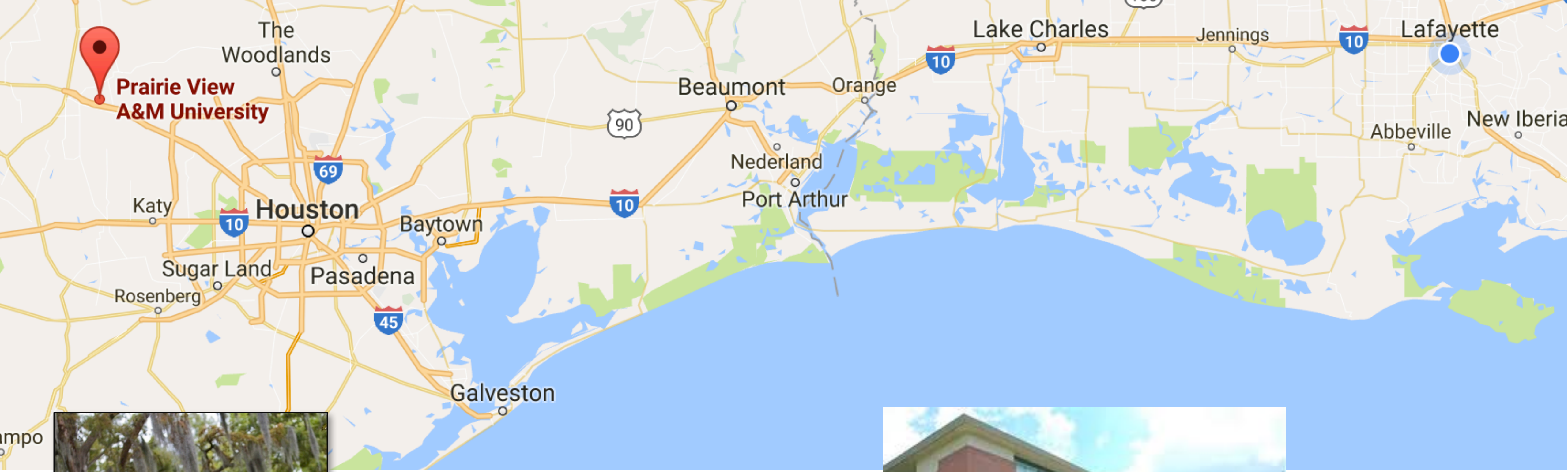
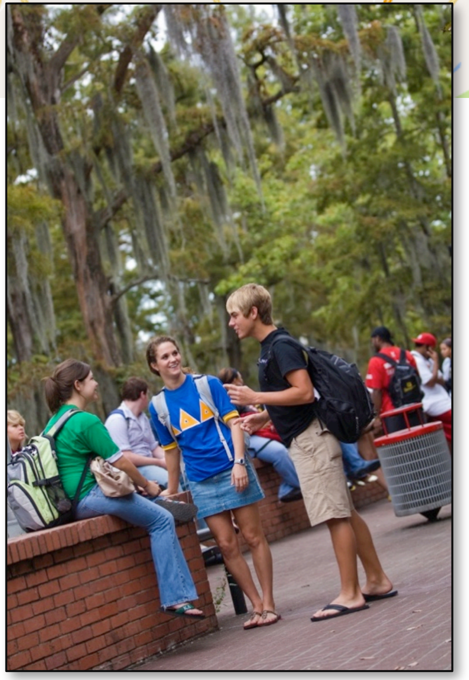




Henry Chu

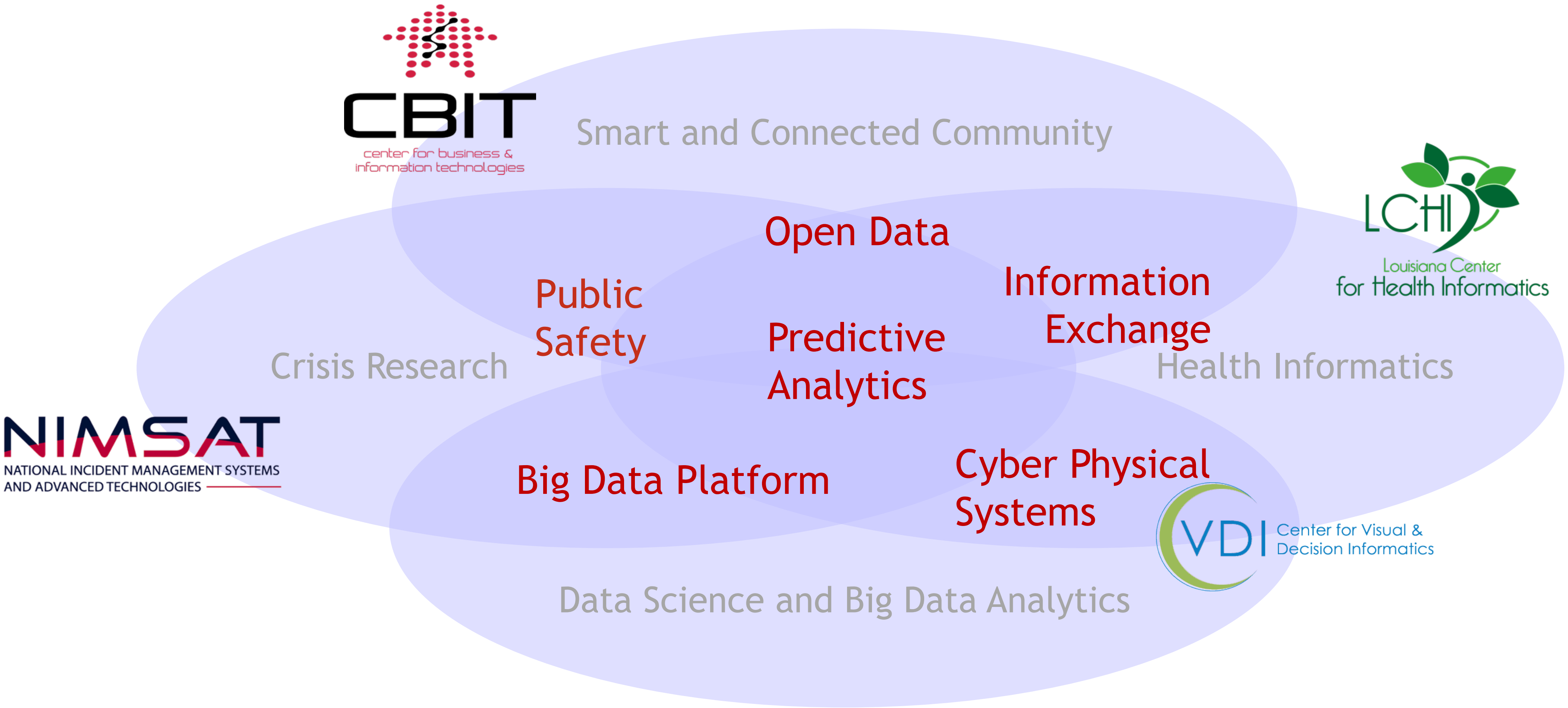
Professor, School of Computing and Informatics

Executive Director, Informatics Research Institute

University of Louisiana at Lafayette

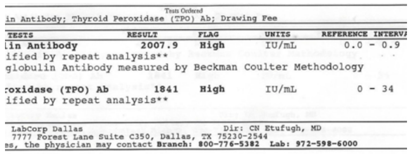
Informatics Research Institute

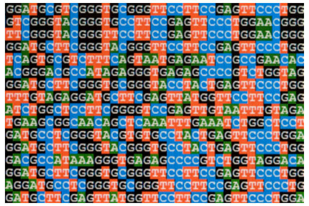
We conduct research in data science to unleash the potential of Big Data for the benefit of society in such areas as health, crisis response, community security & resiliency, and smart & connected community

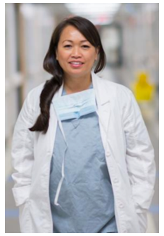


Leveraging Data for Health

twitter







Collect,
Connect,
Aggregrate,
and
Analyze

Clinical
Data
Registry

Clinical Data for
Research Trials

Public Health Data for
Analytics

Intelligent Infrastructure

- Foundation for increased safety and resilience
- Improved efficiencies and civic services
- Broader economic opportunities and job growth

- Deep embedding of sensing, computing, and communications capabilities into traditional urban and rural physical infrastructures such as roads, buildings, and bridges

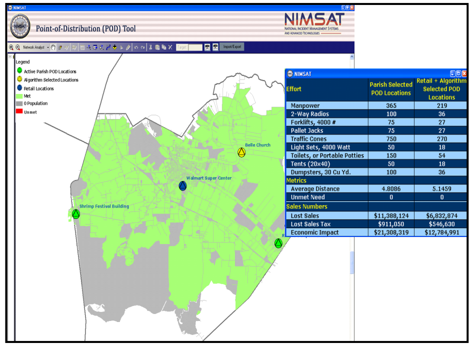
Intelligent Public Safety and Security

- Real time crowd analysis
- Threat detection; dispatch public safety officers
- Anticipate vulnerable settings and events
- New communication and coordination response approaches

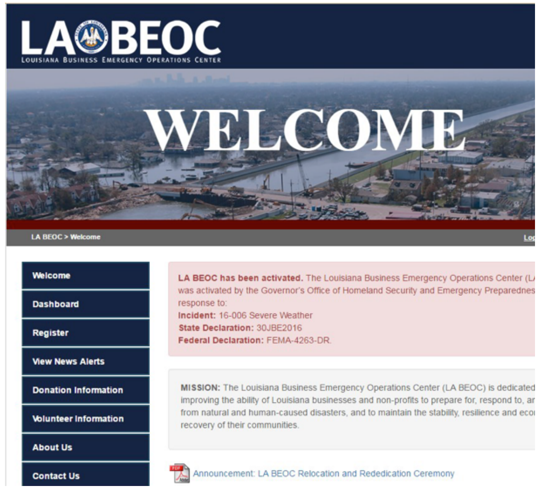
Intelligent Disaster Response

- Real time water levels in flood prone areas
- Timely levee management and evacuations as needed
- Anticipate flood inundation with low-cost digital terrain maps
- Inform vulnerable populations

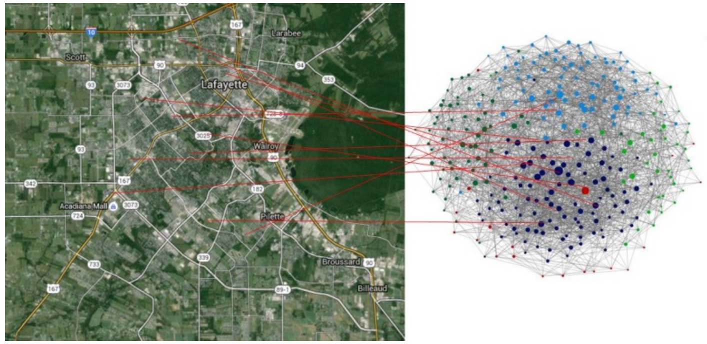
Crisis Research



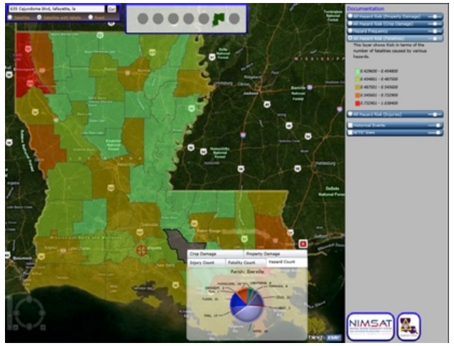
Points of Distribution and Supply Chain Optimization



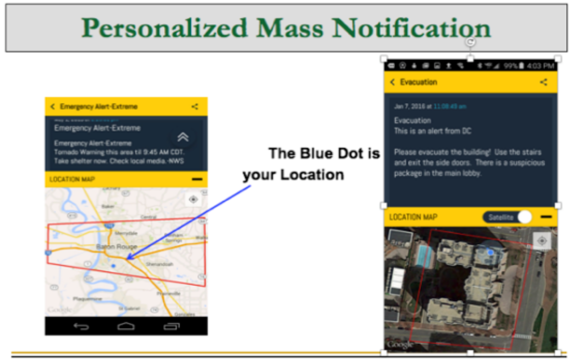
Business Emergency Operations Center



Human Geography Mapping



Louisiana Hazard Information Portal



Geo-Referenced Wireless Emergency Alerting

Consequence Analysis of Natural Gas Pipeline Disruptions

Fuel Demand & Supply Prediction for Regional Evacuation

Big Data Modeling frameworks, Analytics and Tools for Disaster Prediction and Management

- Probabilistic modeling of complex events to develop predictive analytics and enhance the capabilities for appropriate and adaptive response, and to refine response planning.
- Multilevel, multiscale modeling methods for understanding factors that contribute to or undermine community resilience
- Capture and visualize data elements reflecting different aspects of a community, from physical geography to built infrastructure to activities, entities, events, and processes on the infrastructure
- Research into protocols and methods for ensuring both reliability and privacy of data collection and analytics during emergency situations, disasters, and crises.

Virtual Reality Content Creation by Deep Learning of Video Clips

Joe Reed
The NeuMachine

Henry Chu
University of Louisiana at Lafayette

The NeuMachine LLC

UNIVERSITY OF LOUISIANA Lafayette

Motivation

Emergencies that impact buildings

- Fire
- Mass killings
- Floods
- Hurricanes
- Tornadoes
- Toxic gas releases
- Hostage situations
- Chemical spills
- Explosions
- Civil disturbances
- Utility failures
- EMS calls
- Automatic fire/security alarms

Active threat policy/protocol for Dispatch

4.5 <u>Specific Questions for Callers</u>

4.5.1 Specific Location

4.5.1.1 What is the name of the location (school / office / store) as well as address

4.5.1.2 If school, refer to available blueprints (printed or electronic)

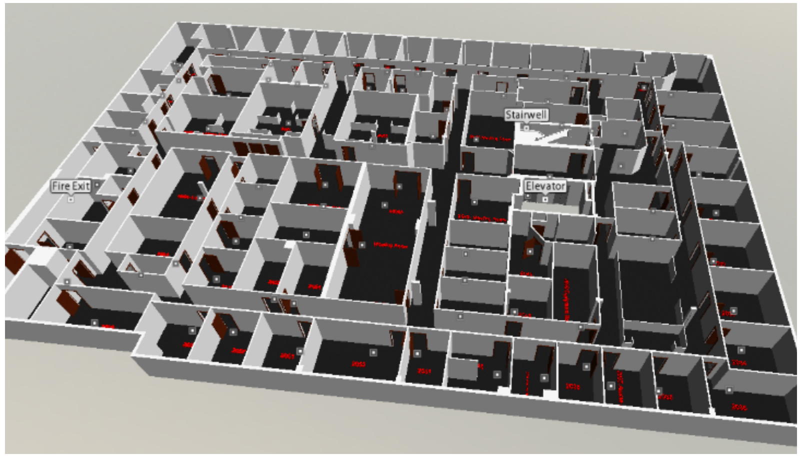
4.5.1.3 If available, utilize school cameras to obtain latest information (once access to school cameras is obtained on scene, Communications staff will discontinue monitoring)

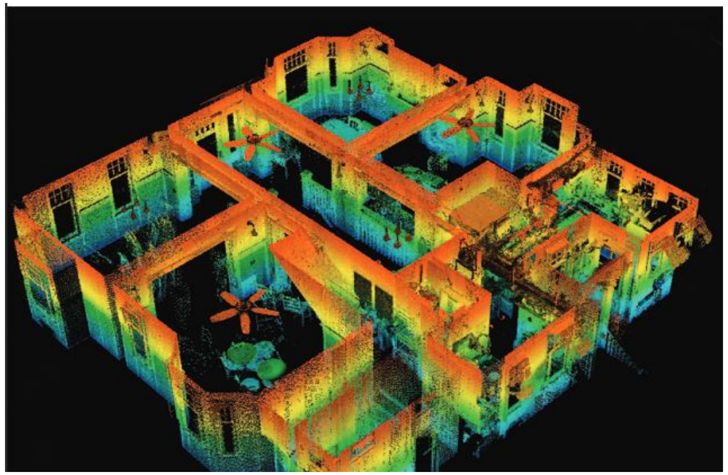
4.5.1.4 Room, office or classroom number

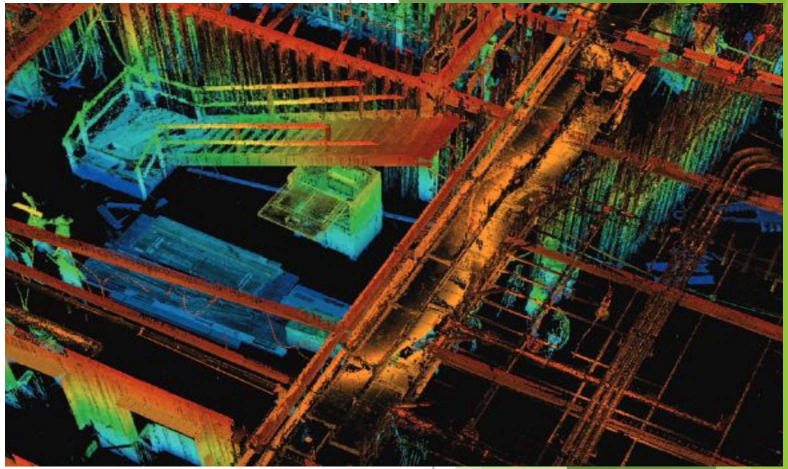
4.5.1.5 <u>**SPECIFIC LOCATION OF THE OFFICE/CLASSROOM**</u> – first floor – second door on right / corner classroom northeast corner of building.

Source: Eagle View Technologies

Motivation: State-of-the-art Solution







- Professional capture of interior imagery and LiDAR, or laser scanning, data
- Post-process data with 360° panoramic imagery and LiDAR data point cloud
- Generation of 3D floor plan models with room attribute data
- Links to MSDS sheets, images and URLs, if available

WHAT IF WE CANNOT DEPLOY A LiDAR UNIT?

Source: Eagle View Technologies

Research challenges in creating virtual objects, humans and environments especially for enhancing physical and interactive realism

High-fidelity, intractable 3D content, such as intelligent virtual humans and interactive virtual environments, drives the creation of compelling graphics innovations such as augmented reality (AR) and virtual reality (VR) applications.

Creating such interactive, smart virtual content goes beyond the traditional graphics goal of attaining visual realism, giving rise to a new wave of exciting opportunities in computer graphics research.

This new research frontier aims to close the loop between 3D scanning and content creation, 3D scene and object understanding, virtual human modeling, physical simulations, 3D graphics researchers, as well as experts in AR/VR, computer vision, robotics and artificial intelligence

With the rapid changes occurring in the field there needs to be a framework for incorporating different modal data into the development pipeline.

To reduce cost and man power, we believe that a tool augmented with Deep Learning can learn tasks needed to create VR content and can learn to do it faster and more efficiently than today’s hand crafted algorithms.

Topics that need to be addressed in the evolution of VR technology

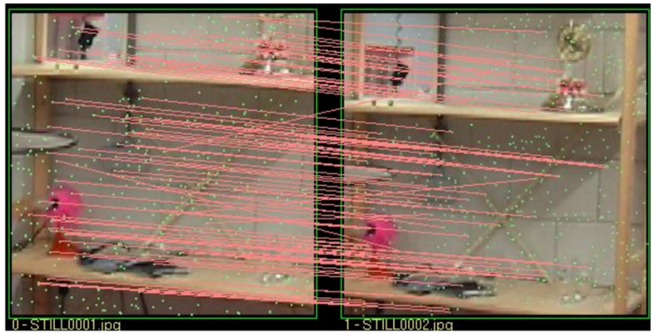
- Affordance analysis of scenes and objects
- Physically-grounded scene interpretation
- Physics-based design of objects cost effectively to provide haptic feel (e.g., 3D printing of special objects, treadmills, moving walls or stairs, terrain like water, rocks, grass, wind)
- Cognitive, perceptual and behavioral modeling of virtual humans
- Virtual human interaction and human perception
- Biomechanics modeling and simulation of human body
- Artificial life and crowd simulations
- Novel applications of AR/VR/haptic devices

3D Scene Reconstruction from Video Clip

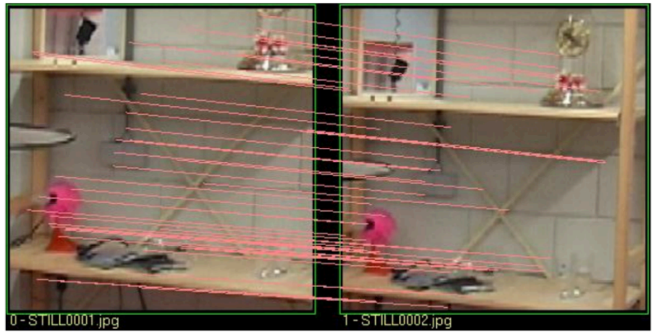
- Handcrafted solutions extensively studied
- Typically rely on
  - feature detection,
  - feature matching (typically poor accuracies),
  - matched pair pruning,
  - solutions of transformation parameters, and
  - stratified reconstruction

3D Scene Reconstruction from Video Clip

- Handcrafted solutions typically based on

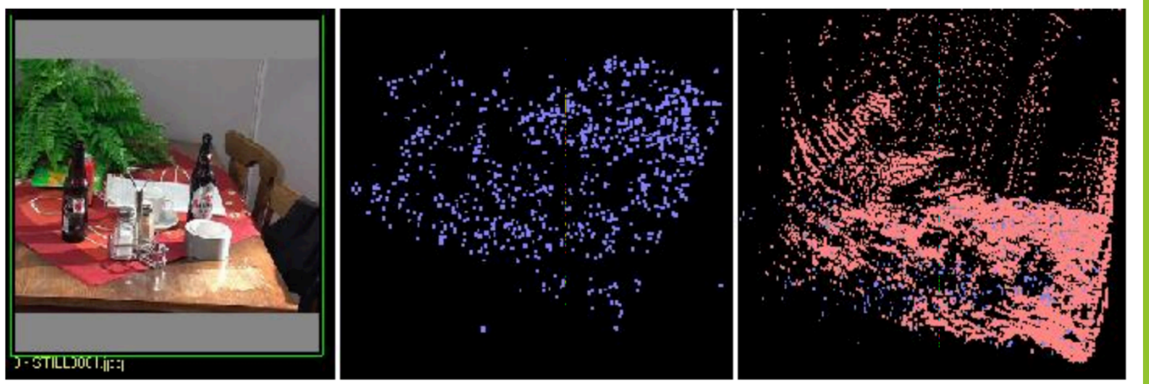


Feature points detection and matching, usually very error-prone

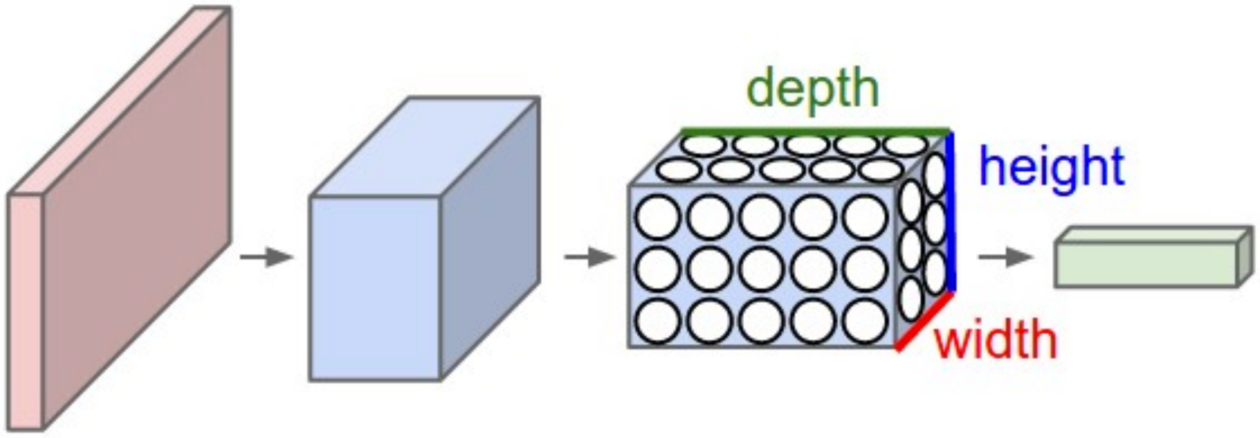


Use 3D parameters to eliminate mis-matched pairs

Stratified reconstruction to create sparse and dense data points



Deep Learning



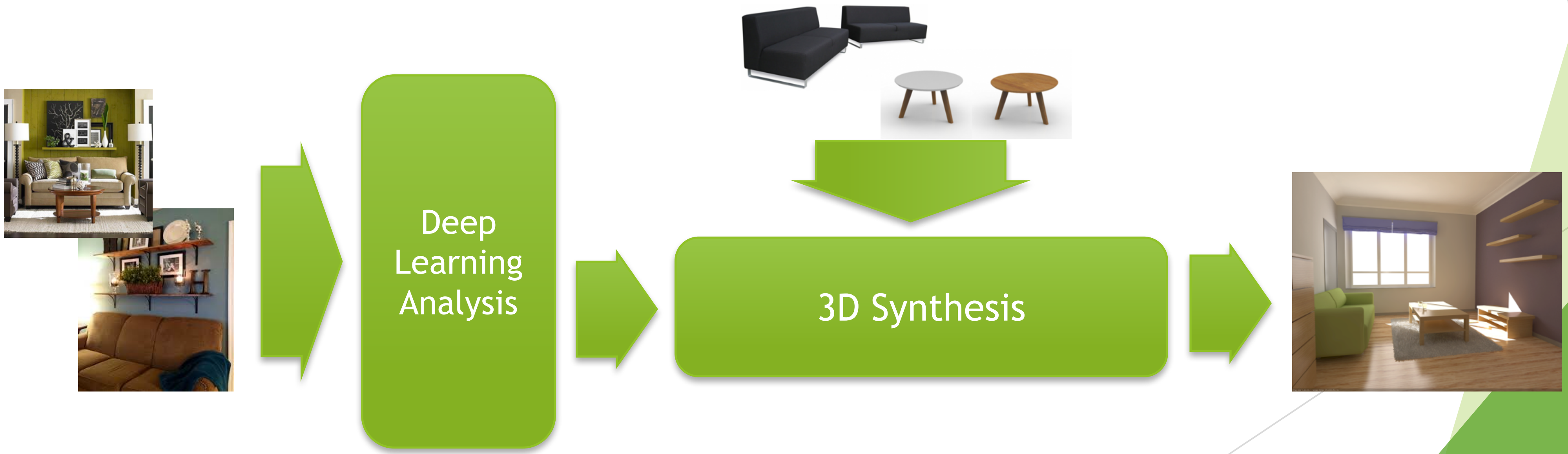
Deep learning supported by GPU processing power has led to classification, detection, and segmentation of image and video data with spectacular results in the past few years

Pilot Work

We hypothesize that using a Deep Learning solution, we can recover sufficient information

- labeled image regions with surface normals and depth information

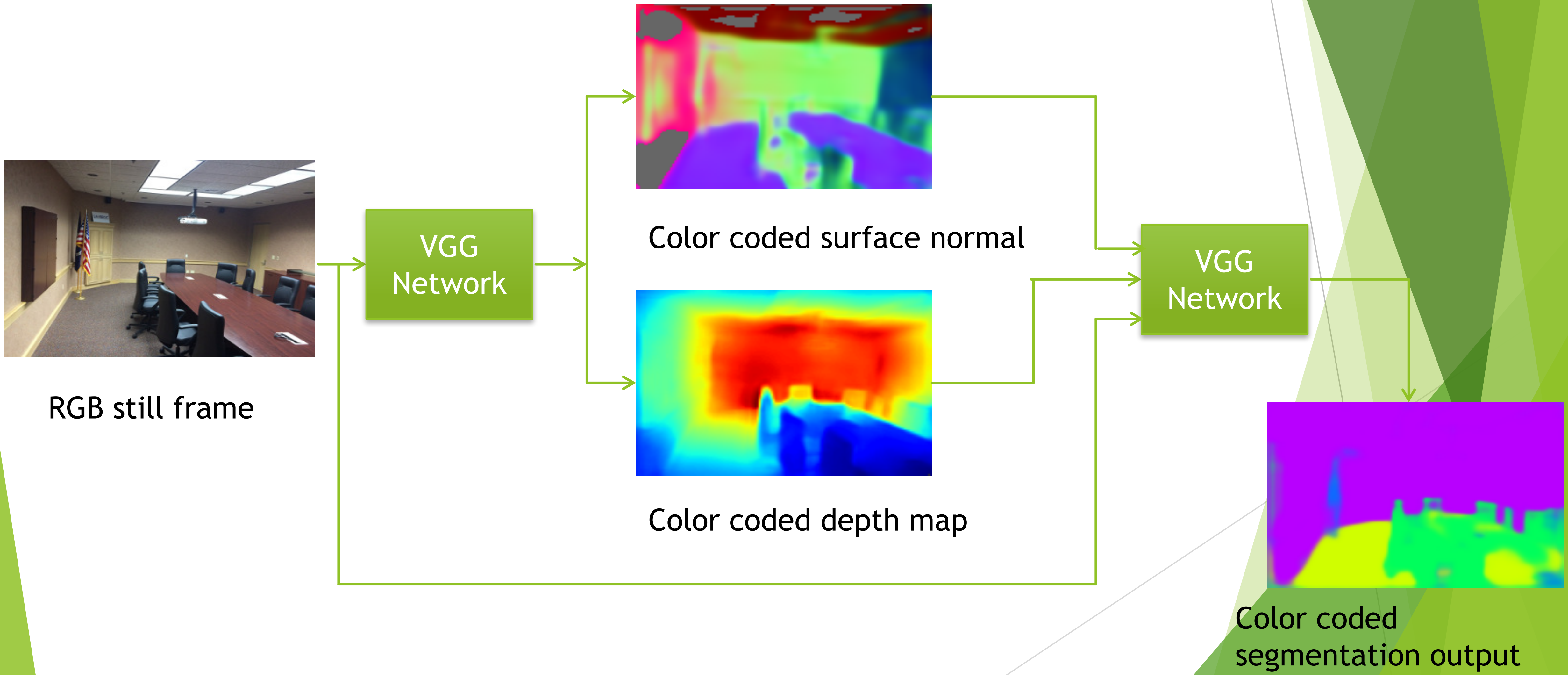
to enable us to recover a 3D scene that can be used in a virtual reality rendering using digital assets



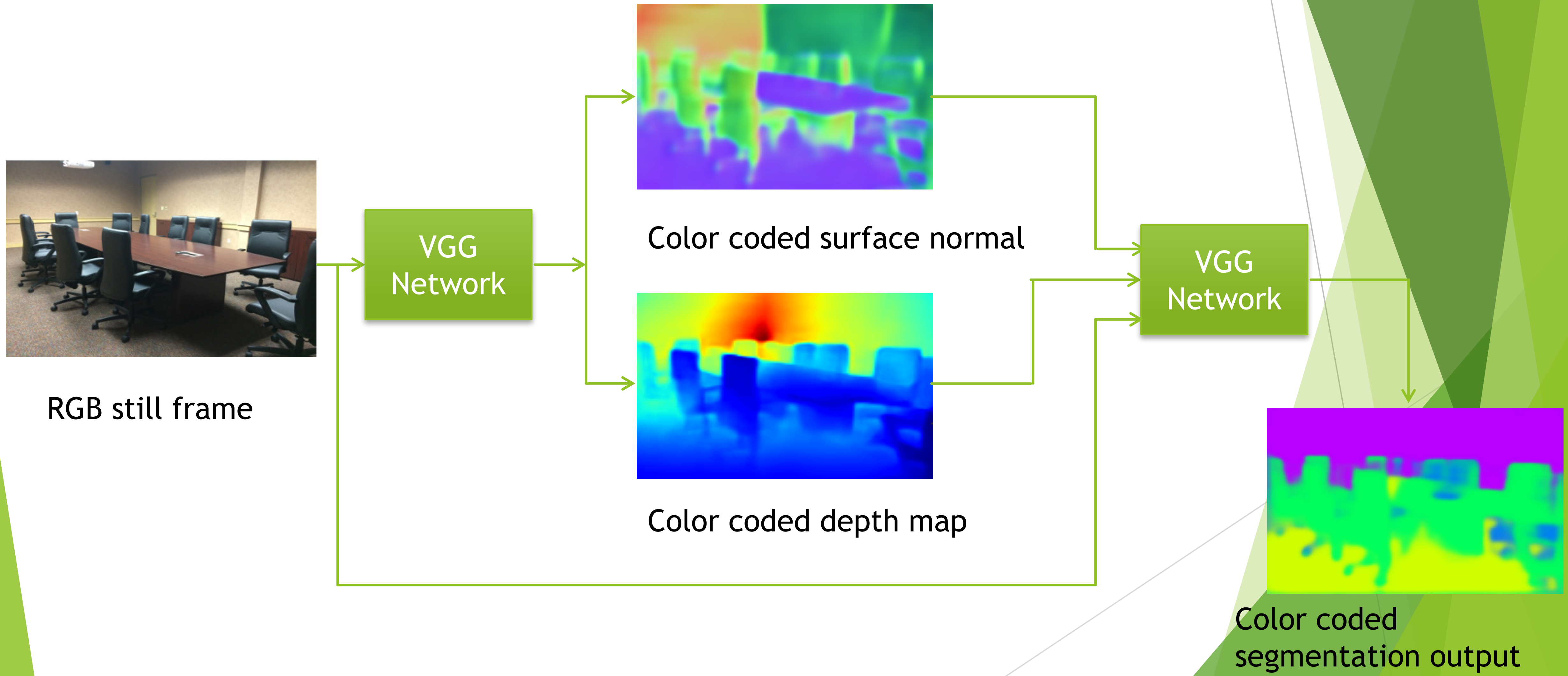
Quick Example

Individual frames are grabbed and resized to 320 by 240 still images

From Deep Learning

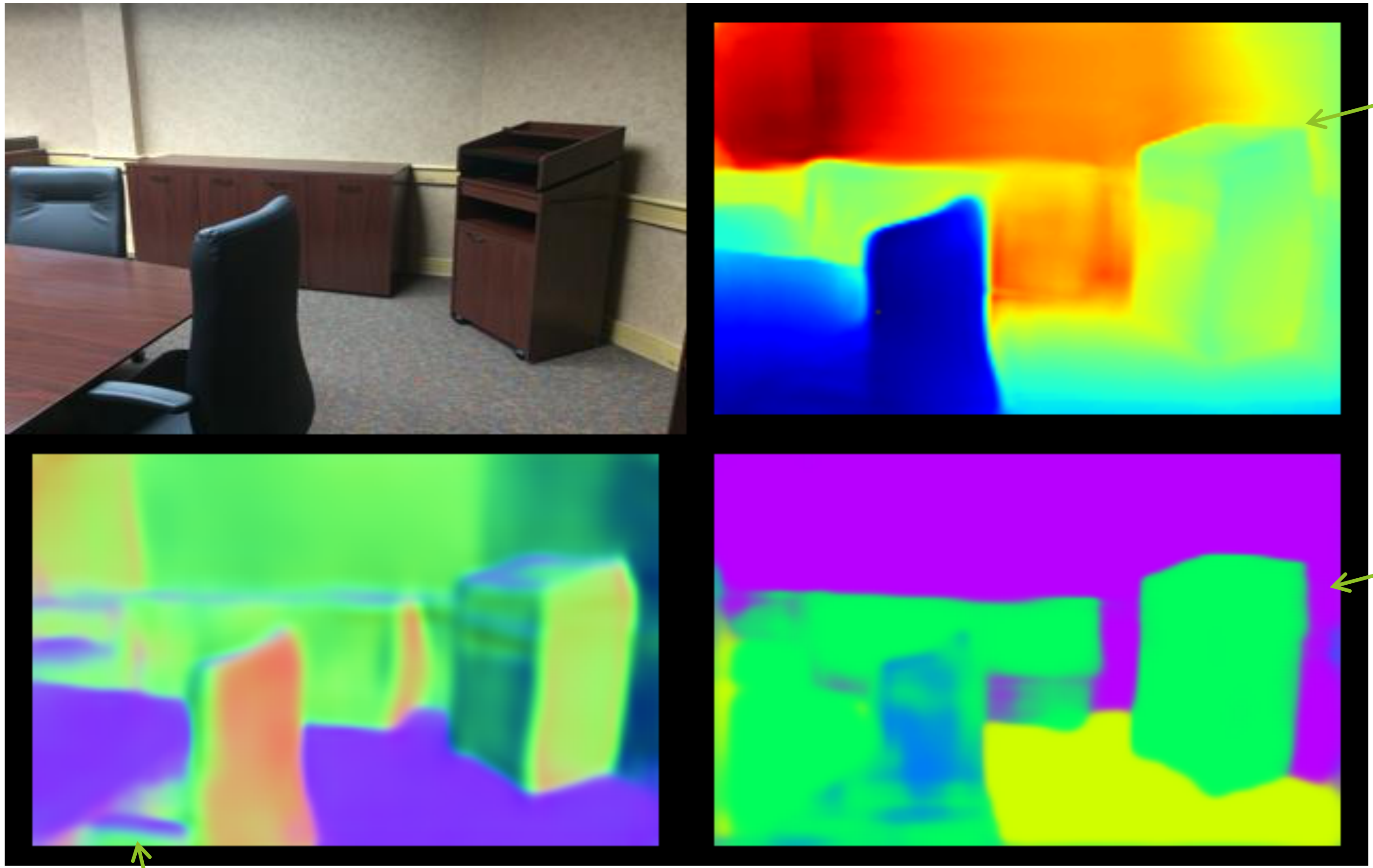


From Deep Learning



Key Frame Video Clip

Sample Frame Output

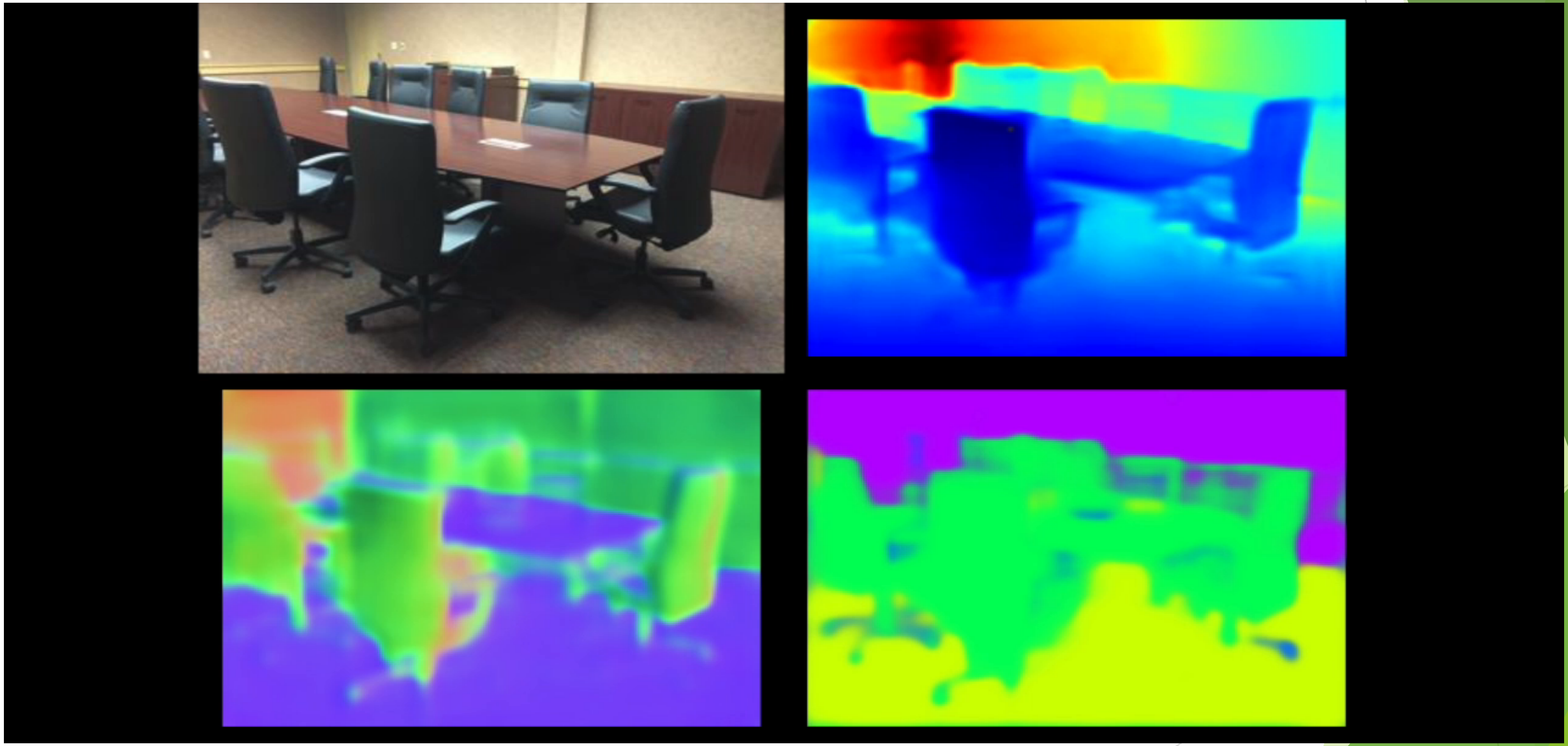


Color-coded distances from camera

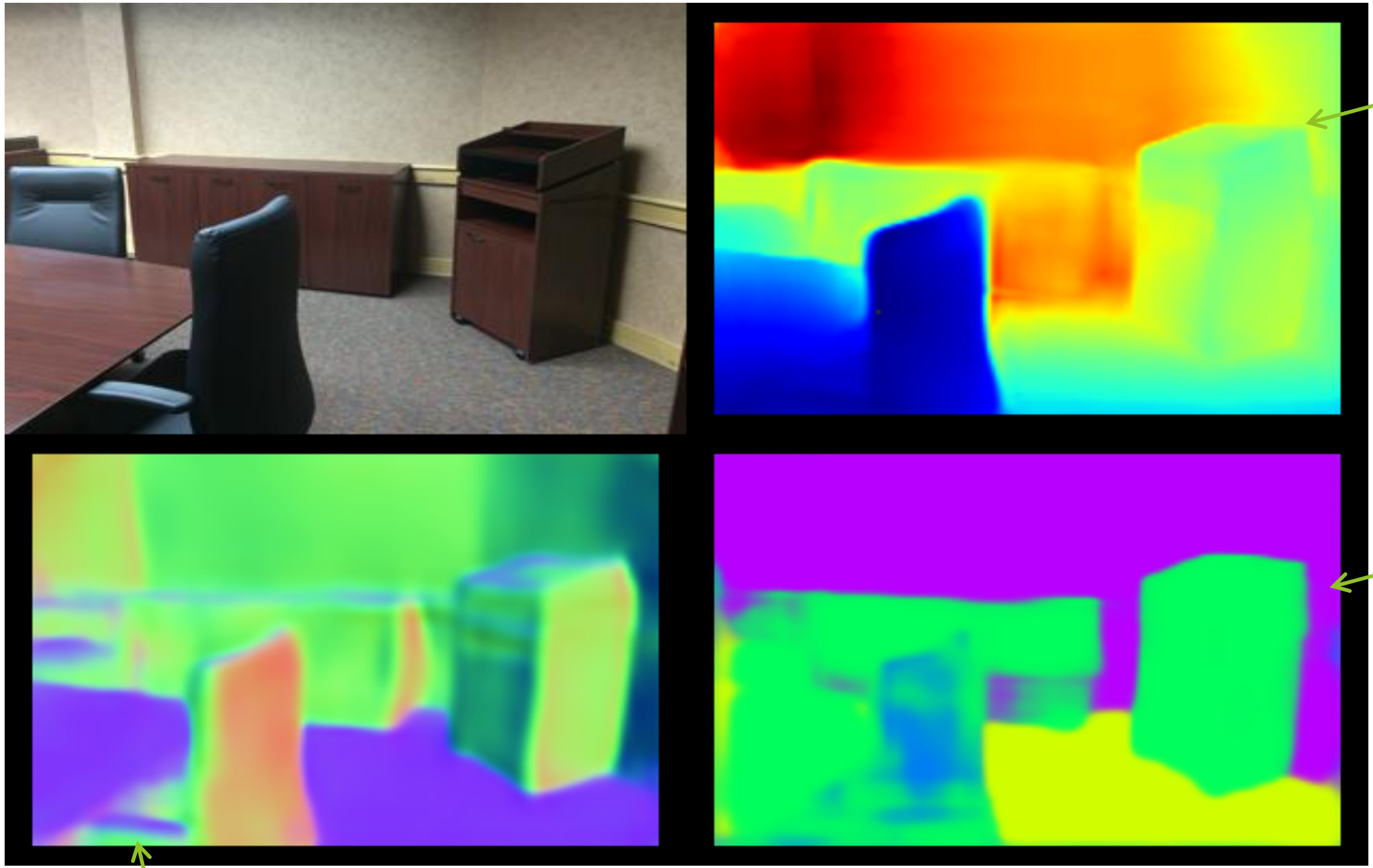
Labels:
- Floor
- Support
- Furniture
- Props

Color-coded surface normal vectors

Key Frame Video Clip Analysis Results



Sample Frame Output

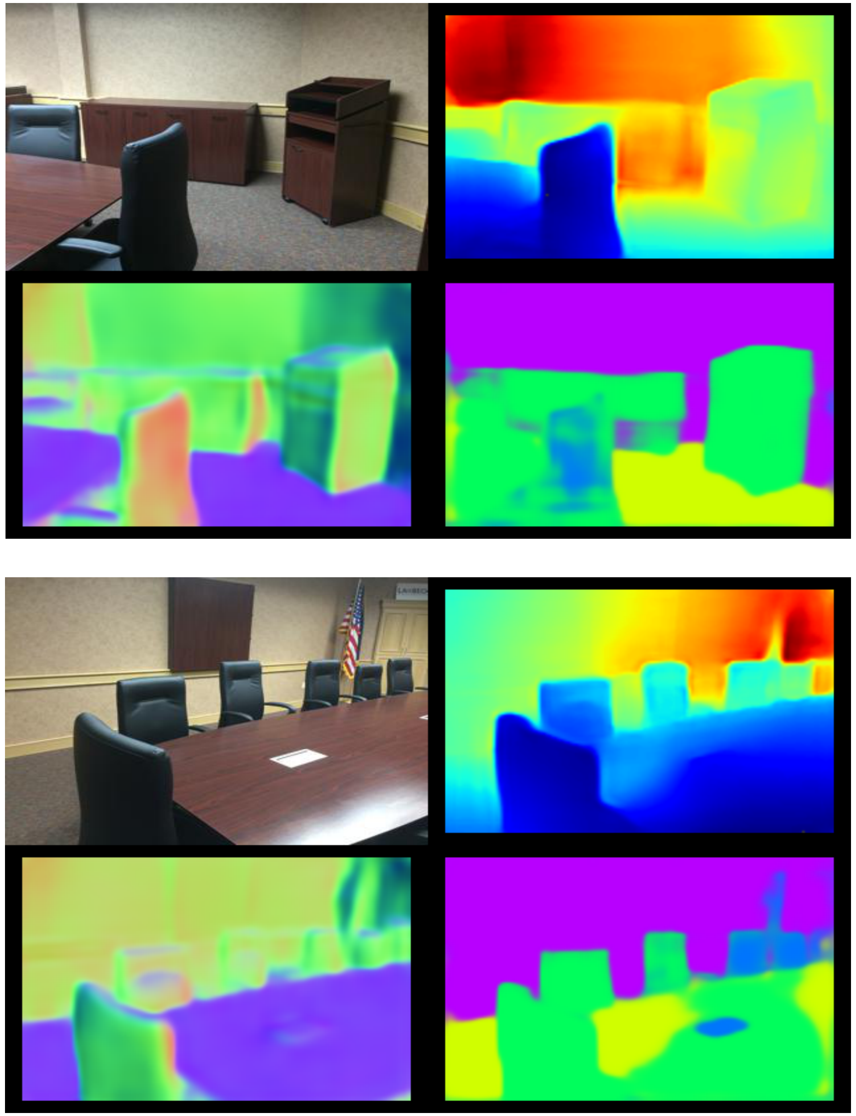


Color-coded distances from camera

Labels:
- Floor
- Support
- Furniture
- Props

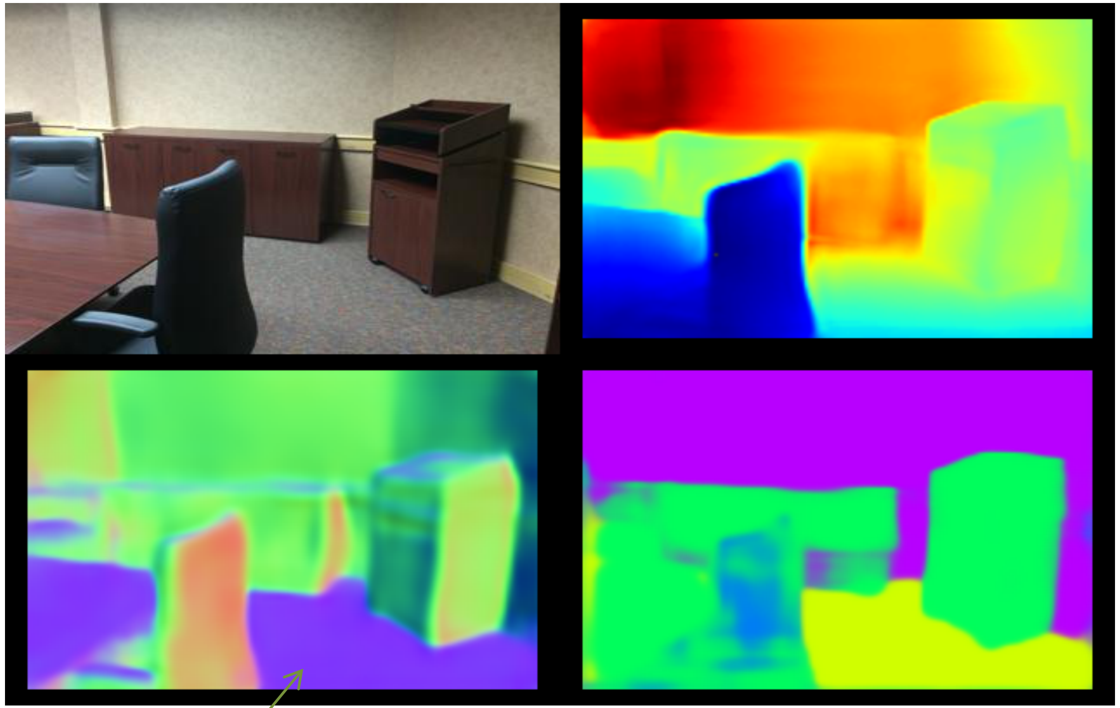
Color-coded surface normal vectors

From Image to 3D Planes



- Surface normals and depth maps are quite accurate
- Labels of floor and support are usually correct
  - Large horizontal surface sometimes mistaken as floor
- Horizontal surface normals seem to be more accurate than those of other surfaces

From Image to 3D Planes

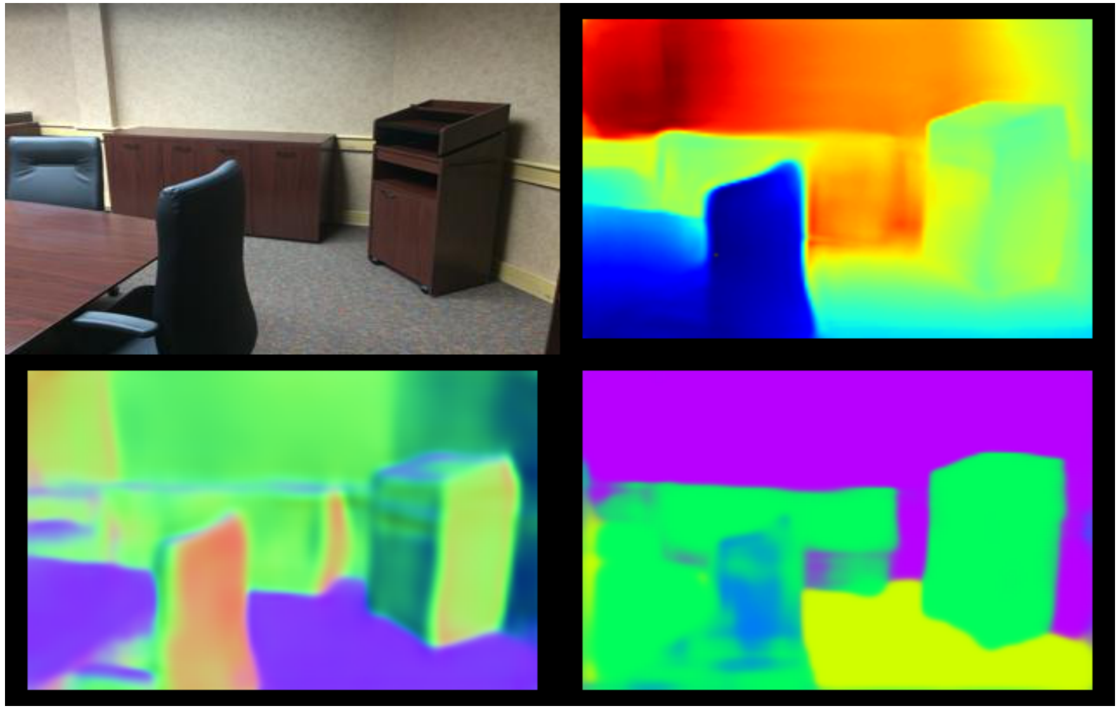


```
[-0.17450304, -0.73930991,  0.57329011],
[ 0.70876521, -0.65309978, -0.18121877],
[-0.66661775, -0.7170617 , -0.09909129],
[ 0.31662983, -0.91938305, -0.03420801],
[-0.1909277 , -0.93683589, -0.1962584 ],
[-0.00796445, -0.32743418,  0.93818361]
```

Horizontal plane

Clustering of all surface normal vectors using *k*-means with *k* = 6

From Image to 3D Planes



Vertical planes →

```
[-0.17450304, -0.73930991,  0.57329011],
[ 0.70876521, -0.65309978, -0.18121877],
[-0.66661775, -0.7170617 , -0.09909129],
[ 0.31662983, -0.91938305, -0.03420801],
[-0.1909277 , -0.93683589, -0.1962584 ],
[-0.00796445, -0.32743418,  0.93818361]
```

Clustering of all surface normal vectors using *k*-means with *k* = 6

From Images to 3D Planes

- We go back to the surface normal map and label each point with the cluster id (0, 1, 2, …, 5) that it belongs to
- Use the cluster id label (“horizontal”, “vertical”, etc) to label each point

3D Planes from Images

Goal is to extract these parameters of each plane in the scene being imaged

- Orientation
- Position
- Scale

Surface normal in world coordinates

Up to scale

Up to scale

3D Planes from Images

Goal is to extract these parameters of each plane in the scene being imaged

- Orientation
- Position
- Scale

We rotate all points (planes) so that the floor (horizontal) plane points up, as the *z*-axis

We can arbitrarily rotate all points so that one of the walls (vertical support) points as the *x*- or *y*-axis

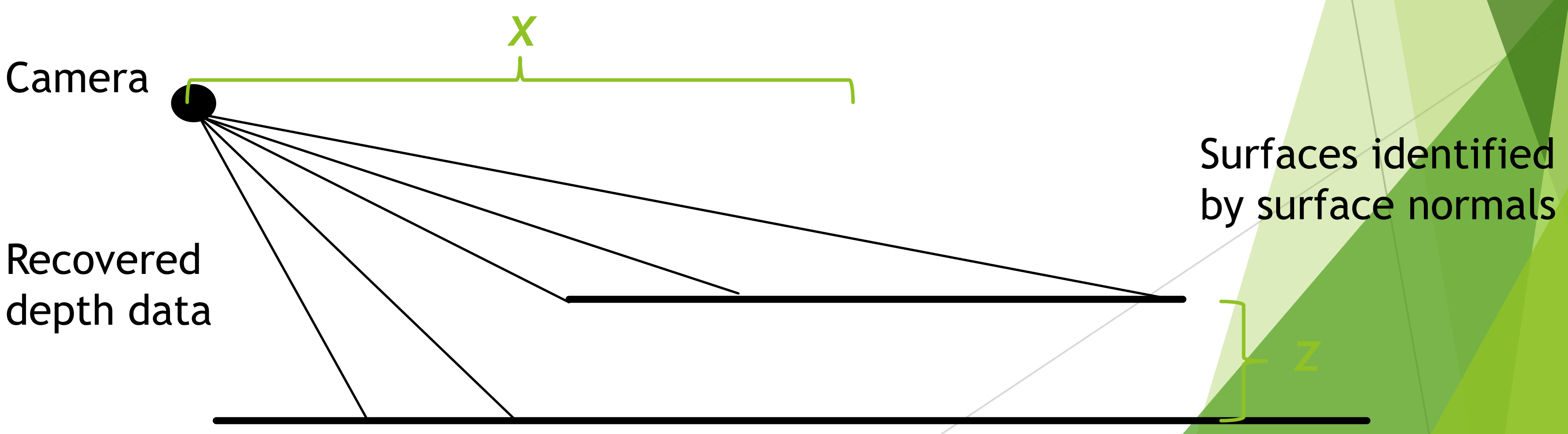
This rotates the scene to align with the world coordinate from the camera coordinate

3D Planes from Images

Goal is to extract these parameters of each plane in the scene being imaged

- Orientation
- Position
- Scale

We use the depth information to position the plane in the scene, up to scale

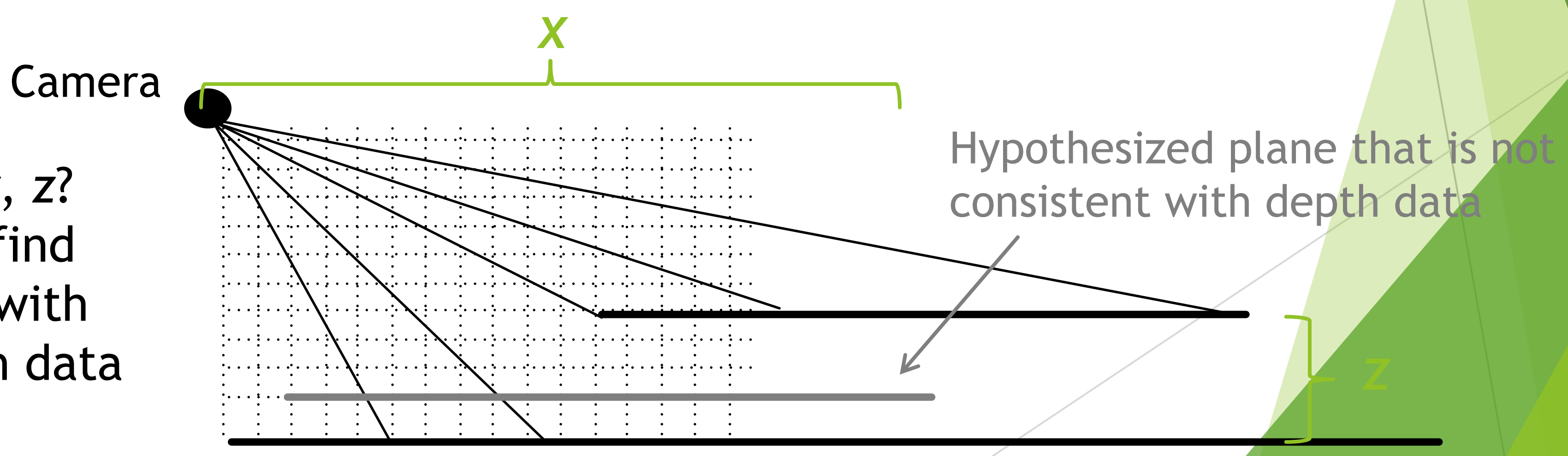


3D Planes from Images

Goal is to extract these parameters of each plane in the scene being imaged

- Orientation
- Position
- Scale

We use the depth information to position the plane in the scene, up to scale



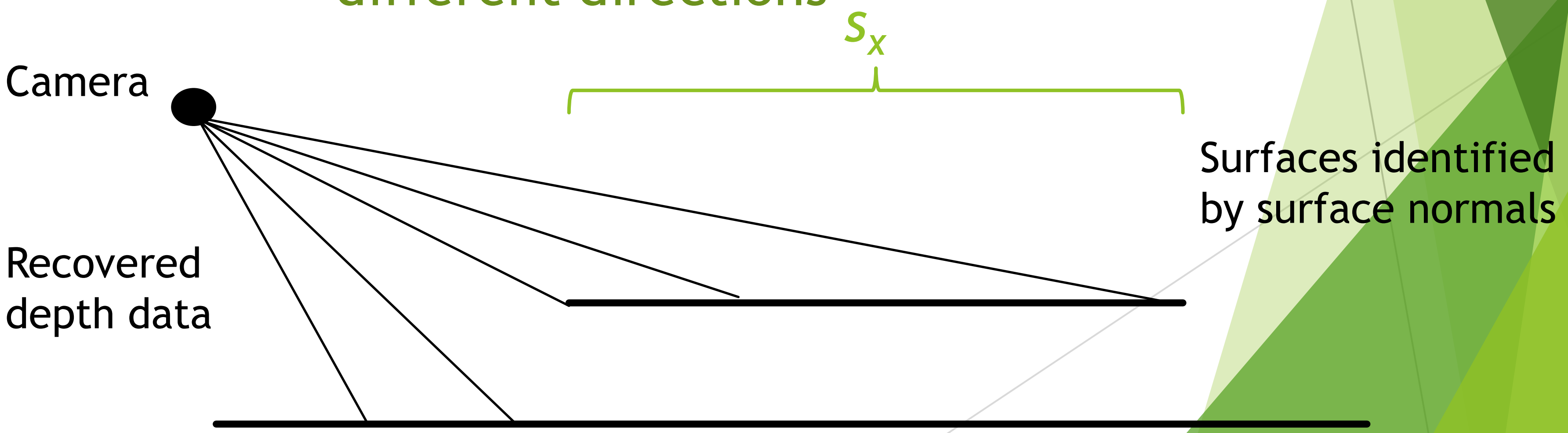
How do we find *x*, *y*, *z*?
Grid the space and find the set that agrees with the recovered depth data

3D Planes from Images

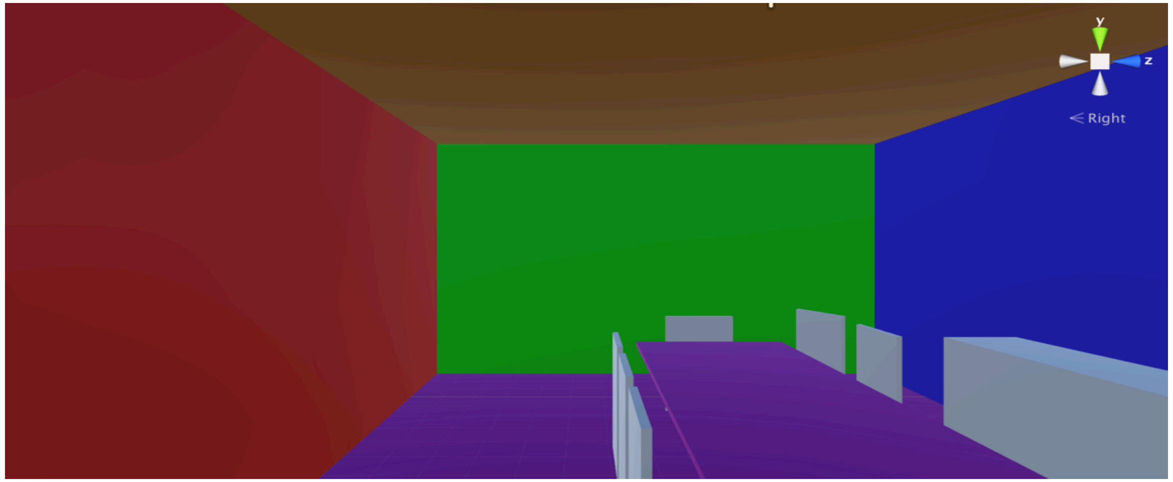
Goal is to extract these parameters of each plane in the scene being imaged

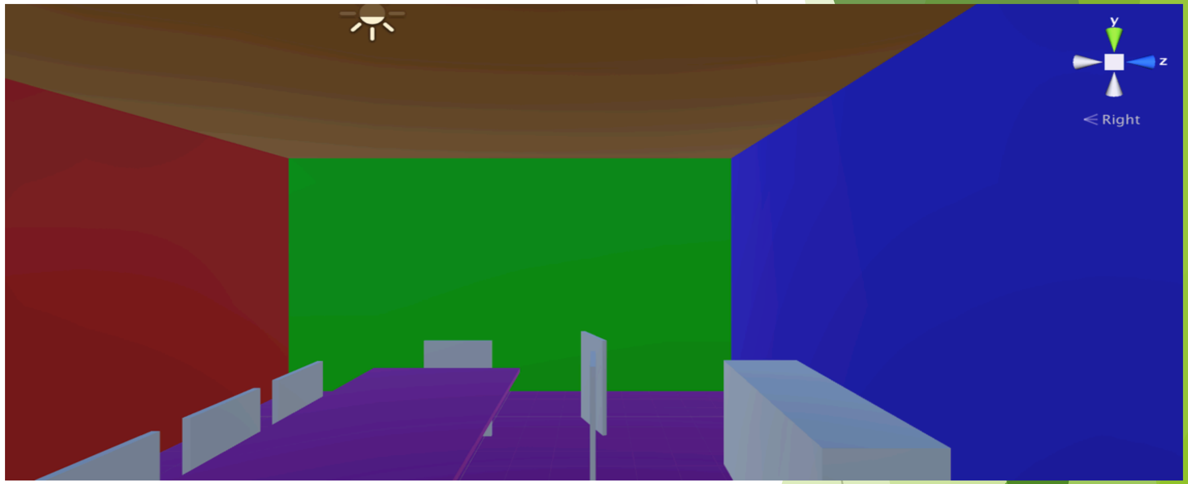
- Orientation
- Position
- Scale

We use the depth information to establish the size of the plane in the scene, up to scale, in different directions



Reconstructed 3D Surfaces



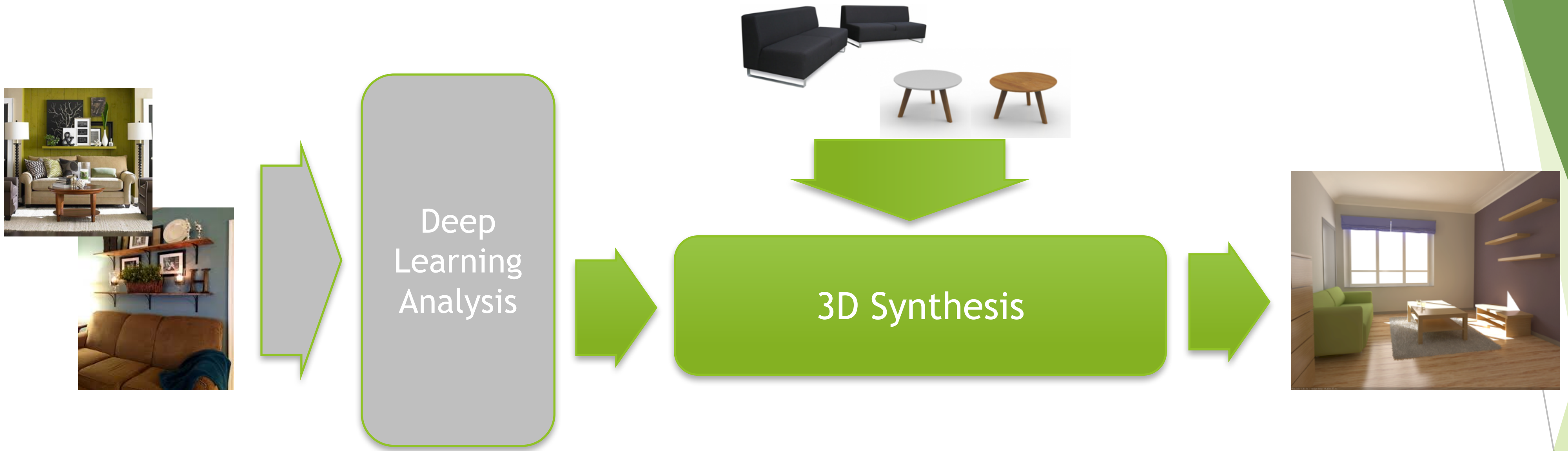


Two views of surfaces in the 3D scene that are consistent with the results obtained from the Deep Learning networks

Ongoing Work

- Fine tune the reconstruction process for one frame
- Rectify reconstruction from different viewpoints obtained by different images
- Connect asset data to insert into the scene (replacing the placeholding surfaces)
- Use an initial classification step to identify scene category (indoor, office, bedroom, etc) to constrain the deep learning network for better accuracy and inference efficiency

How to drive the synthesis of objects?



- Originally we planned to use a classify-and-pick approach
  - Use DL to perform object detection
  - Use object label to search a 3D parts database
- Pivoted to use a Generative Adversarial Network approach

Generative Learning

Generative learning is a theory that involves the active integration of new ideas with the learner's existing schemata. The main idea of generative learning is that, in order to learn with understanding, a learner has to construct meaning actively

-Wikipedia

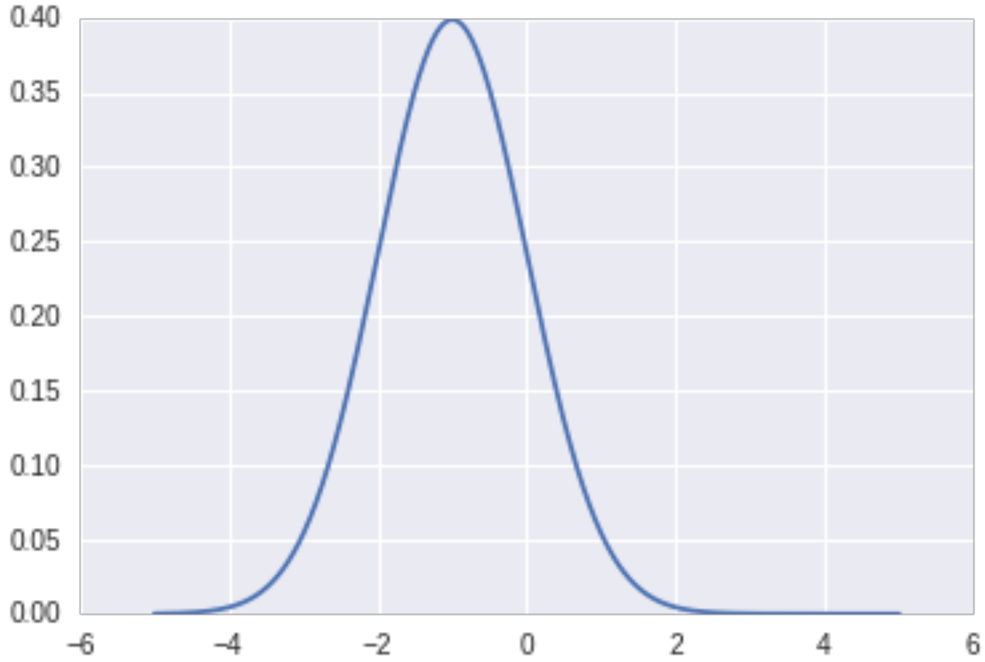
A classifier tries to determine the best $p(y|x)$ where y is the "label" and x is the input

A generative learning system tries to determine the best $p(y,x)$ where y is the "label" and x is the input

Generative Adversarial Network (GAN)

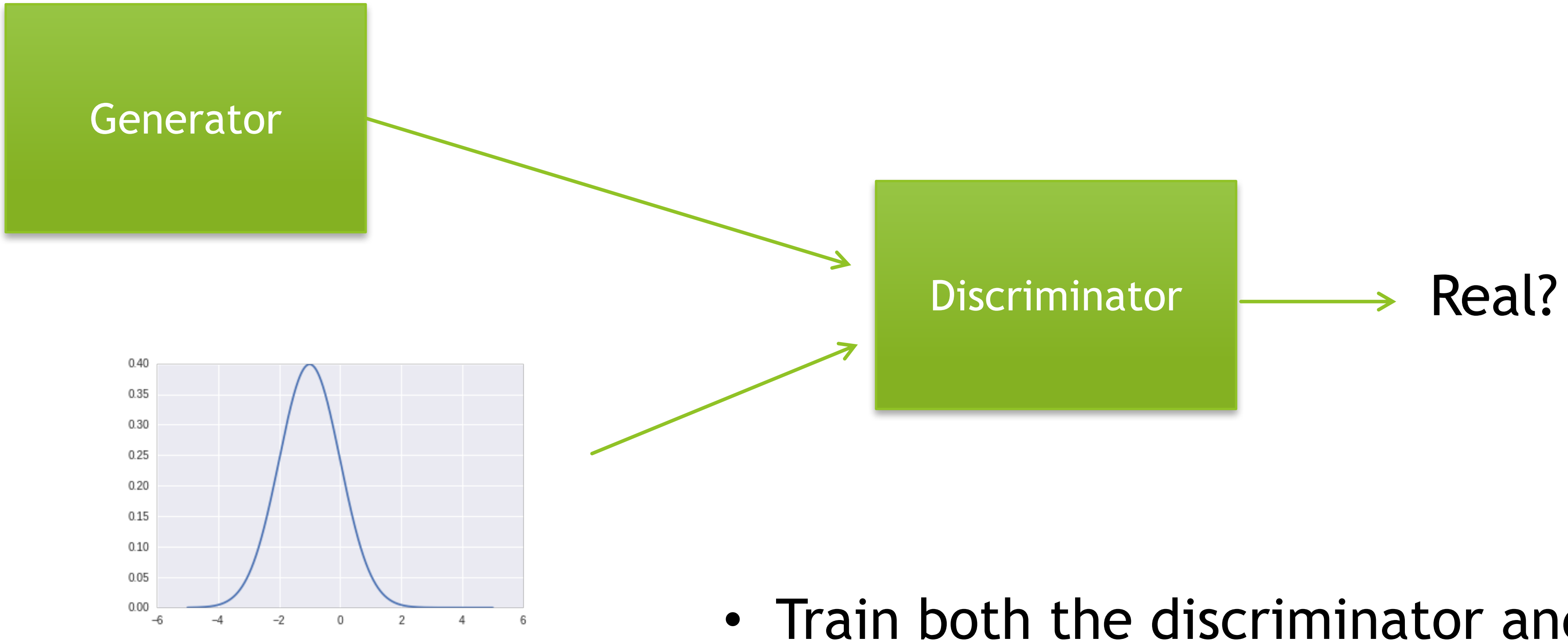
- Goal is to estimate the underlying probability density of p_{data} so that the system can generate any data that are consistent with the original p_{data}
- Image synthesis from an image collection

GAN: Simple Example



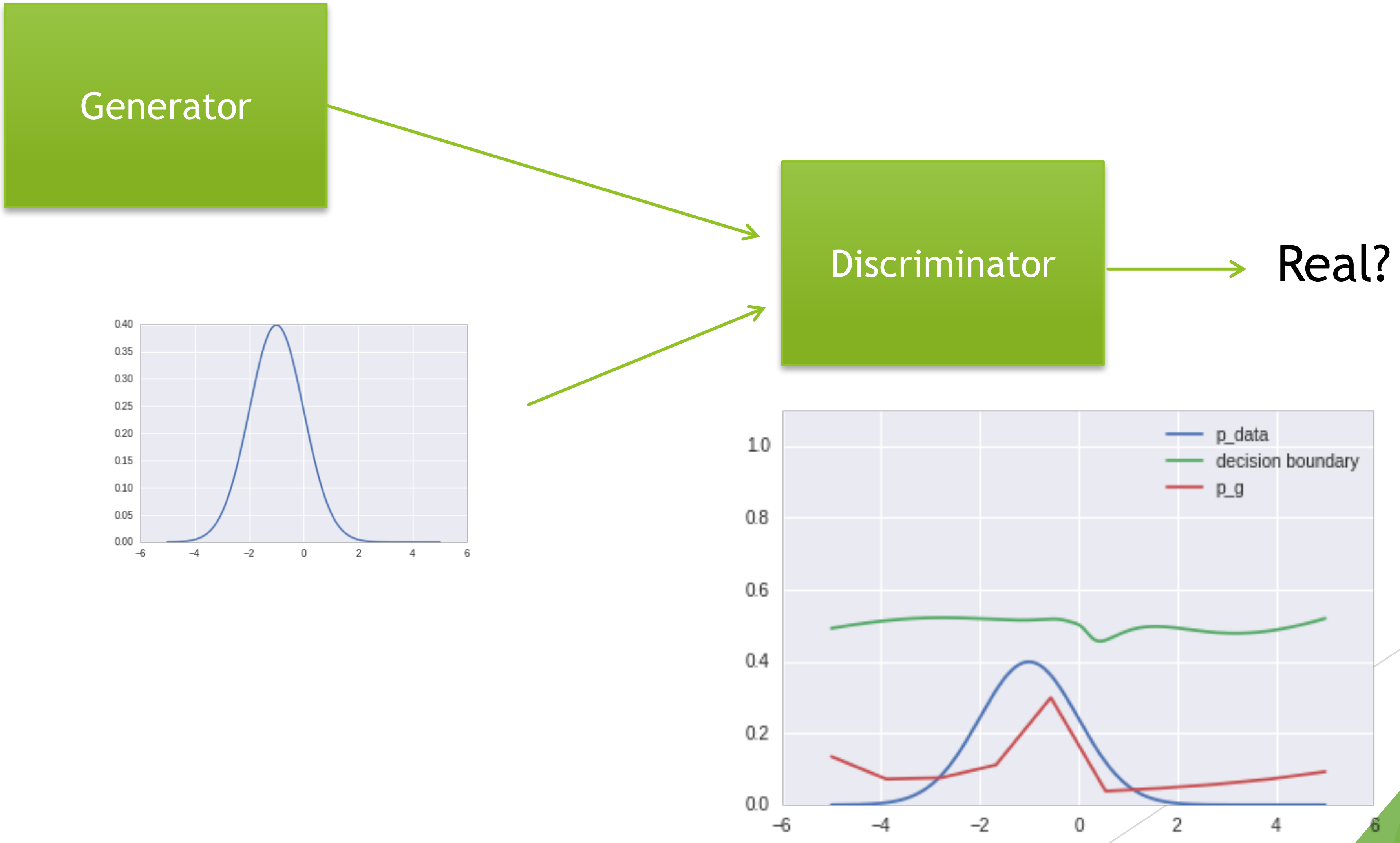
Original data density

GAN: Simple Example

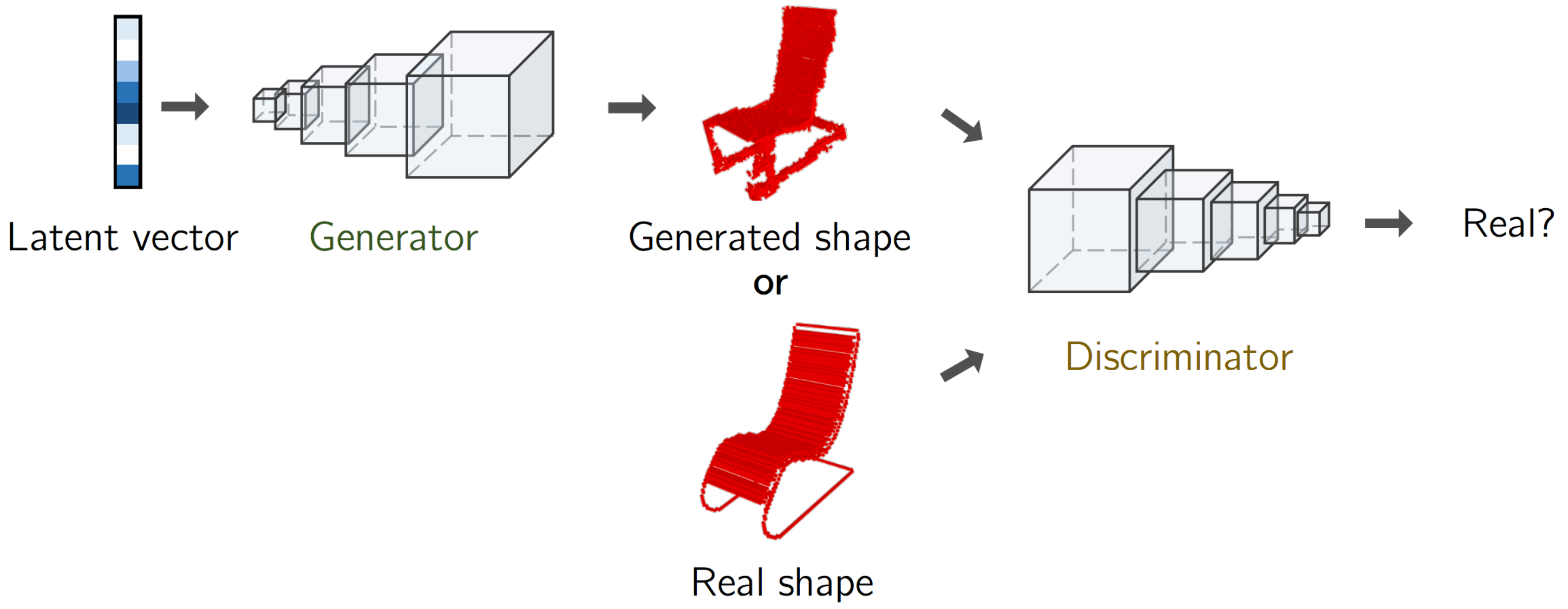


- Train both the discriminator and generator networks together
- Learning converges when the discriminator chooses 0.5

GAN: Simple Example



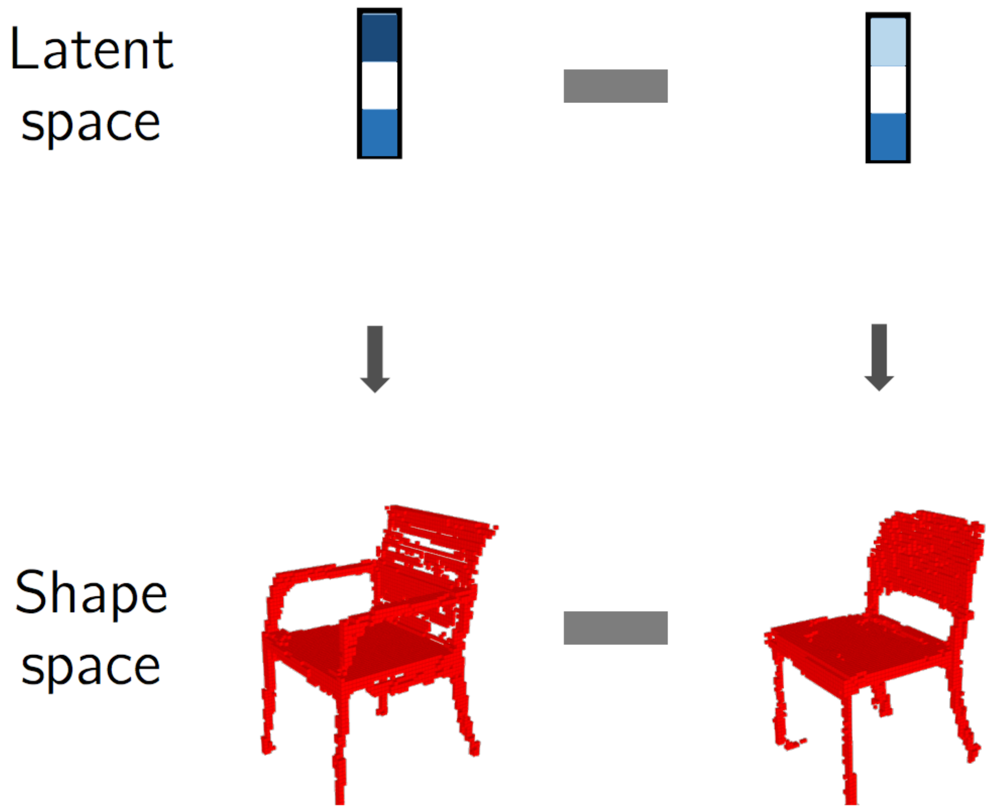
3D Synthesis GAN



3D Synthesis GAN

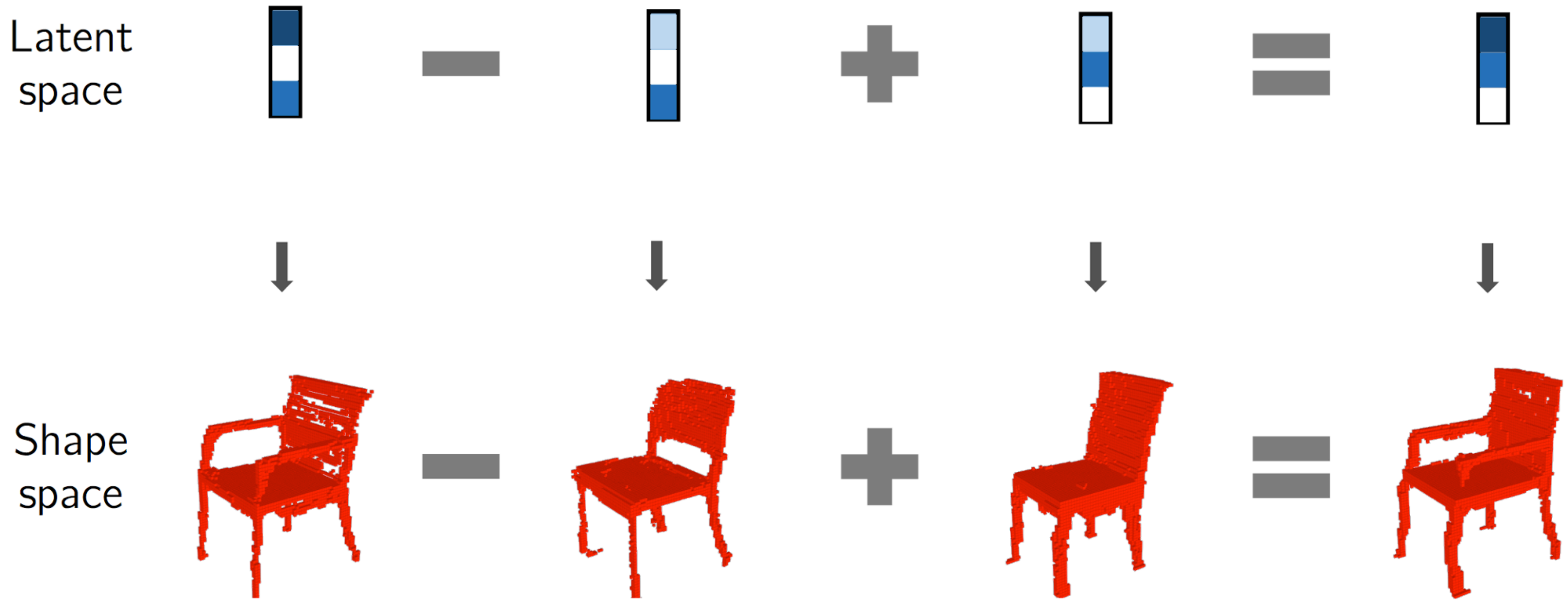
3D Synthesis GAN

Arithmetic in Latent Space

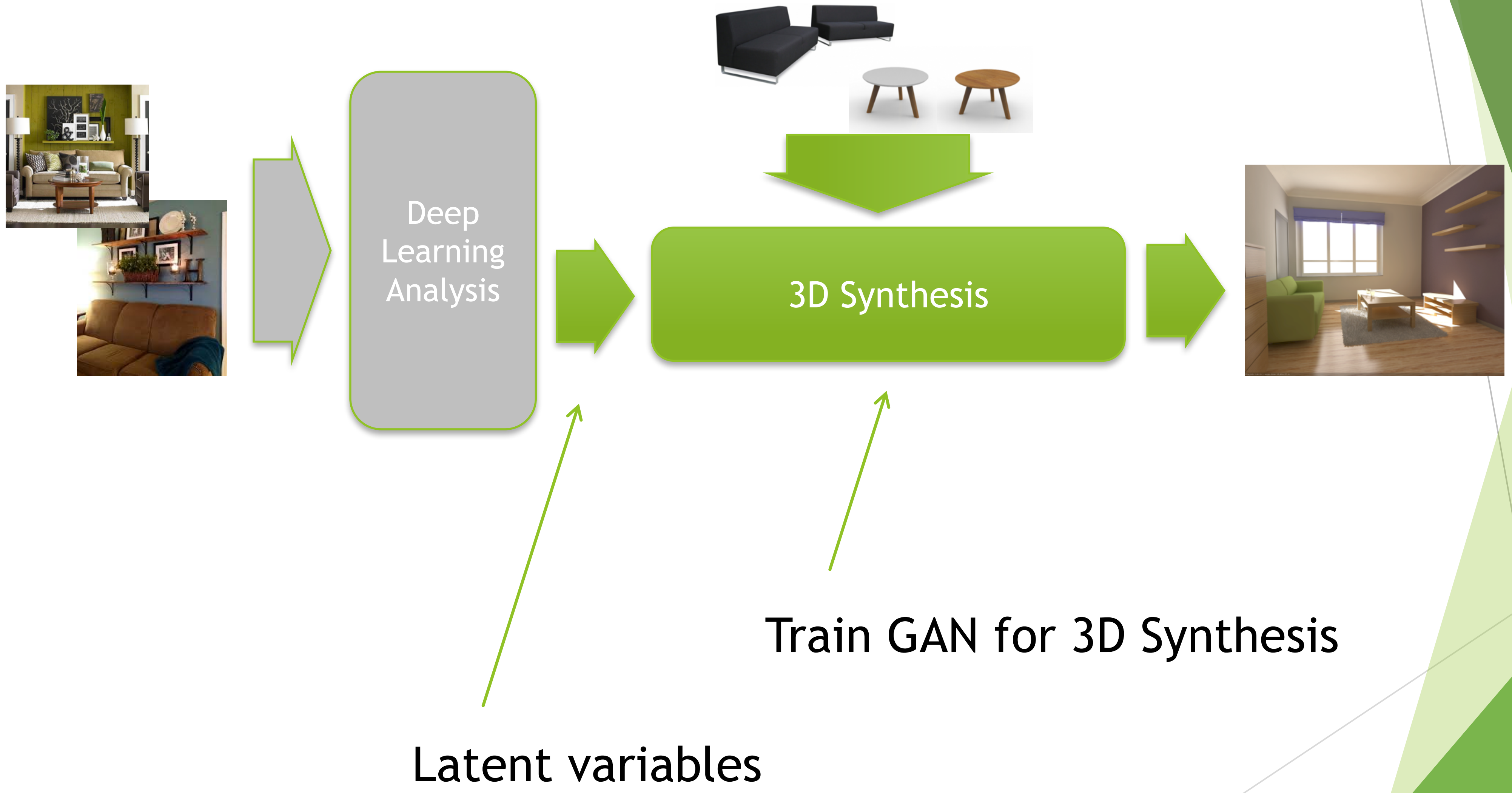


3D Synthesis GAN

Arithmetic in Latent Space



How to drive the synthesis of objects?



Our Ultimate Goal

To establish a framework to be used for a work flow pipeline around creating VR content from 2D legacy movies and their accompanying assets (i.e. script, audio descriptor, closed caption, CG/VFX files)

Use deep learning tools to be more efficient in time and workforce

With more training, adding newer models, and scaling up the GPU compute we can achieve a product solution that can integrate with the existing pipeline used for CG Movies and VR content

The Challenge

- The main challenge is for the computer to recognize low-level and high-level activities in the context of a scene
- Factors that create the challenge are accurate depth estimation, video segmentation into scenes, then into objects both rigid and non-rigid which are further segmented and classified into data structures that can be then used to generate the desired result
- Advances in computer vision and machine learning techniques and algorithms have improved over years, including more accurate eye, face and head tracking and motion capture
- Video recording of human activities can be of use for potential marketing research (for example, how do consumers move in a store and where they stay the most), business surveillance, robotic assembly or as datasets for biologists, sociologists and psychologists to observe human motion and overall action

Our Solution

- An ensemble of neural nets will be used for the production solution of inferencing the immersive experience from the legacy video
- We also use a few commercial software and open-source software applications: Unreal Engine 4, Poser Pro 11, and Blender+Luxrender, & Nvidia's Deep Learning SDK, Auto Desk Maya, Micro Soft Kinect SDK, VR Works, Game Works, all running on Nvidia TITAN X Pascal GPUs.
- We have created a framework for the workflow pipeline: the UI of the system is built from Unreal Engine 4 with the deep learning embedded into the engine pipeline
- The pipeline takes the video and other input in and using well known published algorithms and models for various tasks creates new data structures to be used in the generation of the immersive content
- The workflow consist of a well defined ensemble of neural nets that each produce one set of the data that is needed for the following steps in the process
- After each neural net outputs its data it is sent into a semi-supervised neural net that allows a human to perform a guided quality assurance process to correct any errors with a series of mouse clicks, spoken words, or mouse or pen drawn lines

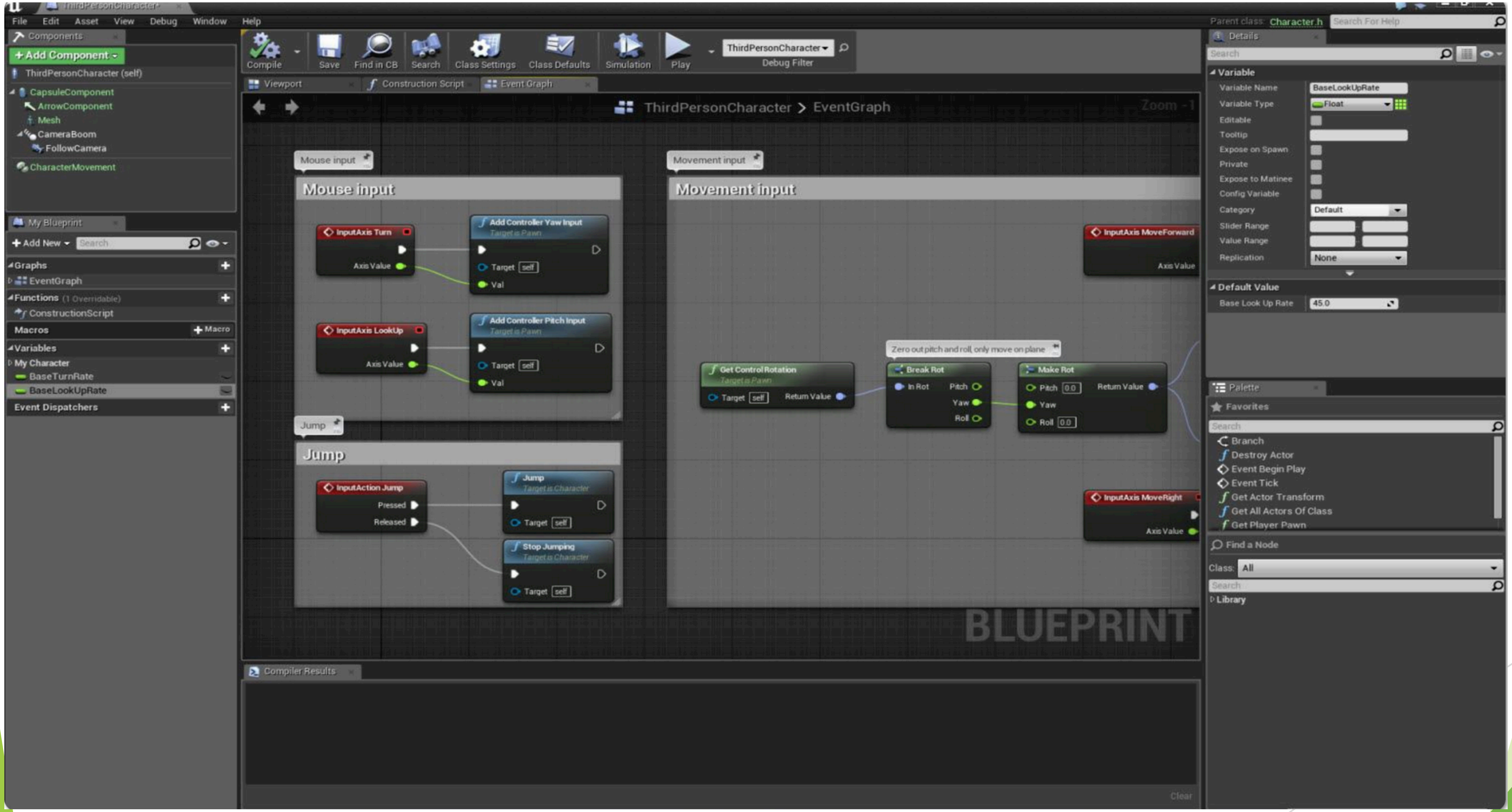
Our Ensemble Applications

- Our ensemble would allow for the creations of VR cinematic experiences and as new VR hardware comes on the market allow for a gamification of VR cinematic content not unlike what is described as Sync Sims in the book “Ready Player One”
- We use the Unreal Engine customized with plugins to enable automation to augment the human work flow pipeline
- We chose UE4 because of its open source and the blue printing, cinematic sequencer, and VR editor
- We chose Poser Pro 11 for its Unreal Portal Development Environment

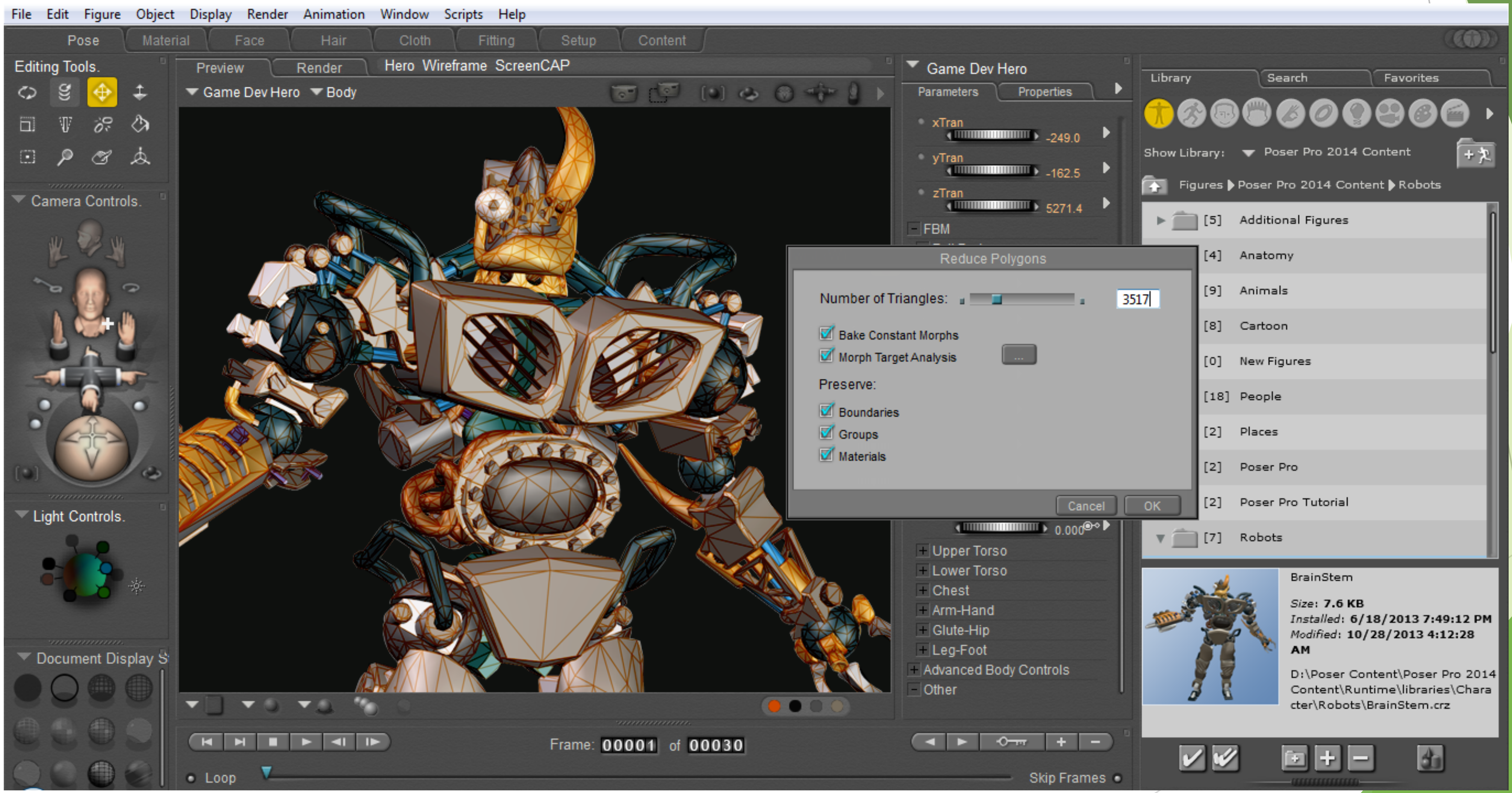
UE4 Pipeline

- FBX is an Autodesk file format that provides interoperability between digital content creation applications such as Autodesk Motion Builder, Autodesk Maya, and Autodesk 3ds Max
  - Autodesk Motion Builder software supports FBX natively, while Autodesk Maya and Autodesk 3ds Max software include FBX plug-ins.
- Unreal Engine features an FBX import pipeline which allows simple transfer of content from any number of digital content creation applications that support the format.
- The advantages of the Unreal FBX Importer over other importing methods are:
  - Static Mesh, Skeletal Mesh, animation, and morph targets in a single file format.
  - Multiple assets/content can be contained in a single file.
  - Import of multiple LODs and Morphs/Blendshapes in one import operation.
  - Materials and textures imported with and applied to meshes.
  - Poser's Unreal Portal Development Environment

Blue Print Editor



Poser's Unreal Portal Development Environment

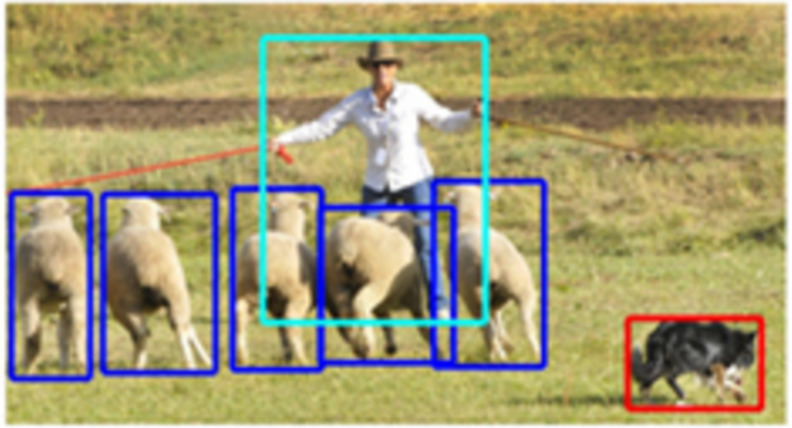


FAIR Facebook AI Research

We were able to replicate this research from FAIR including UR torch and also UnrealCV which both interface with Unreal Engine 4 and serve as an interface to deep learning and open CV libraries and we found that we could achieve similar results which we will provide screen shots of our findings

The algorithm DeepMask[1] segmentation framework coupled with the newwSharpMask[2] segment refinement module. Together, they have enabled FAIR's machine vision systems to detect and precisely delineate every object in an image. The final stage of their recognition pipeline uses a specialized convolutional net, which they call MultiPathNet[3], to label each object mask with the object type it contains (e.g. person, dog, sheep).

(a) classification

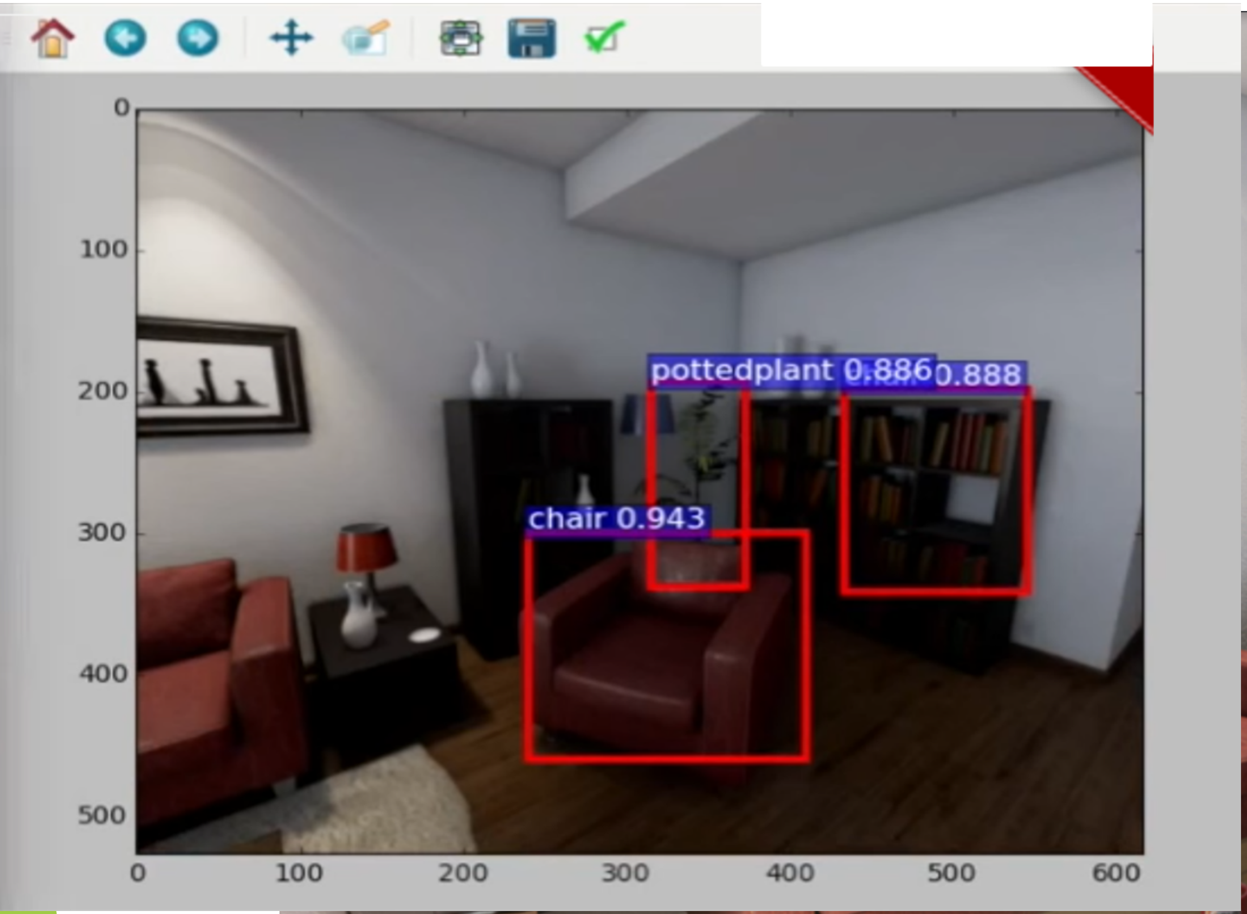


(b) detection

(c) segmentation

Our Ensemble Applications

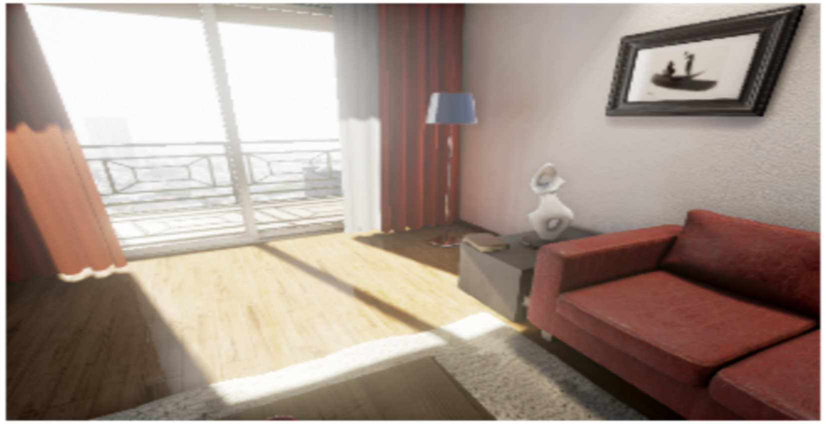
Movie Scene Object Segmentation Captured in Unreal Engine with Embedded Deep Learning

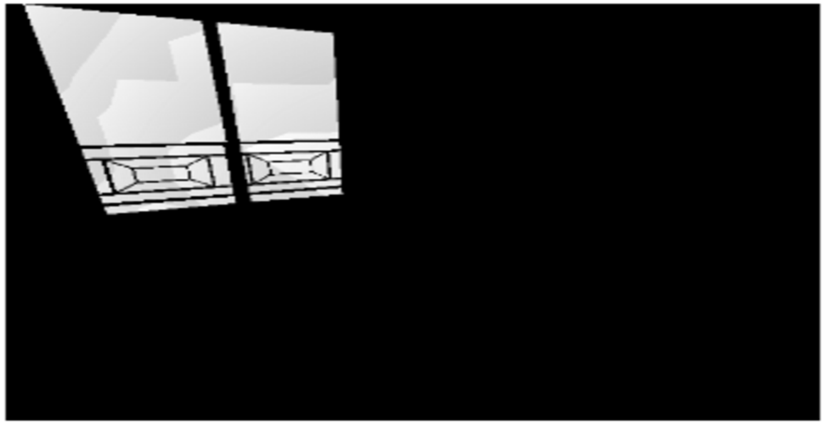


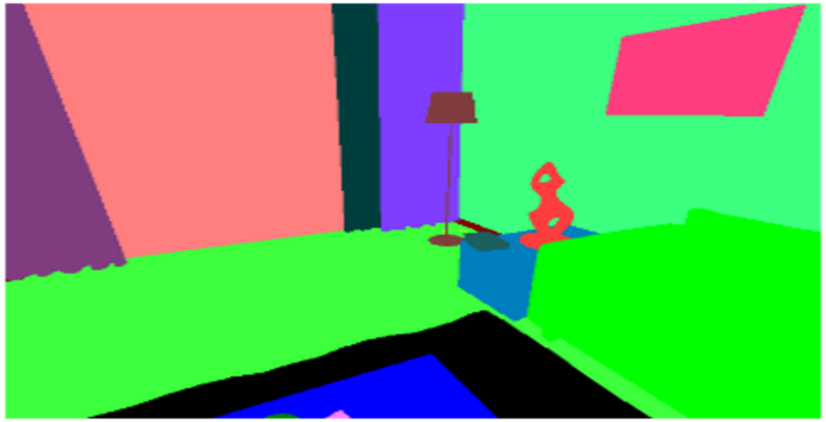
Movie Scene Example

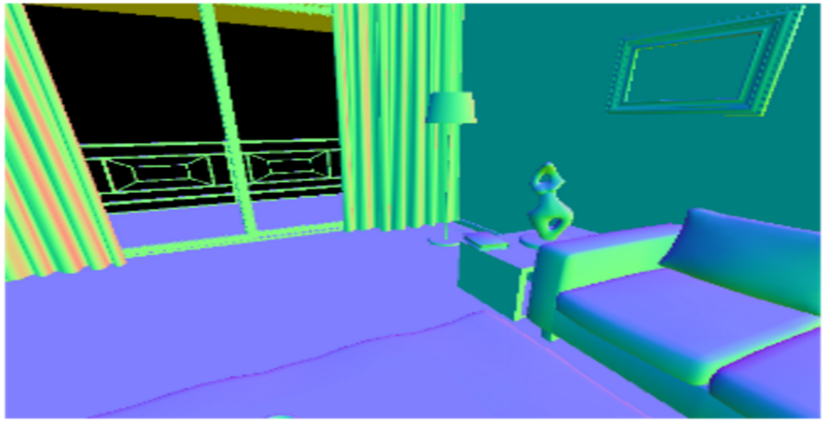
Our Ensemble Application

Reconstructed VR Scene using Unreal Engine with the Torch Plugin with embedded Deep Learning







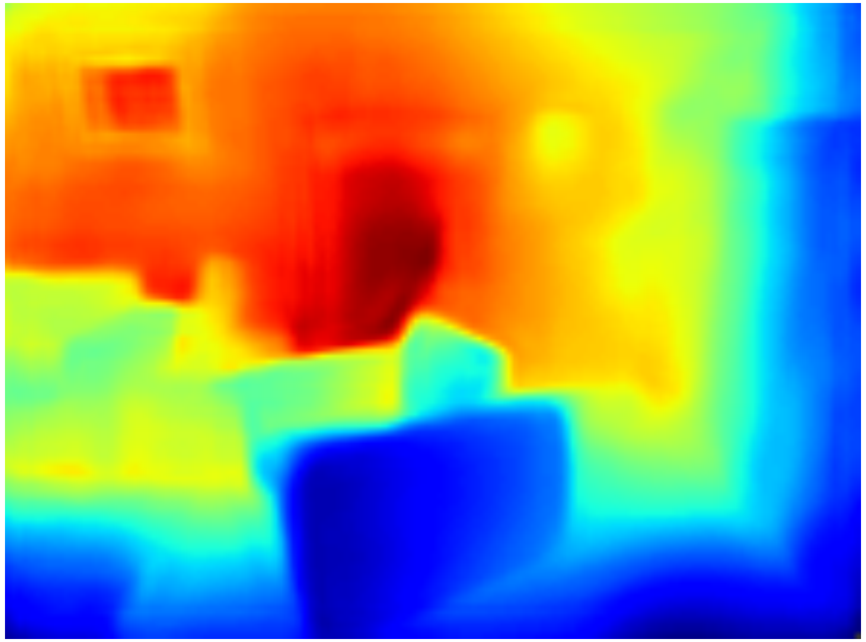


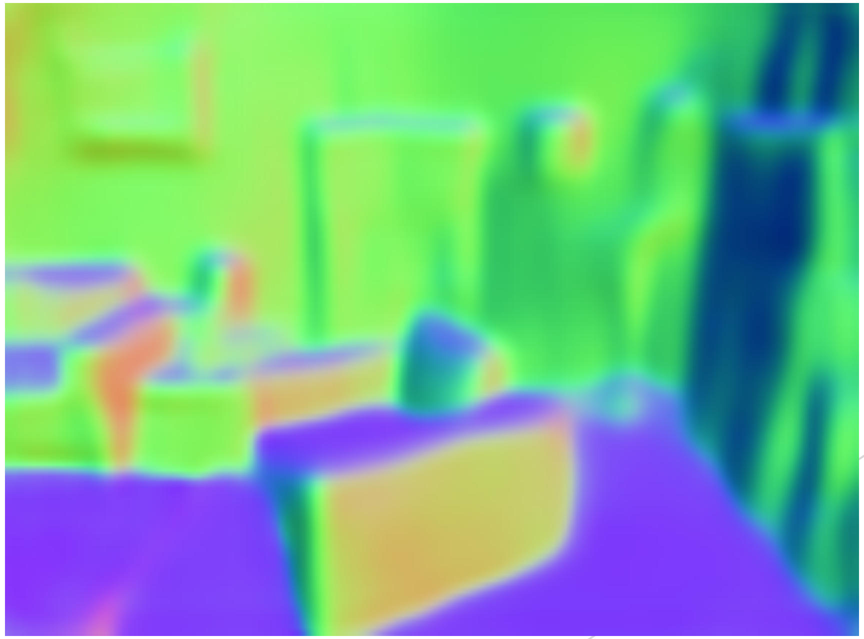
Virtual Reality of Rendered Scene

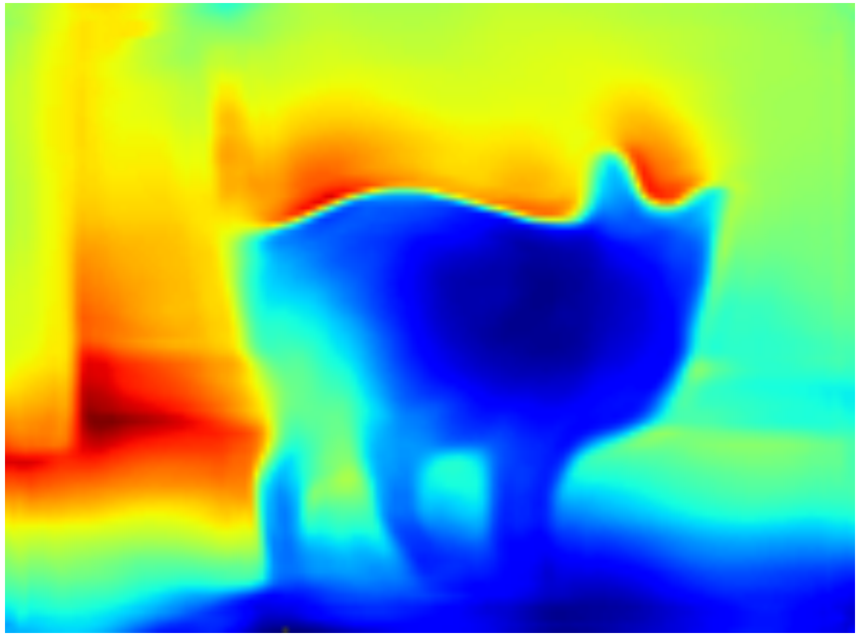
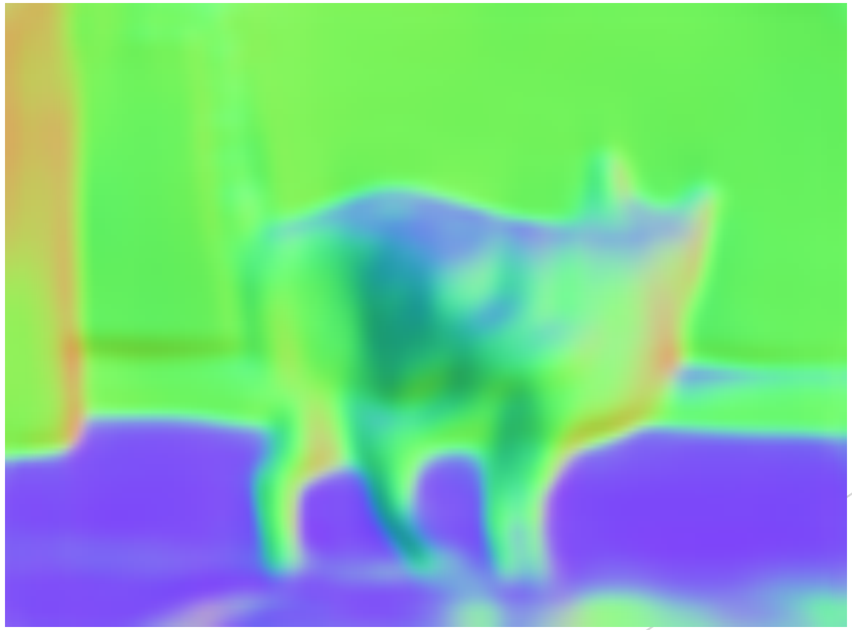
UnrealCV is a project to help computer vision researchers build virtual worlds using Unreal Engine 4 (UE4).

Images generated from the technical demo RelisticRendering

Virtual Reality of Rendered Rendering UE4





Three consecutive frames of a “cat” video: Deepmask from FAIR results

Our Ultimate Goal

- Establish a DL-based framework to be used for a work flow pipeline around creating VR content from 2D video
- With more training, adding newer models, and scaling up the GPU compute we can achieve a product solution that can integrate with the existing pipeline used for CG Movies and VR content
- Although the focus of this presentation is on the input of 2D video into an ensemble of neural nets to create an immersive experience it can easily be adapted for other application
- Some of them are video surveillance, video retrieval and human-machine interaction

For More Information

Henry Chu

Informatics Research Institute
University of Louisiana at Lafayette

Email: chu@louisiana.edu