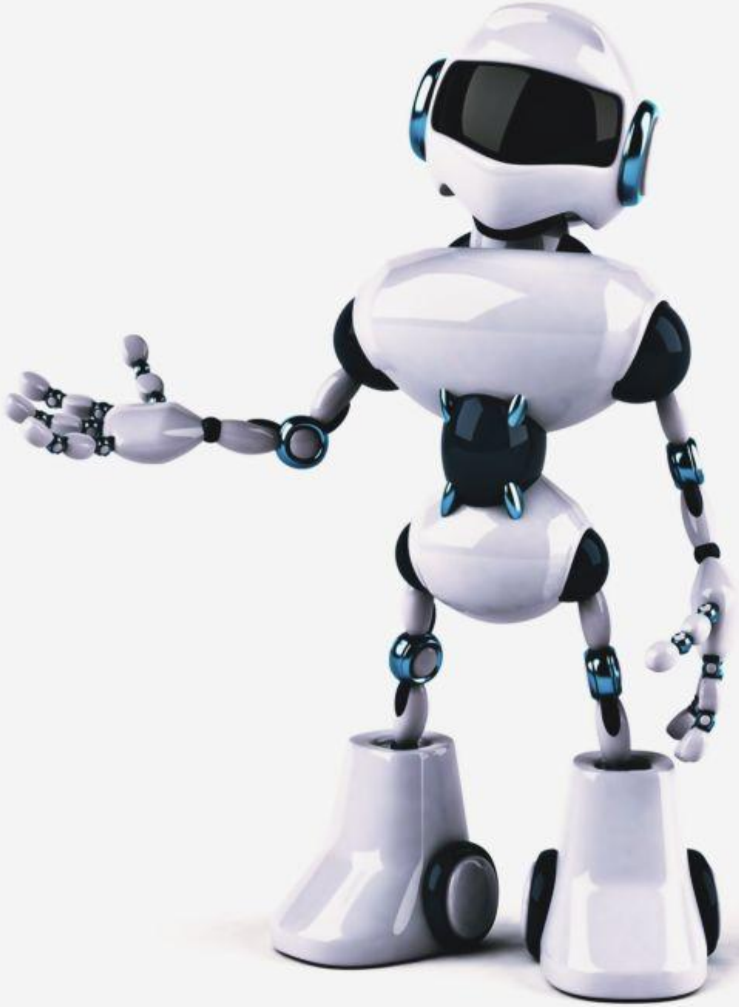
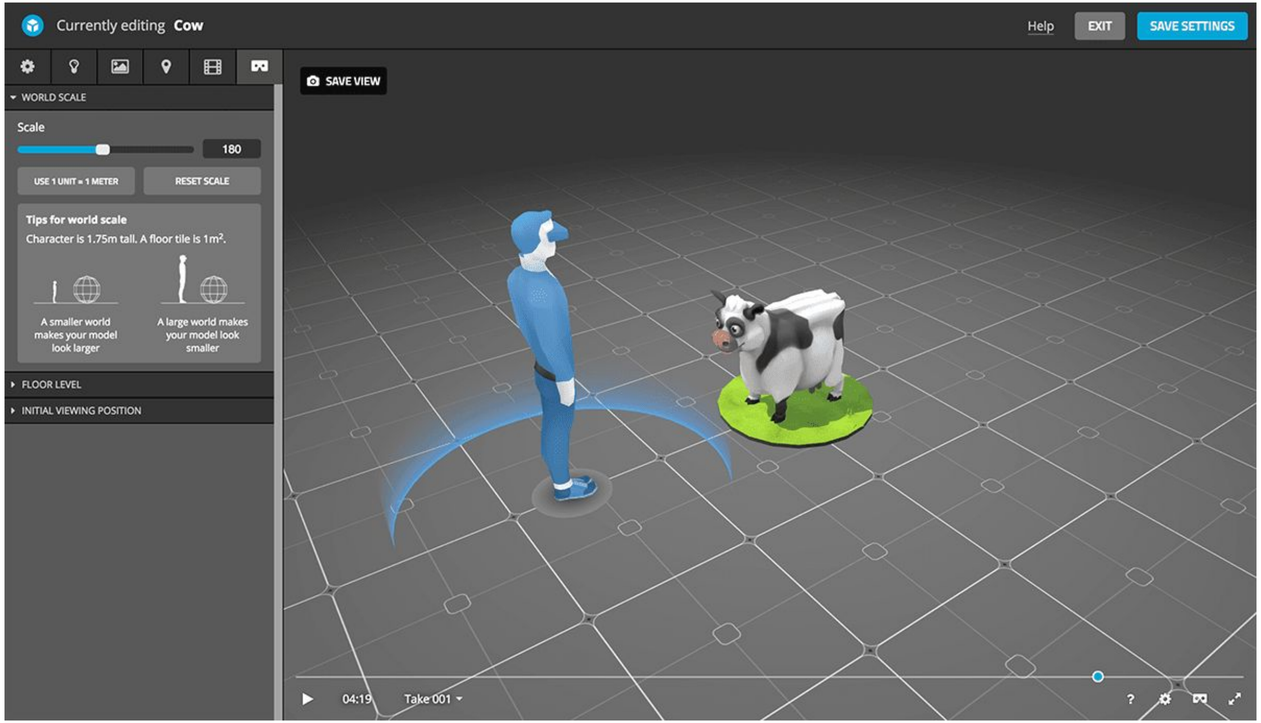
OVA

J.A.R.V.I.S.

- *Akshay Singh Rana*
- *Harmanpreet Singh*
- *Himanshu Arora*
- *Paras Kapoor*



CREATING A WORLD IN VR



Scale 180

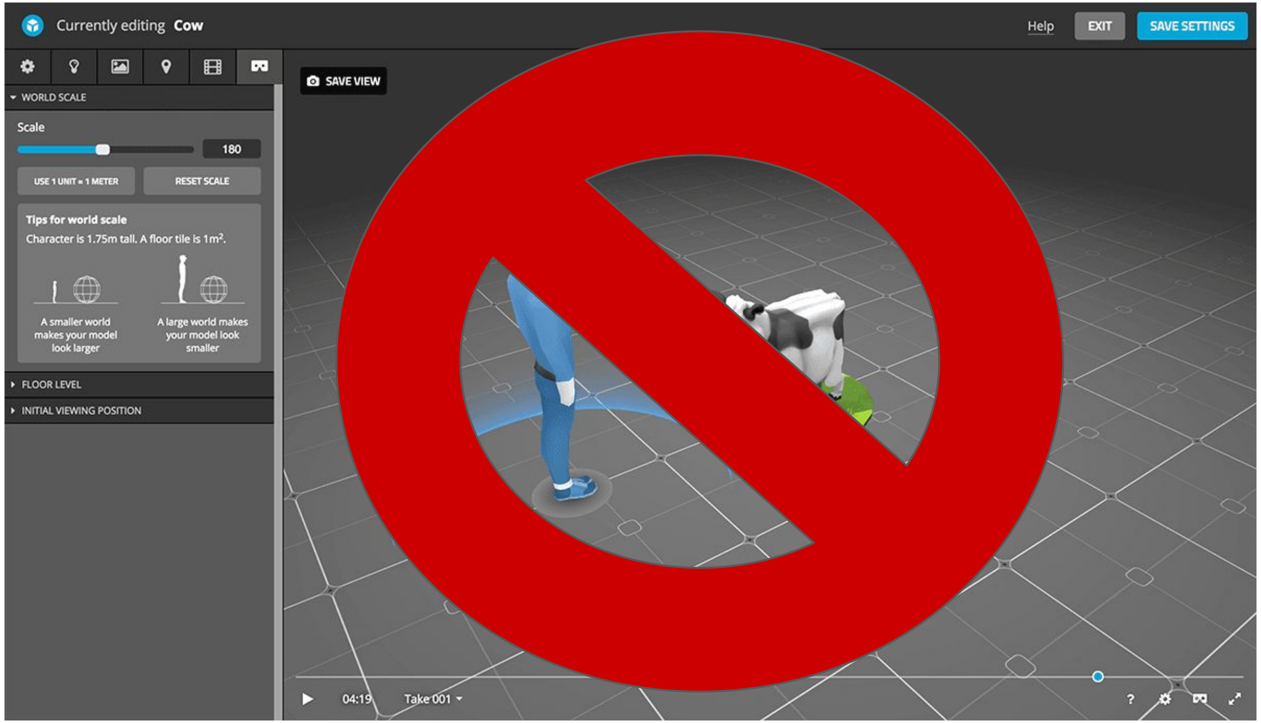
Tips for world scale

Character is 1.75m tall. A floor tile is 1m^2.

A smaller world makes your model look larger

A large world makes your model look smaller

CREATING A WORLD IN VR



Scale 180

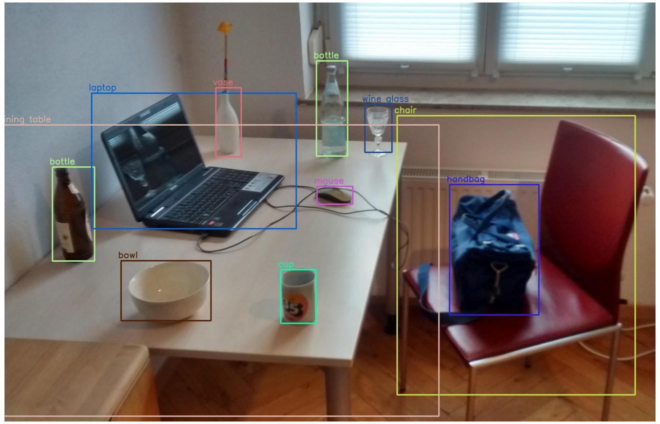
Tips for world scale

Character is 1.75m tall. A floor tile is 1m^2.

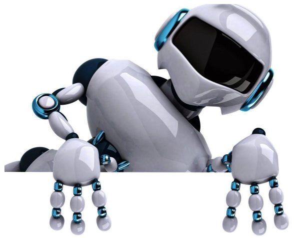
A smaller world makes your model look larger

A large world makes your model look smaller

CREATING A WORLD IN VR USING AI

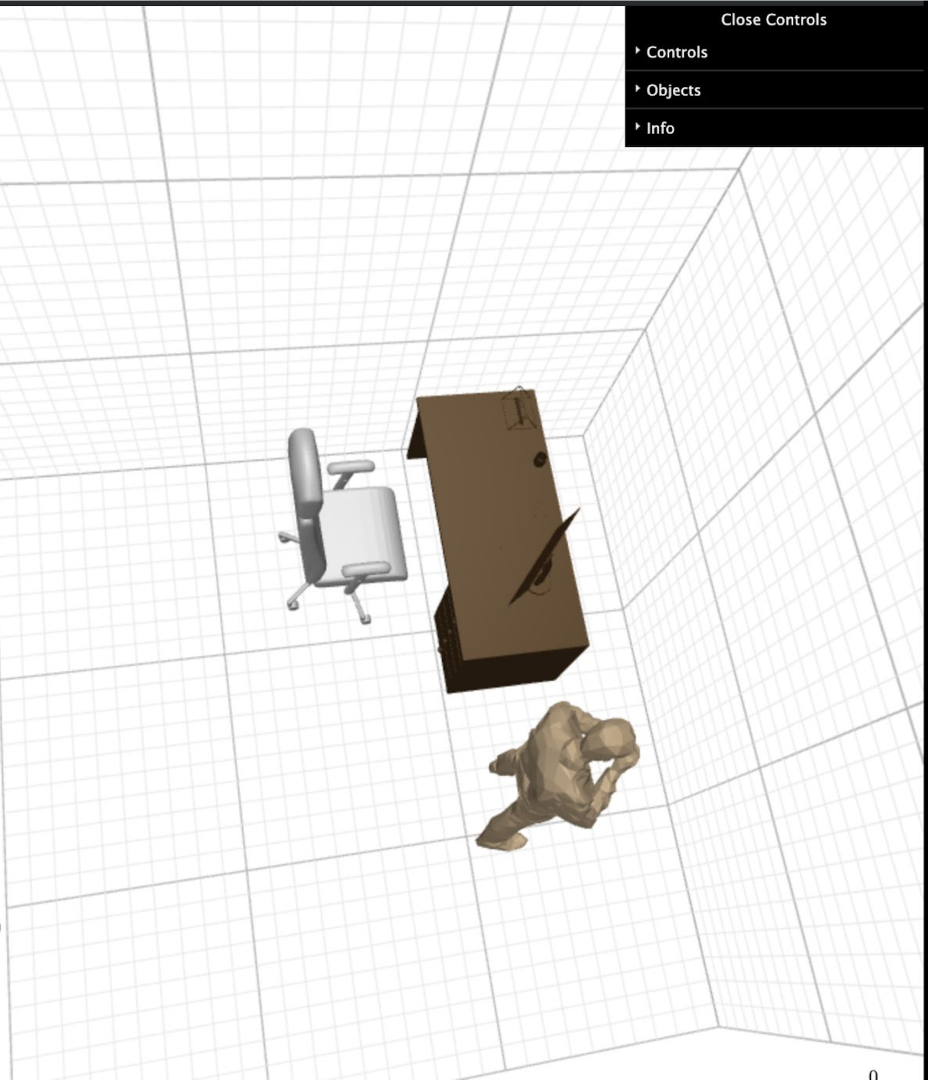


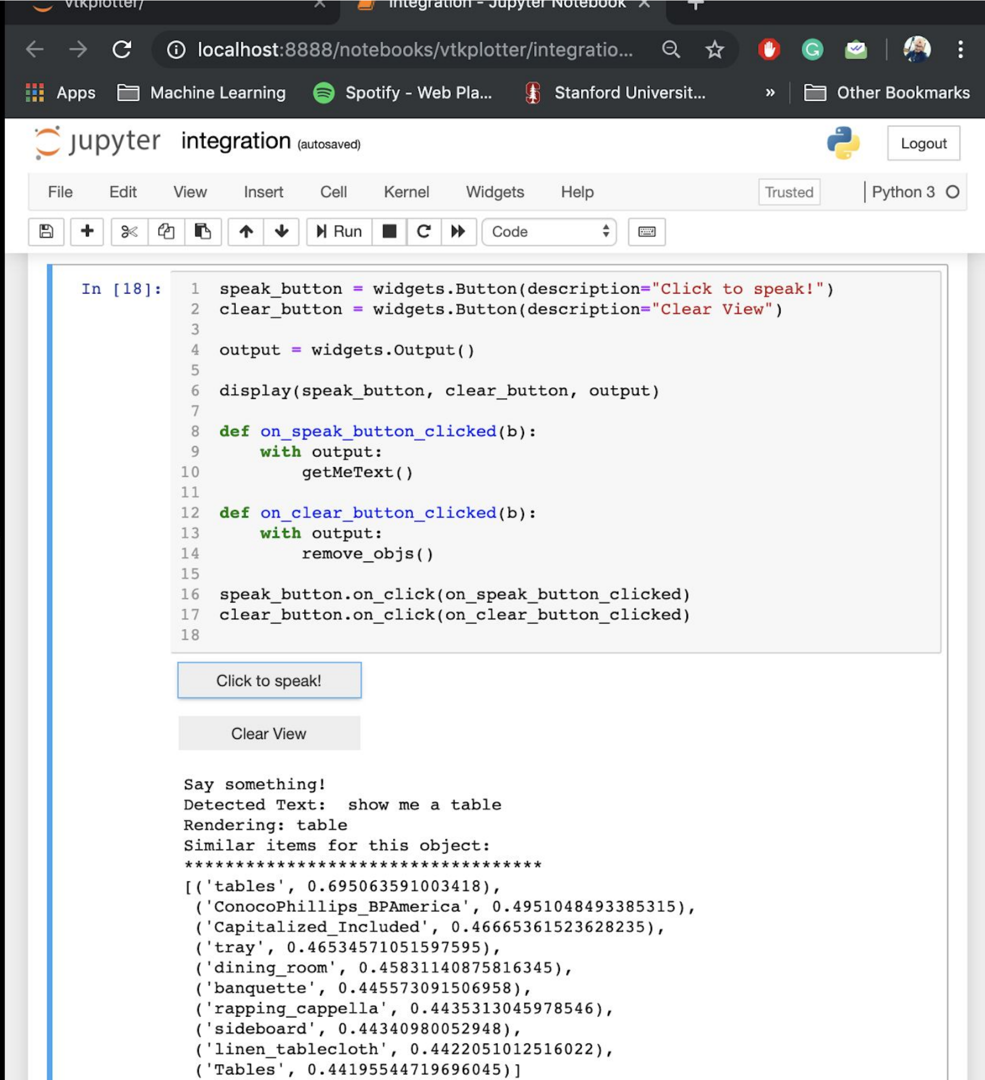
- Selecting and placing 3D objects with voice commands
- Context-based item suggestions



- Creating 3D objects from real-world objects
- Generating similar objects

DEMO

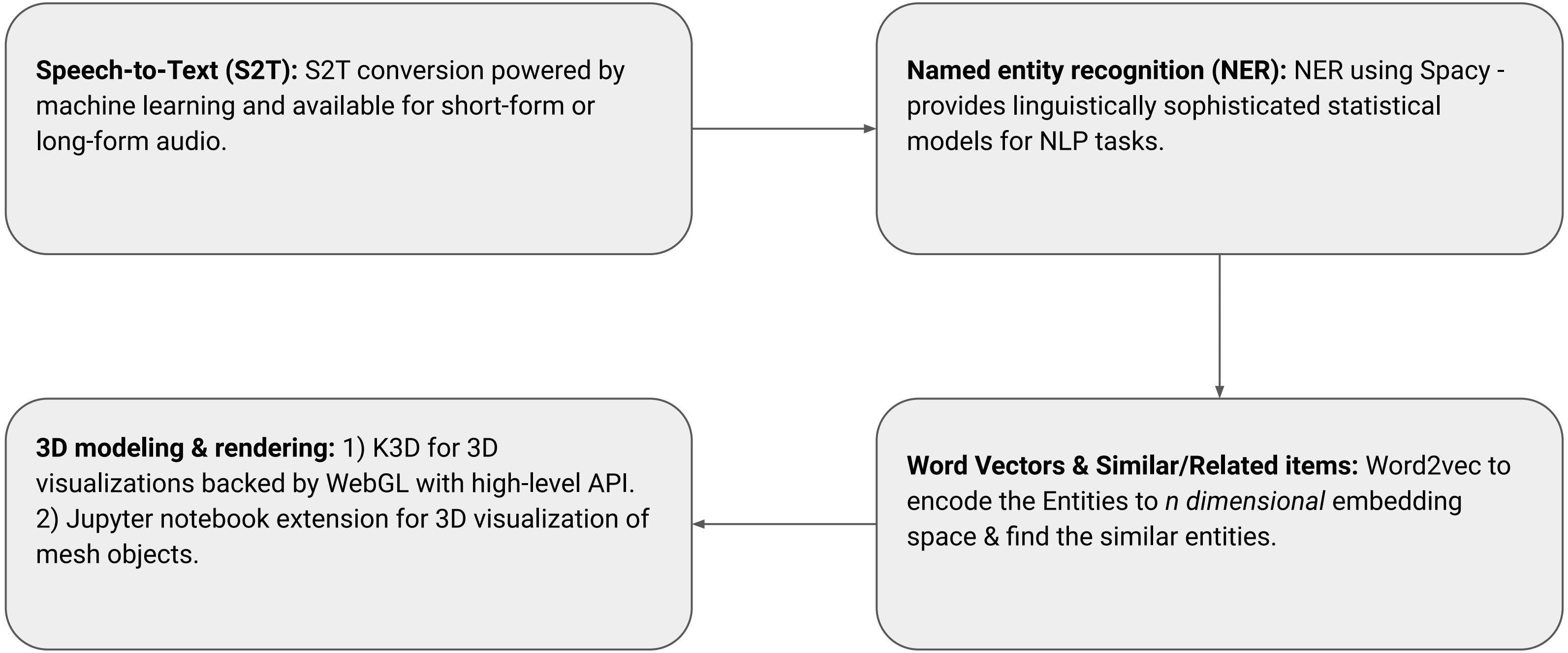



localhost:8888/notebooks/vtkplotter/integratio...
jupyter integration (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
Logout

In [18]:
```python
speak_button = widgets.Button(description="Click to speak!")
clear_button = widgets.Button(description="Clear View")

output = widgets.Output()

display(speak_button, clear_button, output)

def on_speak_button_clicked(b):
    with output:
        getMeText()

def on_clear_button_clicked(b):
    with output:
        remove_objs()

speak_button.on_click(on_speak_button_clicked)
clear_button.on_click(on_clear_button_clicked)
```

```
Say something!
Detected Text:  show me a table
Rendering: table
Similar items for this object:
***********************************
[('tables', 0.695063591003418),
 ('ConocoPhillips_BPAmerica', 0.4951048493385315),
 ('Capitalized_Included', 0.46665361523628235),
 ('tray', 0.46534571051597595),
 ('dining_room', 0.45831140875816345),
 ('banquette', 0.445573091506958),
 ('rapping_cappella', 0.4435313045978546),
 ('sideboard', 0.44340980052948),
 ('linen_tablecloth', 0.4422051012516022),
 ('Tables', 0.44195544719696045)]
```

PIPELINE



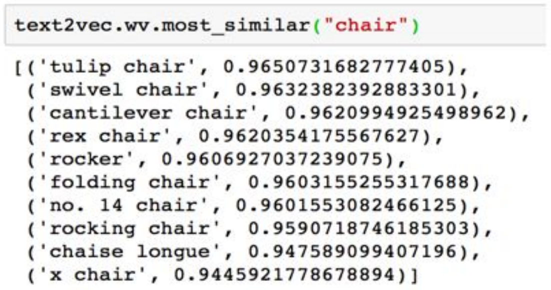
Speech-to-Text (S2T): S2T conversion powered by machine learning and available for short-form or long-form audio.

Named entity recognition (NER): NER using Spacy - provides linguistically sophisticated statistical models for NLP tasks.

Word Vectors & Similar/Related items: Word2vec to encode the Entities to *n dimensional* embedding space & find the similar entities.

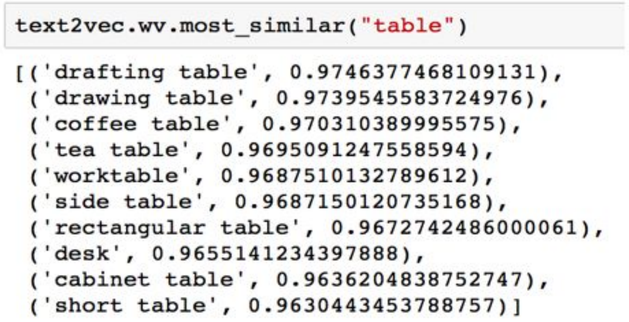
3D modeling & rendering: 1) K3D for 3D visualizations backed by WebGL with high-level API. 2) Jupyter notebook extension for 3D visualization of mesh objects.

Word2vec *- Identifying Most Similar*



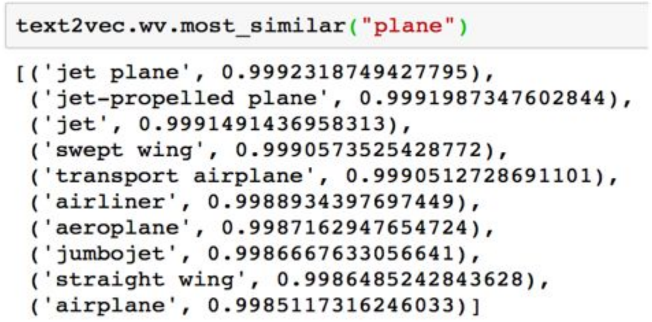
```
text2vec.wv.most_similar("chair")
```

```
[('tulip chair', 0.965073168277405),
 ('swivel chair', 0.9632382392883301),
 ('cantilever chair', 0.9620994925498962),
 ('rex chair', 0.962035417556762),
 ('rocker', 0.9606927037239075),
 ('folding chair', 0.960315525317688),
 ('no. 14 chair', 0.9601553082466125),
 ('rocking chair', 0.9590718746185303),
 ('chaise longue', 0.947589099407196),
 ('x chair', 0.9445921778678894)]
```



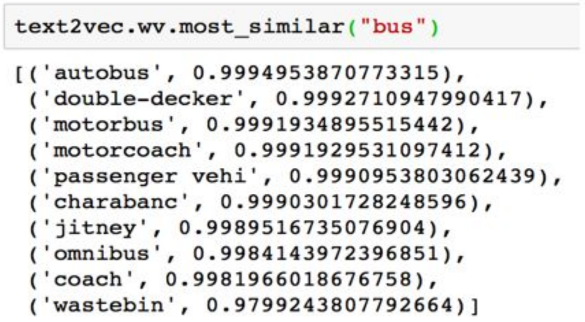
```
text2vec.wv.most_similar("table")
```

```
[('drafting table', 0.9746377468109131),
 ('drawing table', 0.9739545583724976),
 ('coffee table', 0.970310389995575),
 ('tea table', 0.9695091247558594),
 ('worktable', 0.9687510132789612),
 ('side table', 0.9687150120735168),
 ('rectangular table', 0.9672742486000061),
 ('desk', 0.9655141234397888),
 ('cabinet table', 0.9636204838752747),
 ('short table', 0.9630443453788757)]
```



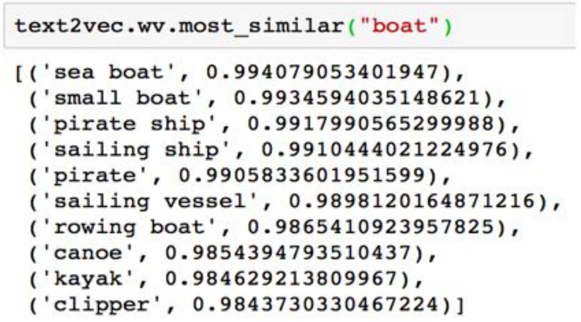
```
text2vec.wv.most_similar("plane")
```

```
[('jet plane', 0.9992318749427795),
 ('jet-propelled plane', 0.9991987347602844),
 ('jet', 0.9991491436958313),
 ('swept wing', 0.9990573525428772),
 ('transport airplane', 0.9990512728691101),
 ('airliner', 0.9988934397697449),
 ('aeroplane', 0.9987162947654724),
 ('jumbojet', 0.9986667633056641),
 ('straight wing', 0.9986485242843628),
 ('airplane', 0.9985117316246033)]
```



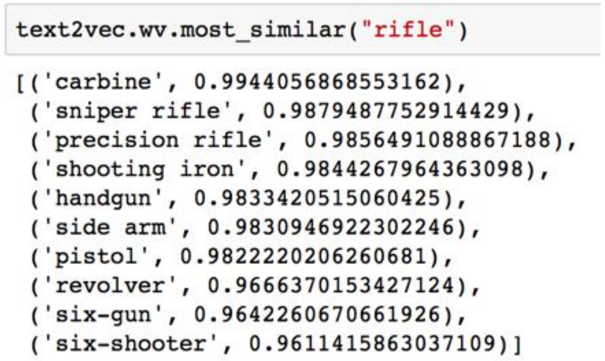
```
text2vec.wv.most_similar("bus")
```

```
[('autobus', 0.9994953870773315),
 ('double-decker', 0.9992710947990417),
 ('motorbus', 0.999193489551544 2),
 ('motorcoach', 0.9991929531097412),
 ('passenger vehi', 0.9990953803062439),
 ('charabanc', 0.9990301728248596),
 ('jitney', 0.9989516735076904),
 ('omnibus', 0.9984143972396851),
 ('coach', 0.9981966018676758),
 ('wastebin', 0.9799243807792664)]
```



```
text2vec.wv.most_similar("boat")
```

```
[('sea boat', 0.994079053401947),
 ('small boat', 0.9934594035148621),
 ('pirate ship', 0.9917990565299988),
 ('sailing ship', 0.9910444021224976),
 ('pirate', 0.9905833601951599),
 ('sailing vessel', 0.9898120164871216),
 ('rowing boat', 0.9865410923957825),
 ('canoe', 0.9854394793510437),
 ('kayak', 0.984629213809967),
 ('clipper', 0.9843730330467224)]
```



```
text2vec.wv.most_similar("rifle")
```

```
[('carbine', 0.9944056868553162),
 ('sniper rifle', 0.9879487752914429),
 ('precision rifle', 0.9856491088867188),
 ('shooting iron', 0.9844267964363098),
 ('handgun', 0.9833420515060425),
 ('side arm', 0.9830946922302246),
 ('pistol', 0.982220206260681),
 ('revolver', 0.9666370153427124),
 ('six-gun', 0.9642260670661926),
 ('six-shooter', 0.9611415863037109)]
```

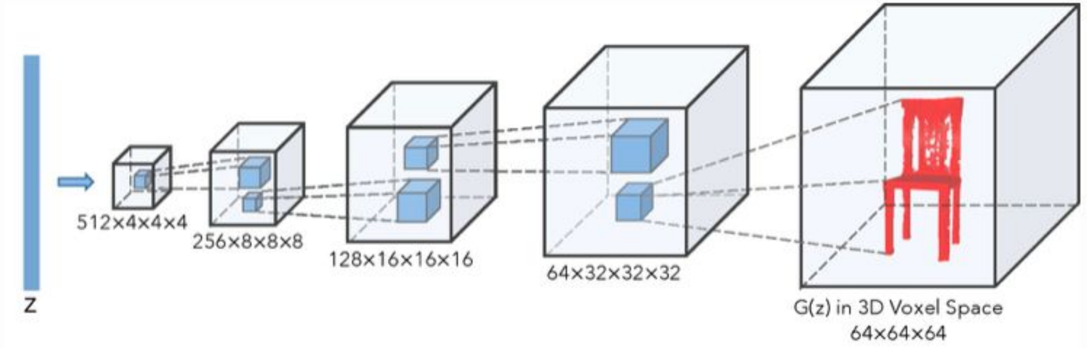
FUTURE WORK

- Training 3D GAN for learning probabilistic latent space
- Scaling objects with voice commands
- Identifying object collisions
- Extend to more scenes

LEARNING A PROBABILISTIC LATENT SPACE OF OBJECT SHAPES VIA 3D GENERATIVE-ADVERSARIAL MODELING

Architecture

The architecture of 3D-GAN is very intuitive with the generator consisting of deconvolutions that upsample high-channeled input feature map to lower channeled output feature map and discriminator just mirrors the generator but consists of strided convolutions. One point to note is that there is not a single fully connected layer in the network, nor at the generator-start nor at discriminator ending. It's fully convolutional in it's true sense.



Using the latent space to generate 3D objects from decoder

Mapping the input text/ relative contextual objects to latent vector space.

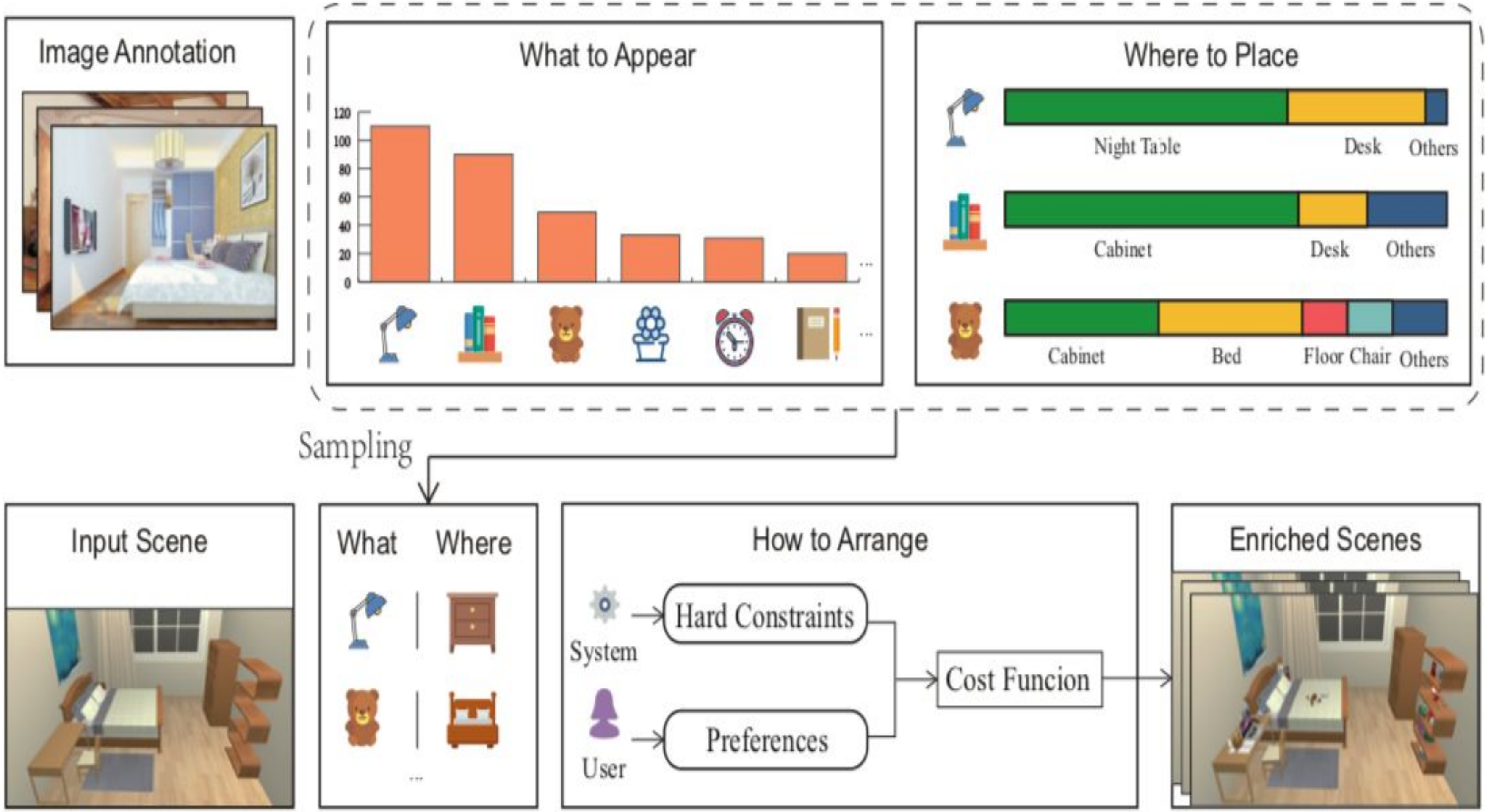


Image Annotation
What to Appear
120
100
80
60
40
20
0
...
Where to Place
Night Table
Desk
Others
Cabinet
Desk
Others
Cabinet
Bed
Floor
Chair
Others
Sampling
Input Scene
What
Where
...
How to Arrange
System
Hard Constraints
User
Preferences
Cost Funcion
Enriched Scenes