3D Asset Generation

Milestone 4

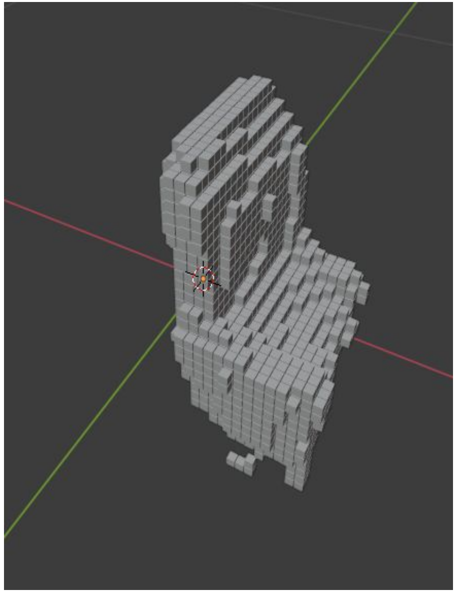
Group A

Aaron Mandeville Justin Chang, Archit Jain, Joshua Benz, Niraj Pandkar, Ji An Lee

Project Idea

- Construct 3D asset from pictures/videos of a real-life object
- Enable these 3D models to be loaded into major game engines for virtual visualization

Dataset



3D model

Use-case (Minecraft)

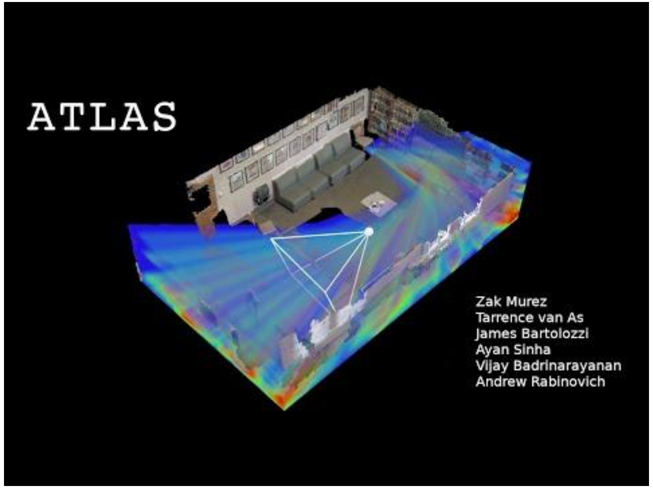
Related Works

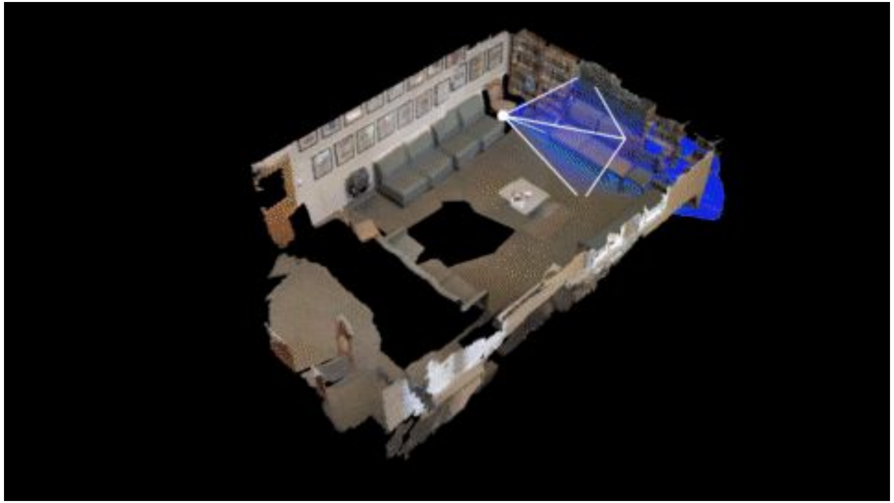
- **ATLAS** : 3D Scene Reconstruction

- **3D-R2N2:** A Unified Approach for Single and Multi-view 3D Object Reconstruction

ATLAS Model

- 3D scene reconstruction
- Input video with corresponding pose information
- Output fine 3D mesh





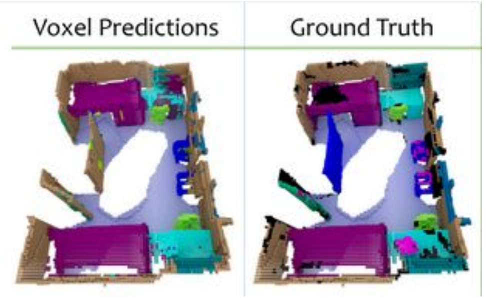
ATLAS Datasets

- **Trained on ScanNet**
  - Contains 2.5 million views from over 1500 scans
  - Created for 3D Scene Understanding
  - Includes:
    - Instance Segmentation
    - camera pose information

- **Tried to convert BigBird Dataset**
  - High quality images of objects
  - Includes:
    - Camera pose information
    - Segmentation masks

http://rll.berkeley.edu/bigbird/

http://www.scan-net.org/

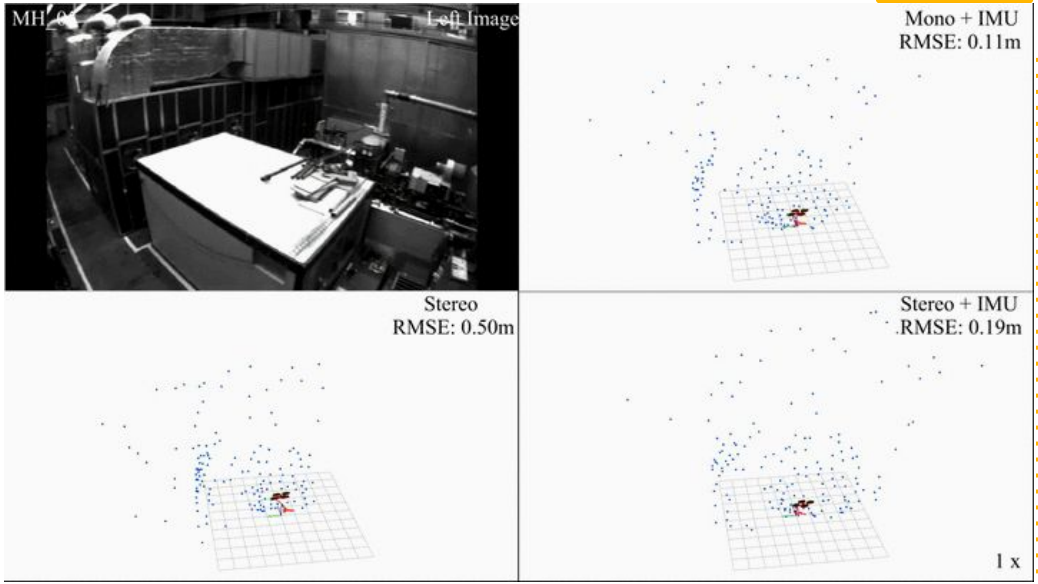


Scannet

ATLAS Data Procurement - VINS-Mono

- **Raw IMU Data**
  - Exponential integration error
  - Measurement drift

- **VINS-Mono (Monocular Visual-Inertial SLAM)**
  - Fuse IMU + Monocular Video
  - Global scale calculation
  - Runs on IOS



ATLAS Shortcomings

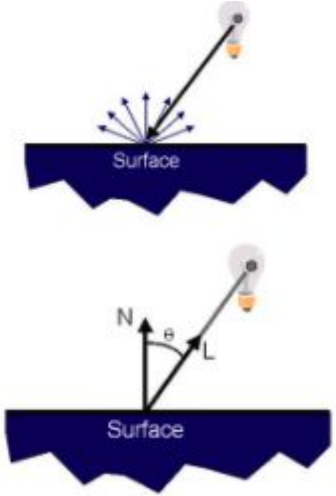
- **Why we had to abandon ATLAS**
- ATLAS focuses on indoor 3D scene reconstruction.
- Scannet is 1.3 TB
- Could not utilize transfer learning because we couldn't replicate scene metadata using our new dataset

id	raw_category	category	count	nyu40id	eigen13id	nyuClass	nyu40class	eigen13class	ModelNet40	ModelNet10	ShapeNetCore55	synsetoffset	wnsynsetid	wnsynsetkey	mpcat40	mpcat40index
1	wall	wall	8277	1	12	wall	wall	Wall					n04546855	wall.n.01	wall	1
2	chair	chair	4646	5	4	chair	chair	Chair	chair	chair	chair	3001627	n03001627	chair.n.01	chair	3
22	books	book	1678	23	2	book	books	Books					n02870526	book.n.11	objects	39
3	floor	floor	1553	2	5	floor	floor	Floor					n03365592	floor.n.01	floor	2
5	door	door	1483	8	12	door	door	Wall	door				n03221720	door.n.01	door	4
1163	object	object	1313	40	7		otherprop	Objects							objects	39
16	window	window	1209	9	13	window	window	Window					n04587648	window.n.01	window	9
4	table	table	1170	7	10	table	table	Table	table	table	table	4379243	n04379243	table.n.02	table	5
56	trash can	trash can	1090	39	6	garbage bin	otherfurniture	Furniture			trash_bin	2747177	n02747177	ashcan.n.01	objects	39
13	pillow	pillow	937	18	7	pillow	pillow	Objects			pillow	3938244	n03938244	pillow.n.01	cushion	8
15	picture	picture	862	11	8	picture	picture	Picture					n03931044	picture.n.01	picture	6
41	ceiling	ceiling	806	22	3	ceiling	ceiling	Ceiling					n02990373	ceiling.n.01	ceiling	17
26	box	box	775	29	7	box	box	Objects					n02883344	box.n.01	objects	39
161	doorframe	doorframe	768	8	12	door	door	Wall	door					doorframe.n.01	door	4
19	monitor	monitor	765	40	7	monitor	otherprop	Objects	monitor	monitor	tv or monitor	3211117	n03782190	monitor.n.04	objects	39
7	cabinet	cabinet	731	3	6	cabinet	cabinet	Furniture			cabinet	2933112	n02933112	cabinet.n.01	cabinet	7
9	desk	desk	680	14	10	desk	desk	Table	desk	desk	table	4379243	n03179701	desk.n.01	table	5
8	shelf	shelf	641	15	6	shelves	shelves	Furniture	bookshelf		bookshelf	2871439	n02871439	bookshelf.n.01	shelving	31
10	office chair	office chair	595	5	4	chair	chair	Chair	chair	chair	chair	3001627	n04373704	swivel_chair.n.01	chair	3
31	towel	towel	570	27	7	towel	towel	Objects					n04459362	towel.n.01	towel	20
6	couch	couch	502	6	9	sofa	sofa	Sofa	sofa	sofa	sofa	4256520	n04256520	sofa.n.01	sofa	10
14	sink	sink	488	34	7	sink	sink	Objects	sink				n04223580	sink.n.01	sink	15
48	backpack	backpack	479	40	7	backpack	otherprop	Objects					n02769748	backpack.n.01	objects	39
28	lamp	lamp	419	35	7	lamp	lamp	Objects	lamp		lamp	3636649	n03636649	lamp.n.02	lighting	28
11	bed	bed	370	4	1	bed	bed	Bed	bed	bed	bed	2818832	n02818832	bed.n.01	bed	11
18	bookshelf	bookshelf	360	10	6	bookshelf	bookshelf	Furniture	bookshelf		bookshelf	2871439	n02871439	bookshelf.n.01	shelving	31
71	mirror	mirror	349	19	7	mirror	mirror	Objects					n03773035	mirror.n.01	mirror	21
21	curtain	curtain	347	16	13	curtain	curtain	Window	curtain				n03151077	curtain.n.01	curtain	12
40	plant	plant	331	40	7	plant	otherprop	Objects	plant				n00017222	plant.n.02	plant	14
52	whiteboard	whiteboard	327	30	7	whiteboard	whiteboard	Objects					n03211616	display_panel.n.01	board_panel	35
96	radiator	radiator	322	39	6	radiator	otherfurniture	Furniture					n04041069	radiator.n.02	misc	40
22	book	book	318	23	2	book	books	Books					n02870526	book.n.11	objects	39
29	kitchen cabinet	kitchen cabinet	310	3	6	cabinet	cabinet	Furniture					n02933112	cabinet.n.01	cabinet	7
49	toilet paper	toilet paper	291	40	7	toilet paper	otherprop	Objects					n15075141	toilet_tissue.n.01	objects	39
29	kitchen cabinets	kitchen cabinet	289	3	6	cabinet	cabinet	Furniture			cabinet	2933112	n02933112	cabinet.n.01	cabinet	7

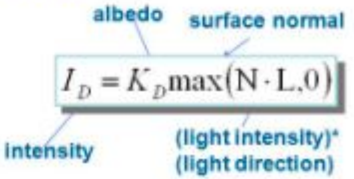
Restrictions of 3D object reconstructions

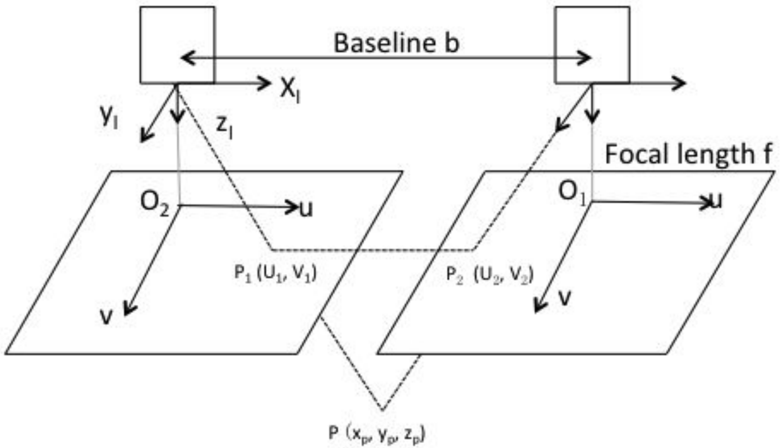
Problems using methods based on SFM or SLAM

1) Objects must be observed from a dense number of views
2) Objects' appearances are expected to be Lambertian (non-reflective)
3) Albedos are supposed to be non-uniform (non-homogenous textures)



- Surface reflects equally in all directions.
  - Examples: chalk, clay, cloth, matte paint
- Brightness doesn't depend on viewpoint.
- Amount of light striking surface proportional to cos θ.





3D-Recurrent Reconstruction Neural Network (R2N2) Model

Unifying single and multi-view 3D reconstruction in a single framework

Key Features

- 3D-R2N2 Model is an extension of 3D convolutional **LSTM**
- Takes in one/more images of an object from various viewpoints and outputs a reconstruction of the object in the form of a **3D occupancy grid** (fig.1, 'Ours')
- **No need** for camera pose **information**, labels, annotations or SLAM
- Use **deep convolutional neural networks** to automatically recover approximated 3D object reconstruction

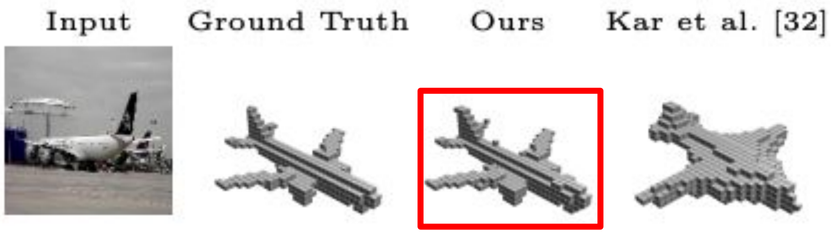
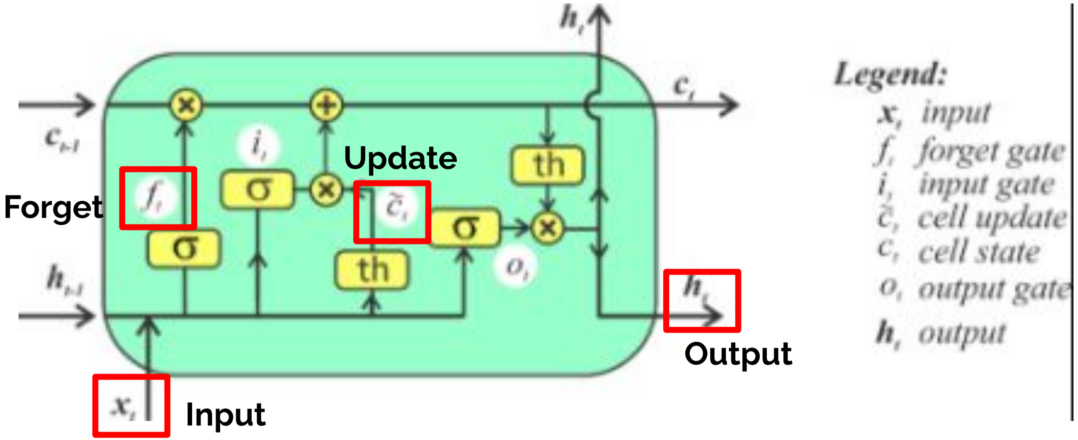


fig.1

3D-Recurrent Reconstruction Neural Network (R2N2) Model

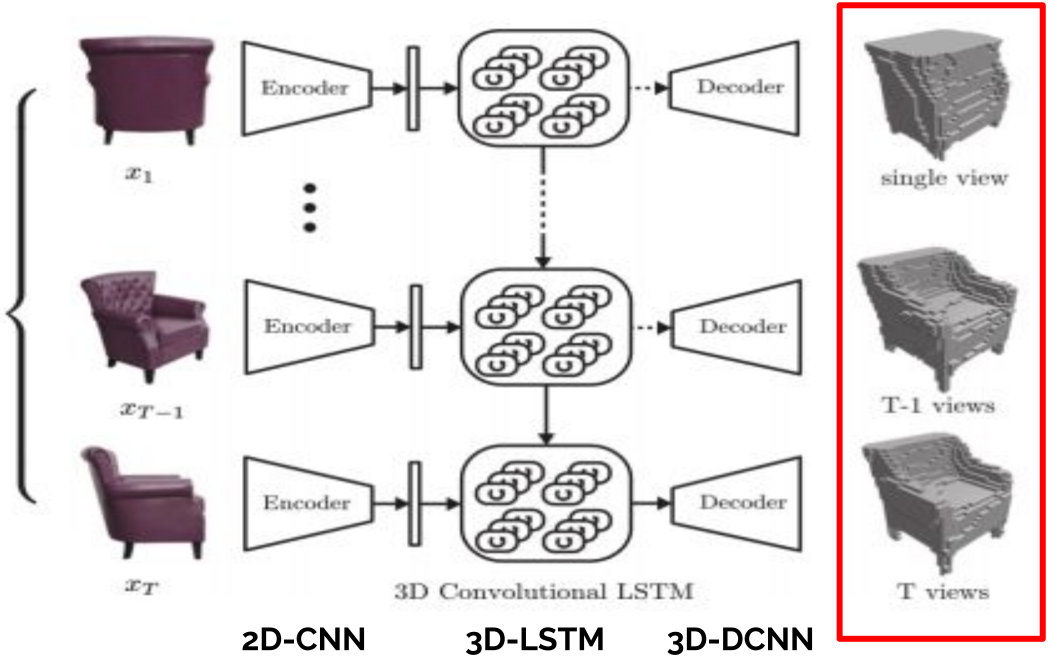
Long Short-Term Memory Unit (**LSTM**)

- Selectively update hidden representations by controlling input and forget gates
- Allows the network to consistently and adaptively learn a suitable 3D representation of an object as information from different viewpoint becomes available



3D-Recurrent Reconstruction Neural Network (R2N2) Model

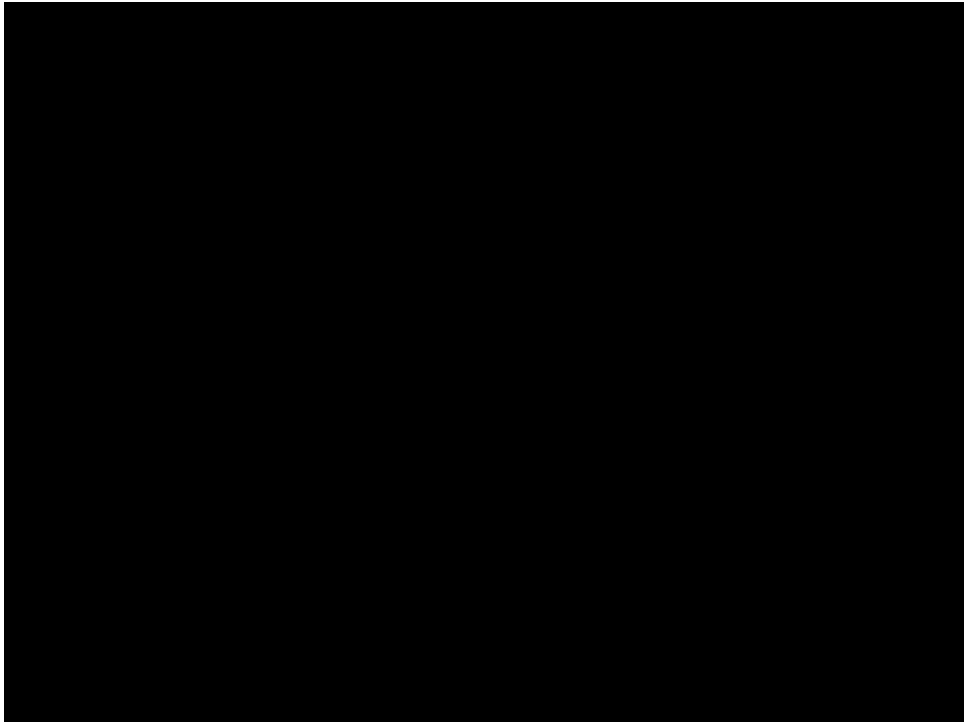
Overall network architecture



3D-R2N2 Datasets

- ShapeNet
  - Semantically annotated 3D CAD models
  - Used to train the model
- PASCAL 3D
- Custom Dataset

3D-R2N2 Demo

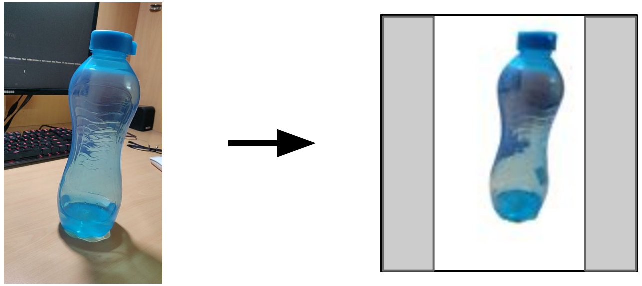


Web Application - Preprocessing

- Flask Web Server (API)
    - Hosted on AWS EC2 instance
    - Serves the deep learning model
- Frame selection from input video
    - Ffmpeg to extract frames
- Removing the background of the object in the image
    - Request to an external service - remove.bg

Web Application - Prediction and Rendering

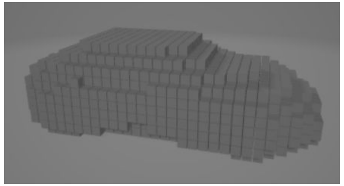
- Dynamic input to the deep learning model
    - Varying number of frames
    - Image dimensions
- Serving the object on the frontend
    - Rendering object (.obj) file
    - Three.js 3D animation library

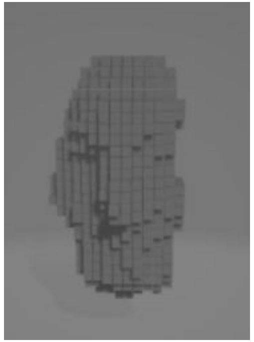


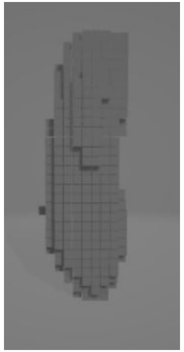
Web Application - Limitations

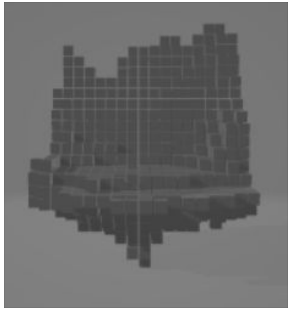
- Limits on external service for removing background from images
- Background must be different from object
- Duration of the video is restricted
- Total inference time needs to be reduced
- Object must be completely visible in all the frames of the video

3D-R2N2 Success







3D-R2N2 Limitations

- Adaptability to objects not included in the Shapenet dataset
- Reconstruction of objects comprising of void is subpar
- Segmentation fails for objects camouflaging with the background

Project Conclusion

- Combination of segmentation and depth information
- 3D reconstruction of objects from 2D videos
- Experiments with ATLAS
- Moving on to 3D-R2N2

Future Scope

- Add more shape priors to the 3D-R2N2 model
- Generate data to fine-tune the ATLAS model
- Generate more finer and accurate results
- Reduce the inference time

Thank you for listening!

Any questions?