- 3D领域
  - 专有名词
    - Talking Head Video:口播视频
      - 基于一张参考照片，和说话音频，生成视频
      - 需要保留身份信息，同时模仿驱动人的面部运动
  - 论文
    - Collaborative Neural Rendering using 2D Anime Character Sheets
      - 使用二维动漫人物表的协作式神经渲染模型
        - 定义
          - 代码：https://github.com/megvii-research/CoNR
          - CoNR：Collaborative Neural Rendering，渲染动漫人物姿势的模型
          - 从一些参考图像（又称角色表）中为指定的姿势创建新的图像
          - 类似模型： SMPL，StyleGAN2，CSE
          - 构建700000张角色表数据集
          - 角色表： 角色表是一个特定角色的图像集合，从不同的视角观察到多种姿态，它涵盖了所有的外观细节，被广泛用于协助动画或其衍生媒体的创作。
          - UDP：Ultra-Dense Pose， 超密集姿势,一种详细的三维姿势表示
        - 模型
          - 输入：角色表I和目标姿势P
          - UDP检测器
            - UDP姿势编码器
            - 多个UDP编码器
          - CINN渲染器
            - 对UDP编码器的的结果进行解码，生成带有姿势的图像
    - OpenPose: 使用部分亲和域的实时多人二维姿势估计
      - OpenPose： Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields
        - 定义
          - 部分亲和域：Part Affinity Fields， PAFs
            - 身体部位与图像中的单个人联系起来的方法
            - 一组二维矢量场，用于编码图像中四肢的位置和方向
            - 例如：结合身体和脚部关键点的检测器，生成PAFs
          - 代码：https://github.com/CMU-Perceptual-Computing-Lab/openpose
          - OpenPose： 开源多人二维姿势实时检测系统
            - 检测器包括：（a）身体+脚的检测，（b）手的检测[63]，和（c）脸部检测。核心模块是身体+脚的组合关键点检测器
          - 人体姿势检测难点
            - 每张图像可能包含未知数量的人，他们可能出现在任何位置或规模
            - 由于接触、遮挡或肢体衔接，人与人之间的互动会引起复杂的空间干扰，使得身体部分的关联变得困难
            - 运行时的复杂性往往随着图像中人的数量而增加，使实时性能成为一个挑战
          - 以前的方法
            - 使用人物检测器首先检测人员，然后在每个检测区域独立估计每个人的姿势
        - 方法（图2）
          - 输入图像，进入CNN模型
            - 由VGG-19的前10层初始化并进行微调
            - 3个连续的3x3卷积核，每一个的输出都被串联起来
          - 预测部分亲和域PAF和预测身体部位置信度
            - 置信度图：如果图像中出现了一个人，那么如果相应的部位是可见的，每个置信度图中就应该有一个峰值， 使用非最大限度的抑制来获得身体部位
            - 对每一对身体部位检测的关联性进行度量，是否是一对，即是否属于同一个人
              - 衡量关联的一个可能的方法是检测肢体上每对部位之间的额外中点，并检查其在候选部位检测之间的发生率，和肢体宽度，长度，方向，都有关系，公式10，
              - 得到候选身体部位对，这是一个NP-Hard 匹配问题（难点）公式13
              - 使用贪婪算法，Hungarian算法来获得最佳匹配的一对身体部位
          - 贪婪解码，输出二维关键点检测结果
    - PIFuHD：用于高分辨率三维人体数字化的多级像素对齐隐式函数
      - PIFuHD： Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization
        - 定义
          - 本文通过高分辨率图片构建更细节的3D立体图像模型， 单一图像中实现高保真的三维重建，分辨率足以恢复手指、面部特征和衣服褶皱等详细信息
          - 这个模型是单视角构建三维人体
            - 参考其它多视角会更好
          - 利用图像到图像的翻译网络来产生人物的背面图像
          - 法线（normal line），是指始终垂直于某平面的直线
          - PIFu：Pixel-Aligned Implicit Function (PIFu) represen-tation ，像素对齐的隐式函数表示法
            - 预测构建的三维空间中的一个点是否在人体内部（二分类） 公式1
          - Multi-Level Pixel-Aligned Implicit Function 多层PIFu
            - 本模型由两个层次的PIFu模块组成
          - 类似模型：DeepHuman，Tex2shape，PIFu
        - 模型（图2）
          - 粗略层
            - 侧重于整合全局几何信息，将降频的512×512图像作为输入，并产生128×128分辨率的主干图像特征
            - 产生三维嵌入做为精细层的输入
          - 精细层
            - 侧重于增加更微妙的细节，将原始1024×1024分辨率的图像作为输入，并产生512×512分辨率的主干图像特征
          - 使用pix2pixHD网络预测图像空间中的背面和正面法线
          - 一组采样点上使用扩展的二分类交叉熵损失(公式5)
    - Depth-Aware Generative Adversarial Network for Talking Head Video Generation
      - 深度感知的生成对抗网络用于口播视频的生成
        - 定义
          - 项目意义： 口播视频生成的目的是产生一个协同的人脸视频，其中包含了来自给定源图像和驱动视频的身份和姿势信息，密集的三维面部几何（如像素深度）对这项任务极为重要，密集的三维几何标注对视频来说成本过高，本文主要从人脸视频中自动恢复密集的三维几何（即深度），而不需要任何昂贵的三维标注数据
          - Depth-Aware Generative Adversarial Network： DaGAN
          - 第一个机制是深度引导的面部关键点检测
            - 由网络估计的面部关键点应该很好地反映面部的结构，因为它们被进一步用于产生特征扭曲的运动场，而深度图则明确地表明面部的三维结构
          - 第二个机制是一个跨模态注意力机制，以指导运动场的学习
            - 运动场可能包含来自杂乱背景的噪声信息，并且不能有效地捕捉与表情有关的微动，因为它们是由稀疏的面部关键点产生的。
          - 数据集： VoxCeleb1，CelebV
        - 模型
          - 生成器
            - 自监督深度学习子网络
              - 首先以自监督的方式，利用人脸视频的两个连续帧学习深度估计。然后，在Fd固定的情况下，整个深度框架被联合训练
            - 深度引导的稀疏关键点检测子网络
              - 给出源图像和来自驱动视频的驱动图像，为每个图像产生深度图，这些深度图及其RGB图像被拼接起来，以学习几何和外观特征，用于检测人脸关键点，生成人脸的相对运动场
            - 特征扭曲模块
              - 接受关键点作为输入，生成运动场
                - 公式8
              - 源深度图和扭曲特征 .进一步学习密集的深度感知注意力图
                - 得到精炼特征Fg
                - 图5
          - 判别器
            - 受到FOMM的启发。输入的图像首先被下采样四次，然后通过核大小为1×1的卷积层，最后我们输出一个大小为512×26×26的预测图。收集中间的特征图
            - 判别器特征匹配
          - 评估指标：
            - 结构相似度（SSIM）和峰值信噪比（PSNR）来评估生成的图像和驱动图像之间的低级别相似度
            - L1、平均关键点距离（AKD）和平均欧氏距离（AED）评估基于关键点的方法