose stick fgures to the target subject.

**Pose detection** Using a pretrained pose detector to get the representation of resuilting pose stick figure.

**Global pose normalization** calculate the scale and translation in the y direction (具体细节? )

## Adversarial Training of image to image Translation

### 1. pix2pixHD framework

D由三个不同规模的描述符构成，其应该趋向于真实的x与y打高分，x与G(x)打低分，即最大化 $L_{GAN}$

G则希望以假乱真，其希望D给G(x)打高分，同时最小化两个损失 $L_{FM}$ ,D给出的一个feature-matching loss,以及一个VGGNET给出的重建误差。

### 2. Temporal smoothing

改进G，其input变为这一次的pose stick figure( $x_{t-1}$ ) + 上次的 $G(x_{t-1})$ 来生成下一帧，而D则两帧一起评分。

### 3.Face GAN

To add more detail and realism to the face region.

有一点类似于残差网络的思想，加一个恒等映射，那么经过与D的对抗，生成的至少不会比映射前的差。

### 1.1.3 Experiments

质量评价中是让生成的图像用Detector检测出关键点，算这些关键点与原始pose的距离，为了避免有些关键点没有检测到的问题，只算全检测到的。

## 1.2 Dense Pose Transfer

### Abstract

Surface-based pose estimation + deep generative models,a two-stream architecture 实现动作的迁移

warping module(基于UV贴图，包含较多纹理细节) + predictive module (Data drive)，具有互补的优点。→ Blending module

两个并行部分最开始都使用一个叫做Densepose的网络，他可以先把pixel分类到24个预先设定的身体表面模块，之后回归精确坐标 •

Densepose 网络结构？

## 1.3 Deep Vidio Portraits

两个视频间实现非常精密的人脸的迁移，不止表情，还包括角度，甚至眨眼。

**OVERVIEW** face reconstruction(使用参数化人脸以及光照模型) → 通过source视频生成这些parametric→use modified parametric synthetic rendering of target → these rendering serve as conditional input to a conditional GAN to get the photo-realistic output.

**亮点** 1. 如何快速重建人脸参数的算法 2. 如何设计的conditional GAN的输入以得到photo-realistic的输出。 3. 网络结构?

## 2 GAN

下面是我自己的一些理解。要求解决的问题是如何找到一个分布模拟真实的分布, 比如生成人脸, 在 $3 \times H \times W$ 如此高维度的空间中, 人脸的分布必然具有稀疏性, 即只有空间中的一小部分可以称之为人脸, 而在高维空间中寻找这样的分布是困难的, GAN其实就是寻找这样一个分布的巧妙结构。

首先我们只有一个数据集, 比如人脸, 也就是说, 我们的任务就是找到这个数据集在高维空间中的分布 ((当然数据集要够大足够好的描述人脸的真实分布), 掌握了这个分布我们就可以自己生成近乎真实的人脸, 实际也就是我们如何从数据集 ((sampling)反解分布的问题。

我们思考GAN的结构,G的目的是将某个分布映射到近乎真实的分布,而A的任务就是告诉G, 你生成的好不好。训练过程中我们是分别训练的, 而最初G并不知道该如何映射, A也只知道数据集中的是好的(也就是真实分布的采样), 他并不知道, 我们先固定G, 他可能映射出一个与真实分布相差很大的分布, 但是就开始告诉A, G生成的不好, 真实从数据集中采样的数据的分布是好的,那么在高维概率空间中, A就知道了此次的G生成的概率分布就是不好的, 然后固定A, 调整G, 使得A无法分辨G与真实数据, 这个过程相当于改变了G的映射方式, 他会避开原来A觉得不好的分布, 生成一个新的分布, 我们再次告诉A这个分布相比真实数据也是不好的, 这样这个过程中, A对整个概率空间的掌握越来越多, 即它不仅知道真实分布是好的的们还知道了许多偏离真实分布的分布是坏的, 在这个过程中, G也不断地调整自己贴进A认为好的分布, 当G近乎真实分布时候, 两者同时收敛。

在此过程中, 很容易产生一个问题, 就是G在调整自己的时候是不是盲目的呢, 也即G在调整自己的映射时候, 是不是确实是向真实分布靠近

呢，在GAN原文中，设计出的loss,在调整完D后，对G来说损失函数正是当前G的映射分布与真实分布的某种距离度量，G最小化这个损失的过程就是与真实分布不断靠近的过程(注意，当G映射分布变化后，A的输出就不再是新分布的距离度量了，因此G的迭代次数一般较少)，这为训练的可收敛性提供了保证。

具体的结构：MMGAN，conditional GAN(条件生成，在loss中加上条件不配对的打低分，相当于G学习一个条件分布？，具体数学？)，Cycle GAN(无监督的学习方法)

### 3 机器学习理论

#### **PAC ((Probably Approximately Correct)Learning**

The learner: the strategy to select hypothesis from the hypothesis set

ML Pipeline: 如何根据Data Sampling 在 Hypothesis Set里面选择一个去逼近target function，这三个因素也是决定机器学习效果的最主要的三个因素。注意越多的采样点可以让我们向target function有一个高概率的逼近。

**PAC learnability** H 是否可以在多项式时间内，找到一个高概率的target function 的高精确度估计。

**Sample complexity** 保证 PAC Solution 的最少训练样本。

**Consistent Learner** Learner 选择的假设完美的符合训练数据。

**PAC Bound - consistent Learner** Version space 里面，至少出现一个在真实分布上的错误率小于 $\epsilon$ 的概率。

**Agnostic Learning** 不知道target function 是否在hypothesis set里面的情况下使用最小训练误差学习。

**PAC Bound - Agnostic Learner** 讲述了训练误差与泛化误差在概率上相近的水平与样本数的关系。

**VC-dimension** 对于 infinite Hypotheses Space 的复杂度的度量，使用VC-dimension的度量在许多情况下可以得到sample复杂度的一个更紧致的下界，其定义为H能shatter的最大样本数。

**Growth Function** H空间可以给与sample的label种类个数最大值。

**shattering** the instances 所有的label可能都会有一些hypothesis与之对应。

### **Convex Analysis**

several important concept

凸包，凸锥，凸集，

凸集分离定理等