RO014 | Embodied AI for Computational Perception and Understanding of Spatial Designs

Karimi Zayan and Prannaya Gupta

Introduction

- **Semantic Segmentation**
  - Identify Specific Regions of Interest
  - Each Pixel assigned a Class
  - Highly detailed algorithms → huge models
- **Our Task**
  - Segment HDB Apartment Interiors (**Task 1**) and Exteriors (**Task 2**)
  - Gain Computational Perception of Architecture of such Designs

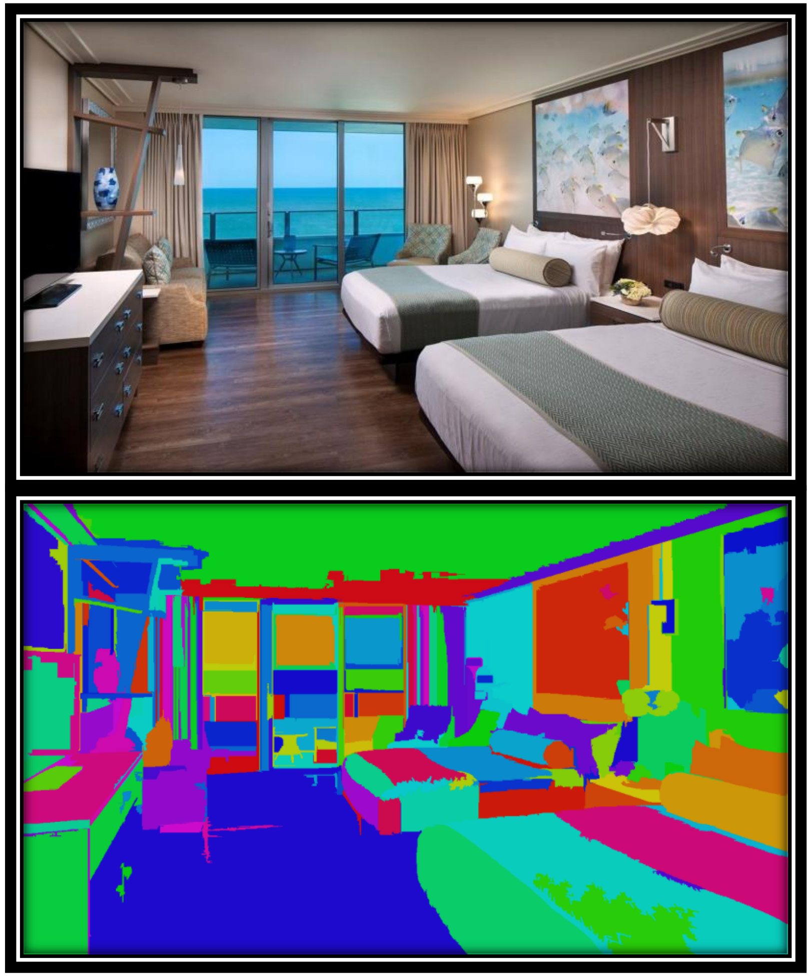


FIGURE 1: An example of Semantic Segmentation Applied on a Hotel Room from ADE20K

Methodology

- Datasets
  - **Task 1**: MIT ADE20K Dataset
    - 3000 Classes, 150 Classes in Apartment Interiors
  - **Task 2**: SUTD *Artificial-Architecture* Laboratory HDB Facade Dataset
    - 3D Exterior Facades, snapshots to create 2D images
    - 4 classes (Background, Wall, Void and Window)
- Algorithms
  - **Transfer Learning**: knowledge gained in one task is used for a similar second task.
  - A **pretrained** Semantic Segmentation model used and
  - The **weights in the last few layers** of the model trained
  - Good results with **limited data** and **limited training**.
  - Google Colaboratory for Training

FIGURE 2: A sample image and it's annotation from the ADE20K dataset

FIGURE 3: A sample image and it's annotation from the HDB Facade Dataset

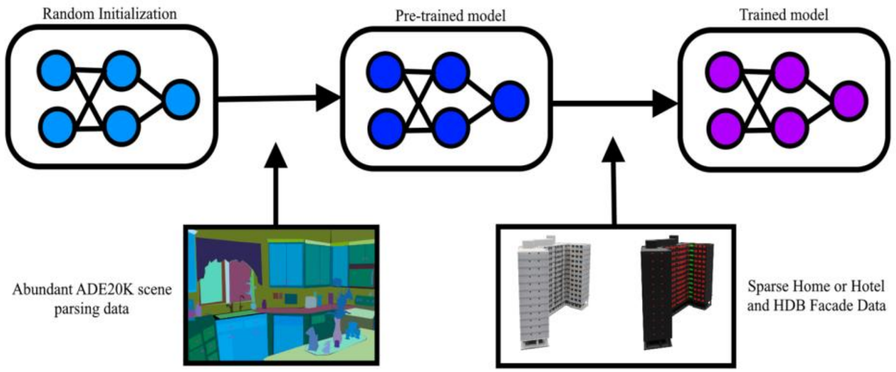


FIGURE 4: Illustration of how Transfer Learning works (partial image from HDB Façade dataset and ADE20K)

Interior Task Methodology

Step I: Data Collection
- ADE20K dataset: images and semantic annotations
- Only images of housing an hotel interiors

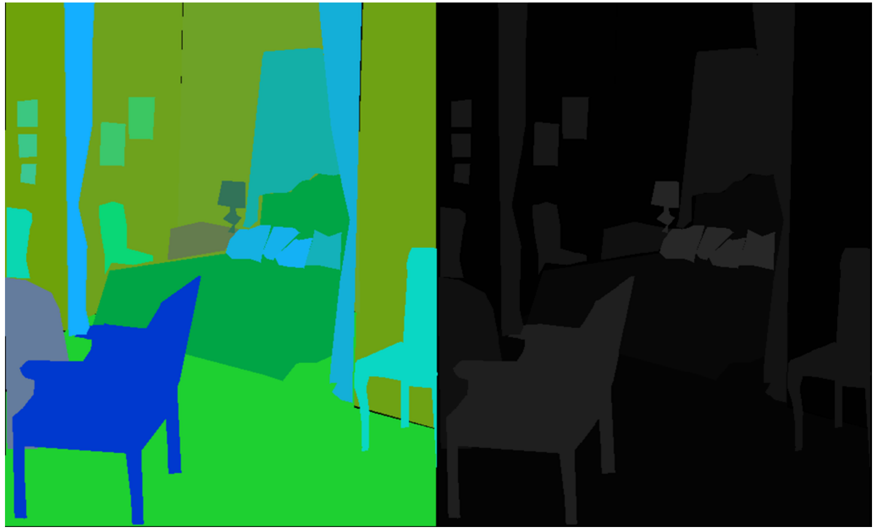


FIGURE 5: Sample of original image and image with minimised classes from ADE20K

Step II: Data Preprocessing
- Minimised the standard 3000 classes to about 150
- hotel and home interior items

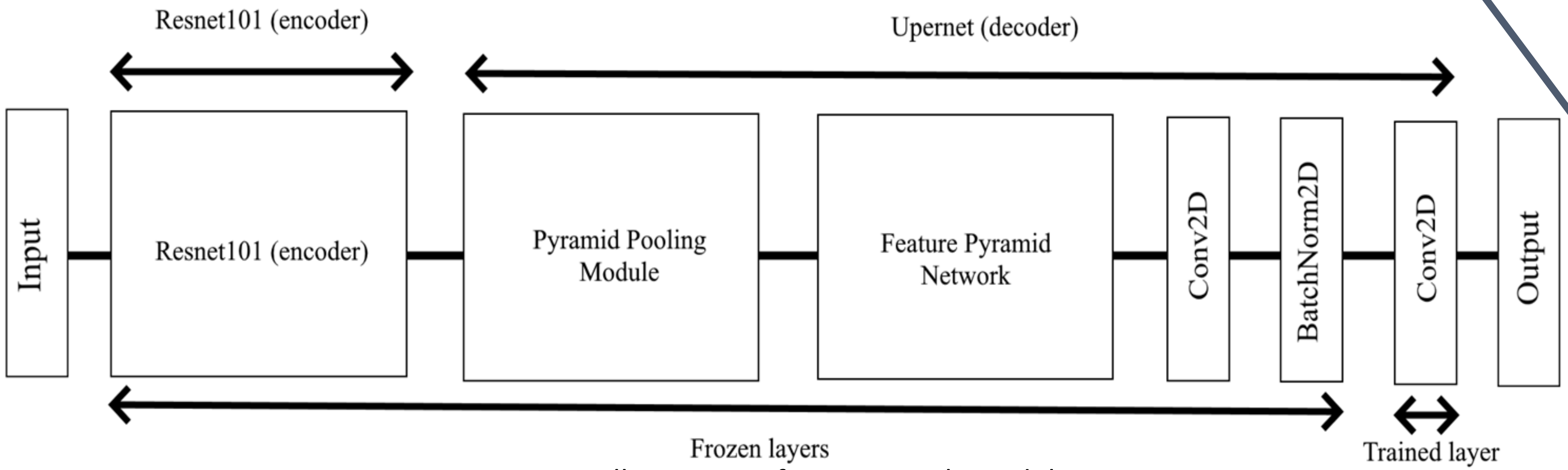


FIGURE 6: Illustration of Interior Task Model

Step III: Pretrained Model
- MIT SemSeg's semantic segmentation model
- trained on dataset like ADE20K

Step IV: Transfer Learning
- Froze all the layers except last convolutional layer
- 3 input channels (RGB)
- 150 output channels → objects in the scene

Step V: Training
- 5400 images
- 10 epochs, 120 iters
- 87.89% training accuracy

Exterior Task Methodology

Step I: Data Collection
- HDB building facade dataset
- buildings and annotations
- 4 classes → background, walls, windows and voids

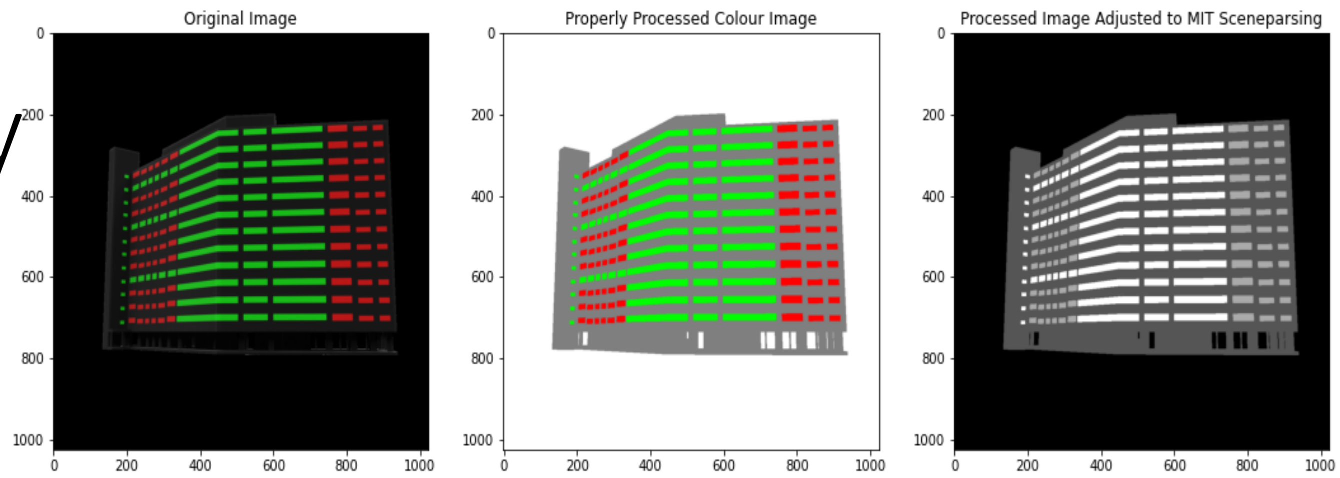


FIGURE 7: A display of our Data Processing Technique (Image from HDB Façade Dataset)

Step II: Data Preprocessing
- Annotation colours not standardised
- Uniformisation (Expert Systems)

Step III: Pretrained Model
MIT SemSeg's semantic segmentation model

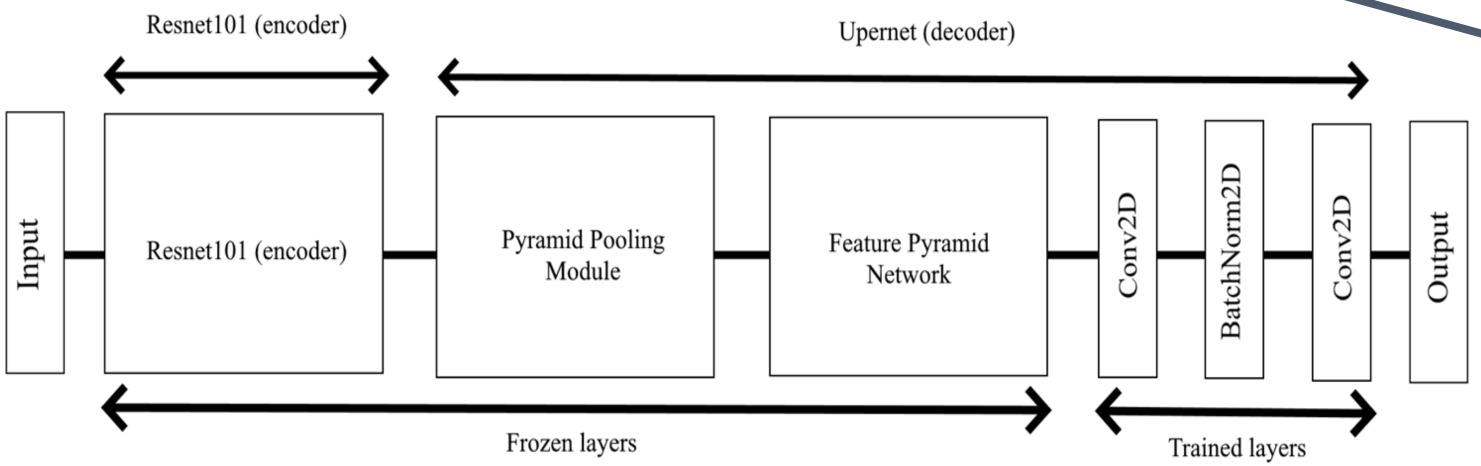


Figure 8: Illustration of Exterior Task Model

Step IV: Transfer Learning
- Froze all the layers except last 2 convolutional layers+ last batch normalization layer
- 3 input channels
- 4 output channels

Step V: Training
The model was run for 10 epochs for 120 iterations each. We achieved 98.28% training accuracy on a 2000 image dataset.

Results

	Interior Task Model	**Exterior Task Model**
IOU	46.35%	59.79%
PA	78.34%	95.20%

TABLE 1: Table of Results for Both Models

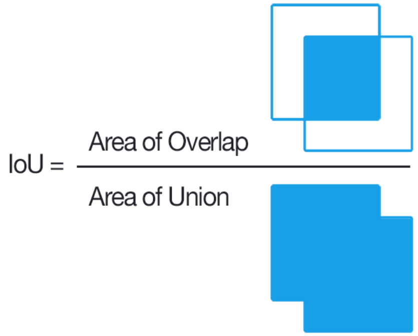
Industry-standard models have
- IoU ≈ 40-45%
- PA ≈ 75-80%

→ Our models performed quite better.

Validation Metrics

- **Intersection-over-Union (IoU)**
  - Average (per class) measure of the overlap between the predicted segmentation and the ground truth

FIGURE 9: Illustration of IoU (Image from Wikimedia Commons)



- **Pixel Accuracy (PA)**
  - Ratio of predicted pixels properly classified to match ground truth against the total number of pixels in each image

Conclusion and Discussion

- Interior Task
  - High Amount of Noise
  - Difficulty segmenting partial objects
- Exterior Task
  - Difficulty differentiating between Voids (Corridors, Void Decks etc) and Windows
  - Higher accuracy than Interior Task: due to low class count
  - Backgrounds not developed, doesn't work well on real-life images

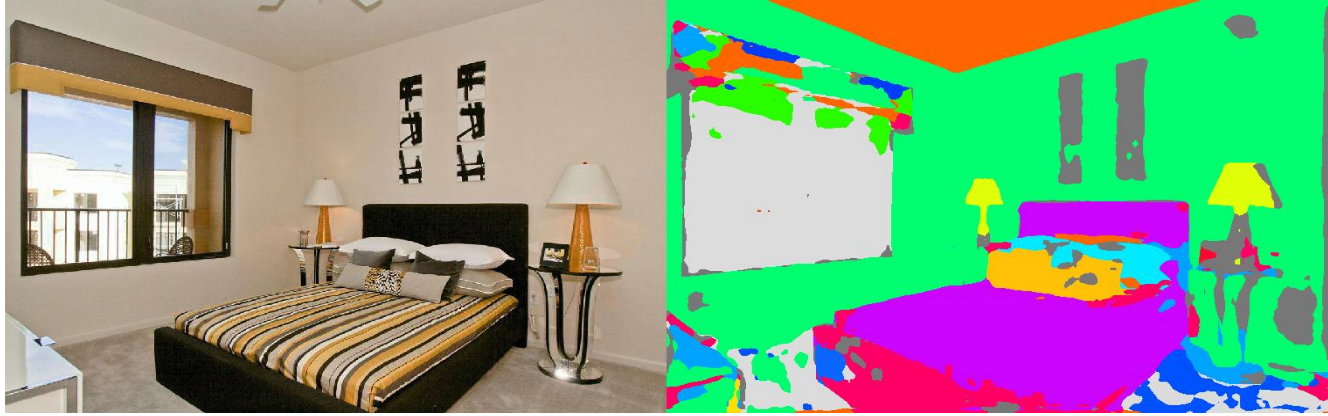


FIGURE 11: Sample Result from Interior Task (Partial Image from ADE20K)

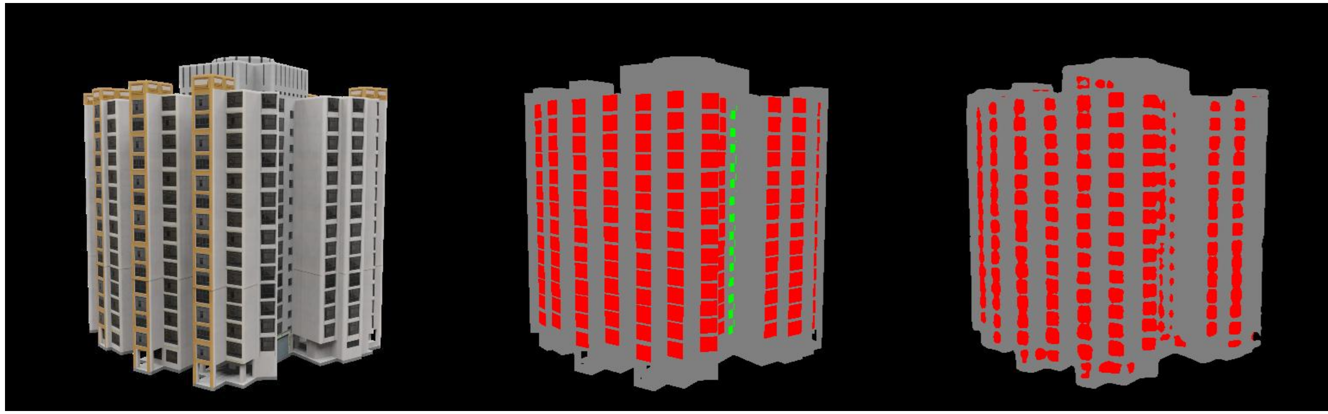
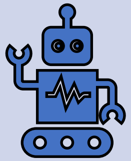


FIGURE 12: Image, Annotation and Predicted Annotation from Exterior Task
(partial image from HDB Façade Dataset)

- Transfer Learning on UperNet Architecture (not been done largely before)
- Large number of layers do not have to be trained to achieve a quality model.

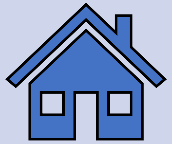
Relevance



Can be used by "agents" for **navigation**

Drones can identify windows for deliveries

Home Robots can identify regions of interest



Architects can use this to design better **blueprints** for housing models

FIGURE 13: Drone Delivery System (Image taken from Wikimedia Commons)

Future Work

- **Material Type Detection**
  - Classifying not only the object but also the object material

FIGURE 14: Example of Material Type Detection
(Partial Image taken from Wikimedia Commons)

- **HDB buildings with backgrounds**
  - Extending the problem to HDB images with backgrounds

FIGURE 15: Possible Image to Train On
(Image taken from the Housing Development Board)

RO014 | Embodied AI for Computational Perception and Understanding of Spatial Designs

Karimi Zayan and Prannaya Gupta

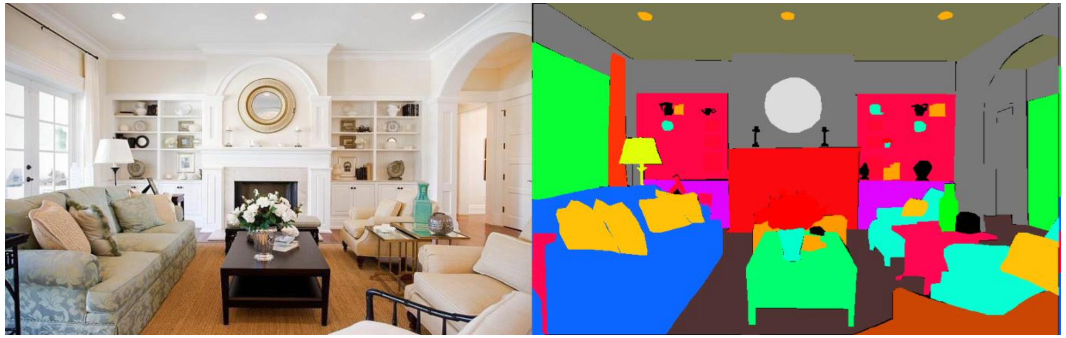


FIGURE 16: A sample image and it's annotation from the ADE20K dataset

FIGURE 17: A sample image and it's annotation from the HDB Facade Dataset

Semantic Segmentation used to identify specific regions of interest, specifically by assigning a class to every pixel of a given image

Perform semantic segmentation on HDB apartment and interiors and exteriors, Gain Computational Perception of Architecture of such Designs

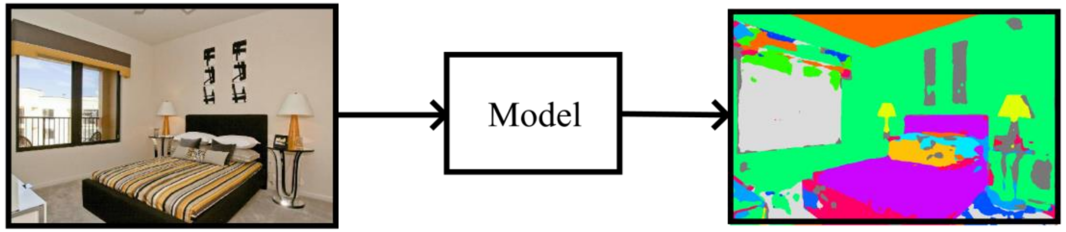


FIGURE 18: An illustration of the Semantic Segmentation Task

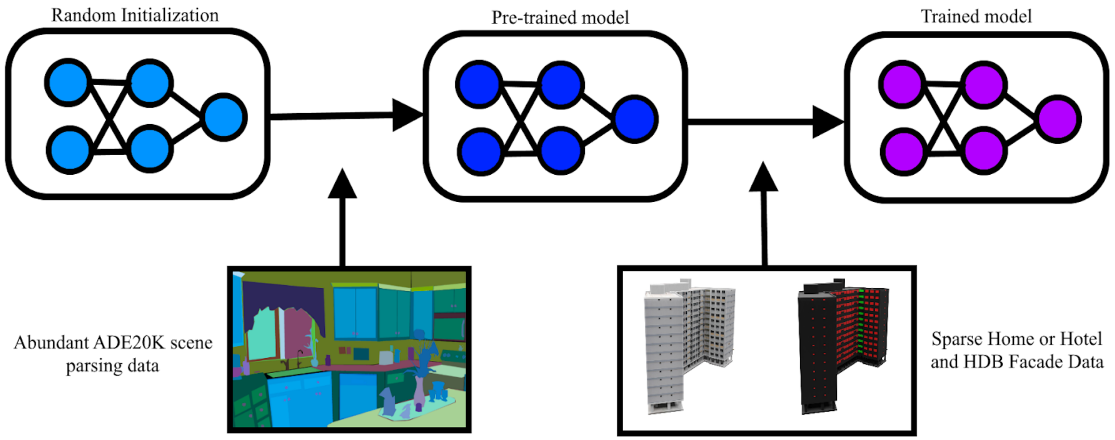


FIGURE 19: Illustration of how Transfer Learning works

Transfer Learning is a Machine Learning technique where knowledge gained in one task is used for a similar second task.
In our case, a pretrained Semantic Segmentation model was used and only the weights in the last few layers of the model were trained.
This allowed us to achieve good results with limited data and limited training.

Step I: Data Collection

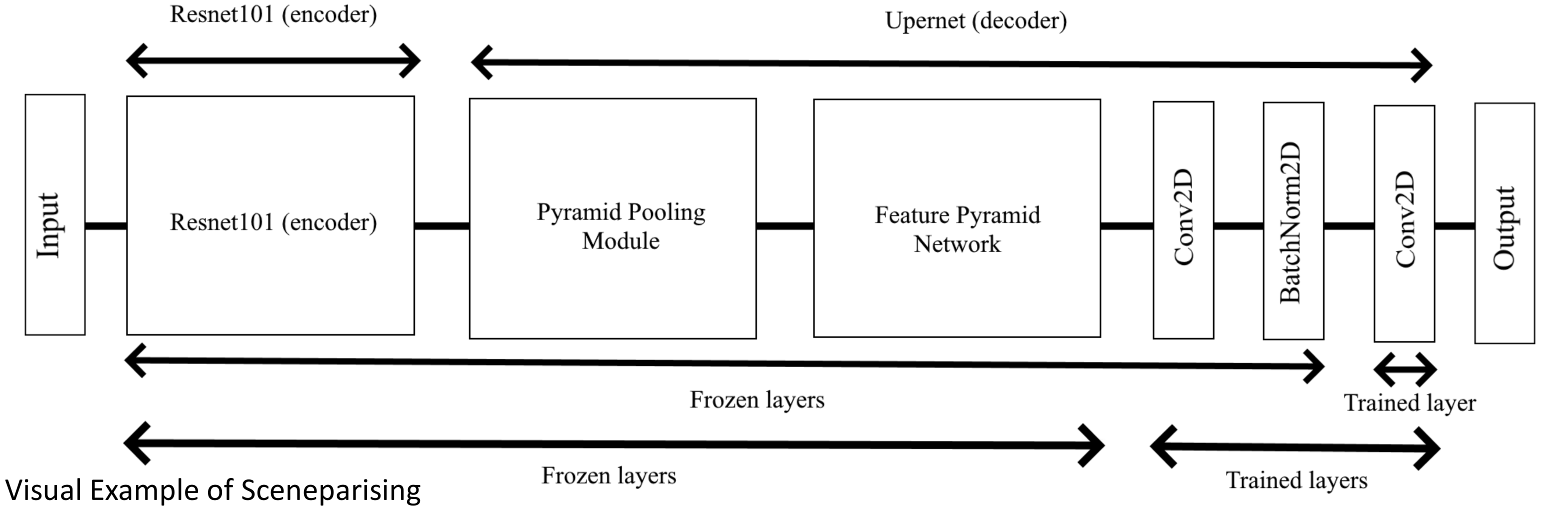
- Images of homes or hotels from the ADE20K dataset
- HDB building facade dataset, buildings and their annotation, 4 classes → background, walls, windows and voids

Step II: Data Preprocessing

- Minimising the number of classes from 3000 to 150 (essential classes such as chair, wall)
- Standardising colours in the images as colours were not standardised

Step III: Pretrained Model

- MIT SemSeg's semantic segmentation model
- trained on a similar dataset to ADE20K



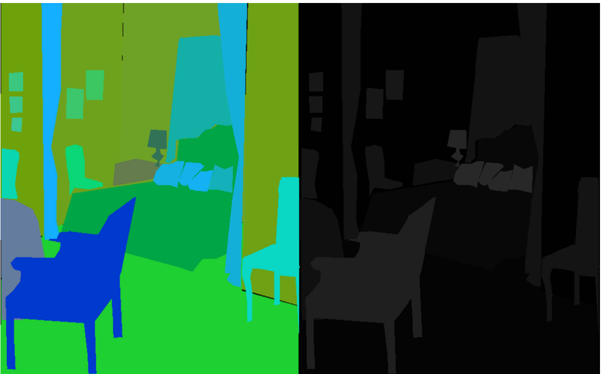
Step IV: Transfer Learning

- Last layer
- Last 3 layers

Step V: Training

10 epochs, 120 iterations

FIGURE 20: Visual Example of Sceneparising



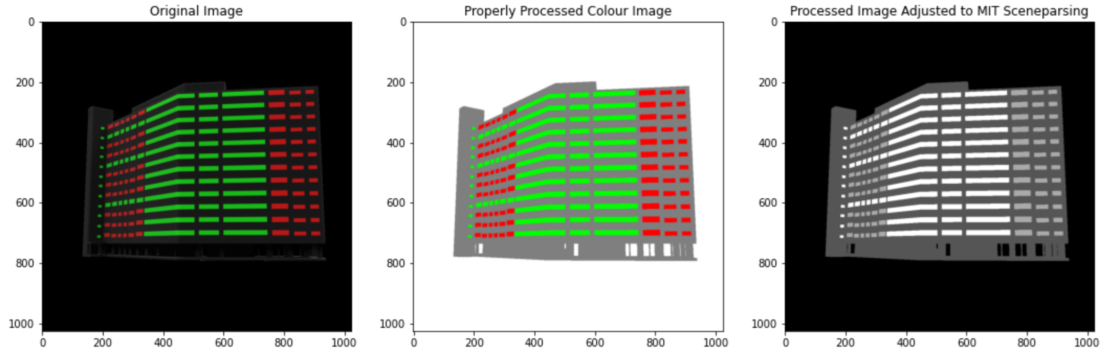


FIGURE 21: A display of our Data Preprocessing Process

	Interior Task Model	Exterior Task Model
IOU	46.35%	59.79%
PA	78.34%	95.20%

Interior Task

High Amount of Noise, Difficulty segmenting partial objects

Exterior Task

Difficulty differentiating between Voids (Corridors, Void Decks etc) and Windows, Higher accuracy than Interior Task: Likely due to low class count, Backgrounds not developed, doesn't work well on real-life images

Transfer Learning on UperNet Architecture is possible
Large number of layers do not have to be trained to achieve a quality model.

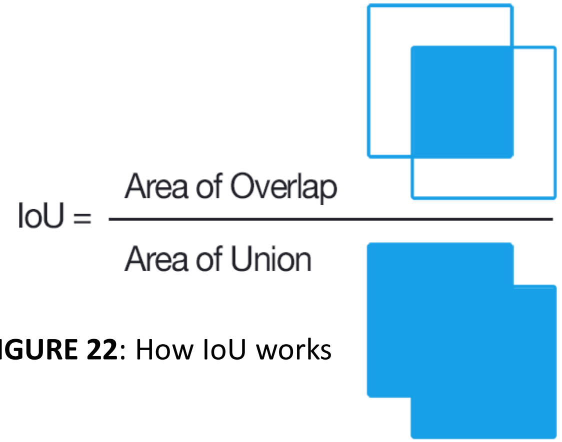
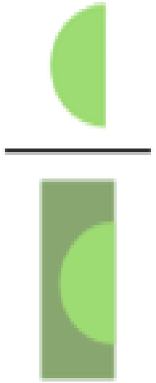


FIGURE 22: How IoU works

FIGURE 23: How PA works

Pixel Accuracy (PA)
Ratio of predicted pixels properly classified to match ground truth against the total number of pixels in each image



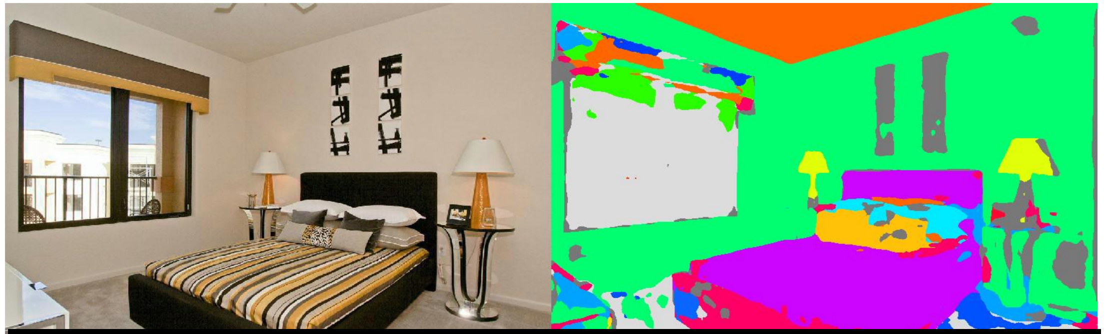


FIGURE 24: Interior Task I/O

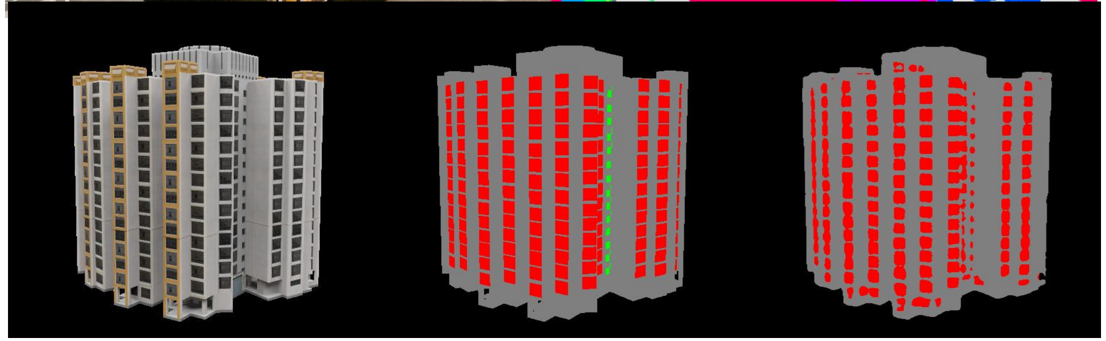


FIGURE 25: Exterior Task I/O

Can be used by agents for navigation, (e.g., drone delivery), Architects/Interior designers can use this to design better blueprints for housing models

References

[1] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. 2015. arXiv: 1411.4038 [cs.CV].

[2] Stevo Bozinovski. “Reminder of the First Paper on Transfer Learning in Neural Networks, 1976”. In: Informatica 44.3 (Sept. 2020). DOI: 10 . 31449 / inf . v44i3 . 2828. URL: https://doi.org/10.31449/inf.v44i3.2828.

[3] MIT. ADE20K dataset. URL: https://groups.csail.mit.edu/vision/ datasets/ADE20K/.

[4] Jianxiong Xiao et al. “SUN database: Large-scale scene recognition from abbey to zoo”. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010, pp. 3485–3492. DOI: 10.1109/CVPR.2010.5539970.

[5] Bolei Zhou et al. “Places: A 10 million Image Database for Scene Recognition”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).

[6] Bolei Zhou et al. “Scene Parsing through ADE20K Dataset”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[7] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: Advances in Neural Information Processing Systems 32. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers. neurips.cc/paper/9015- pytorch- an- imperative- style- highperformance-deep-learning-library.pdf.

[8] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[9] Tete Xiao et al. Unified Perceptual Parsing for Scene Understanding. 2018. arXiv: 1807. 10221 [cs.CV]

[10] Shervin Minaee et al. Image Segmentation Using Deep Learning: A Survey. 2020. arXiv: 2001.05566 [cs.CV].

[11] Irem Ulku and Erdem Akagunduz. A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images. 2022. arXiv: 1912.10230 [cs.CV].

[12] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. 2016. arXiv: 1511 . 00561 [cs.CV].

[13] Changing Colorspaces. URL: https://docs.opencv.org/4.x/df/d9d/ tutorial_py_colorspaces.html.

[14] Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. Neurocomputing, 406, 302–321. https://doi.org/10.1016/j.neucom.2019.11.118