# 虚拟人产品调研

## Avatar and Its Applications:

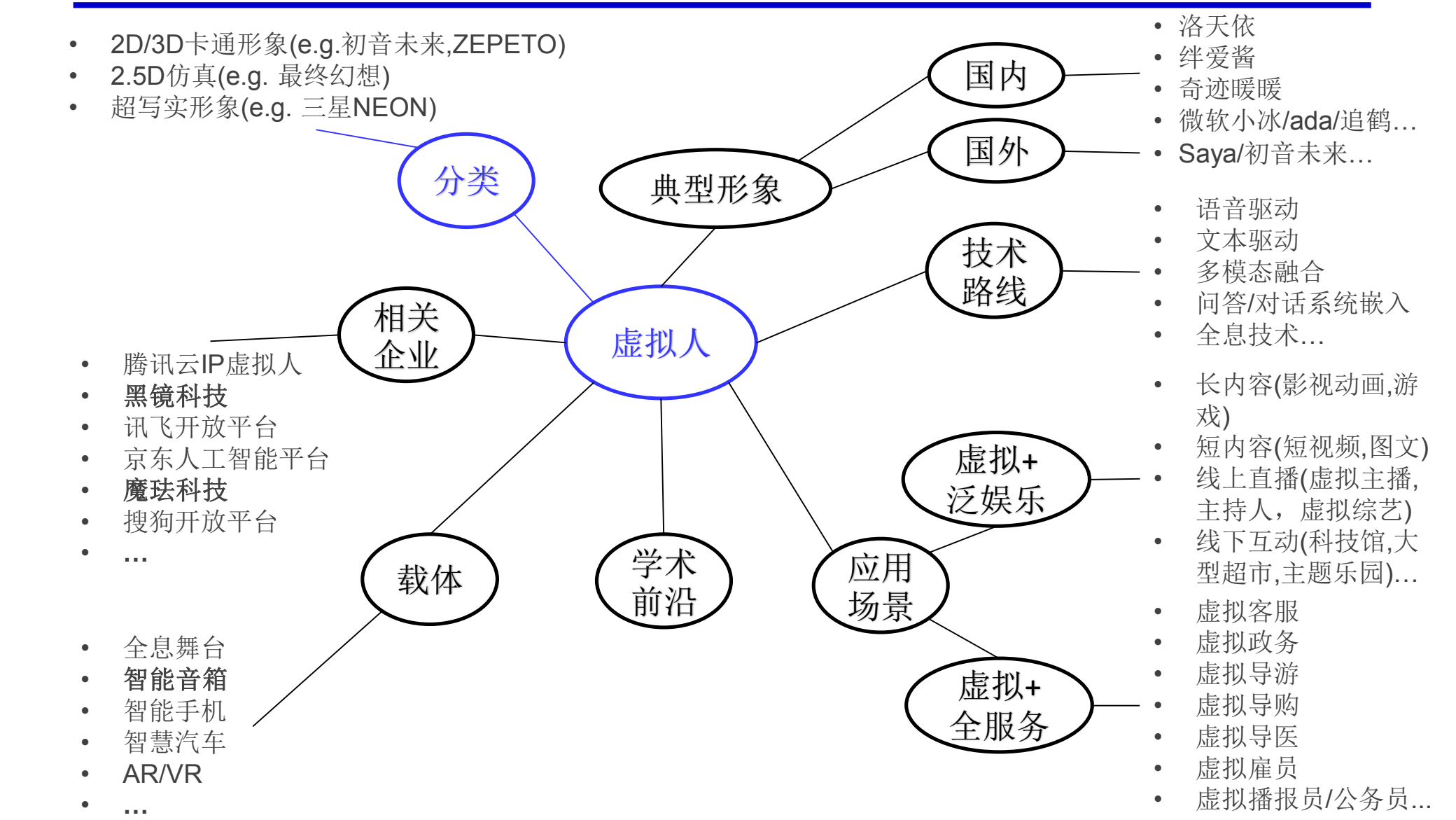
## A Survey

钱 锐 Rui Qian,

PCG平台与内容事业群 社交基础技术部

ruiqian@tencent.com

# 虚拟人产品调研

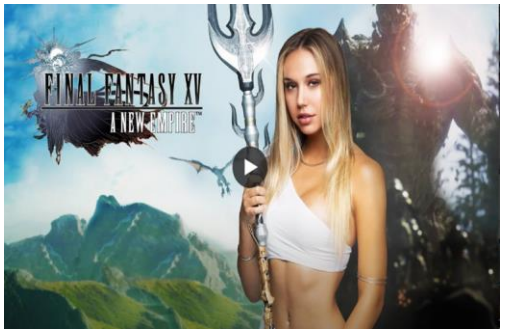

虚拟人
分类
典型形象
国内
国外
技术路线
相关企业
载体
学术前沿
应用场景
虚拟+泛娱乐
虚拟+全服务
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
洛天依
绊爱酱
奇迹暖暖
微软小冰/ada/追鹤…
Saya/初音未来…
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...

# 虚拟人分类

- **2D/3D卡通形象 (e.g.初音未来,ZEPETO)**
- **2.5D仿真 (e.g. 最终幻想)**
- **超写实形象 (e.g. 三星NEON)**

ZEPETO

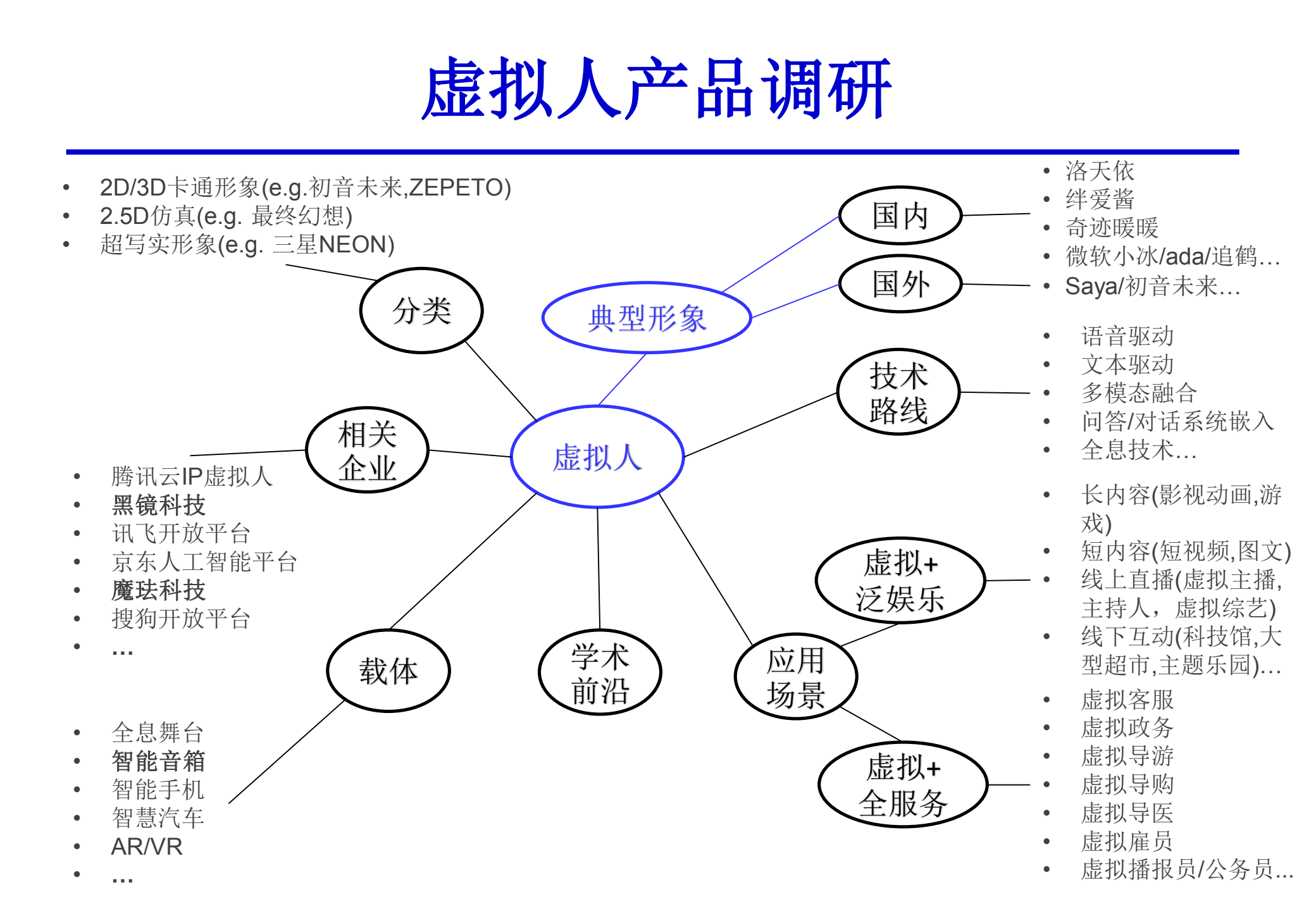
初音未来




最终幻想


三星NEON

# 虚拟人产品调研


虚拟人
典型形象
国内
国外
分类
技术路线
相关企业
载体
学术前沿
应用场景
虚拟+泛娱乐
虚拟+全服务
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
洛天依
绊爱酱
奇迹暖暖
微软小冰/ada/追鹤…
Saya/初音未来…
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...

# 虚拟人典型形象

**目前市场上有名的虚拟人:**

- **国外：初音未来, Saya,Lil Miquela,imma,noonoouri,shudu…**
- **国内：洛天依, 绊爱酱, 奇迹暖暖, 微软小冰, 鹤追, Ada,Siren…**

初音未来

日本高中生Saya

shudu

Lil Miquela

imma

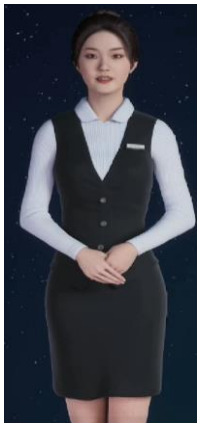
…

洛天依


绊爱酱

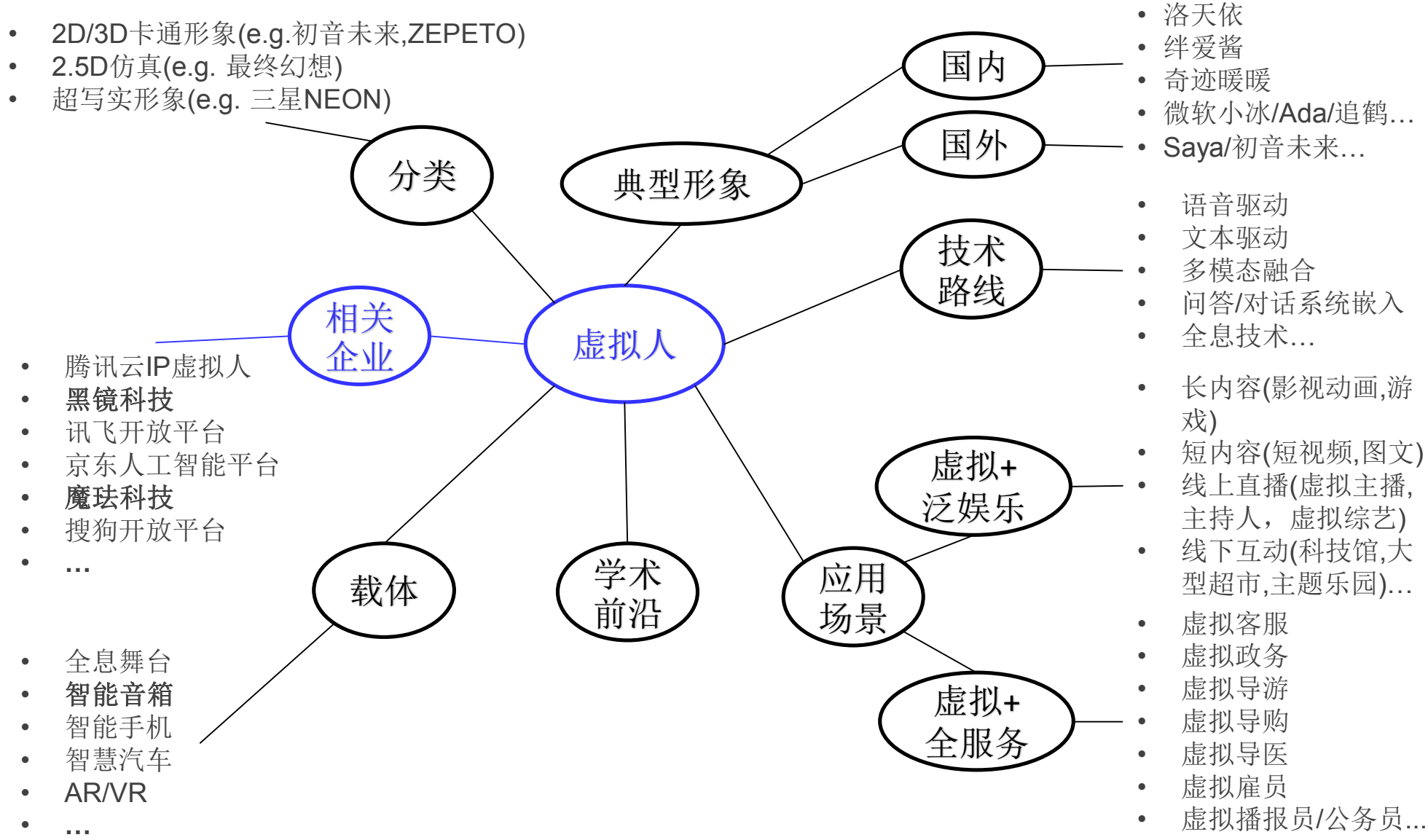
微软小冰

鹤追



Ada

# 虚拟人产品调研


虚拟人
分类
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
典型形象
国内
洛天依
绊爱酱
奇迹暖暖
微软小冰/Ada/追鹤…
国外
Saya/初音未来…
技术路线
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
相关企业
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
应用场景
虚拟+泛娱乐
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
虚拟+全服务
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...
学术前沿
载体
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…

# 虚拟人相关企业

**魔珐科技: 赋能虚拟+X全新生态**

- **实现制作技术智能化、流程管理标准化、资产运营标准化、角色打造规模化**
- **虚拟内容制作:**

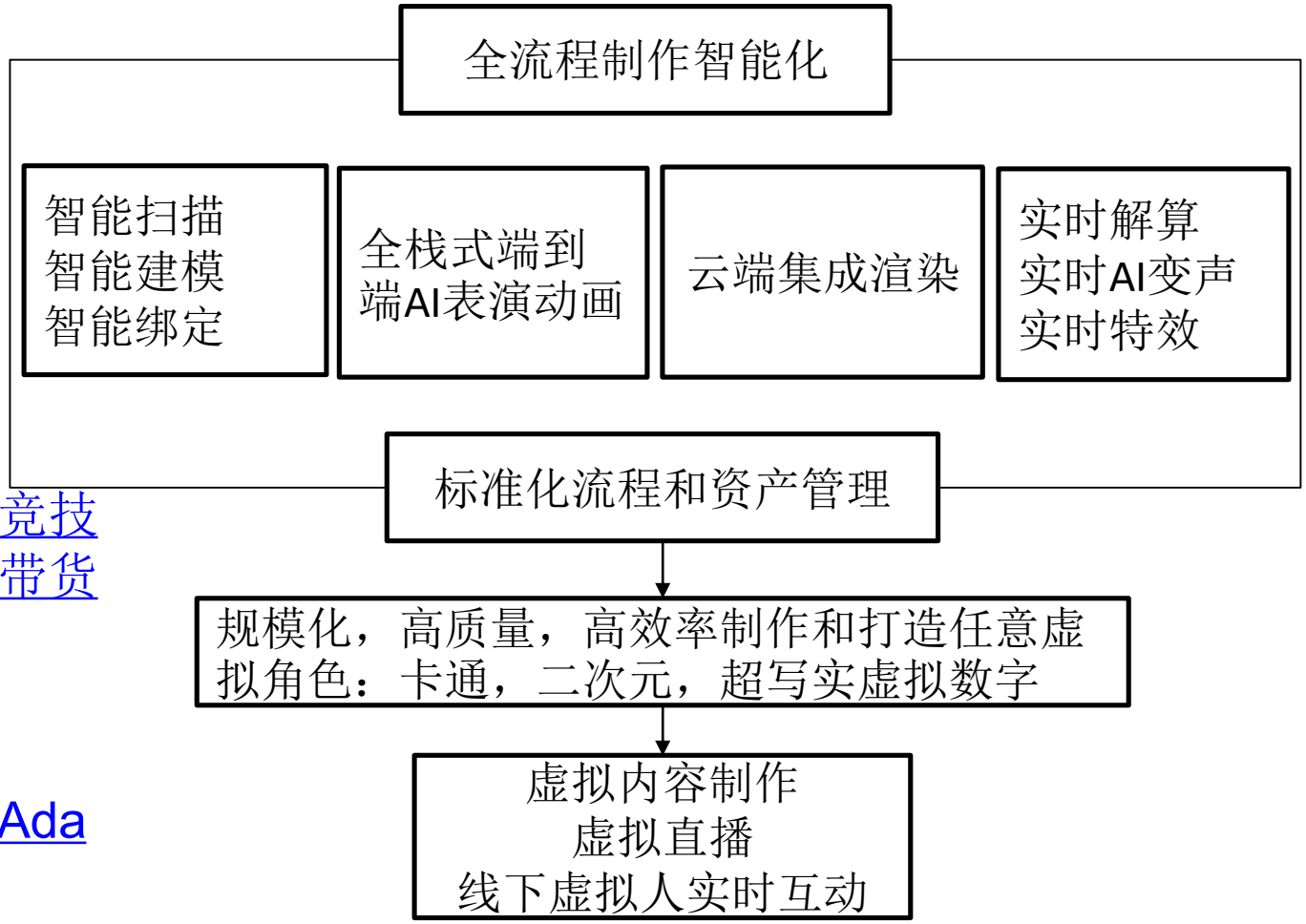
王者荣耀xMeco果汁茶
穿越火线十一周年
虚拟鹤追MV
鹤追采访VCR

- **虚拟IP打造**

灵狐姐三连拍
打造虚拟KOL - 翎Ling
虚拟灵狐和真人主播同台竞技
“二次元少女”齐麟直播带货

- **全智能虚拟数字人**

光大银行—阳光小智
中国联通虚拟智能助手
多模态全智能虚拟数字人Ada



# 虚拟人相关企业

**黑镜科技**: **真人 AI 形象引擎**和**AR 引擎**

- **幻脸**
- 人脸重建
- **儿童虚拟服饰重建与试装系统**
- **基于虚拟 AI 形象的车载智能娱乐系统**
- 让相册“活”起来
- AvatarMoji

给定一张正面人脸照片，一键生成个性化表情包


# 虚拟人相关企业

京东人工智能平台: NeuHost虚拟主播

基于3D图形学建模、音素识别以及生成式对抗网络模型技术，结合语音合成，语音识别和多轮对话能力，实现中英双语的新闻自动播报以及带有虚拟形象的智能对话服务机器人，支持视频和文本的自动化输出。

- 产品特点:

语音&形象定制

语音驱动口型匹配

表情实时驱动

产品价值

- 探索形象变现新模式

创造新奇的交互体验

24小时值守

高时效性

- 产品对象

地方融媒体

综艺节目主持

网红直播

智能客服

# 虚拟人相关企业

搜狗AI开放平台：基于少量音视频数据快速、低成本生成主播形象，只需输入文本即可生成AI合成主播播报的实时音视频流，主播的表情、唇动保持自然一致。

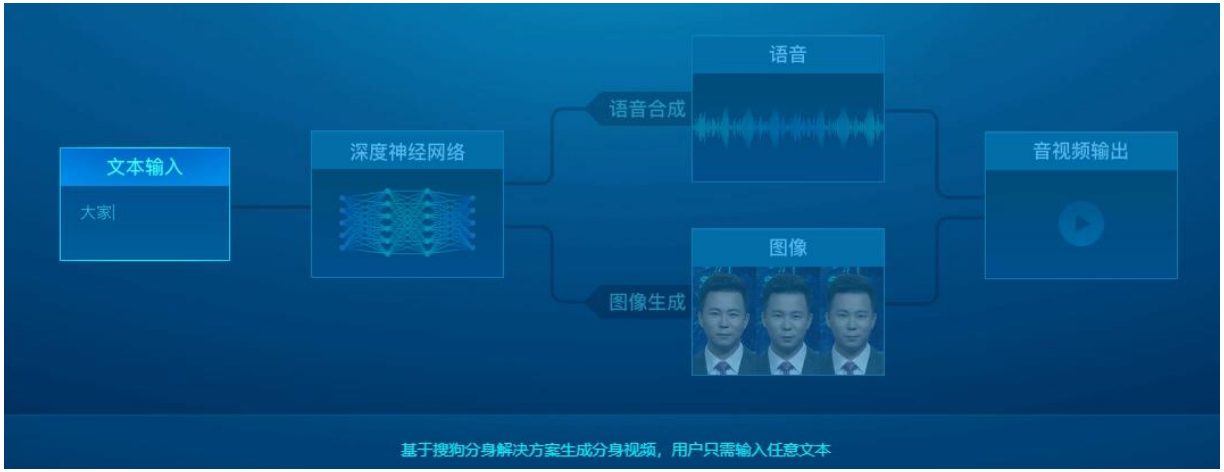
- 产品特点：

形象效果逼真
定制成本低
解决方案成熟
定制化灵活

- 面向对象

新闻类播报

- 技术方案
- 合作案例



# 虚拟人相关企业

讯飞AI开放平台:利用讯飞的语音合成、语音识别、语义理解、图像处理、机器翻译等多项人工智能技术，实现了多语言的新闻自动播报，并支持文本到视频的自动输出。

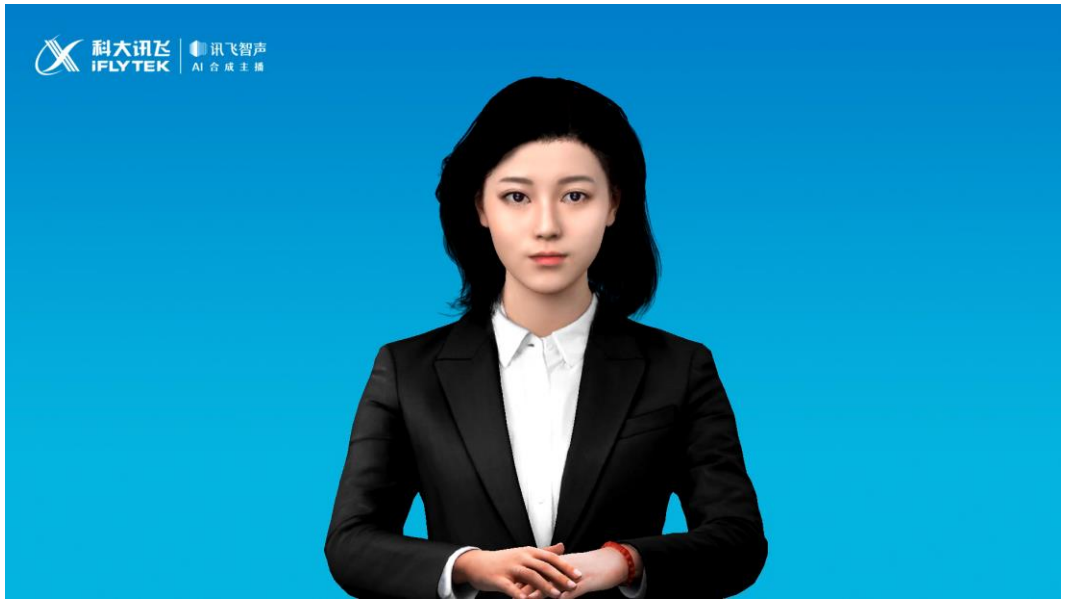
- 产品特点:

多语言播报
实时合成
支持多种情绪
声音/形象定制

- 应用场景

面试/面审场景
虚拟客服场景
迎宾导览场景
媒体新闻生产
培训课件生产
广告视频生产

- 合作案例



# 虚拟人相关企业

腾讯云IP虚拟人:利用语音交互、虚拟形象生成等 AI 技术，赋予文娱 IP 角色多模态交互的能力，助力媒体、教育、会展、客服等行业的智能娱乐化双升级。

- 产品特点：

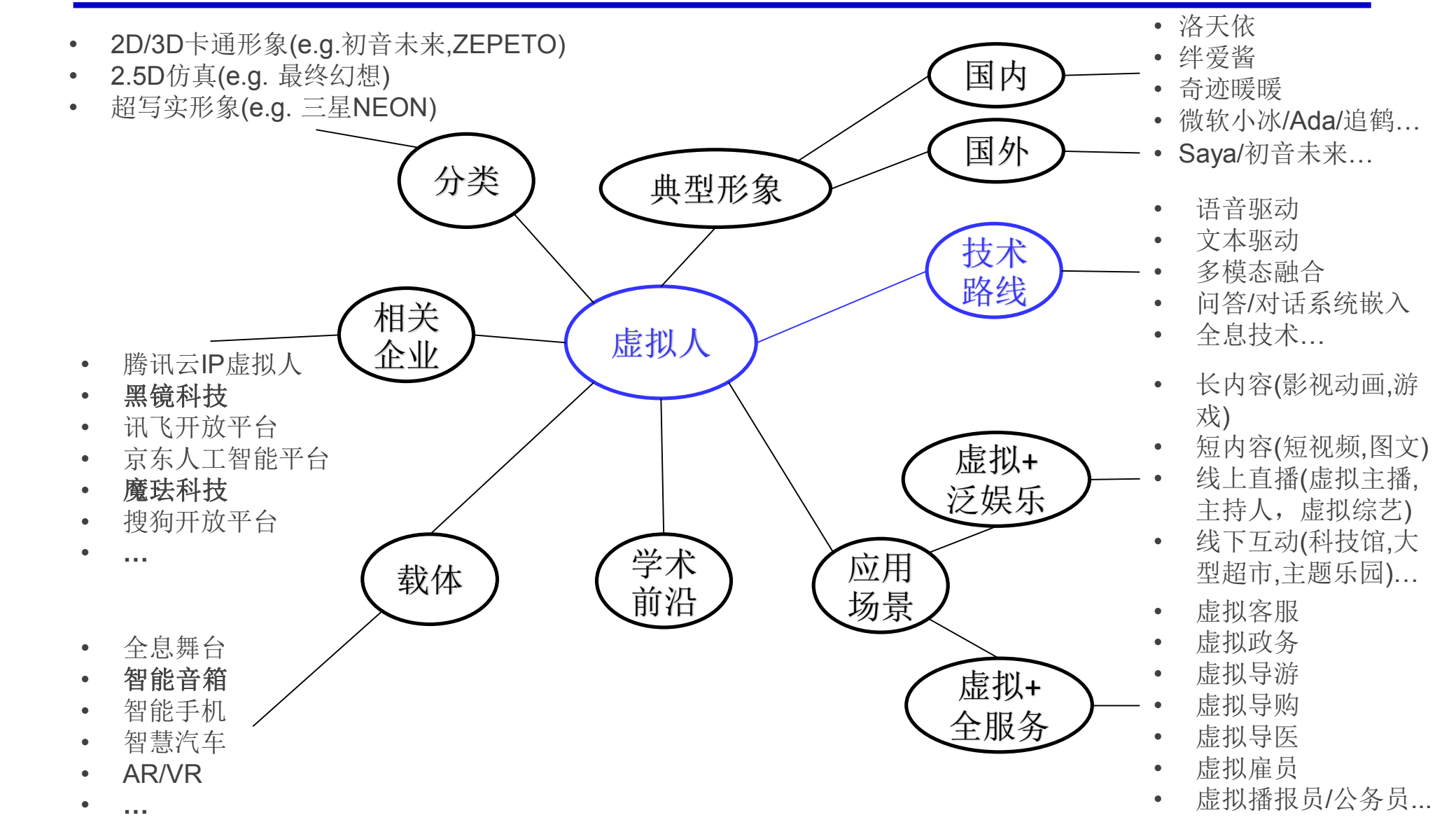
企业品牌IP化
体验升级
投入成本低
解决方案成熟
海量IP授权
软硬件一体化

- 应用场景

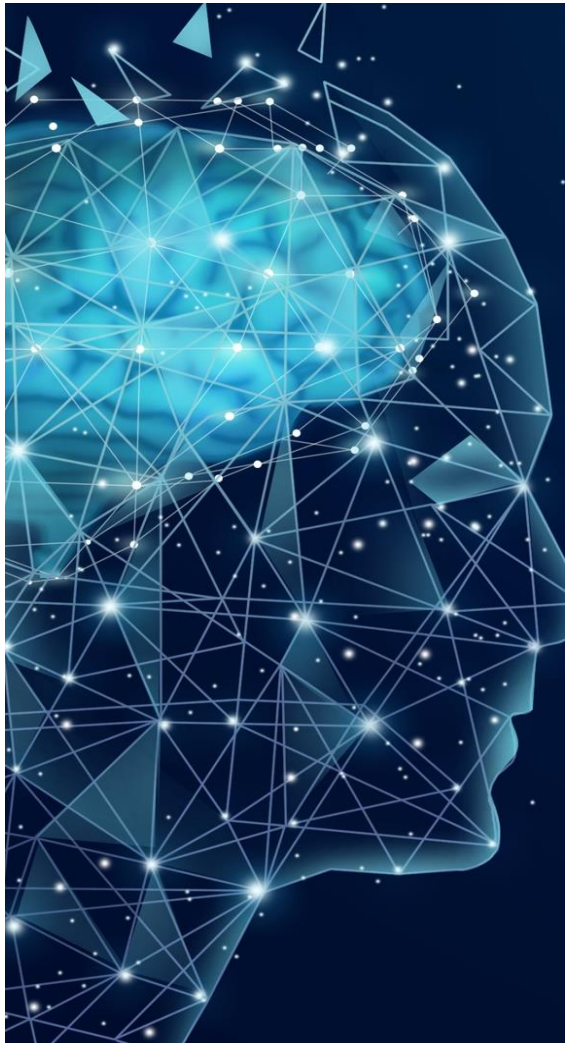
虚拟主播
虚拟教师
虚拟客服
虚拟助手
虚拟导游
IP品牌营销

- 合作案例

# 虚拟人产品调研


虚拟人
分类
典型形象
技术路线
相关企业
载体
学术前沿
应用场景
国内
国外
虚拟+泛娱乐
虚拟+全服务
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
洛天依
绊爱酱
奇迹暖暖
微软小冰/Ada/追鹤…
Saya/初音未来…
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...

# 虚拟人技术路线



## “有形象”–智能扫描/智能建模/智能绑定

- 智能扫描、建模和绑定，全流程智能化完成面部及身体扫描、建模、绑定、动画、解算和渲染等

## “会表演”– 实时动画生成/解算/AI变声/特效

- **基于端到端AI表演动画技术：**通过演员表演实时捕捉演员的面部表情、眼神、身体动作、手指动作等，并实时驱动到3D虚拟角色上，生成高质量的表演动画。
  **基于语音驱动动画技术：**通过语音对话即可快速生成高质量的虚拟数字人动画。
  **基于指令生成动画技术：**通过固定指令智能化、高效地生成动画，例如输入一段音乐虚拟人就会跳舞、输入路线指令虚拟人即可自然行走等。

## “能说会道” – 语音合成

- 实时TTSA（Text to Speech and Animation）技术, 基于文本实时自动生成高质量的虚拟数字人的语音和动画数据。

## “能听懂” – 语音识别和语义理解

## “能看到” – 视觉感知/情绪识别

- 人脸/人体检测、人脸表情/人体动作识别、手势识别、人脸/人体属性识别，单相机实时三维人体重建与三维姿态估算。

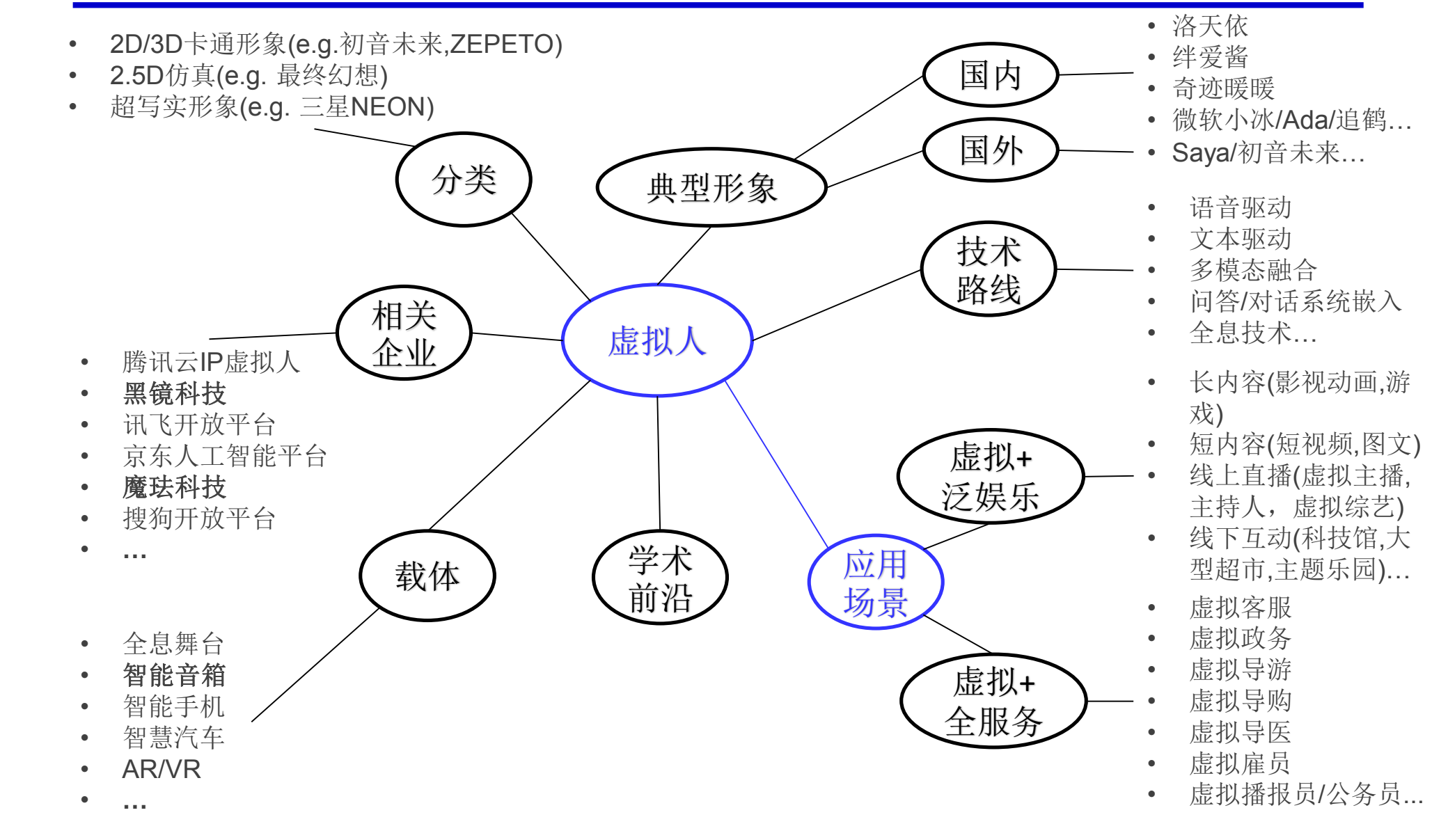
## “可被看到” – 渲染

- 离线渲染和实时渲染，实现虚拟数字人高质量的渲染效果。

## 全息技术

## AR/VR

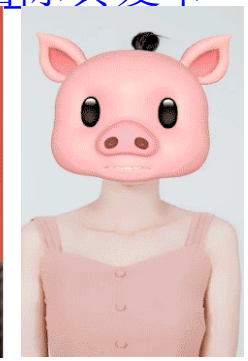
# 虚拟人产品调研


虚拟人
分类
典型形象
国内
国外
技术路线
相关企业
载体
学术前沿
应用场景
虚拟+泛娱乐
虚拟+全服务
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
洛天依
绊爱酱
奇迹暖暖
微软小冰/Ada/追鹤…
Saya/初音未来…
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...

# 虚拟人应用场景

**虚拟+泛娱乐：**

- 长内容(影视动画,游戏): 虚拟偶像/虚拟IP/虚拟KOL
  - 番剧PV制作《镜·双城》
  - 热门番剧《灵笼》
- 短内容(短视频,图文):
  - 王者荣耀xMeco果汁茶
  - 让相册“活”起来
- 线上直播(虚拟主播,主持人，虚拟综艺)
  - 虚拟灵狐和真人主播同台竞技
  - “二次元少女”齐麟直播带货
  - 萌音歌姬24h在线卖唱
- 线下互动(科技馆,大型超市,主题乐园)
  - 虚拟主持人惊艳亮相9·9国际真爱节




iPhone Animoji


Facebook MSQRD


* KOL: 关键意见领袖（Key Opinion Leader,简称KOL）是营销学上的概念.
* 虚拟偶像/IP/KOL 为同类别

# 虚拟人应用场景

**虚拟+全服务：**

- 虚拟客服
- 虚拟政务
- 虚拟导游
- 虚拟导购
- 虚拟导医
- 虚拟雇员
- 虚拟播报员
- 虚拟公务员
- 金融助手
- 虚拟老师
- 市政助手
- 虚拟律师
- 虚拟伴侣
- 虚拟试衣
- 虚拟管家/电脑管家
- …

光大银行—阳光小智
中国联通虚拟智能助手
多模态全智能虚拟数字人Ada


百度&浦发虚拟员工


谷歌教育

3D虚拟试衣


商汤科技AI医生:小糖

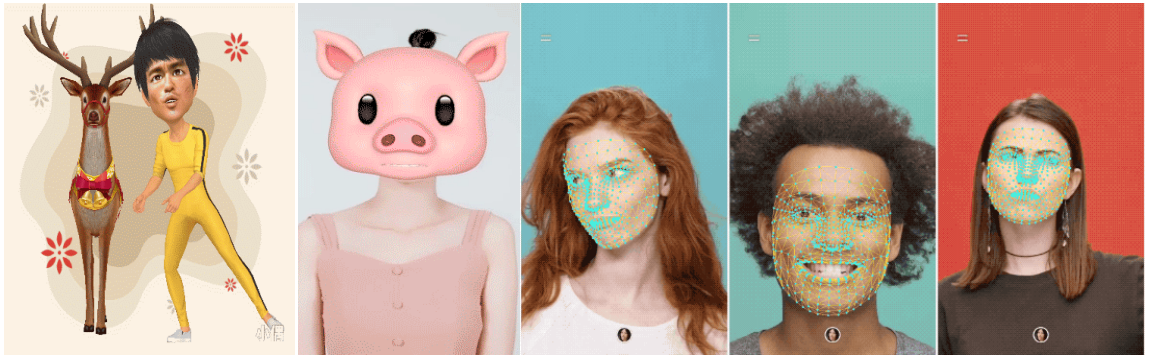

扎克伯格的人工智能管家：贾维斯



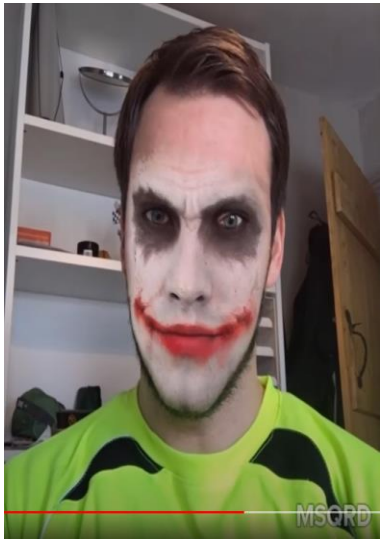
日本Gatebox利用全息技术推出的Azuma Hikari

# 虚拟人QQ场景

- 长内容(影视动画,游戏): 虚拟偶像/虚拟IP/虚拟KOL
  - 番剧PV制作《镜·双城》
  - 热门番剧《灵笼》
- 短内容(短视频,图文):
  - 王者荣耀xMeco果汁茶
  - 让相册“活”起来
  - 幻脸
- 线上直播(虚拟主播,主持人，虚拟综艺)
  - 虚拟灵狐和真人主播同台竞技
  - “二次元少女”齐麟直播带货
  - 萌音歌姬24h在线卖唱



Facebook MSQRD

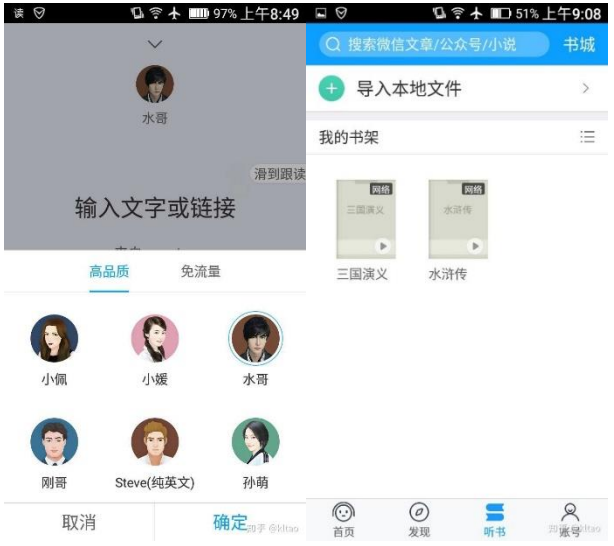



iPhone Animoji

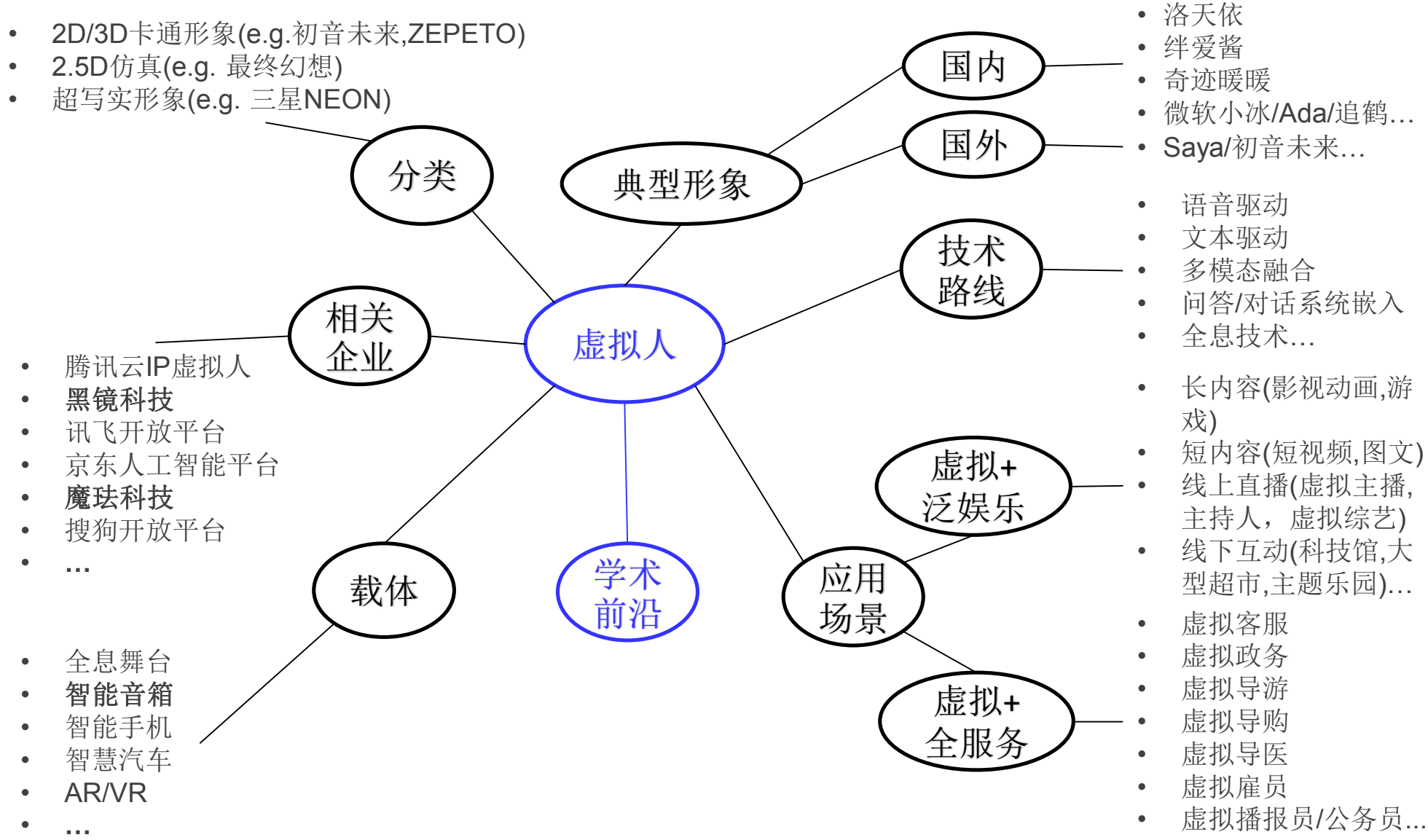
# 虚拟人QQ场景

- 有声读书插件
  - 喜马拉雅(听书), Eotu(听新闻及小说、在线翻译)、妙读和樊登读书(书籍精华解读)、一席（视频类）、科大讯飞。
  - 微信读书&QQ读书
    - 录入自己的声音，用自己的的声音朗读(回避版权问题);
    - 服务于小说迷、快餐式、碎片化阅读;
    - 语音交互下的查询和解析;
    - 支持导入和网页阅读;
    - 依托QQ的海量文学资源(解决源问题);
    - 情绪模拟(https://tencent-ailab.github.io/durian/);
    - 减少屏幕使用时间;
    - 作为类似于siri中的子模块

| 产品名称 | Slogan | 产品定位 | 衍生产品 |
| --- | --- | --- | --- |
| 微信读书<br>（版本2.2.3） | 让阅读不再孤独 | 一款基于微信关系链的社交型阅读产品。<br>【官方主打】<br>个性化阅读风格；<br>和好友一起发现优质好书；<br>和好友交流阅读感想；<br>和好友比拼阅读排行 | 【微信公众号】<br>微信读书（服务号）<br><br>【小程序】<br>微信读书电台<br>微信读书排行<br>微信读书书城 |
| QQ阅读<br>（版本6.5.61） | A.海量原著，想读就读<br>B.阅读自由主义<br>（胡歌代言） | 主打网络文学，以资源取胜<br>打造全阅读内容入口，千万级作品储备，400万签约作家，拥有海量热门影视原著 | 【微信公众号】<br>QQ阅读服务号<br>QQ阅读订阅号<br><br>【小程序】<br>QQ阅读 |

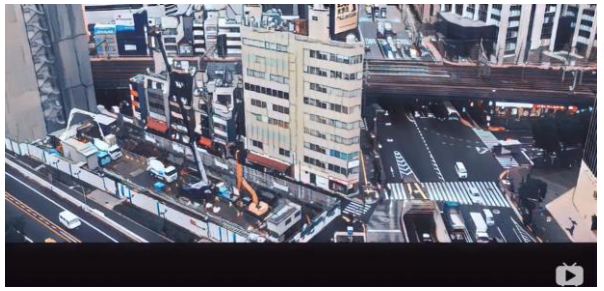


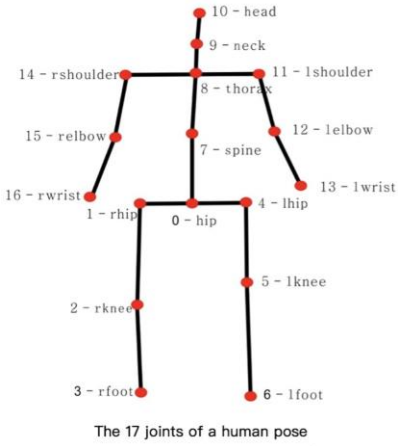
# 虚拟人产品调研


虚拟人
典型形象
国内
国外
洛天依
绊爱酱
奇迹暖暖
微软小冰/Ada/追鹤…
Saya/初音未来…
分类
2D/3D卡通形象(e.g.初音未来,ZEPETO)
2.5D仿真(e.g. 最终幻想)
超写实形象(e.g. 三星NEON)
技术路线
语音驱动
文本驱动
多模态融合
问答/对话系统嵌入
全息技术…
相关企业
腾讯云IP虚拟人
黑镜科技
讯飞开放平台
京东人工智能平台
魔珐科技
搜狗开放平台
…
应用场景
虚拟+泛娱乐
长内容(影视动画,游戏)
短内容(短视频,图文)
线上直播(虚拟主播,主持人，虚拟综艺)
线下互动(科技馆,大型超市,主题乐园)…
虚拟+全服务
虚拟客服
虚拟政务
虚拟导游
虚拟导购
虚拟导医
虚拟雇员
虚拟播报员/公务员...
学术前沿
载体
全息舞台
智能音箱
智能手机
智慧汽车
AR/VR
…

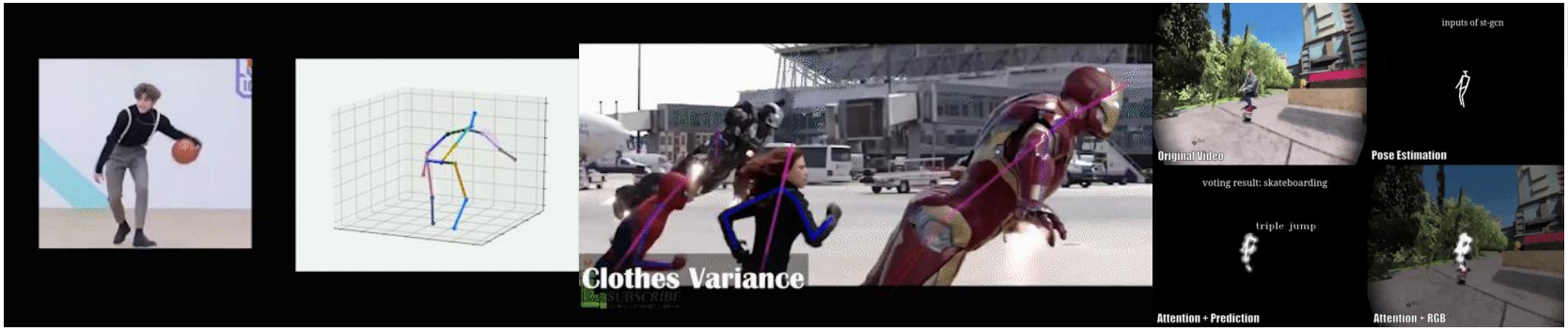
# 虚拟人学术前沿

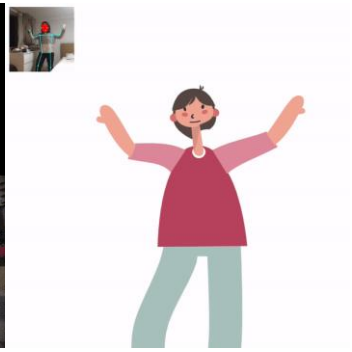
Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition[ arXiv, Code]
Talking Head Anime from a Single Image[ Project, Code]
Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs[Project,arXiv,Code]
Learning to Cartoonize Using White-box Cartoon Representations[Project,arXiv,Code]
DurIAN : Duration Informed Attention Network For Multimodal Synthesis[Project,arXiv]
Neural Voice Puppetry: Audio-driven Facial Reenactment[Project,arXiv,Code]





The 17 joints of a human pose



# 总结

**在虚拟数字1.0时代，虚拟人主要以TO B端为主，赋能泛娱乐行业及服务行业：**

- 虚拟人以虚拟主播、虚拟助手、虚拟IP为主，角色和任务存在差异。
- 虚拟人突破影视游戏二次元壁，进入金融医疗市政体系。

个人想法：

1 语音驱动下基于图像/视频内容的跨模态检索.

2 基于虚拟人的有声阅读器，服务对象为小说迷,长时间看新闻的人群.

3 全自动体育赛事播报系统，即从短文本生成到自动播报一体化，能进行简单的知识推理和问答，这是新闻类主播(e.g.搜狗，讯飞)目前不具备的.

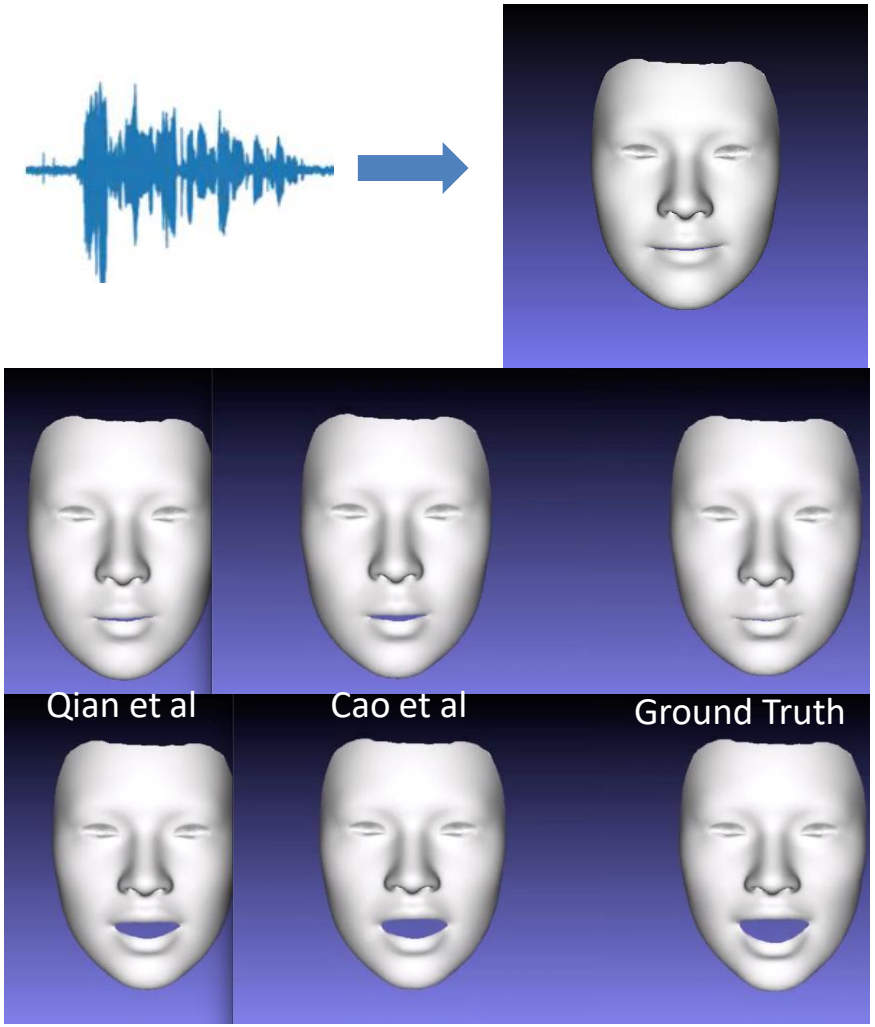
语音, 图像，NLP结合案例: 萌音歌姬24h在线卖唱

PCG
平台与内容事业群

# 实习生留用答辩汇报

社交基础技术部-智能应用中心-语音研究组-钱锐（ruiqian）

# 腾讯实习工作总结-3DMM虚拟人口型合成




## 工作内容概述及最终成果

难点在于语音驱动(audio-driven)背景下的图像帧合成。
原项目**存在的问题是局部图像帧口型张开和闭合的幅度与ground truth差距明显，参数调优出现瓶颈。**
主要贡献:

- **改进原项目的优化目标**:拟合基底系数　拟合基底点位
  抠取并拟合嘴部相关基底点位，**嘴部闭合明显改善，并消除原项目过拟合问题**;
- 利用局部图像帧的连续性，借助attention机制进行图像帧的平滑，**消除嘴型过渡中的颤抖现象**;
- **贡献一套完整的训练、测试、可视化代码，简称NVP**，与原项目baseline对齐，系统显存(992M vs 22G)和内存(2.66G vs 90G)友好。

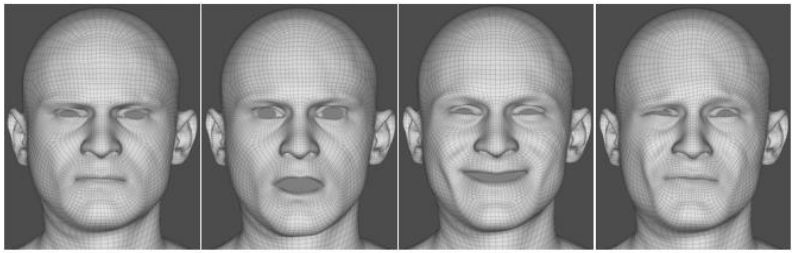
项目不足点：

- **嘴部张开的幅度还没有ground truth明显。**

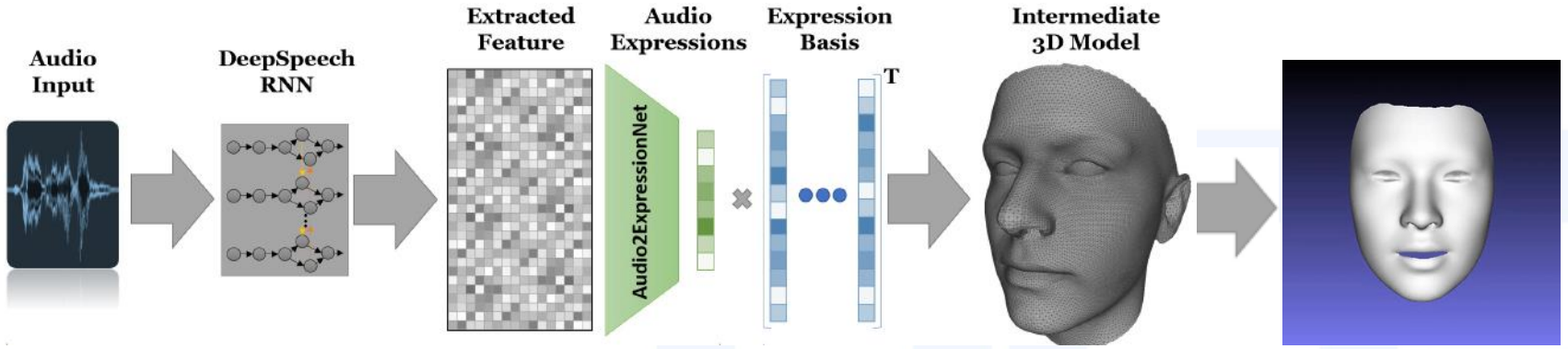
# 3DMM虚拟人口型合成-项目背景

**虚拟人口型合成难点**——语音驱动(audio-driven)背景下的图像帧合成

- 跨模态: 语音和图像处于不同的特征空间;
- 音素与表情是多对多关系;



**虚拟人口型合成技术方案**




输入语音 → Deep Speech / Kaldi 生成特征 → Audio2ExpressionNet (CNN+FC) → 表情基系数 → 3D脸部动画

# 3D虚拟人口型合成-优化目标函数

**虚拟人口型合成的优化目标**:

拟合基底系数(27维) →拟合基底点位(3448维度) →抠取并拟合嘴部相关基底点位(904维)

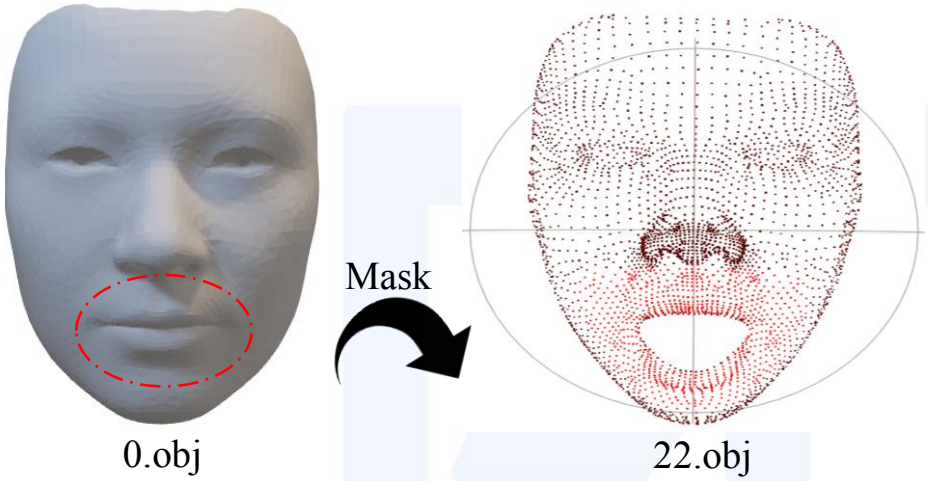
具体地，基底系数到基底点位的转换公式为:

$$(x, y, z) = \text{mean} + (\text{blendshape} - \text{mean}) * \text{weight}$$

blendshape为基底，weight为模型预测的系数。**嘴部闭合明显改善，并消除原项目过拟合问题。**

**虚拟人口型合成抠取嘴部相关点位:**

Meshlab手动抠取平均脸(i.e., 0.obj)嘴部相关点位(x, y, z), 利用这些相关点位建立反向索引index = f(x, y, z)。由于每个点的index具有固定语义，以index为掩码mask施加在(20-46).obj共27个基底上，即可得到27个基底与嘴部相关的mask。

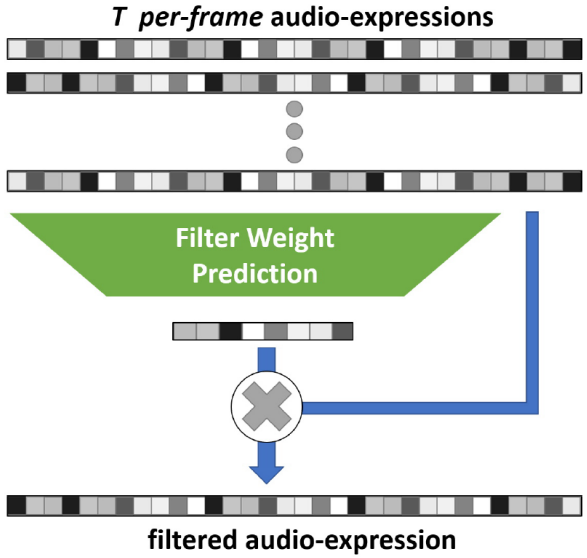



PCG
平台与内容事业群

# 3DMM虚拟人口型合成-图像帧平滑

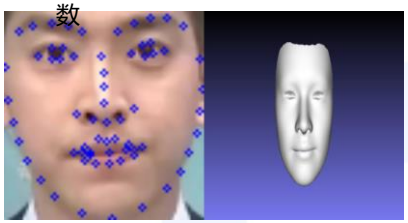
利用**局部图像帧的连续性**，借助Attention机制进行图像帧的平滑，**消除嘴型过渡中的颤抖现象。**

- 由于数据筛选过程破坏了帧的连续性(筛选过后，有1/3-1/2的图像帧被剔除)，因此原项目采取独立预测每一帧的基底系数的方式；
- 通过统计观察发现，被剔除的帧通常是整段连续缺失，局部依然连续。

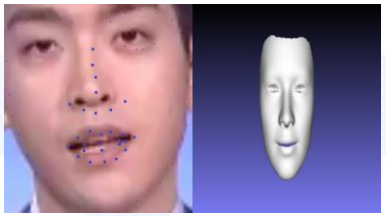



Attention机制

点位标记，输出图像帧系数



对于点位标记不当的帧进行人工筛选，e.g.，把牙齿的位置当做嘴唇的边缘

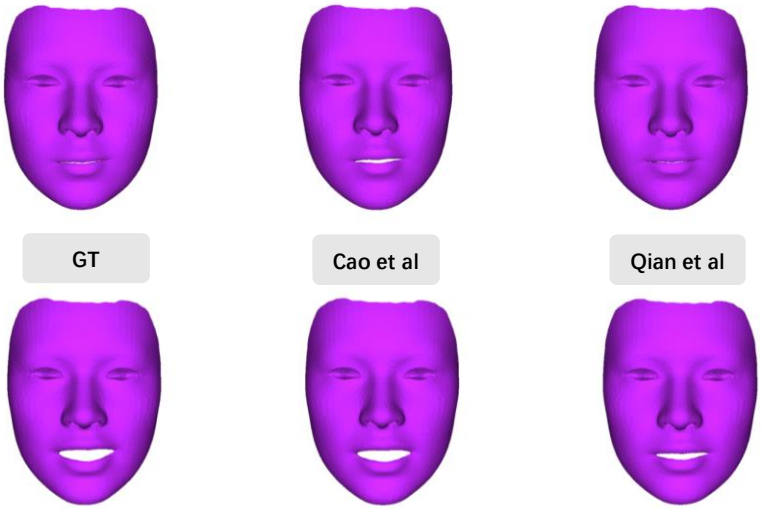


数据筛选

# 3DMM虚拟人口型合成-优化效果

**NVP(Qian et al)与baseline(Cao et al)定性分析**

- 相比于baseline，**NVP在嘴部闭合的时候更加接近ground truth;**
- 相比于baseline，NVP在嘴部张开的时候幅度不够;

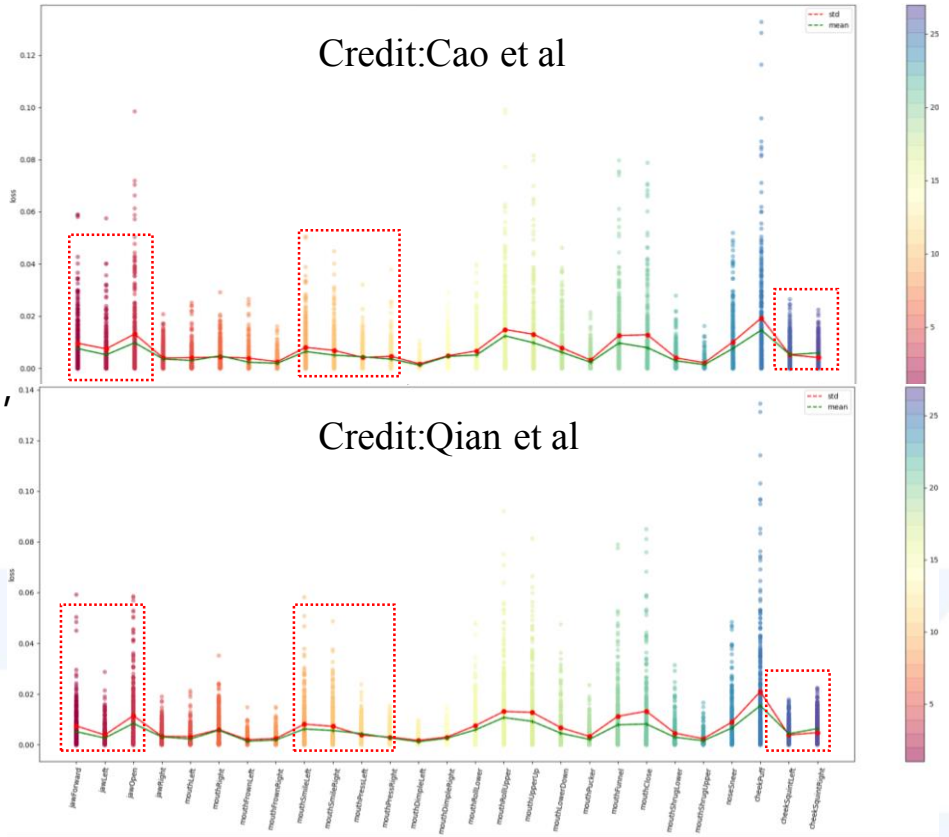



PCG
平台与内容事业群

# 3DMM虚拟人口型合成-优化效果

**NVP(Qian et al)与baseline(Cao et al)定量分析**

- 可视化baseline (credit:Cao et al)与ground truth(GT)之间的loss在每个维度上的分布差异，同理可视化NVP(credit:Qian et al)的loss差异。从对比图可以看出，NVP在某些维度上得到改善, (e.g. mouthPressright etc)。



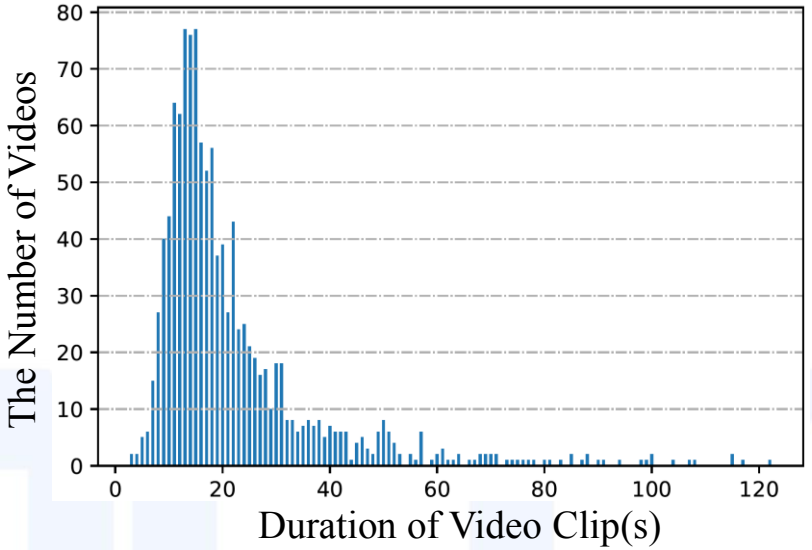

NVP(Qian et al)目前在嘴型张开上与ground Truth还有差距，改进方向如下:

- 目前的网络比较浅，batch size=64, seq_len=8, cuda_mem=992M,考虑加深网络;
- 目前的attention机制只考虑了帧的连续性，没有考虑图像帧的先后顺序，考虑使用单向或者双向的LSTM进行序列建模;
- 考虑使用3D重构中的chamfer loss,对重构得到的基底做几何约束。

# 3DMM虚拟人口型合成-NVP项目

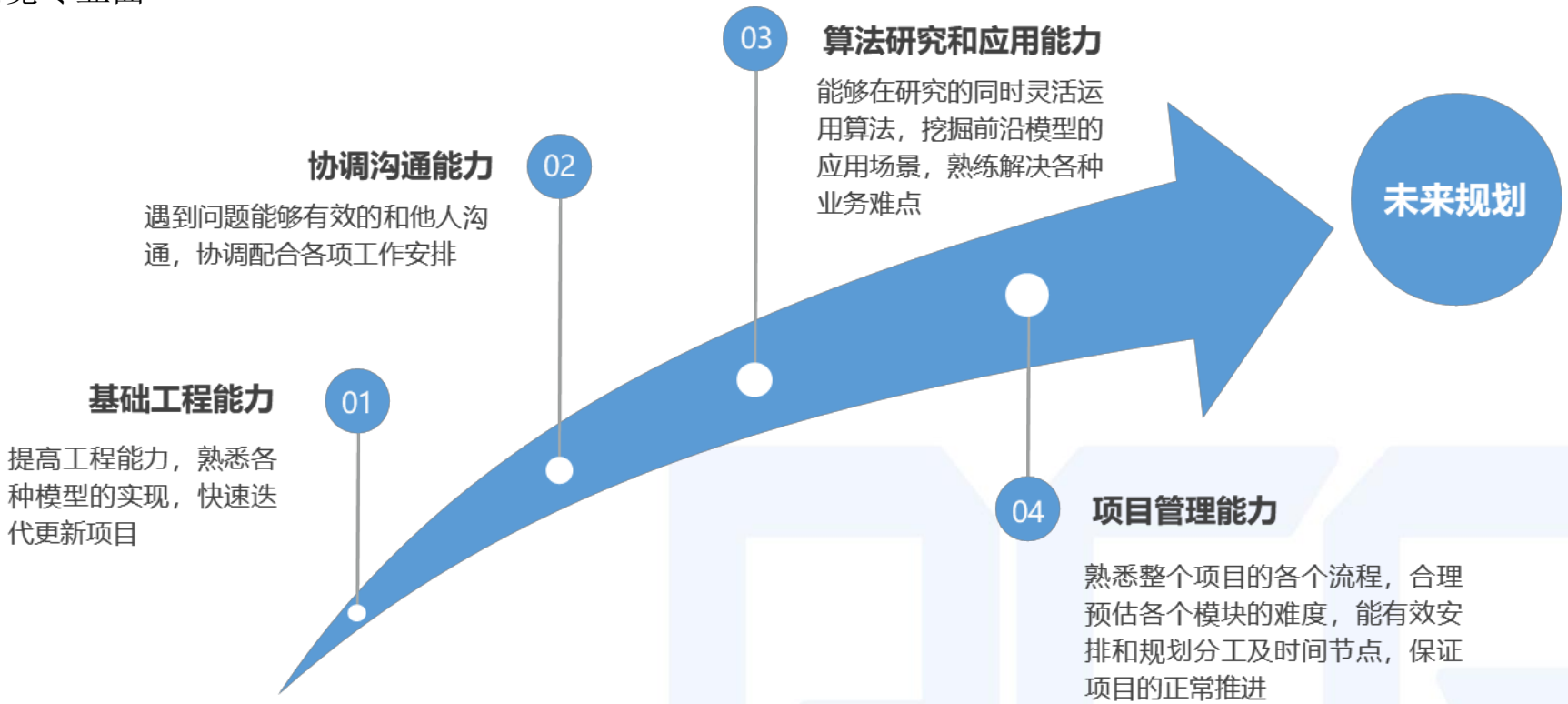
**贡献一套完整的训练、测试、可视化代码，简称NVP.**

- 与原项目baseline对齐，系统显存(992M vs 22G)和内存(2.66G vs 90G)友好，原项目以视频为单位进行输入，将所有视频padding成等长序列，造成大量的无效计算;
- NVP支持tensorboard训练过程可视化和模型对比分析。

# 个人未来工作畅想

- 规范思维习惯
- 拓宽专业面




PCG
平台与内容事业群

PCG
平台与内容事业群

# Thank you