Paper #24186

Generative AI-Powered 3D-Content Creation for Military Training

Deepak Haste[1], Michael Renda[1], Sudipto Ghoshal[1], Eduardo Barrera[2], Jason Wong [3]

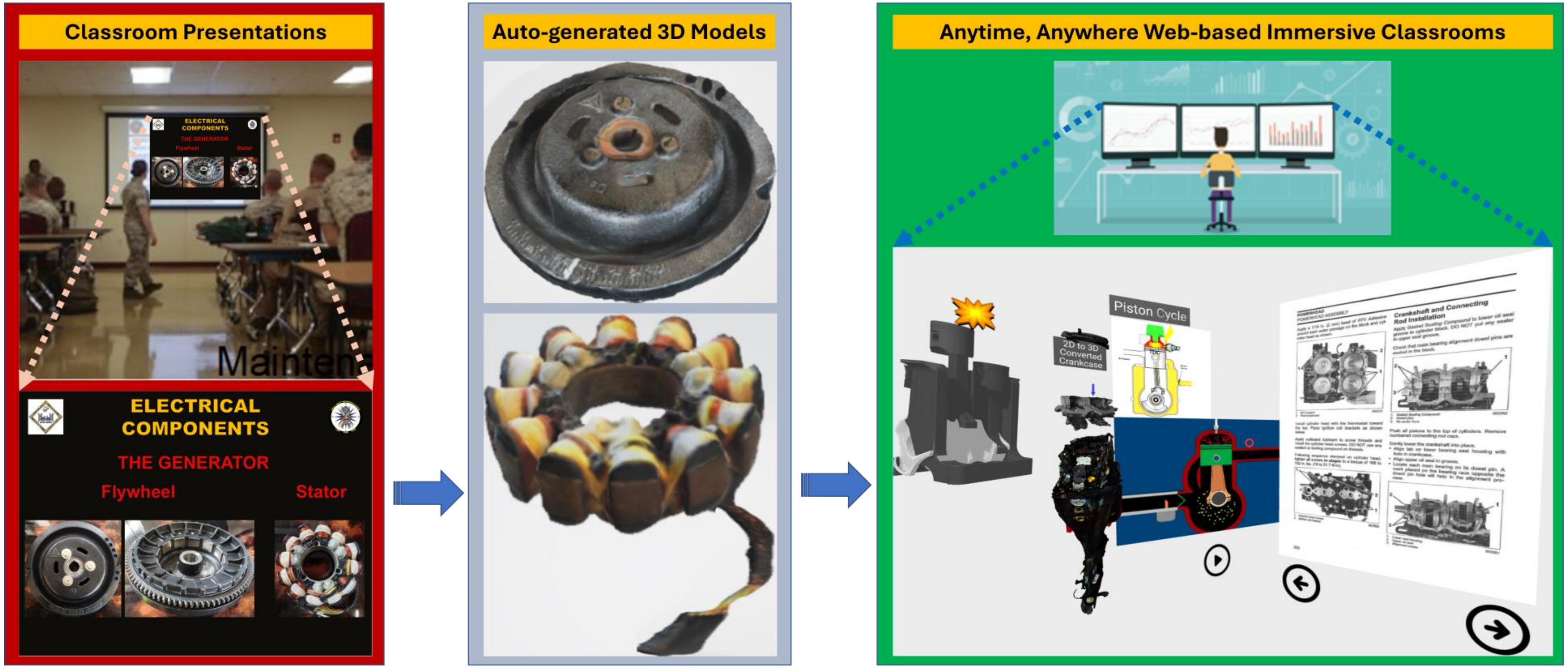
Qualtech Systems, Inc.[1], Charles River Analytics[2], Naval Information Warfare Center Pacific[3]

Background – Challenges

- USMC's immersive training initiative
  - Enterprise Ground Maintenance Training Simulator (EGMTS)
  - AR/VR for improved readiness of warfighting equipment
  - Enhancing technician skills and abilities
- Challenges in generation of immersive environments
  - Lack of readily available 3D models
  - Need for CAD and graphics software
  - Need for specialized, expensive hardware
- Need for a 3D model repository
  - Use generative AI to create 3D models from 2D pictures of USMC equipment
  - Reduces program costs and accelerates student training

Immersive Content Generation - Approach

- Enterprise training platform
  - 3D scans from 2D pictures using photogrammetry on cellphones cameras
  - Vision-based generative AI techniques for 2D-to-3D conversion



  - Reusable 3D object libraries for use in schoolhouse

Immersive Content Generation - The Framework

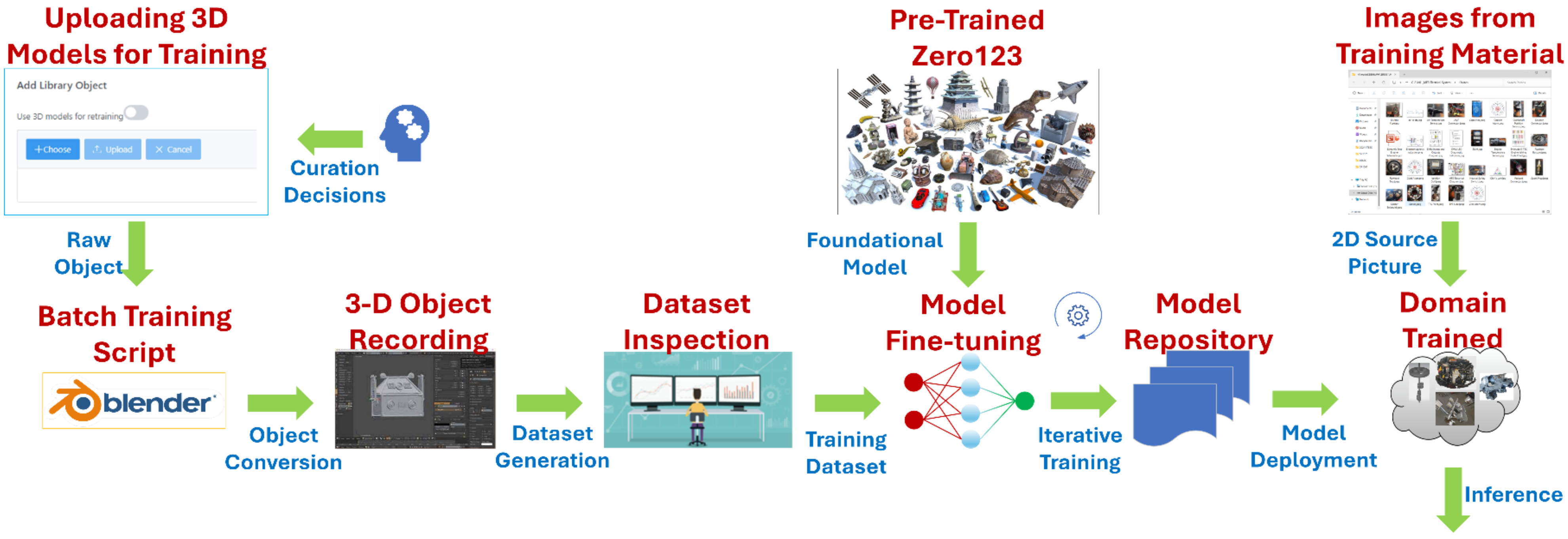
- State-of-the-Art Zero-1-to-3
  - Predecessor NeRFs and GNNs lacked critical features such as fidelity and realism
  - 2D image to 3D model
  - Zero-1-to-3 leverages geometric priors from large-scale diffusion models to convert 2D images to 3D models and retain original textures

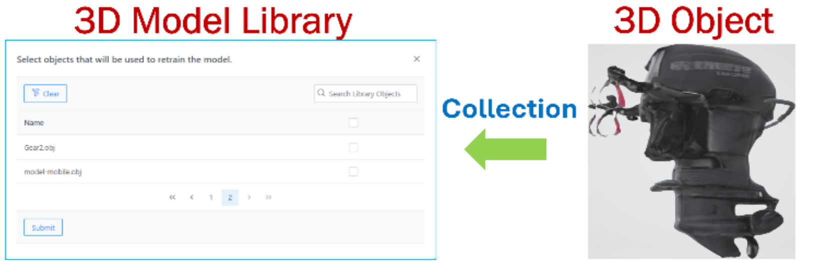
Immersive Content Generation - USMC Schoolhouse Considerations

- Automated content transformation with updateable 3D content
  - Immersive content hosted on LMS such as Moodle
- 3D model training challenges
  - Models not pre-trained for USMC electromechanical domain
    - E.g., Marine engines at MCES, combat vehicles at MCCSSS, combat radio equipment at MCCES, etc.
    - Need for domain-specific fine training using human-in-the-loop
  - Classified material taught at schoolhouses
    - Using cloud-based generative AI for 3D object creation violates USMC cybersecurity policies
    - Utilize local server hardware for on-prem generative AI

Concept of Operations (CONOPS)

- Data curation for target domain
  - Convert relevant 3D objects for training and fine-tuning Zero-1-to-3

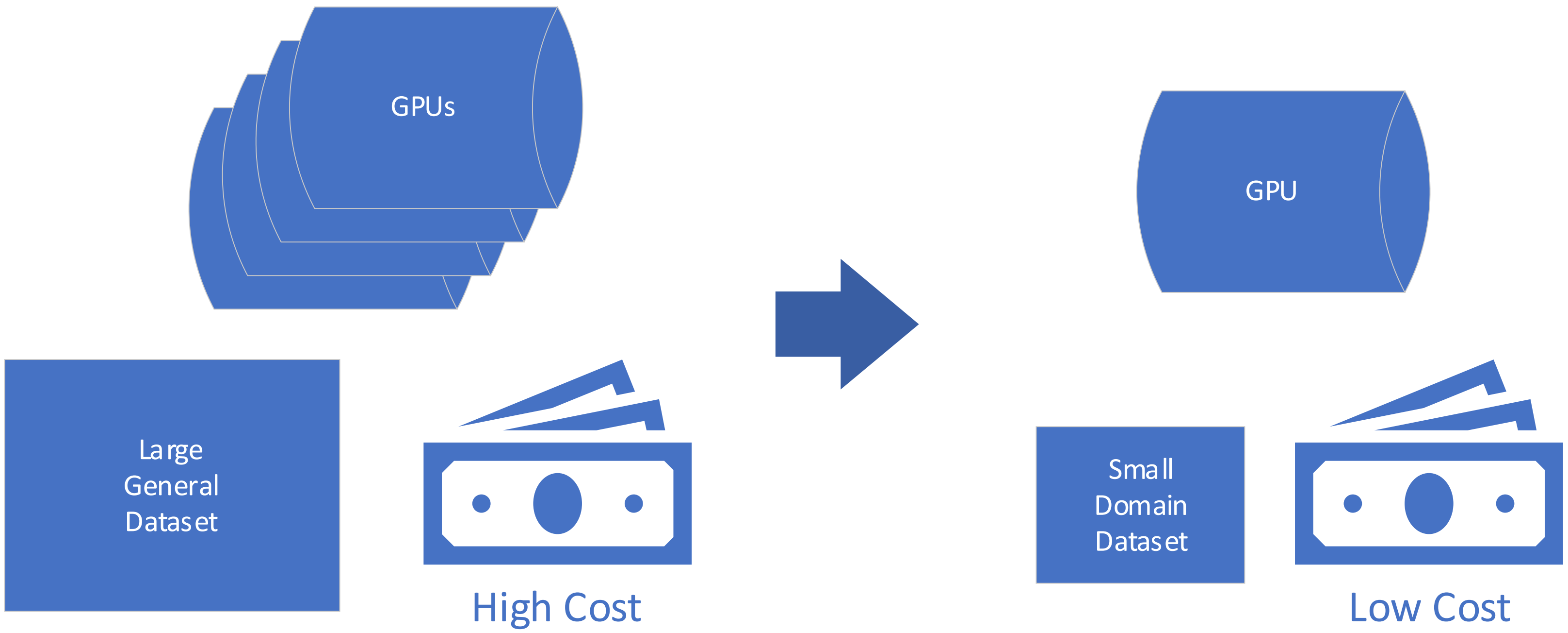




- Model fine-tuning for target domain
  - Using curated dataset
- 3D object generation from domain-trained model
  - Converts 2D image to 3D object

Experimental Setup

- Tailored Zero-1-to-3 for use in USMC schoolhouse environments, modifying model size, efficiency, and target domain
- Training requirements lowered from *eight A100 80GB* GPUs to one *A100 40GB* GPU
  - After adjustments, model training cost is lowered to $25K



- Research team fine-tuned three popular pre-trained model weights for Zero-1-to-3
  - Zero123-Base: original author weights, fine-tuned to *Zero123-Base_USMC*
  - Zero123-XL: updated weights with larger non-USMC dataset, fine-tuned to *Zero123-XL_USMC*
  - Zero123-Stable: StabilityAI improved weights, fine-tuned to *Zero123-Stable_USMC*

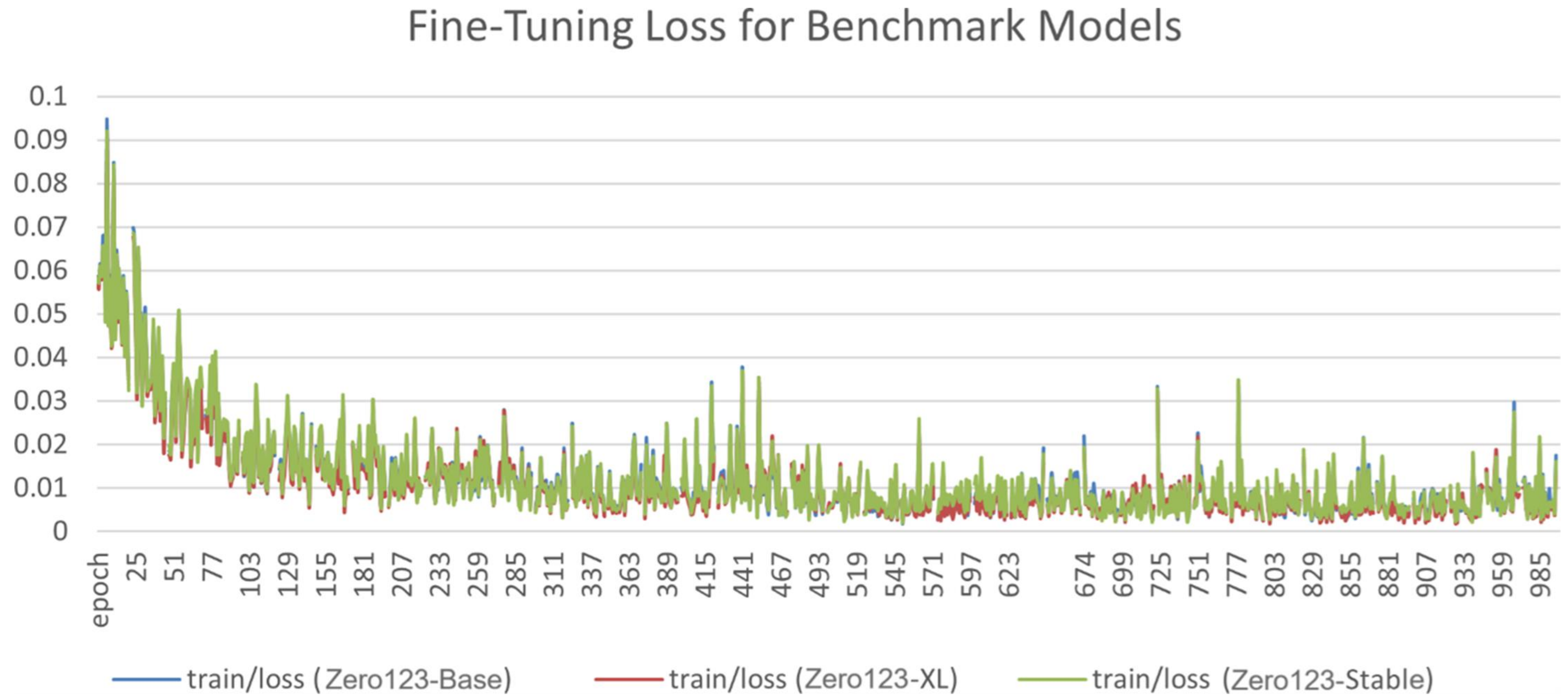
Experimental Setup (Cont.)

- Performance analysis of different models for generating 3D content
- Benchmarking study includes model performance metrics
  - Comparison shows the model training methodology and required training conditions

	Zero123-Base (Liu, 2023)	Zero123-XL (Updated Objaverse)	Zero123-Stable (StabilityAI Weights)
Training Dataset	The standard Obajverse dataset of 800,000 objects	Updated 2024 ObjaverseXL dataset of 12 million objects	Updated 2024 ObjaverseXL dataset of 12 million objects
Fine-Tuning Dataset	In-House Electro-mechanical objects dataset	In-House Electro-mechanical objects dataset	In-House Electro-mechanical objects dataset
Specialized Equipment/Environment	NVIDIA A100 40 GB Tensor Core GPU	NVIDIA A100 40 GB Tensor Core GPU	NVIDIA A100 40 GB Tensor Core GPU
Training Time on 100 Objects	1 – 2 hours	12 – 24 hours	12 – 24 hours
Object Generation Speed	10 - 30 minutes	10 - 30 minutes	10 - 30 minutes
Output Quality	Low	Medium	Medium

Experimental Results

- Fine-tuning through iterative adjustments to learning rate, batch size, training steps, etc.
- Evaluated model's convergence, performance, and loss reduction using three different pretrained weights benchmarks
- Fine-tuning process led to improvement in loss reduction across all three benchmarks

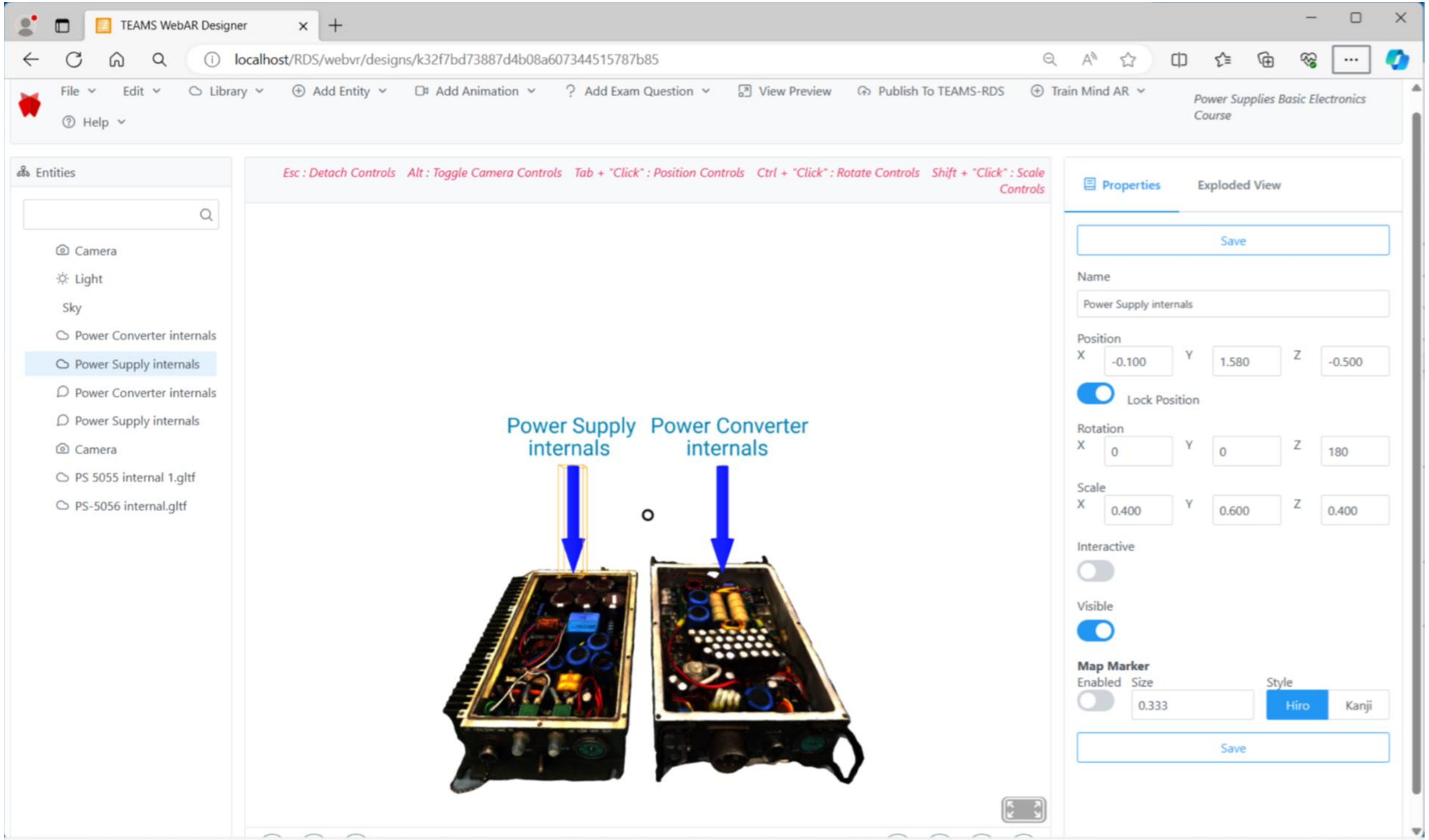


Experimental Results

- Progression from original image to 3D reconstruction using six different models
  - Fine-tuned models show improved quality, textures, and detail in the target domain

Integration of 3D Models with the Immersive Content Creation Pipeline

- AR/VR Designer
  - Supports PNG, JPEG, SVG file formats
  - Reusable library
  - No specialized software/hardware
  - Supports animations, component identification tests and practical applications
  - Moodle compliant



Cost Benefit Analysis of 3D Model Creation Pipelines

- Cost benefit analyses of 3D model creation pipelines
  - Generative AI
  - Photogrammetry
  - Vendor-sourced

	Generative AI-powered 2D image to 3D model conversion	3D Scanning/Photogrammetry	Outside contractors specializing in 3D Modeling
Source material	Common images found in USMC schoolhouse material such as presentations and student handouts	Physical training objects found in a schoolhouse	Thousands of high-quality photos taken over several days. Graphic designers process photos into a scale-model.
Specialized equipment/environment	NVIDIA A100 40 GB Tensor Core GPU	iPhone 13 Pro or later	Unity
Cost	Initial investment of $25K for Deep Learning server. $0 subsequently.	$1K for iPhone 13 Pro or later. $0 subsequently.	Hundreds of thousands of dollars to implement each high-complexity procedures
Time to create 3D model for use in a library	10 - 30 min	15 min	1 month of development time per equipment
3D object quality	Low-Medium	Medium	High
Scalability	High. Ability to batch-covert several images at once.	Medium. Proportional to the extent of the effort put in 3D-scanning the equipment.	Low. Vendor will have to be contracted for each equipment.
Customization potential	Medium. Users can pick and choose pictures of desired physical objects from course material.	High. Ability to organically generate 3D models from desired physical objects.	Low. Limited ability to supply own 3D content (mostly CAD). Tens of thousands of dollars and up to weeks and months for modifications.

Current Limitations and Future Enhancements

- Model training limited to single objects against plain backgrounds
  - Scenes with complex backgrounds not supported
- One-shot conversion of 2D photos
  - Multiple photos from different angles not supported
- Future enhancements in 3D object generation
  - Support for “multi-shot” 2D to 3D conversion
  - Support for scenes with complex backgrounds
  - Combine text-to-image conversion diffusion models with image-to-3D models, to enable Text-to-3D model using “prompts”

Conclusions

- Automated web-based 3D content-generation pipeline
  - Zero-1-to-3 model fine-tuned with minimal domain training rapidly produces 3D models
  - 3D scanning using photogrammetry
  - Integrates with AR/VR content library for generation of immersive content
- Provides time savings from manual creating 3D creation
- Does not require 3D modeling or graphical design skills
- Saves expenses on professional CAD modeling licenses and contracting services
- Minimizes physical training assets and saves schoolhouse instructors' time

Acknowledgements

- Thank-You!
  - Dr. Peter Squire from ONR for supporting and funding the research
  - Dr. Jason Wong for his technology guidance
  - Mr. James (Scotty) Moore, Capabilities Branch Head at MCES, for access to the SCMC training material, and for facilitating touchpoints with the Marines at the MCES
  - GySgt Craig Ruhnke at the MCES for providing usage and valuable feedback
  - Dr. Crystal Maraj, for valuable guidance as our Birddog through the paper submission process

Questions?

- Contact Information
  - Mr. Deepak Haste
    Qualtech Systems, Inc.
    100 Corporate Place Suite 220
    Rocky Hill, CT 06067
    deepak@teamqsi.com
  - Dr. Jason Wong, Ph.D.
    Naval Information Warfare Center Pacific
    San Diego, California
    jason.h.wong.civ@us.navy.mil
  - Mr. Eduardo Barrera
    Charles River Analytics
    625 Mt Auburn St #3
    Cambridge, MA 02138
    ebarrera@cra.com