

# 智慧教育-多媒体学习环境构建方向

## 网络多媒体小组101

导入学习资料

本PPT用于小组同学科研导入及有意向同学的选拔。

对于推免学生，作为小组（夏令营）的选拔依据，请据此了解实验室的主要方向，根据自己的兴趣在**最后一页挑选1篇代表性文章进行复现并介绍**（复现且实验分析）。

### 考核形式

- ① 准备时间约1个月（具体以大数据工程实验室的通知为准）。
- ① 独立进行 20min（10min 汇报+10min 提问）的PPT考核在线答辩。

有意参与考核的同学也请与感兴趣的导师提前取得联系：王志文 [wzw@xjtu.edu.cn](mailto:wzw@xjtu.edu.cn); 闫彩霞 [yancaixia@xjtu.edu.cn](mailto:yancaixia@xjtu.edu.cn) ; 龚铁梁 [gongtl@xjtu.edu.cn](mailto:gongtl@xjtu.edu.cn)> ; 张未展 [zhangwzh@xjtu.edu.cn](mailto:zhangwzh@xjtu.edu.cn);

# 通用知识与技能推荐

- 基础课程

- 李宏毅深度学习 (b站)
- 李沐 (b站相关论文分享视频)
- [斯坦福 CS231](#)
- [现代计算机图形学入门](#)

- 编程工具

- Python
- Unity (C#)
- C/C++

- 论文阅读

- 流媒体实时传输与分析方向

- "Neural adaptive video streaming with pensieve" - SIGCOMM - 2017

- ["Server-driven Video Streaming for Deep Learning Inference"](#) - SIGCOMM - 2020

- 沉浸式交互技术与应用方向

- "Collaborative Work in Augmented Reality: A Survey" - IEEE TVCG - 2020

- "NeRF--: Neural radiance fields without known camera parameters"

- "From Capture to Display: A Survey on Volumetric Video"

- 模型架构搜索与轻量化方向

- "AutoFormer: Searching Transformers for Visual Recognition" - ICCV - 2021

- "A Survey of Methods for Low-Power Deep Learning and Computer Vision"

- "A Survey on Deploying Mobile Deep Learning Applications: A Systemic and Technical Perspective"

- Diffusion-based AIGC

- "Controllable Generation with Text-to-Image Diffusion Models: A Survey"

# 流媒体实时传输与分析方向

# 流媒体传输优化

## • 自适应流媒体传输ABR

- Pensieve<sub>[SIGCOMM'17]</sub> --> PiTree<sub>[MM'19, 网络可解释性与轻量化]</sub> --> GENET<sub>[SIGCOMM'22, 课程学习]</sub>
- 说明：完整项目和数据集均开源，论文、代码对应度很高，复现性较好，涉及模型训练、强化学习、点播流媒体传输优化等领域。Pensieve目前仍然是ABR领域最具复现价值和实用性的baseline
- 相关任务：Pensieve项目提供了除Pensieve外众多启发式方法的实现和评估测试，借助该项目可进一步挖掘在学习算法、应用场景等方面的设计空间，具体可参考国内黄天驰等人的论文，如QARC<sub>[MM'18]</sub>, Comyco<sub>[MM'19]</sub>, Stick<sub>[MM'20]</sub>等
- 链接：[Pensieve官方文档](#), [Pensieve-CS244课程](#), [PiTree](#), [GENET](#), [黄天驰在ABR方向的相关论文](#)

## • 实时通信

- [Concerto](#)<sub>[MobiCom'19]</sub>, [OnRL](#)<sub>[MobiCom'20]</sub>, [Loki](#)<sub>[MobiCom'21]</sub> (强化学习驱动的低延迟视频传输)
- 云游戏:
  - Pudica: Toward Near-Zero Queuing Delay in Congestion Control for Cloud Gaming
  - Enabling High Quality Real-Time Communications with Adaptive Frame-Rate
  - ZGaming: Zero-Latency 3D Cloud Gaming by Image Prediction

## • 大语言模型在网络领域的应用:

- NetLLM: Adapting Large Language Models for Networking

# 流媒体传输优化

- **推荐阅读论文**

- 会议：SIGCOMM, NSDI, INFOCOM, MM
- 期刊：ToN, TPDS, TMC, JSAC

- **相关技能学习**

- 流媒体传输：MPEG-DASH, RTMP, WebRTC (RTP/RTCP)
- 相关工具：
  - FFmpeg (提供了音视频录制、转换以及流化音视频的完整解决方案)
  - Wireshark (网络封包分析)
  - Linux tc, mahimahi, NS-3 (网络环境仿真)
- 编程语言：Python, C/C++

- **学习和实践流程**

- 点播 (Pensieve) --> 直播 (RTMP/HTTP-FLV协议, SRS服务器架构) --> RTC (WebRTC)
- 基于论文表述查看并复现开源代码, 跑通完整实验
  - 在复现代码过程中, 根据需求去学习和掌握相关工具和技能
  - 探索新场景、新需求、新方法, 在该方向上的实践和应用

# 视频实时传输与分析(Live video streaming and analytics)

- 相关研究

- 自适应流媒体传输ABR (Human-oriented streaming): RobustMPC -- **Pensieve** -- PiTree
- 视频实时传输与分析 (Machine-centric streaming):
  - Configuration selection: VideoStorm -- AutoML -- CASVA
  - Pipeline optimization: **Elf**, AccDecoder, AccMpeg, RECL
  - summary link: [Video analytics literature review \(notion.site\)](https://arxiv.org/pdf/1810.07862.pdf), survey: <https://arxiv.org/pdf/1810.07862.pdf>
  - 推荐会议: NSDI, Mobicom, Sigcomm, Infocom, MLSys, OSDI, ACM MM
- 强化学习 (Reinforcement learning)
  - 入门课程: Open AI (spinning up !), 台大李宏毅
  - RL x streaming: Pensieve, Stick, Loki, GENET, Swift
  - 推荐会议: ICLR, ICML, NeurIPS, AAAI

- 部分推荐文章(文章标题):
  - Live video analytics
    - A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP
    - PiTree: Practical Implementation of ABR Algorithms Using Decision Trees
    - Genet: Automatic Curriculum Generation for Learning Adaptation in Networking
    - From Ember to Blaze: Swift Interactive Video Adaptation via Meta-Reinforcement Learning
    - CASVA: Configuration-Adaptive Streaming for Live Video Analytics
    - Server-Driven Video Streaming for Deep Learning Inference
  - Cloud-edge collaborative Learning
    - AdaptiveNet: Post-deployment Neural Architecture Adaptation for Diverse Edge Environments
    - DUET: A Tuning-Free Device-Cloud Collaborative Parameters Generation Framework for Efficient Device Model Generalization
    - RECL: Responsive Resource-Efficient Continuous Learning for Video Analytics
- 推荐复现项目:
  - Pensieve: <https://web.mit.edu/pensieve/>
  - RECL (暂无代码)



# 沉浸式交互技术与应用方向

# 体积视频

- **三维信息捕获和场景重建:**

- 显式格式: Point Clouds, Mesh, Voxels --> 隐式格式: NeRF
- 文章推荐: Livescan3d(三维信息捕获)<sub>[1]</sub>, 点云配准综述<sub>[2]</sub>, NeRF<sub>[3]</sub>, Hyperreel(动态NeRF)<sub>[4]</sub>
- 工作实践: 传统三维信息捕获依赖设备, 难以凭借普通设备实现。建议可以尝试复现一下NeRF<sub>{1}</sub>或者Hyperreel<sub>{2}</sub>, 相关的开源工作链接放在页面下方

- **体积视频流:**

- 主要方法: 3D Tiling(显著性检测, 视频编码), 6DoF prediction(时间序列预测), Adaptive Streaming(决策)
- 文章推荐: Vivo<sub>[5]</sub>, Yuzu<sub>[6]</sub>, Vues<sub>[7]</sub>, Saliency-based tiling<sub>[8]</sub>
- 说明: 此方向处于起步阶段, 暂时没有可用的开源代码, 但相信很快这部分论文和代码会有爆发式出现。

体积视频演示视频: [https://www.youtube.com/watch?v=iwUkbi4\\_wW](https://www.youtube.com/watch?v=iwUkbi4_wW)

{1}.<https://github.com/bmild/nerf> {2}.<https://github.com/facebookresearch/hyperreel>

# 体积视频 - 文献推荐

- [1]. Kowalski M, Naruniec J, Daniluk M. Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors[C]//2015 international conference on 3D vision. IEEE, 2015: 318-325.
- [2]. Huang X, Mei G, Zhang J, et al. A comprehensive survey on point cloud registration[J]. arXiv preprint arXiv:2103.02690, 2021.
- [3]. Wang Z, Wu S, Xie W, et al. NeRF--: Neural radiance fields without known camera parameters[J]. arXiv preprint arXiv:2102.07064, 2021.
- [4]. Attal B, Huang J B, Richardt C, et al. HyperReel: High-fidelity 6-DoF video with ray-conditioned sampling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 16610-16620
- [5]. Han B, Liu Y, Qian F. ViVo: Visibility-aware mobile volumetric video streaming[C]//Proceedings of the 26th annual international conference on mobile computing and networking. 2020: 1-13.
- [6]. Zhang A, Wang C, Han B, et al. {YuZu}:{Neural-Enhanced} volumetric video streaming[C]//19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22). 2022: 137-154.
- [7]. Liu Y, Han B, Qian F, et al. Vues: practical mobile volumetric video streaming through multiview transcoding[C]//Proceedings of the 28th Annual International Conference on Mobile Computing And Networking. 2022: 514-527.
- [8]. Li J, Zhang C, Liu Z, et al. Optimal volumetric video streaming with hybrid saliency based tiling[J]. IEEE Transactions on Multimedia, 2022.

# NeRF and 3D Gaussian Splatting

- **NeRF (Neural Radiance Field) :**

- 静态NeRF: NeRF, InstantNGP, Plenoxels, TensorNeRF
- 动态NeRF: D-NeRF, HyperReel,
- 工作实践: NeRF是使用神经(神经网络训练)辐射场(物理上的光学模型)对三维场景进行建模的隐式表征方法。在光线变化、视角合成上比传统显示方法更好。

- **3D GS (3D Gaussian Splatting) :**

- 静态GS: 3D Gaussian Splatting
- 动态GS: 4D Gaussian Splatting, 4K4D
- 工作实践: GS在光学模型上使用和NeRF一样的辐射场模型, 在训练上则使用三维高斯分布+机器学习的过程, 同时使用了点云分布作为高斯分布的中心位置的先验, 使得GS在训练的速度快于NeRF, 硬件要求更低。

NeRF: <https://www.matthiewtancik.com/nerf>

3D Gaussian Splatting: <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>

# NeRF and 3D Gaussian Splatting - 文献推荐

## 论文阅读 (标题)

综述 **NeRF** *Nerf: Neural radiance field in 3d vision, a comprehensive review*

**3DGS** *A survey on 3d gaussian splatting*

基石 **NeRF** *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*

*Plenoxels: Radiance Fields without Neural Networks*

**3DGS** *3D Gaussian Splatting for Real-Time Radiance Field Rendering*

## 代码尝试

**NeRF** <https://github.com/bmild/nerf>

**3DGS** <https://github.com/graphdeco-inria/gaussian-splatting>

# AR 远程会议 (Teleconference/Telepresence)



- 基础领域

- 计算机图形学 (Computer Graphics)
- 人机交互 (Human Computer Interaction)
- 计算机视觉 (Computer Vision)

- 主要技术

- 三维角色动画 (3D Character Animation)
- 三维场景理解 (3D Scene Understanding)
- 三维交互技术 (3D Interactive Techniques)
- (360) 视频处理 (360 Video Processing)

- 其他关键词

- Telepresence, Teleconference, Telecollaboration, Avatar,  
Point Cloud, Mesh, NeRF, Co-Presence

- 推荐技能、工具学习

- Unity 3D: 用于基于头戴式显示器的AR, VR应用开发与研究
  - 组内主要使用 Microsoft HoloLens 2 ([对应的 Unity 教程](#))
  - 也可以使用 Unreal, 但组内大多使用 Unity
- ARKit 相关开发: 用于基于 iOS 设备的 AR 应用开发与研究 ([教程](#))
- 任意语言的网络编程: 用于会议各方之间进行数据与媒体的通信
  - 组内大多使用Python, 因为可以更灵活得整合深度学习处理模块, Unity 内使用 C#

- 目标会议与期刊

- [IEEE ISMAR](#), [IEEE VR](#), [ACM UIST](#), [IEEE TVCG](#), [ACM SIGGRAPH](#) ([Asia](#))

- 推荐课程

- [现代计算机图形学入门](#)
- [基于物理的计算机动画入门](#)

Microsoft HoloLens 2



Meta Quest Pro



Apple Vision Pro



更多背景知识和学习资料可按以上关键词搜索 (按英文关键词搜索会更准确, 资料更丰富)



# AR 远程会议 (Teleconference/Telepresence)

- 主要研究目标

- 在不同的场景摆放下, 进行 AR 会议中虚拟化身动作与位置的语义同步

- 探索新颖直观的可视化方法来增强 AR 会议的注意力感知, 提升共存感与参与感

- 从用户表示方法, 光照阴影效果, 动作与行为合理性等方面对 AR 会议体验进行优化

- 推荐阅读 (文章标题)

- Placement Retargeting of Virtual Avatars to Dissimilar Indoor Environments

- Interactive Augmented Reality Storytelling Guided by Scene Semantics

- Virtual Agent Positioning Driven by Scene Semantics in Mixed Reality

- Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration

- GazeChat: Enhancing Virtual Conferences with Gaze-aware 3D Photos

- OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings

- Holoportation: Virtual 3D Teleportation in Real-time

- Augmented Virtual Teleportation for High-Fidelity Telecollaboration

- MR360: Mixed Reality Rendering for 360° Panoramic Videos

- Live 3D Portrait: Real-Time Radiance Fields for Single-Image Portrait View Synthesis

- Real-time Shadow-aware Portrait Relighting in Virtual Backgrounds for Realistic Telepresence

模型架构搜索和轻量化方向



# 神经网络架构搜索 (NAS)

以点带面，以下任务周期因人而异，一般为一到两月。

- **蒙**：读经典NAS论文 “AutoFormer: Searching Transformers for Visual Recognition” 。
- **渐**：以AutoFormer为点，自行查阅相关文献与资料，解决不明白的NAS概念。
- **履**：复现AutoFormer的Test，验证作者预训练权重结果。  
<https://github.com/microsoft/Cream/tree/main/AutoFormer#test>
- **损**：修改AutoFormer源码，屏蔽搜索时的模型验证过程，用新的简单易得的指标（比如随机数）替代原有的top-1指标，使得搜索过程在一分钟内完成。

# 神经网络架构搜索 (NAS)

论文复现建议与  
环境搭建建议

- **蒙**: 熟悉Python语言, PyPi换源, Python包安装与调用。
- **渐**: 熟悉Git操作, Python中requirements.txt的使用。
- **履**: 熟悉Docker操作, Docker镜像构建 (Dockerfile), Docker容器运行。
- **恒**: 使用Dockerfile构建AutoFormer运行环境 (包含CUDA), 通过docker挂载的方式(docker run -v xxx:yyy), 向容器挂载AutoFormer代码与Imagenet数据集, 最后指定使用的GPU(docker run --gpus)。完成能够成功运行AutoFormer的容器。具体细节注意查阅Docker官方文档。(也可以不使用Docker, 但是你搭建的环境很可能会难以迁移与复现。)

# 模型轻量化和高效微调

- 模型轻量化方法
  - 参数量化 (Parameter Quantization)
  - 参数剪枝 (Parameter Pruning)
  - 卷积核压缩及矩阵分解 (Compressed Convolutional Filters and Matrix Factorization)
  - 神经网络架构搜索 (Network Architecture Search)
  - Survey: <https://arxiv.org/pdf/2003.11066.pdf>
- 模型高效微调方法基础：
  - LoRA, QLoRA
  - Adapter Tuning
  - Prompt Learning
- 背景知识
  - Transformer: <https://openreview.net/pdf?id=2q334K-GjaL>
  - Vision Transformer: <https://arxiv.org/pdf/2010.11929.pdf>
  - DETR: <https://arxiv.org/pdf/2005.12872.pdf>
  - Segment Anything: <https://github.com/facebookresearch/segment-anything>
  - PEFT: Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey
- 推荐会议
  - ICLR ICML NIPS CVPR ECCV ICCV
- 推荐文章及仓库
  - When Object Detection Meets Knowledge Distillation: A Survey
  - Which Tokens to Use? Investigating Token Reduction in Vision Transformers
  - <https://github.com/huggingface/peft>

Diffusion-based AIGC

# Diffusion-based AIGC

基本原理：Diffusion；开源模型：StableDiffusion

- 前置知识

- 贝叶斯定理、后验分布、变分推理、ELBO估计、马尔可夫链

- 基石文章

- Diffusion Models Beat GANs on Image Synthesis
- Denoising Diffusion Probabilistic Models
- Denoising Diffusion Implicit Models
- High-Resolution Image Synthesis with Latent Diffusion Models
- Classifier-Free Diffusion Guidance

- 推荐学习

- DDPM <https://zhuanlan.zhihu.com/p/661279401>
- DDIM <https://zhuanlan.zhihu.com/p/639540034>
- Generative models <https://zhuanlan.zhihu.com/p/558937247>
- Conditional guidance <https://zhuanlan.zhihu.com/p/675387002>
- Diffusion (李宏毅) [https://www.bilibili.com/video/BV14c411J7f2?p=1&vd\\_source=ff583ac40f420a3ea244761fa70cbc15](https://www.bilibili.com/video/BV14c411J7f2?p=1&vd_source=ff583ac40f420a3ea244761fa70cbc15)
- StableDiffusion [https://zhuanlan.zhihu.com/p/634646929?utm\\_id=0](https://zhuanlan.zhihu.com/p/634646929?utm_id=0)

- 推荐会议

- ICLR ICML NIPS CVPR ICCV ECCV AAAI

- 综述文章

- Diffusion models in vision: A survey
- Controllable Generation with Text-to-Image Diffusion Models: A Survey

- 代码复现

- 初始：学会使用Diffusers包调用StableDiffusionXL生成图片  
(<https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0>)
- 进阶：在Diffusers版本的StableDiffusionXL中复现Training-Free Layout Control with Cross-Attention Guidance论文（无参考代码）

# Continual/ Lifelong Learning

机器学习算法 & 理论

# Continual/lifelong learning

- 前置知识
  - 增量学习、在线持续学习、零样本学习, 泛化分析
- 基石文章
  - Gradient based sample selection for online continual learning.
  - Learning fast, learning slow: A general continual learning method based on complementary learning system.
  - Rainbow memory: Continual learning with a memory of diverse samples.
- 推荐会议
  - ICLR ICML NeurIPS CVPR
- 推荐期刊
  - JMLR, AI, IEEE TIT
- 综述文章
  - A Comprehensive Survey of Continual Learning: Theory, Method and Application
  - Continual Learning of Large Language Models: A Comprehensive Survey
- 代码复现
  - <https://github.com/xialeiliu/Awesome-Incremental-Learning>(各类CL的代码, 可选)
  - <https://github.com/GMvandeVen/continual-learning> (Nature Machine Intelligence, 建议)

# 适合复现的项目

- 流媒体实时传输与分析方向： Pensieve (<https://web.mit.edu/pensieve/>)
- 沉浸式交互技术与应用方向：
  - 3D Gaussian Splatting: <https://github.com/graphdeco-inria/gaussian-splatting>
  - NeRF: <https://github.com/bmild/nerf>
- 模型轻量化和高效微调： SAM以及SAM-adapter (<https://github.com/KidsWithTokens/Medical-SAM-Adapter>)
- 神经网络架构搜索： Autoformer (<https://github.com/microsoft/Cream/tree/main/AutoFormer>)
- Diffusion-based AIGC：
  - 初始：学会使用Diffusers包调用StableDiffusionXL生成图片 (<https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0>)
  - 进阶：在Diffusers版本的StableDiffusionXL中复现Training-Free Layout Control with Cross-Attention Guidance论文（无代码）
- Continual/ Lifelong Learning
  - <https://github.com/xialeiliu/Awesome-Incremental-Learning>(各类CL的代码，可选)
  - <https://github.com/GMvandeVen/continual-learning> (Nature Machine Intelligence, 建议)