Kartaverse Research

December 2023 Scrapbook of Gen AI/ML/NeRFian Research Papers

Doc Compiled By: Andrew Hazelden (andrew@andrewhazelden.com)
Doc Created: 2023-12-06
Doc Updated: 2023-12-12 07:51 PM (AST)

First Draft

Overview

This document is a summary of notable computer vision and ML research released, promoted, or discussed in December 2023.

Starting this month, I hope to compile a monthly scrapbook, with one new overview document released each month. The goal is to document the evolution of these novel computational imaging and machine learning ideas and techniques. Let's see how long I can keep up with the release pace of this seemingly unending 24x7 stream of ML research before it becomes too much for even me to track!

The selected research materials were filtered using the perspective of what info would be relevant to artists and creative technologists who work in the immersive media sector (XR/VP/VFX). This was the key deciding factor of what gets included in the monthly scrapbook list or skipped. There are easily 100x more innovative things that happened inside this "monthly snapshot" timescale but it becomes overwhelming to sift through it all at a certain point.

Special Thanks

Special thanks go out to contributions from friends, Rob Sloan, Deborah Stone, Peter Morse, Scott Broock, Andrew Cochrane, Josiah Reeves, Jared Sandrew, Alexis Haggar, Benjy von Cramon, Jacob Danell, Dunn Lewis, Maciej Zemojcin, and industry contacts on LinkedIn/Slack/Discord/Instagram.

Without the combined effects of all your guidance, none of this information would have been collected together for a unified scrapbook experience. 📔

Run-Anywhere Code Development

The Cosmopolitan compiler extension provides an interesting idea for better Windows/Linux/macOS portability of c-code language-based runtimes. Cosmopolitan makes accessible ML runtimes like llamafile possible.

Project Page:

- https://justine.lol/cosmopolitan/index.html
- https://justine.lol/cosmopolitan/howfat.html

GitHub Repo:

- https://github.com/jart/cosmopolitan

Depth Estimation

PatchFusion

PatchFusion is an interesting ML paper that offers improved depth estimation workflows. The input to the PathFusion library is 2D monoscopic source imagery.

This approach could be used in virtual production-centric VAD (Virtual Art Departments) to "dimensionalize" 2D source footage.

The ML synthesised depth data allows basic RGBD workflow including 6DoF camera motion, bokeh lens defocus simulation, Z-depth keying. The Z-depth information can help art direct the output from Stable Diffusion "ControlNet" based generative AI usage.

GitHub Repo:

- https://github.com/zhyever/PatchFusion
- https://zhyever.github.io/patchfusion/

Research Paper:

- https://arxiv.org/abs/2312.02284

LinkedIn Shared Example Depth Output:

- https://www.linkedin.com/posts/thejab_seriously-impressed-with-patchfusion-high-activity-7139485729683902464-TnSD

Marigold (MIT licensed)

Marigold is a permissively licensed depth estimation toolset.

GitHub Repo:

- https://marigoldmonodepth.github.io/
- https://github.com/prs-eth/marigold

Research Paper:

- https://arxiv.org/abs/2312.02145

Hugging Face:

- https://huggingface.co/spaces/toshas/marigold

LinkedIn Post:

- https://www.linkedin.com/posts/thejab_the-first-double-image-is-stable-diffusion-ugcPost-7140202979588030464-Puto?utm_source=share&utm_medium=member_desktop

Machine Learning

Books and Guides

To get started with ML, here are several helpful learning guides for writing and using algorithms. These resources can help one make sense of the rapidly changing ML/Computer vision technologies by providing a firm grasp of the technology that underpins algorithms.

Introduction to Algorithms Book

- Amazon | Introduction to Algorithms, 3rd Edition
- Wikipedia | Introduction to Algorithms

TBD

- More learning resources to follow

Runpod

RunPod GPU services can provide the compute cloud needed for custom 3rd party model training

- https://www.runpod.io/

MLX

MLX is an interesting PyTorch-style NN framework optimized for Apple Silicon. It can easily run on low resource laptops and other "home office" level computers.

- https://github.com/ml-explore/mlx
- https://github.com/ml-explore/mlx-examples

llamafile

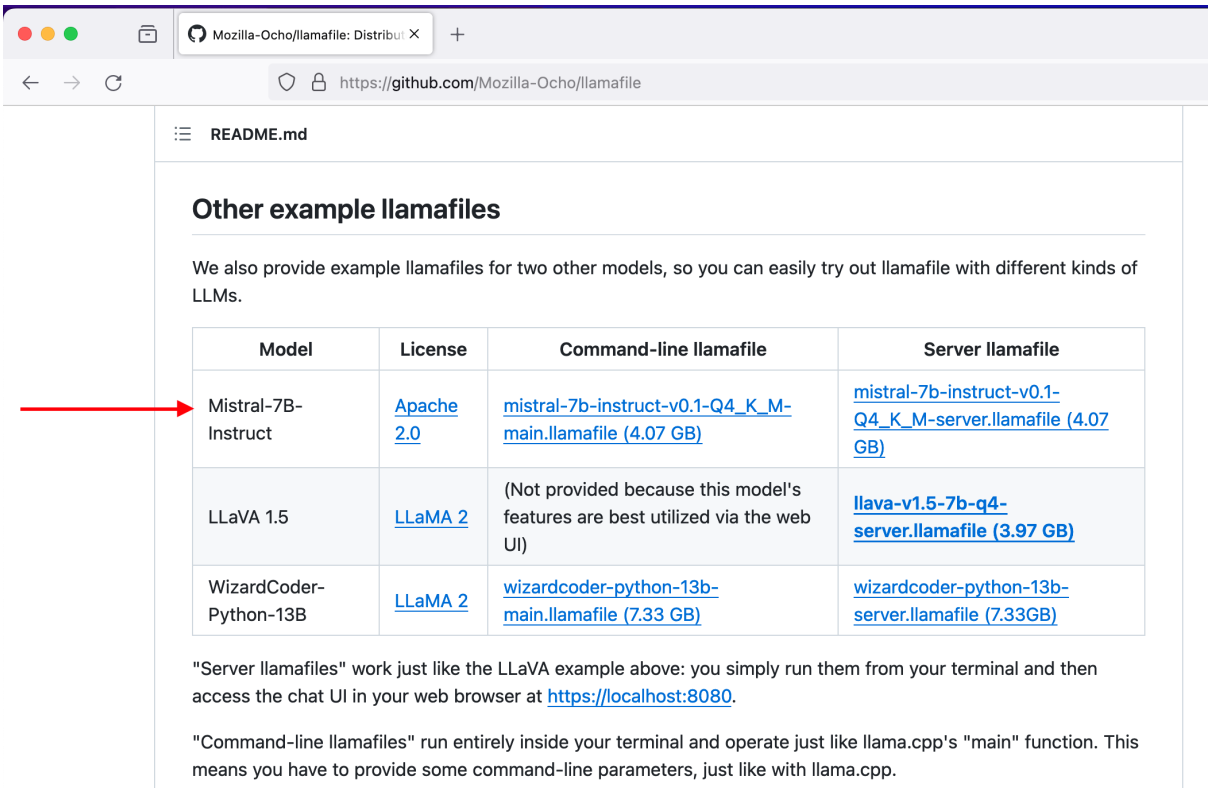
llamafile is a cross-platform compatible local on-premise LLM, that can run on a CPU or a GPU. llama file is powered by the Cosmopolitan technology for a single run-anywhere executable.)

- https://github.com/Mozilla-Ocho/llamafile
- https://simonwillison.net/2023/Nov/29/llamafile/
- YouTube | Igor Ričanović | llamafile | Local LLMs Made Easy

Mistral Model Tuning:

- https://www.linkedin.com/posts/drjimfan_mistral-founded-1-yr-ago-closed-a-400m-activity-7140021470549733376-nnlG?utm_source=share&utm_medium=member_desktop
- https://mistral.ai/news/mixtral-of-experts/

llamafile has a Mistral 7B model in the bundled examples one can download and run:



Rob Sloan Provided Tip: There are several firms working on finetunes or alternatives to existing ML models. These companies are trying to perfect the existing models for use in different applications and use cases.

- https://www.linkedin.com/company/nousresearch/
- https://www.linkedin.com/company/mosaicml/

A* Pathfinding

This pathfinding approach is great for street map route planning. Side note: A* Pathfinding concepts can also (potentially) extended over to explaining part of OpenAI Q*bert claims of using a Q* (Q star)

deep reinforcement learning solver idea to help with chain of thought reasoning for general purpose AI systems too.

OpenAI Q*

- The Verge | OpenAI Breakthroughs
- Arstechnica | OpenAI Q*bert
- OpenAI Form | Q*
- Reuters | Q*
- Reddit | Q*

General Purpose Pathfinding

- YouTube | A* (A-Star) Pathfinding Algorithm Visualization on a Real Map
- GitHub | A* Pathfinding C++ Library
- Game Engine Integrations
    - Unity | Aron Granberg | Aline Debugging
    - Unity | Aron Granberg | A* Pathfinding Pro
    - YouTube | Aron Granberg | A* Pathfinding Pro
    - Aron Granberg Website | A* Pathfinding Pro

City Route Analysis:

- QGIS | Vector Analysis | Network Analysis
- OpenStreet Map | Route Analysis
- Open Street Map | Open Source Routing Machine
- PostGIS - If one is exploring pathfinding approaches, for route planning on GIS city-level street data, a good companion tool is PostGIS/PostgreSQL for geographic data centric database storage tech.

NeRF Projects & Tech

Point2CAD

Convert point cloud data into CAD based solids models

Project Page:

- https://www.obukhov.ai/point2cad

GitHub Repo:

- https://github.com/YujiaLiu76/point2cad

Research Page:

- https://struco3d.github.io/cvpr2023/papers/11.pdf

LinkedIn Page:

- https://www.linkedin.com/posts/robcsloan_cad-3dmodeling-computervision-ugcPost-7140470053648424960-owDf

Adaptive Shells for Efficient Neural Radiance Field Rendering

Project Page:

- https://research.nvidia.com/labs/toronto-ai/adaptive-shells/

Video:

- https://research.nvidia.com/labs/toronto-ai/adaptive-shells/assets/suppl_video.html

Research Page:

- https://arxiv.org/abs/2311.10091

LinkedIn Page:

- https://www.linkedin.com/posts/alexandrecarlier_adaptive-shells-was-just-awarded-best-paper-ugcPost-7140391780662824961-k1nx

NeRF Studio | Blender VFX Addon

This is a previously released toolset that is getting renewed attention for allowing VFX scene integration like work to happen for mixing CG elements with live action background plate footage.

Project Page:

- https://docs.nerf.studio/extensions/blender_addon.html

GitHub Repo:

- https://github.com/nerfstudio-project/nerfstudio/

LumaAI WebGL (threeJS gSplat) Library

Project Page:

- https://lumalabs.ai/luma-web-library
- https://neuralradiancefields.io/luma-web-webgl-library-from-luma-ai/

GitHub Repo:

- https://github.com/lumalabs/luma-web-examples

LinkedIn:

- https://www.linkedin.com/posts/gravicle_threejs-gaussiansplatting-3dgs-activity-7138218988097982464-4tsV?utm_source=share&utm_medium=member_ios

Gaussian Splatting SLAM

This project shows an implementation of real-time SLAM mapping (Simultaneous localization and mapping) approaches via gSplat camera alignment and interactive 3D reconstruction.

Project Page:

- https://rmurai.co.uk/projects/GaussianSplattingSLAM/

YouTube Videos:

- YouTube | Gaussian Splatting SLAM

Research Paper:

- https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/dyson-robotics-lab/hide-et-al_GaussianSplattingSLAM_Dec2023.pdf

via Bilawal Sidhu on LinkedIn

- https://www.linkedin.com/posts/bilawalsidhu_holy-crap-real-time-3d-gaussian-splatting-ugcPost-7140049792079970305-HxWQ?utm_source=share&utm_medium=member_ios

GPS-Gaussian: Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis

GitHub Repo:

- https://shunyuanzheng.github.io/GPS-Gaussian

NVIDIA EmerNeRF

LinkedIn Post:

- https://www.linkedin.com/feed/update/urn:li:activity:7137916219646509057/

Relightable Gaussian Codec Avatars

GitHub Repo:

- https://shunsukesaito.github.io/rgca/

Research paper:

- https://arxiv.org/pdf/2312.03704.pdf

SuGaR (Surface-Aligned Gaussian Splatting)

LinkedIn Post:

- https://www.linkedin.com/posts/robcsloan_unrealengine-gaussiansplatting-mesh-ugcPost-7138981748717051905-T6KU?utm_source=share&utm_medium=member_ios

GitHub Repo:

- https://anttwo.github.io/sugar/

Research Paper:

- https://arxiv.org/abs/2311.12775

NeRFiller: Completing Scenes via Generative 3D Inpainting

LinkedIn Post:

- https://www.linkedin.com/posts/petermorse_nerfiller-completing-scenes-via-generative-activity-7138755796657442816-vK9e

Project Page:

- https://ethanweber.me/nerfiller/

Research Paper:

- https://arxiv.org/abs/2312.04560

Material Palette: Extraction of Materials from a Single Image

GitHub Repo:

- https://astra-vision.github.io/MaterialPalette/

Research Paper:

- https://arxiv.org/abs/2311.17060

PanoHead | Geometry-Aware 3D Full-Head Synthesis in 360°

GitHub Repo:

- https://sizhean.github.io/panohead
- https://github.com/sizhean/panohead

Research Paper:

- https://arxiv.org/abs/2303.13071

ReconFusion

ReconFusion represents a transformative approach in the field of 3D reconstruction, particularly with Neural Radiance Fields (NeRFs). This novel technique efficiently reconstructs real-world scenes using just a few photos, overcoming the challenges of the traditionally time-consuming process that requires extensive image datasets. ReconFusion leverages a diffusion model for novel view synthesis, enhancing the quality and practicality of 3D model generation.

Project Page:

- https://reconfusion.github.io/

Research Paper:

- https://arxiv.org/abs/2312.02981

PyNeRF

PyNeRF: Enhancing Neural Radiance Fields with Pyramidal Structure, presented at NeurIPS 2023, introduces a significant enhancement in Neural Radiance Fields (NeRFs). This innovative approach addresses the limitations of traditional NeRFs in handling scenes captured at different camera distances, a challenge that has been a bottleneck for realistic 3D rendering and reconstruction.

Project Page:

- https://haithemturki.com/pynerf/

GitHub Repo:

- https://github.com/hturki/pynerf

Hugging Face:

- https://huggingface.co/papers/2312.00252

Image and 3D object Selection and Segmentation

Faster Region CNN

A fresh implementation of the Faster R-CNN object detection model in both PyTorch and TensorFlow 2 with Keras, using Python 3.7 or higher. Although several years old now, Faster R-CNN remains a foundational work in the field and still influences modern object detectors.

GitHub Repo:

- https://github.com/trzy/FasterRCNN

Research Paper:

- https://arxiv.org/abs/1506.01497

LinkedIn Post:

- https://www.linkedin.com/posts/arkadiuszszadkowski_deeplearning-geoai-3dgis-activity-7138788174192975872-I2Bn

OmniSeg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning

An omniversal 3D segmentation method, which (a) takes as input multi-view, inconsistent, class-agnostic 2D segmentations, and then outputs a consistent 3D feature field via a hierarchical contrastive learning framework. This method supports (b) hierarchical segmentation, (c) multi-object selection, and (d) holistic discretization interactively.

GitHub Repo:

- https://oceanying.github.io/OmniSeg3D/

Research Paper:

- https://arxiv.org/pdf/2311.11666.pdf
- https://github.com/trzy/FasterRCNN/blob/master/docs/publications/fast_rcnn.pdf

EfficientSAM (Efficient Segment Anything)

GitHub Page:

- https://yformer.github.io/efficient-sam/

Hugging Face:

- https://huggingface.co/spaces/yunyangx/EfficientSAM

GluonCV

GluonCV is a flexible and powerful computer vision toolset that allows object segmentation, action detection and more.

Project Page:

- https://cv.gluon.ai/

GitHub Repo:

- https://github.com/dmlc/gluon-cv

Gen AI

Gen AI Image/Video

RunwayML

Product Page:

- https://runwayml.com

Company News:
- https://runwayml.com/blog/runway-partners-with-getty-images/

SDXL / SDXL Turbo

Project Pages:
- https://stablediffusionweb.com/StableDiffusionXL
- https://stability.ai/news/stability-ai-sdxl-turbo

GitHub Repo:
- https://github.com/Stability-AI/generative-models

Hugging Face Page:
- https://huggingface.co/stabilityai/sdxl-turbo

Stability AI ClipDrop Page:
- https://clipdrop.co/stable-diffusion-turbo

Reddit Pages:
- https://www.reddit.com/r/StableDiffusion/comments/186496i/introducing_sdxl_turbo_a_realtime_texttoimage/
- https://www.reddit.com/r/StableDiffusion/comments/15aapcb/sdxl_10_is_out/

Automatic 1111 WebUI

GitHub Repo:
- https://github.com/AUTOMATIC1111/stable-diffusion-webui
- https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features
- https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Optimum-SDXL-Usage

ComfyUI

ComfyUI provides a procedural AGen AI based node graph interface. This is initially scary for a new user but it allows one to create more complex editable workflows that are parametric.

YouTube Videos:
- YouTube | STABLE VIDEO DIFFUSION | COMFYUI
- YouTube | ComfyUI AnimateDiff Prompt Travel: Runpod.io Cloud GPUs Tutorial

GitHub Repo:
- https://github.com/comfyanonymous/ComfyUI

LinkedIn Post:

- https://www.linkedin.com/posts/brian-bullock-9aa0512_stabilityai-comfyui-machinelearning-ugcPost-7136018046904635392-E0PR

Pika Labs

- https://pika.art/
- https://pika.art/launch

Google Gemini

YouTube Videos:

- YouTube | Gemini: Google's newest and most capable AI model

Gemini News:

- https://blog.google/technology/ai/google-gemini-ai/
- https://deepmind.google/technologies/gemini/#introduction
- https://www.theverge.com/2023/12/7/23992737/google-gemini-misrepresentation-ai-accusation

Visual Electric

- https://visualelectric.com/

DreaMoving: A Human Dance Video Generation Framework based on Diffusion Models

YouTube Videos:

- #DreaMoving: A Human Dance Video Generation Framework based on #diffusion Models #videogeneration

GitHub Repo:

- https://github.com/dreamoving/dreamoving-project/

Research Paper:

- https://arxiv.org/abs/2312.05107

Hugging Face page:

- https://huggingface.co/papers/2312.05107

Gen AI 3D Geometry Modelling

XCube | Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies

Project Page:

- https://research.nvidia.com/labs/toronto-ai/xcube/

Research Paper:

- https://arxiv.org/abs/2312.03806

Audio AI

Training Custom Voice Models

Kartaverse research notes about training a personal voice model from custom audio recordings can be viewed here:

- Kartaverse Research | Voice Talent Trained and Owned AI Voice Models

Meta AudioBox

- https://ai.meta.com/blog/audiobox-generating-audio-voice-natural-language-prompts

Meta SeamlessExpressive

- https://seamless.metademolab.com/expressive

Virtual Location Scouting

LinkedIn Post:

- https://www.linkedin.com/posts/saultoresan_virtualscouting-vr-visualization-activity-7138867327957127168-cd7K

What's coming in 2024

The next wave of R&D in 2024 will likely include Gmesh (Gaussian Mesh) concepts, better temporal stability from generative AI, and more data processing tools adding multi-view support.

Also, we will start to see an industry-wide move to have a richer 3D scene graph that allows for hybrid volumetric data types that are intermixed during scene reconstruction, volumetric data editing, and interactive rendering workflows.