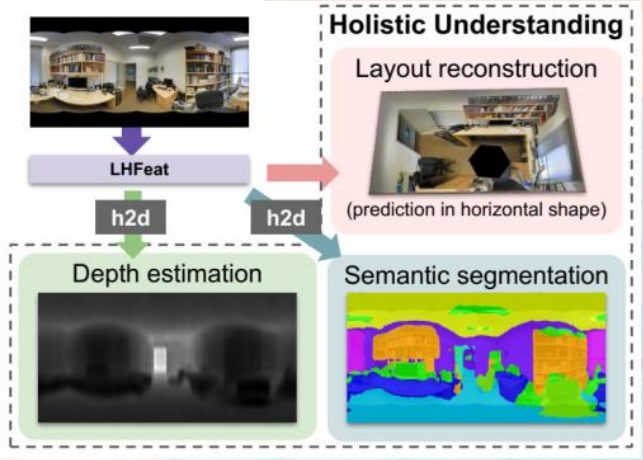
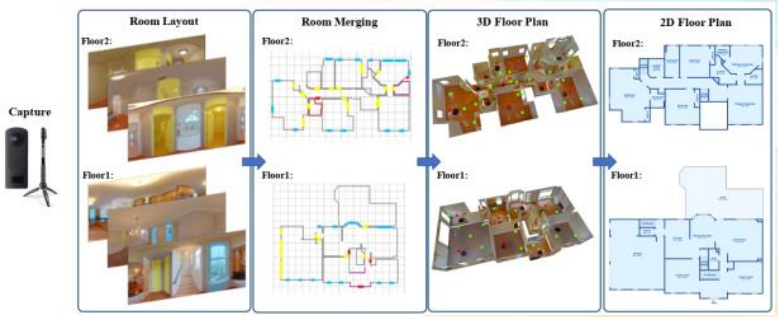
SESSION 4: INTEGRATED INDOOR MODEL

Introduction

- Input: single and multi-view information
  - 3D room model and/or pixel-wise information
  - Camera positions and/or multi-view features
- Output: permanent structure scene
  - Single or multi-room scene
  - Structured floorplan with registered panoramas
  - Objects: not covered in this course…
    - Total scene understanding is a topic itself
- Pre-requisite: images registration
  - Not strictly



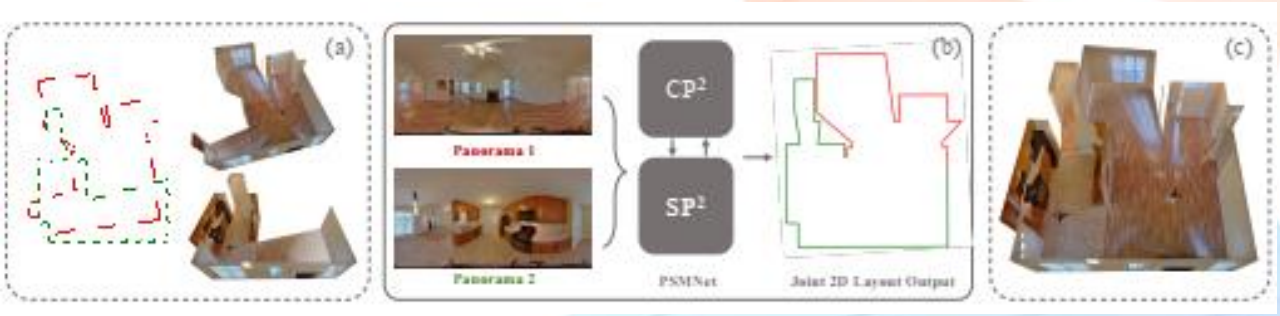
HoHoNet - Sun CVPR2021



ZInD – Cruz CVPR2021

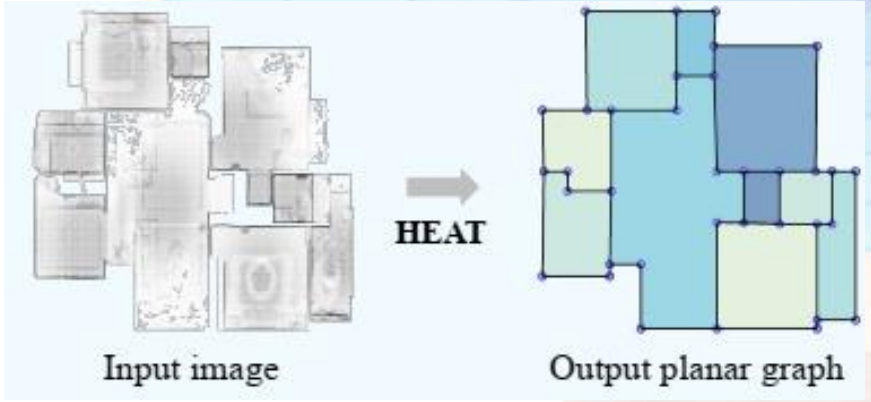
Shabani et al. ICCV2021

Tasks

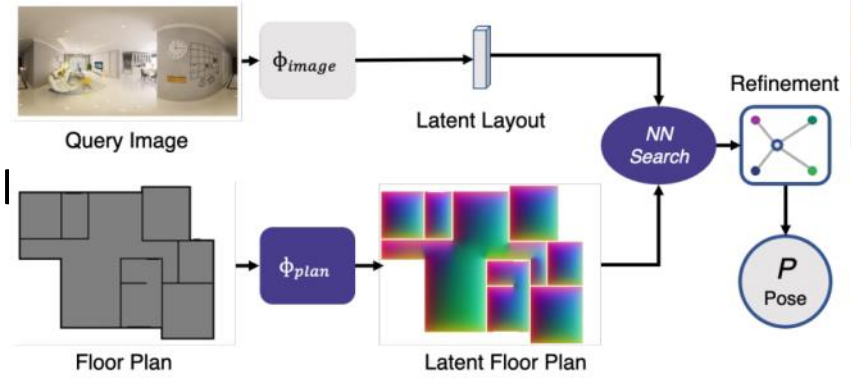


PSMNet - Wang CVPR *2022*

- Multi-view layout estimation
  - Integrating multiple single-view analysis
  - Sparse input: common case
  - Single or multi-room target
- Structured floorplan reconstruction
  - Multi-room segmentation
  - Dense input: professional capture
  - Walls, door, etc. identification
- 3D scene reconstruction and view localizatio
  - Sparse and dense input: specific cases
  - Combining multi-modal data for a 3D model



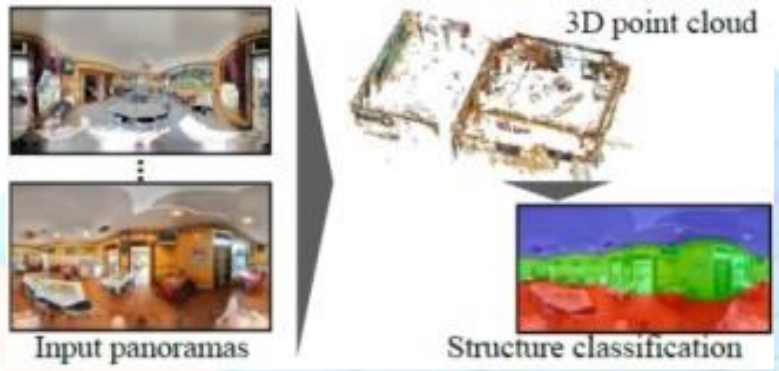
HEAT – Chen CVPR *2022*

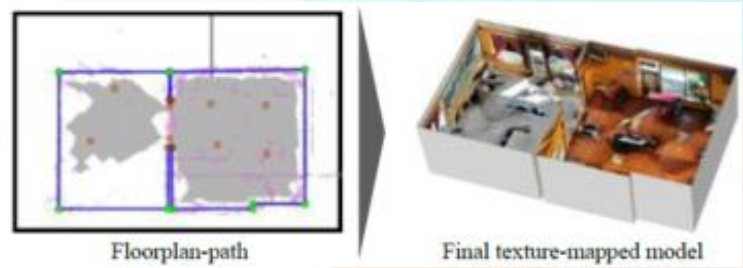


LalaLoc++ - Howard *ECCV 2022*

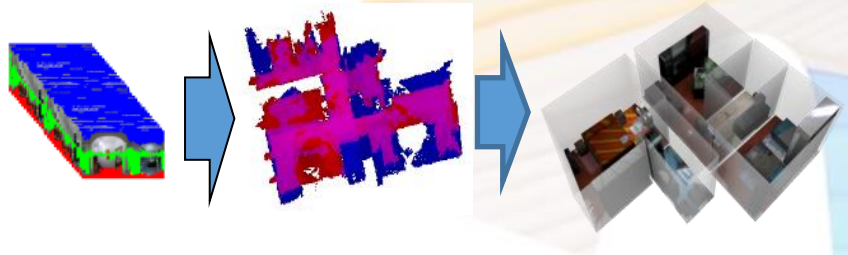
Multi-view layout estimation

- Early approaches
  - Exploiting multi-view registration
    - World reference frames
    - Sparse 3D information
  - Cabral et al 14: panorama analysis to complete 3D data
    - Externally calculated point cloud from MW-MVS
    - Labeled superpixels
  - Pintore et al 19: 3D facets from multiple panoramas
    - Assuming VW (vertical walls): less restrictive than MW
    - E2P transform locally applied to each super-pixel
      - 2D super-pixel +sparse MV features -> 3D facet
      - 3D facets from multiple images joined to identify layout





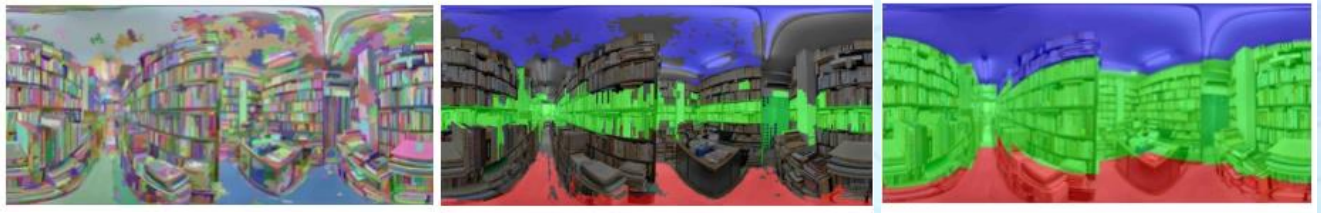
Cabral et al. CVPR2014



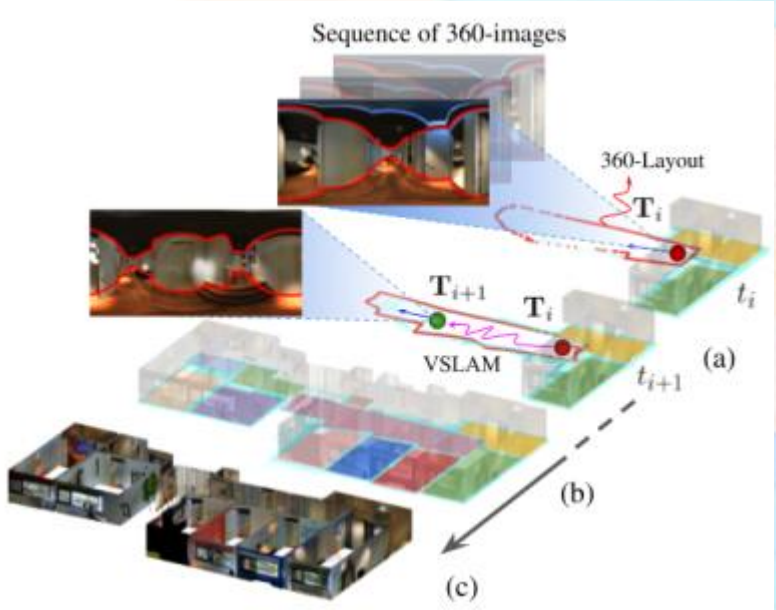
Pintore, Ganovelli, Jaspe, Gobbetti CGF 2019

Multi-view layout estimation

- Early approaches limitations
  - Image segmentation not robust
    - Hand-crafted features
    - Empirical criteria and thresholds
  - 3D data quality leads reconstruction
    - Dense images coverage needed
- Data-driven techniques
  - Boosted the computer vision approaches
  - Effective with sparse images coverage
  - Single-view predictions fusion



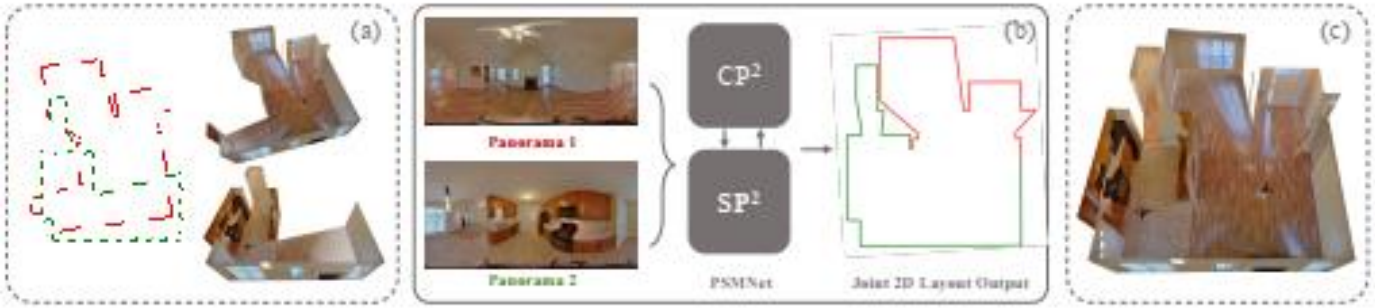
Cabral 2014: labeling propagation



360DFPE Solarte RAL 2022

Multi-view layout estimation

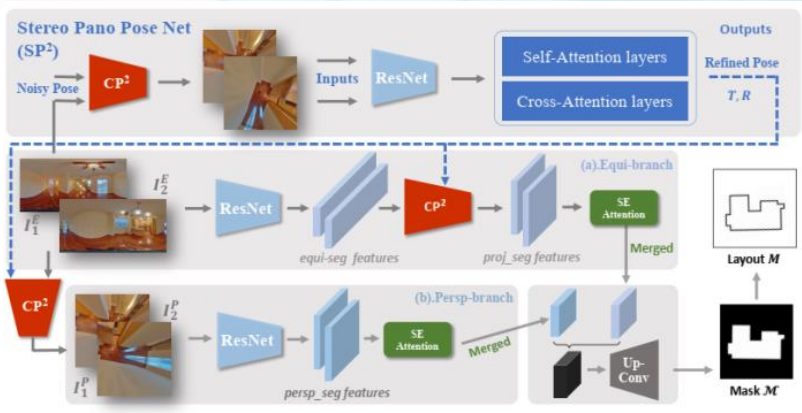
- Single image limits
  - > 10 corners – multi-purpose environments
- End-to-end joint layout-pose estimator
  - Input: pair of panoramic images
    - Usually wide baseline, noisy alignment – incomplete layouts
- NB. Single image layouts usually have different scale
  - Common using same camera height as scale factor



PSMNet - Wang CVPR *2022*

End-to-end joint layout-pose estimator 1

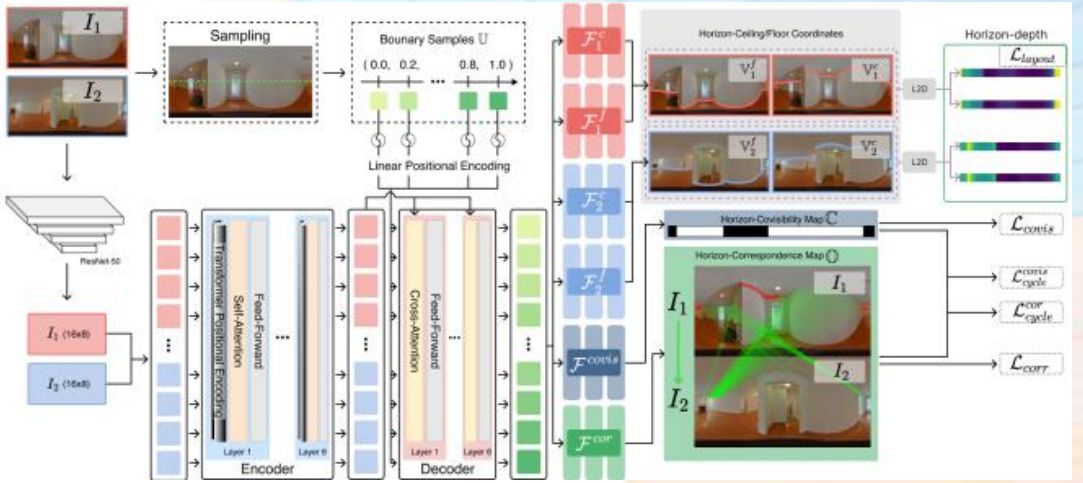
- Stereo pose network: Img1 and Img2 mutual pose
  - Computed in E2P space (AtlantaNet, Dula-Net)
- Equirectangular branch
  - ResNet on E1 and E2 to extract features
  - E2 features projected to E1 + cross-attention joining
  - E2P on equi feats: output floorplan space
- E2P branch
  - Images projection P1 and P2, P2 image projected to P1
  - ResNet on P1 and P2 to extract features
  - Cross-attention joining
- Cross attention joining: equi + E2P
- Decoding all to merged footprint mask

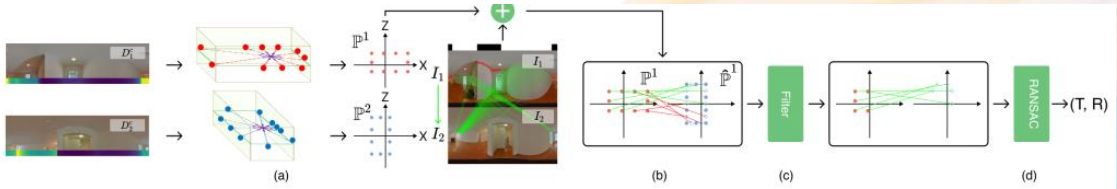


PSMNet - Wang CVPR *2022*

End-to-end joint layout-pose estimator 2

- Single equirectangular branch
  - ResNet features from I1 and I2
  - Single transformer – multihead output
    - Horizon ceiling/floor coordinates
      - 2 layouts (Led2Net – Wang CVPR2021)
    - Horizon covisibility and correspondences maps
- Geometry-aware registration
  - Covisibility and correspondences maps
  - Registration pipeline (RANSAC)
- Layout direct fusion

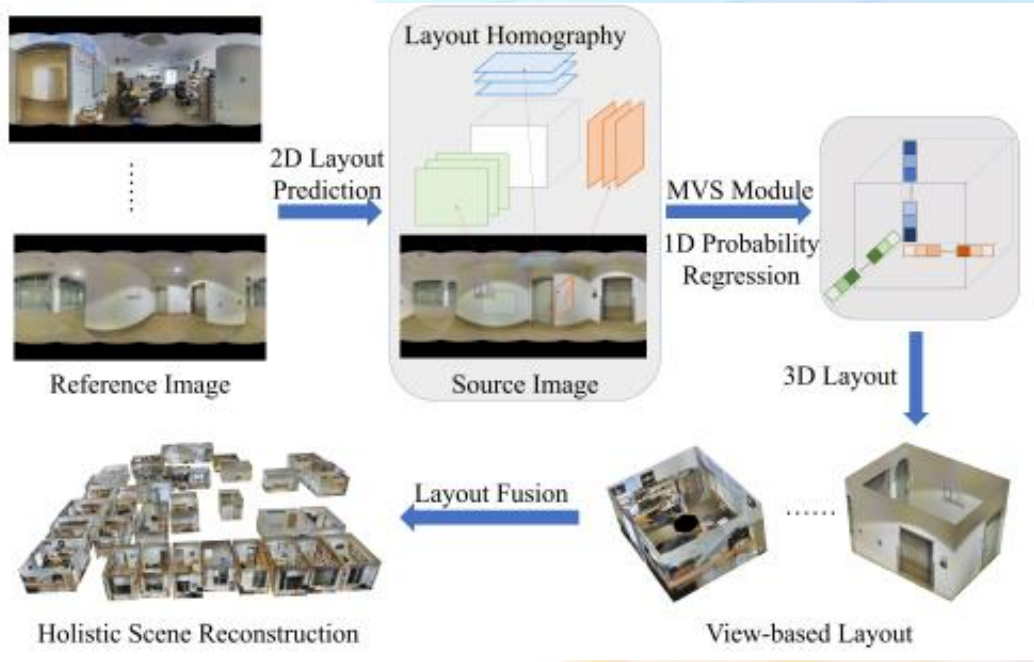




GPR-Net - Su CoRR *2022*

Multi-view layout estimation with MVS

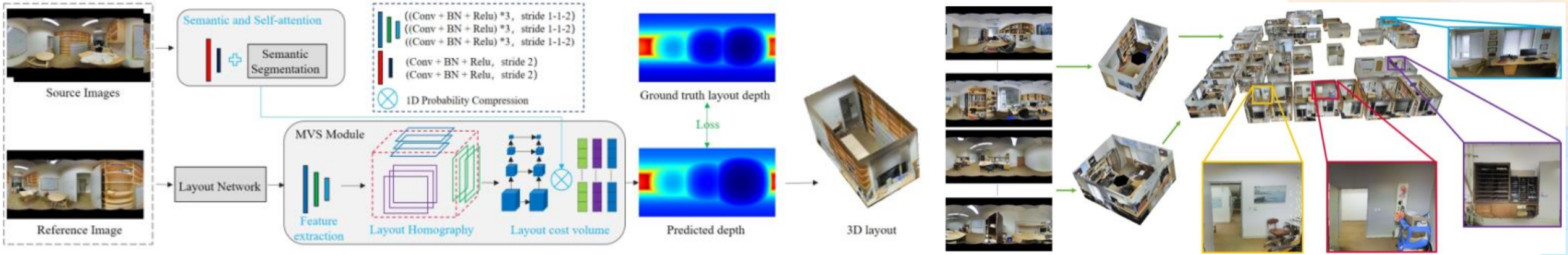
- Combining single image layout and multi-view stereo (MVS)
  - Each panorama treated as reference view with a set of associated source views
  - Layout as a set of 3D planar elements
  - Semantics and self-attention to enforce structural analysis



MVlayoutNet – Hu ACM MM2022

Multi-view layout estimation with MVS

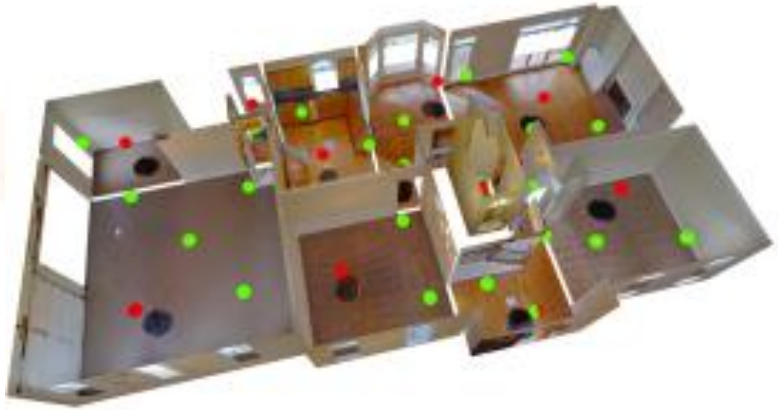
- 2D boudaries prediction for reference and source images
  - 3D elements fitting into 2D layout and aggregated as cost volumes
    - 1 D probability map for each layout element
      - Depth of the reference layout image
- Fusion on each reference room layout at the same scale



MVlayoutNet – Hu ACM MM2022

Layout estimation from sparse images

- Previous: 2<= images per room
  - Professional capture (eg. Zillow indoor dataset)
  - Easy-moderate challenge
- More common
  - Non-professional capturing
  - Very wide baseline
  - Sparse coverage
  - Hard registration and reconstruction

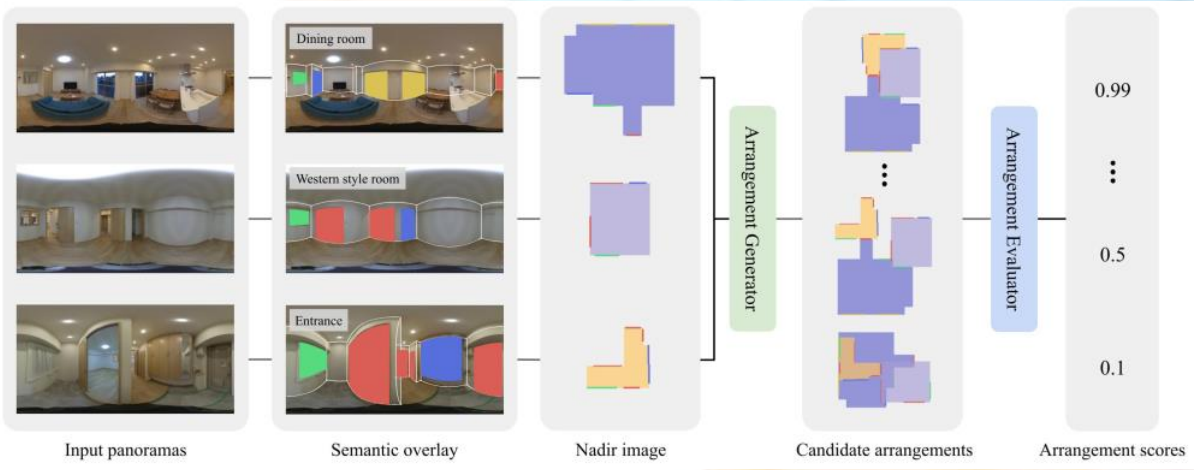


ZInD - Cruz CVPR *2022*

Shabani et al. ICCV 2021

Floorplan estimation from wide baseline

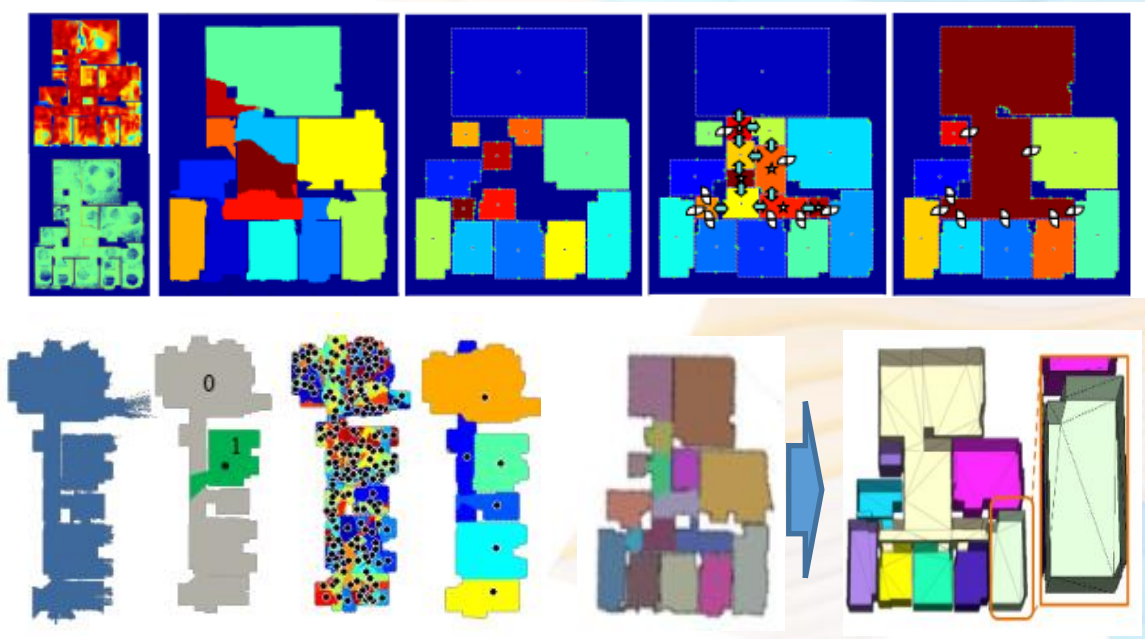
- Input: indoor panoramas with little to no visual overlaps
- Pipeline
  - For each panorama
    - Layout estimation
    - doors/windows detection
    - Top-down nadir view (256x256 16 channels)
  - Arrangement generator
    - Floorplan candidates
    - Graph of nadir images
  - Arrangement evaluator
    - Output: 2D relative camera pose for each panorama



Shabani et al. ICCV 2021

Floorplan segmentation with small baseline

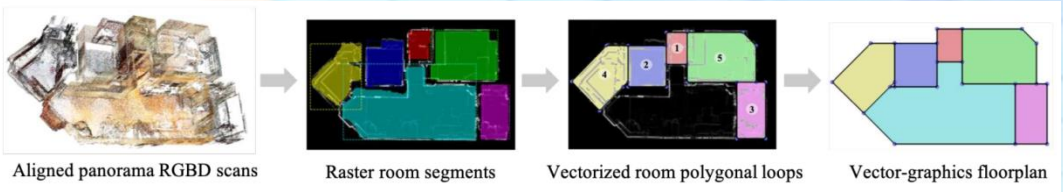
- Input: Registered RGBD panoramas-> point cloud -> density map
  - RGB+dense depth: from instruments, MWS or direct prediction
  - Smaller baseline: SfM or IPC allowed
  - Ouput: top-down maps of interior space
    - 3D rooms can be extruded
  - Base for indoor structured graph
    - Walls, objects, connections as nodes
- Early heuristic approaches
  - Room segmentation as space clustering
    - Free space evidence

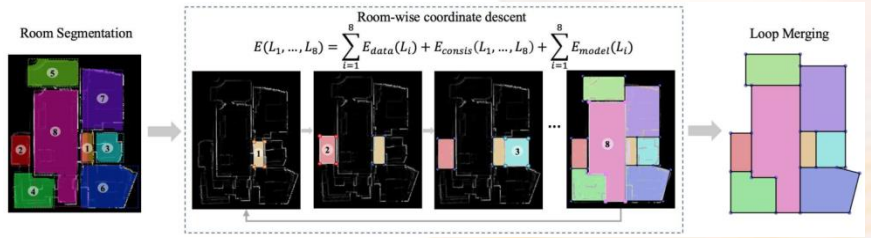


Ikeata et al. ICCV 2015

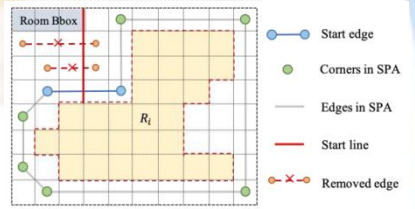
Data-driven floorplan segmentation

- Hybrid approach: example 1
  - Input: 4 channels density map
    - Density+average 3D normal
  - Instance semantic segmentation technique
    - Mask-RCNN
  - Floorplan graph inference
    - Reconstruction of multiple polygonal loops
    - Room-wise coordinate descent
  - Loop merging



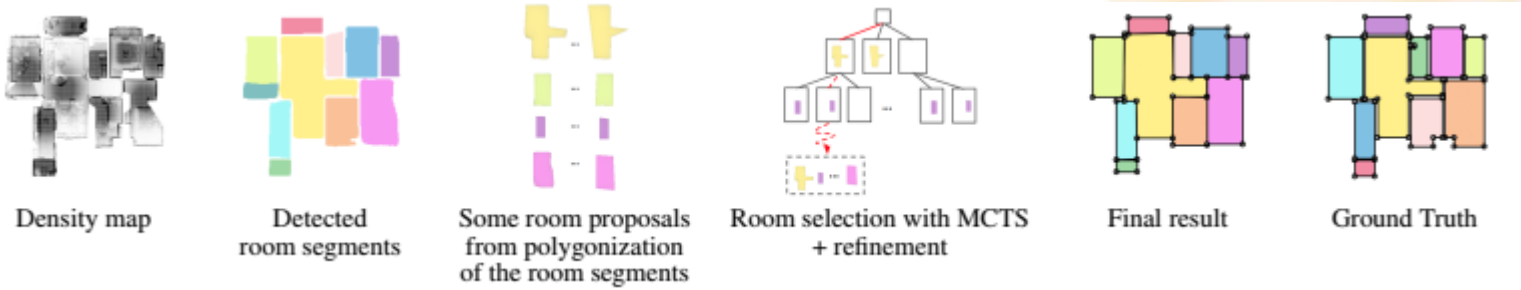


FloorSP ICCV 2019



Data-driven floorplan segmentation

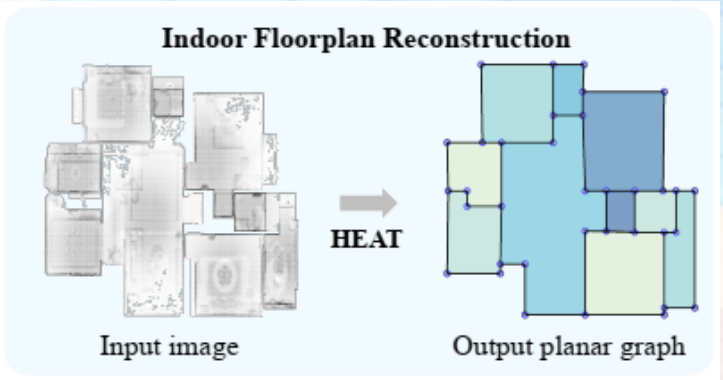
- Hybrid approach: example 2
  - Input: Mask-RCNN room proposals
  - Room shapes jointly while adjusting their locations
    - Monte Carlo Tree Search (MCTS) algorithm
      - guided by a learned scoring function
        - Density map and proposed shape image
  - Differentiable refinement step

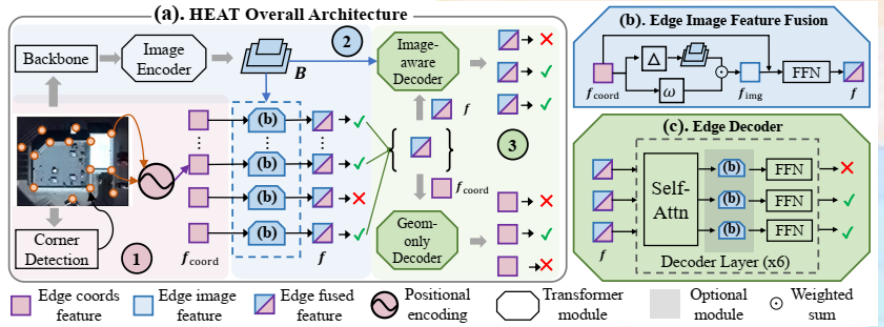


MonteFloor ICCV 2021

Data-driven floorplan segmentation

- Fully data-driven
  - Multi-stage pipelines
  - Es. Holistic edge attention transformer (HEAT)
    - Input: density map (same of MonteFloor, etc.)
      - i.e., **3D points accumulate on vertical walls, etc.**
    - DETR corner detector
      - Edges are nodes
    - 64x64 feature candidates-> 256x256confidence map
    - Transformers
    - End-to-end training data generated on the fly
      - From detected edges vs. GT
      - Output: floorplan edges

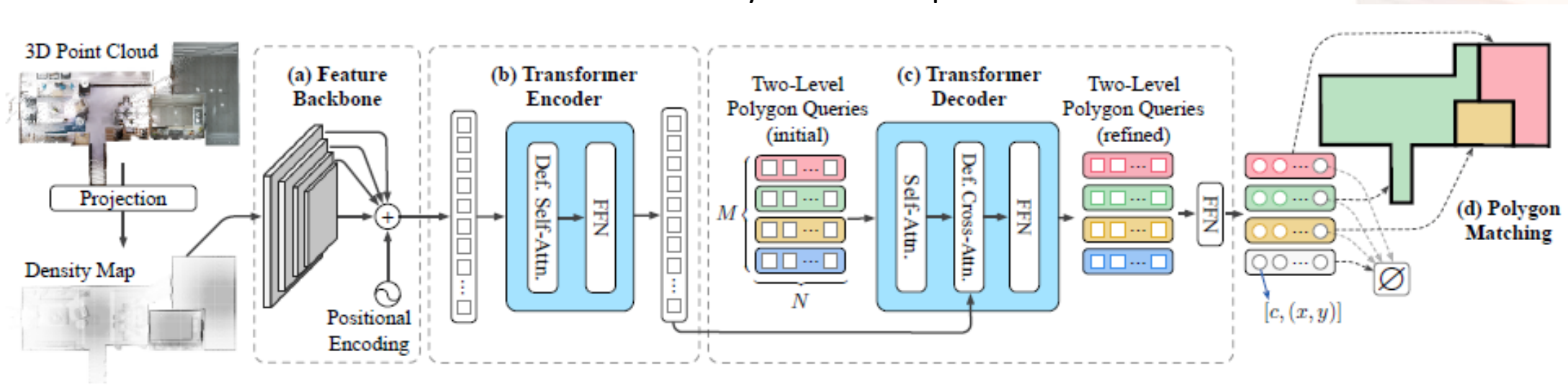




HEAT CVPR 2021

Data-driven floorplan segmentation

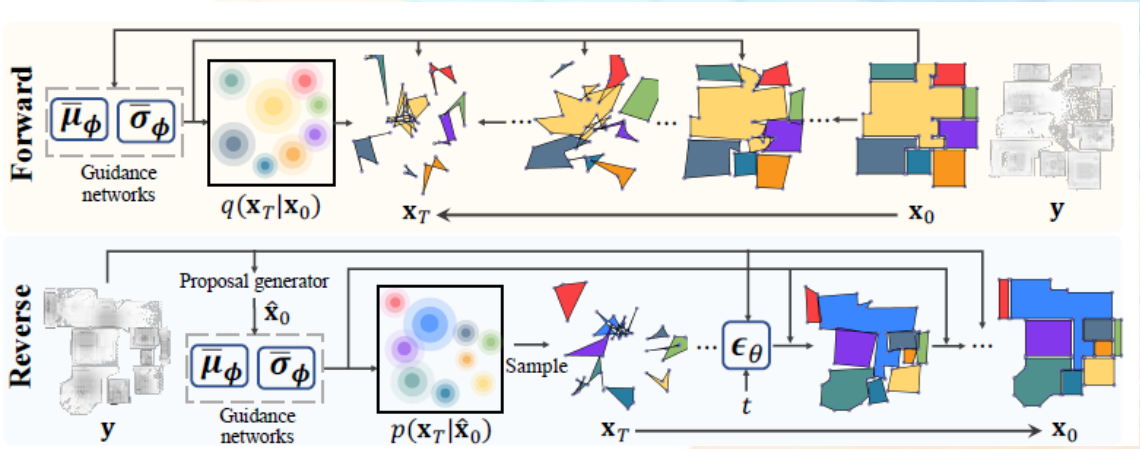
- End-to-end pipelines
  - directly maps a density image to a set of room polygons
    - E.g., RoomFormer, SLIBO-Net
    - Two-levels queries formulation
      - Fixed size: M rooms – N corners – binary mask to map valid room and corners

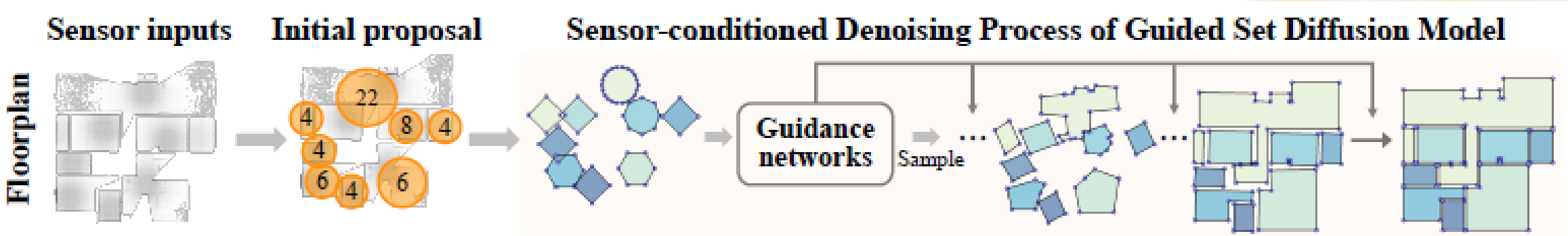


RoomFormer CVPR 2023

Data-driven floorplan segmentation

- Diffusion-based approaches
  - PolyDiffuse
    - Polygonal shape reconstruction via Guided Set Diffusion Models (GS-DM)
    - RoomFormer as proposal and denoising network
    - Guidance network does not use images
    - Basically used as refinement

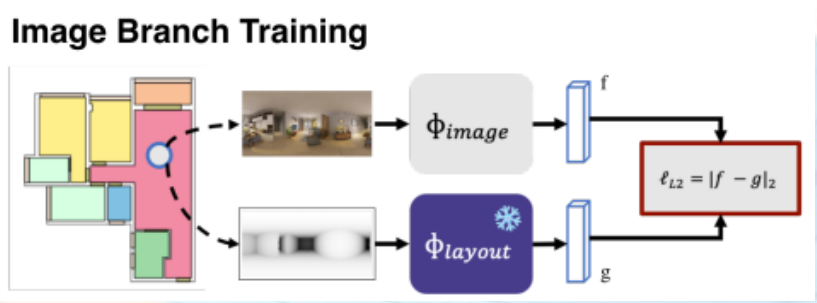


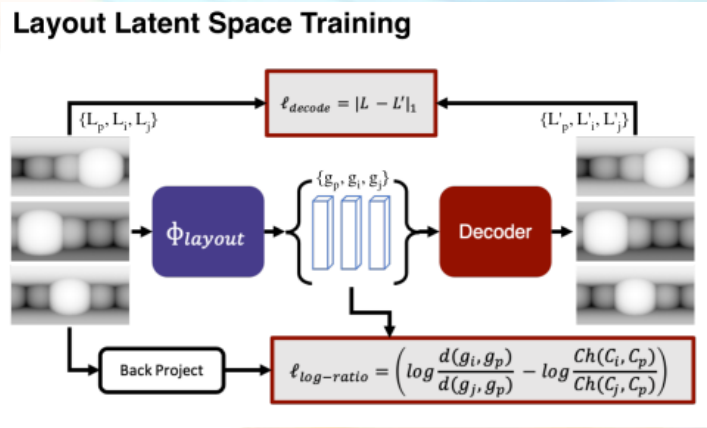


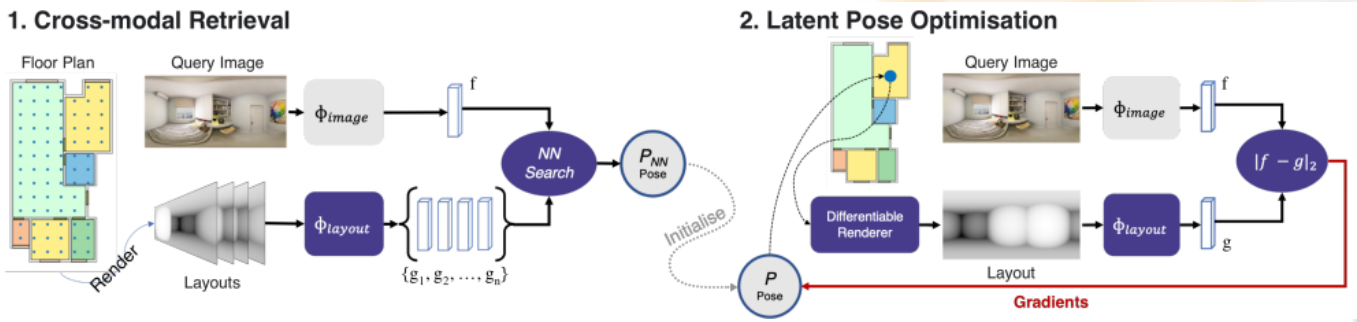
PolyDiffuse NeurIPS *2024*

Reconstruction and localization

- Enhanced panoramic image integration
  - Exploiting latent features
- Hybrid example 1: align the floor plan to a panorama
  - 2D sampled positions – rotation in assumed known
  - 3D floor plan extrusion
  - rendering of 3D rooms as panoramic layout
  - Floorplan latent representation
  - Single image latent representation
  - NN search and refinement



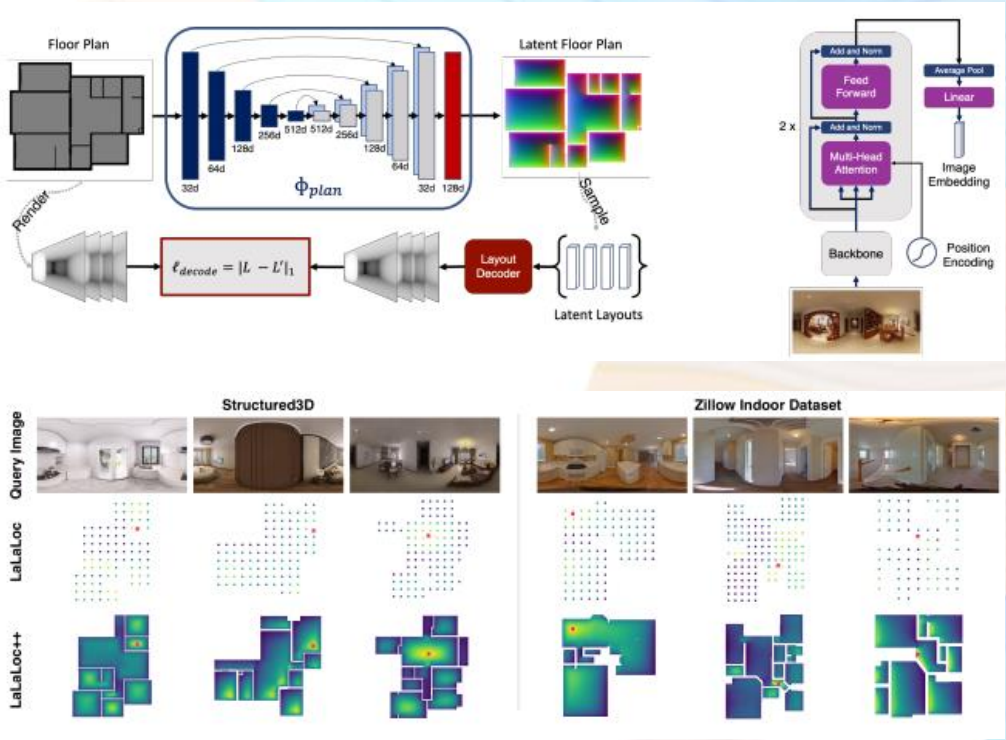




LaLaLoc ICCV 2021

Reconstruction and localization

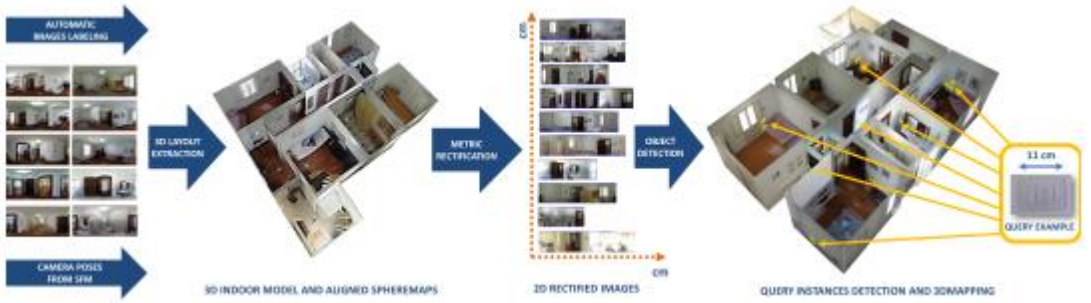
- Full data-drive example 2: LaLaLoc++
  - Full data-driven
  - Latent floorplan instead of individual rooms
    - Rendering only to train latent floorplan estimator
    - Prediction directly in latent space
    - Recovered position not only in a fixed grid
  - Shared latent space between image and floorplan
    - image layout similar to latent floorplan sampled layouts
  - Gradient refinement for sub-pixel refinement
  - Rotation can be estimated



LaLaLoc++ ECCV 2022

3D floorplan reconstruction

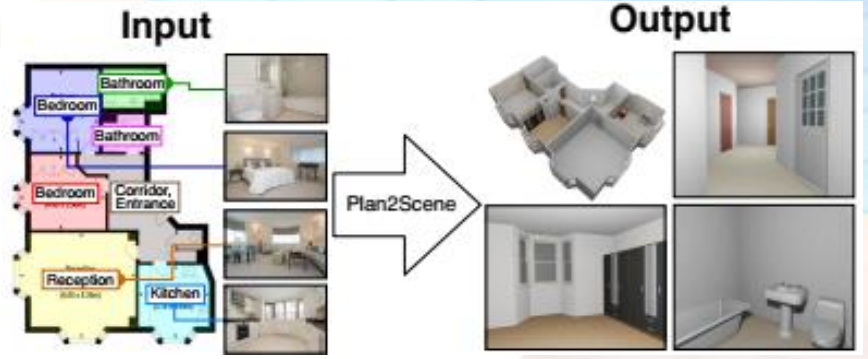
- Input: single room or floorplan layout and associated panoramas
- Early approaches
  - SfM + super-pixels + geometric reasoning
    - Recovering floorplan and registered images
  - Simple texturing by splatting input images
  - Problems
    - Low adaptability and robustness
    - Cluttered images are splatted on walls
      - Many visual artifacts

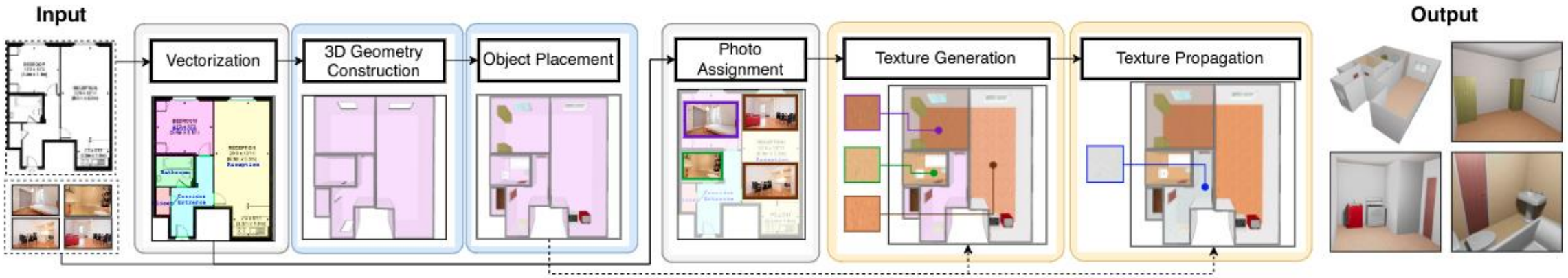


Pintore, Pintus, Ganovelli, Scopigno, Gobbetti. CAG 2018

3D floorplan reconstruction

- Data-driven solutions
  - Floorplan to 3D scene becomes a specific task
  - Layout from reconstruction or CAD blueprint
  - Objects from recovery or CAD atlas
  - Photo assignment: fine alignment not necessary

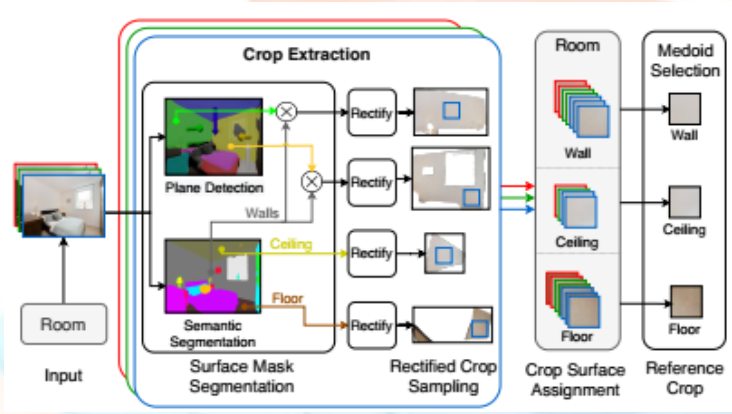


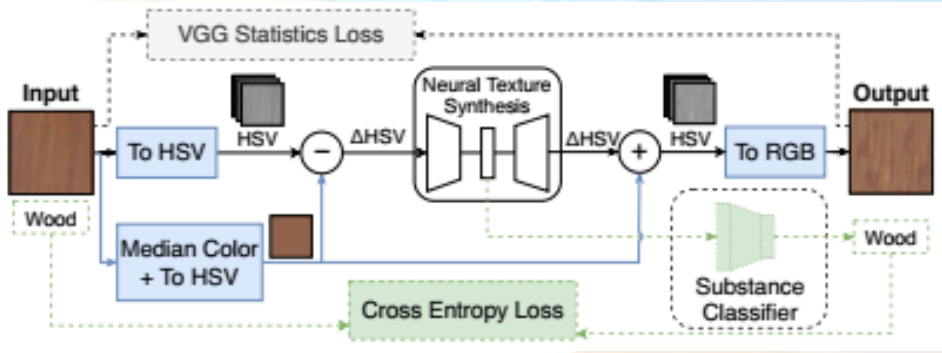


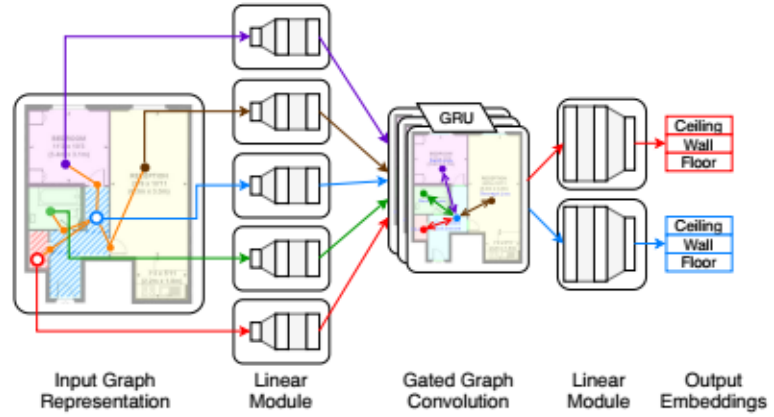
Plan2Scene CVPR 2021

3D floorplan reconstruction

- Main focus on:
  - Texture generation for observed surfaces
    - Semantic matching
    - Encoder-decoder network for synthesis
      - stationary statistics
  - Texture propagation for unobserved surfaces
    - Occlusion or missing images
    - Room-door-room connectivity to propagate
      - GCN network
        - rooms are nodes and edges are doors



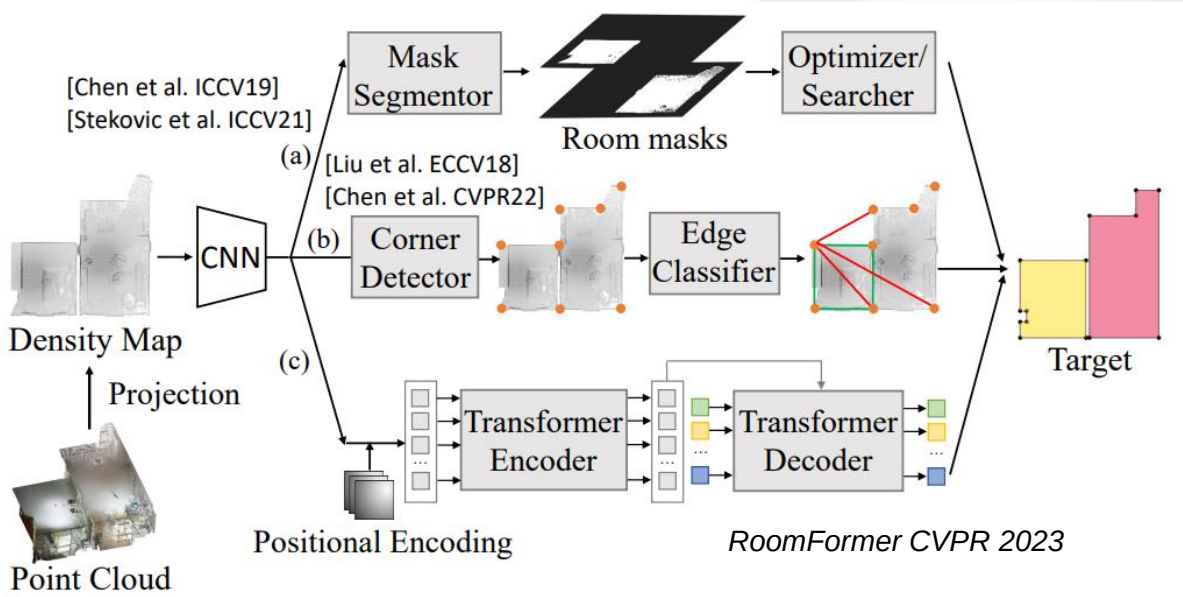




Plan2Scene CVPR 2021

Example of floorplan reconstruction pipeline

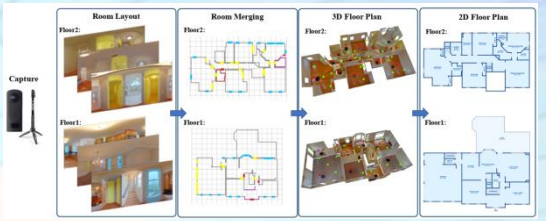
- Input: a set of equirectangular images
  - Dense coverage -> point cloud from SfM -> **density map**
  - Sparse coverage -> registration + predicted depths -> **density map**
    - Opz: composition of individual rooms layout
- Core task: floormap segmentation
- Output: multi-room segmentation
  - **Set of 2D polygons**
  - Opz: heights of floors and ceilings



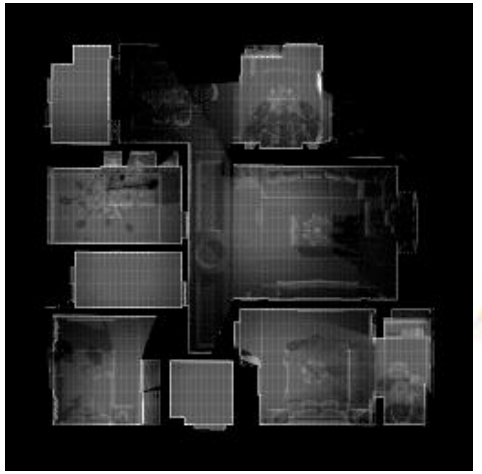
Floorplan reconstruction pipeline 1/4

- Build a floormap from images
  - Dense coverage
    - Positions from SfM
    - Densified depth maps
  - Sparse coverage
    - Positions
      - Professional instruments – ZIndoor, etc.
      - Extreme SfM – ICCV2021, etc.
    - **Predicted mono depth maps (see previous talk)**
- Floormap example: 2D occurrences map
  - 3D points accumulation on vertical walls
  - Scaled to a fixed-size map

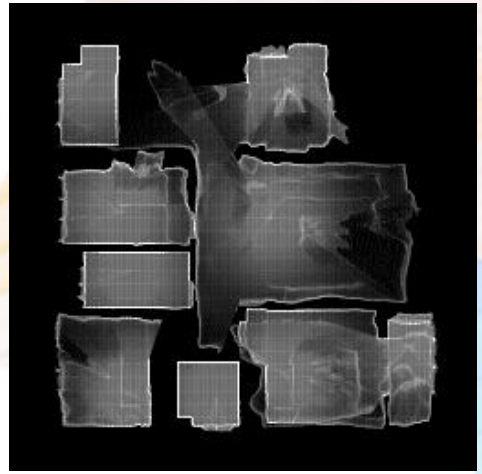
equirectangular images + RT



ZInD – Cruz CVPR2021



Dense coverage

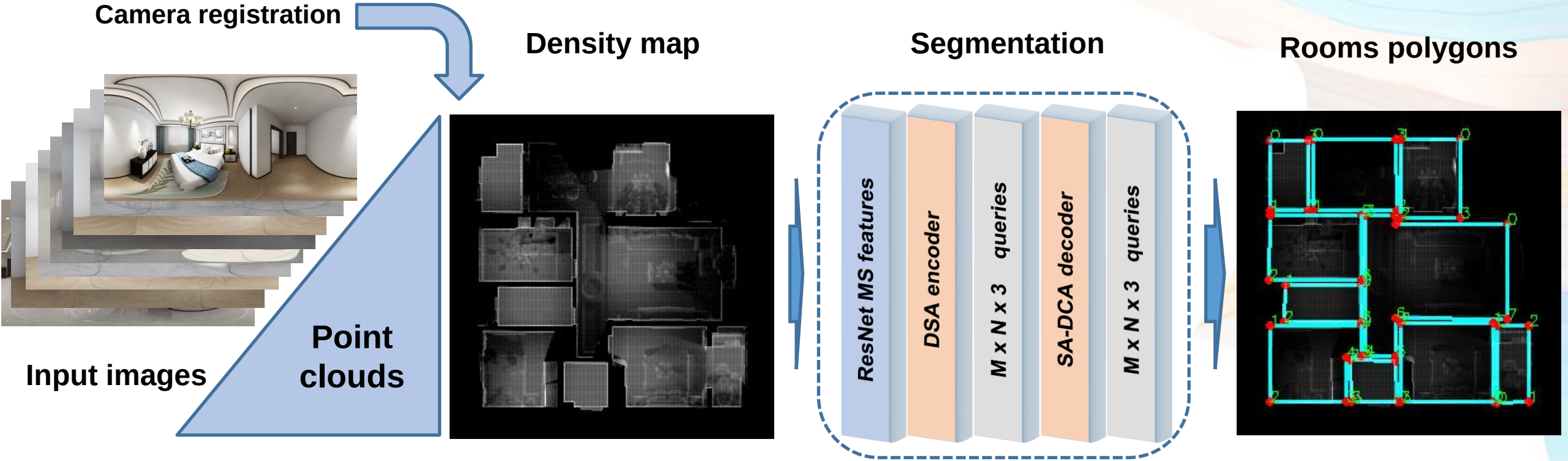


Sparse-predicted coverage

Floorplan reconstruction pipeline 2/4

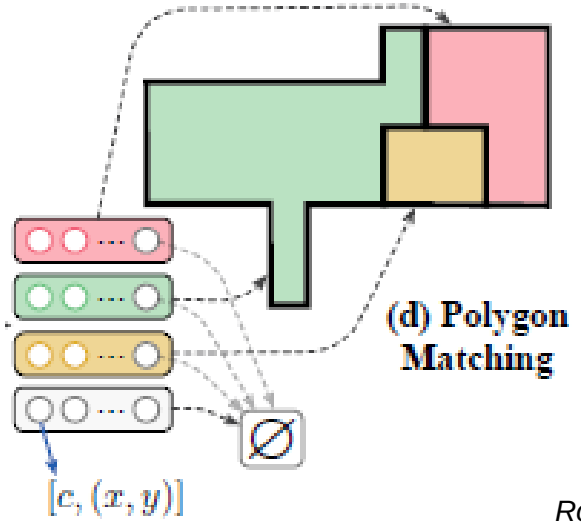
- Floormap segmentation
  - MxNx2 (2D corners) + MxNx1 (valid rooms-corners)

Implementation available at:
https://github.com/ywyue/RoomFormer



Floorplan reconstruction pipeline 3/4

- Supervised training
  - Density maps annotated with 2D polygons
  - Hungarian matching



Match the fixed-number predictions with the arbitrary-number ground truth

$$\hat{\sigma} = \sum_{m=1}^{M} \arg\min_{\sigma} \mathcal{D}(V_m, \hat{V}_{\sigma(m)}) \qquad \mathcal{D}(V_m, \hat{V}_{\sigma(m)}) = \lambda_{\text{cls}} \mathcal{D}(c, \hat{c}) + \lambda_{\text{coord}} \mathcal{D}(p, \hat{p})$$

Loss: $$\mathcal{L} = \sum_{m=1}^{M} (\lambda_{\text{cls}} \mathcal{L}_{\text{cls}}^{m} + \lambda_{\text{coord}} \mathcal{L}_{\text{coord}}^{m} + \lambda_{\text{ras}} \mathcal{L}_{\text{ras}}^{m})$$

RoomFormer CVPR 2023

Floorplan reconstruction pipeline 4/4

Integrated indoor model: summary

- Target: permanent structure representation
  - Multi-view layout estimation
  - Multi-room segmentation
  - 3D scene reconstruction
- Open problems
  - Reconstruction with minimal number of images is common, but hard
  - Multi-room scenes are recovered by state-of-the-art solutions, but still limited by heavy priors
  - Multi-story buildings, pillars, stairs do not fit well with current layout approaches
  - 3D structural models must be enriched with geometric details or photorealism for a number of applications -> next session!

HoHoNet - Sun CVPR2021

NEXT SESSION:
VISUAL REPRESENTATION GENERATION AND EXPLORATION