# Point Rend : Image Segmentation as Rendering

(2020, Facebook AI Research)

Smart Convergence

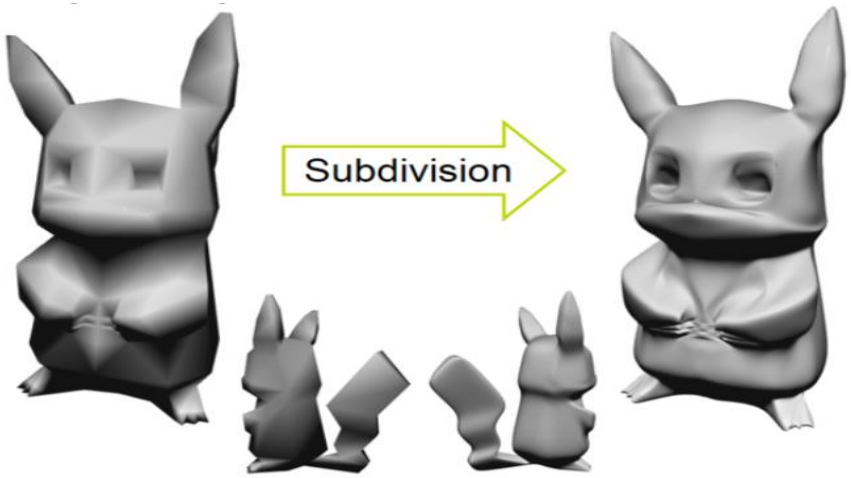
Moon Goo Young

고려대학교 KOREA UNIVERSITY

# Rendering

‘to cause (someone or something) to be in a specified condition’

**3D → 2D Rendering**




**Coarse → Fine Rendering**

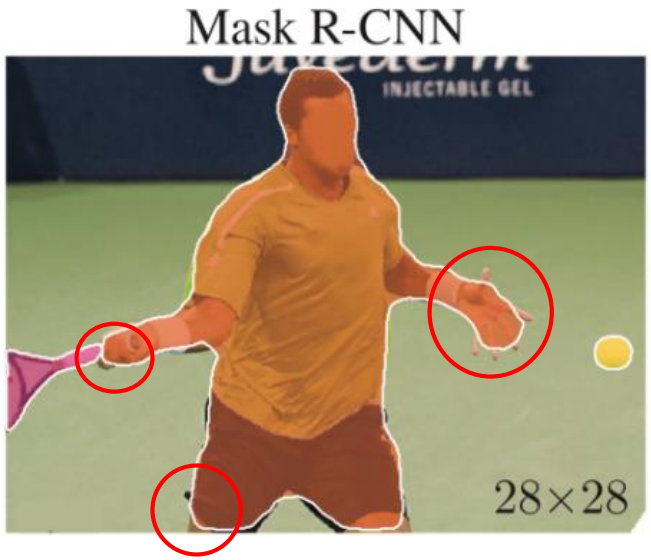
**Rendering ↔ Image Synthesis**

# Limitation of Mask R CNN

## 1. Uniform representation

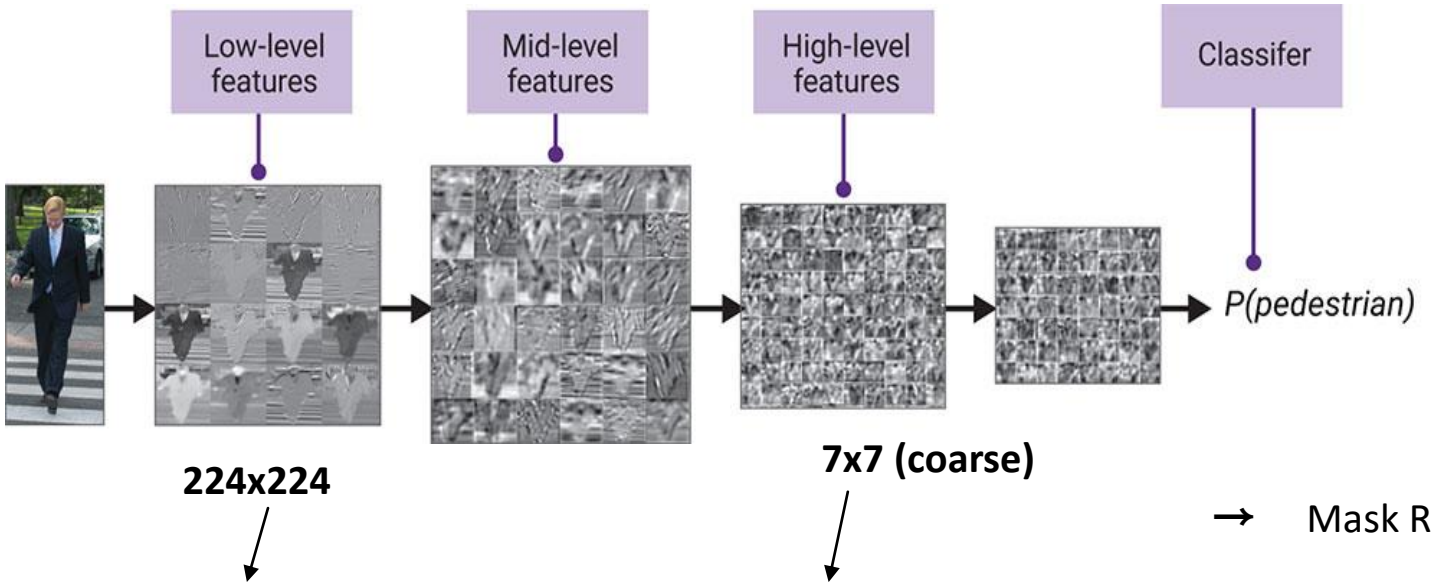
- ‘Mapping pixels sampled on a **regular grid** to a label map, or a set of label maps on the same grid.’

  → computation is allocated uniformly to all pixels (uniform representation)


- Regular grid will unnecessarily over-sample the smooth areas (easy tasks) while simultaneously under-sampling object boundaries (hard tasks).

## 2. Prediction mask with fixed size (28x28)




Fine detail but hard to compute

Easy to compute but lacks in detail

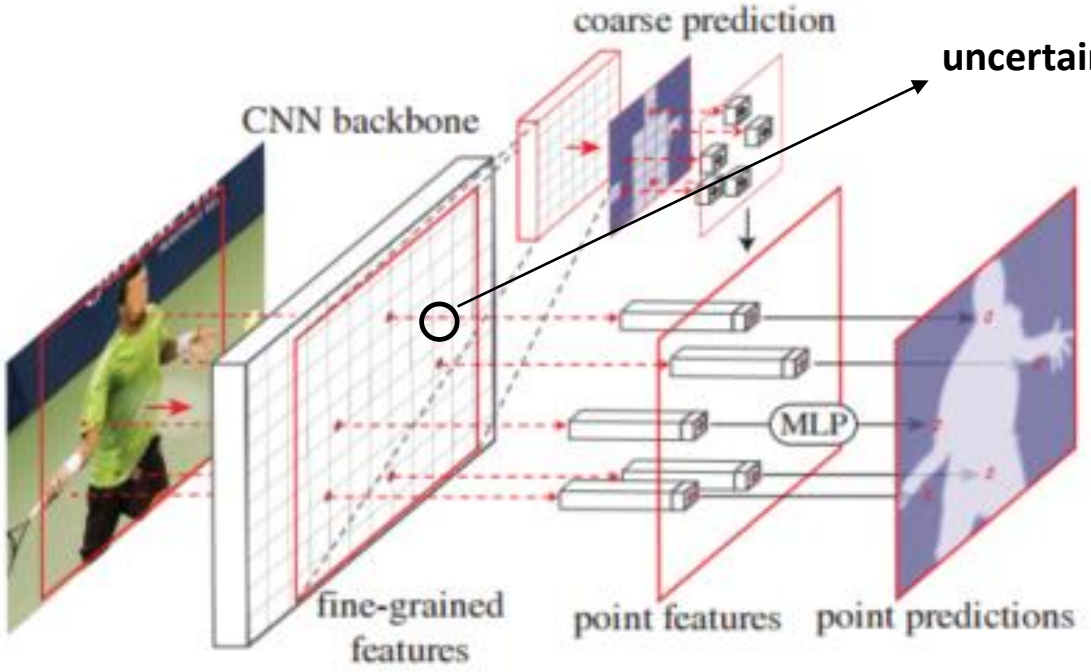
→ Mask R CNN use 28x28 feature **maps irrespective of object size**

- For large objects it produces undesirable blobby output that over-smooths the fine-level details of large objects.

# Point Rend

## Point-based Segmentation

- **Point based** segmentation predicts at **adaptively selected locations** based on an **iterative subdivision algorithm** (rendering algorithm).




**uncertain pixels** in the semantic split (hard task)

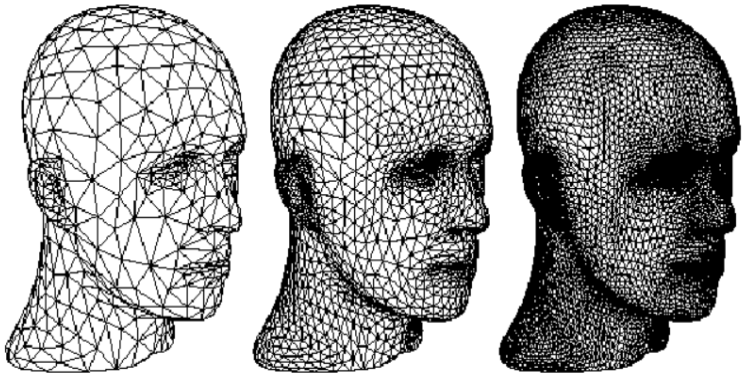
1. Yields a coarse(7x7) mask prediction for each detected object
2. PointRend selects a set of points and make prediction
3. MLP uses **interpolated features** to make final predictions

   → coarse prediction mask + fine-grained features

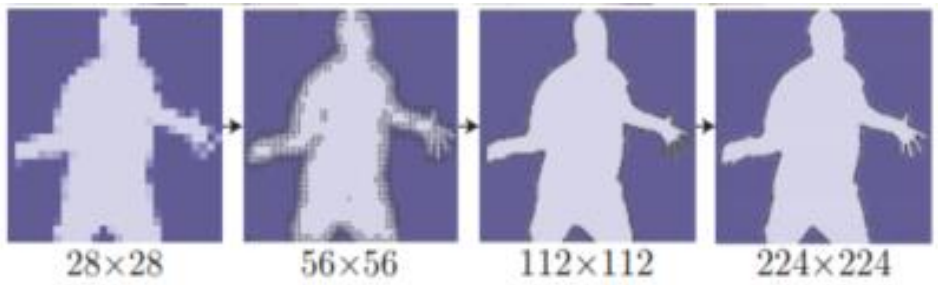
# Point Rend

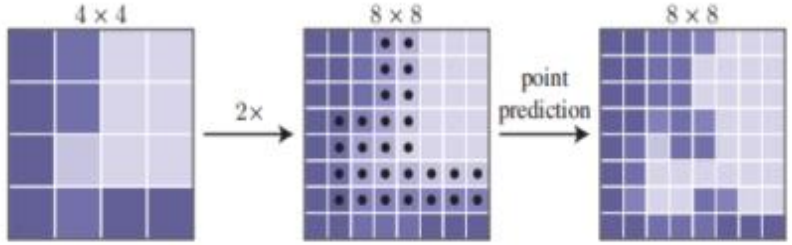
- **Image segmentation as rendering problem**

  → **Adaptive Subdivision** of C.G rendering  (coarse → fine-grained)



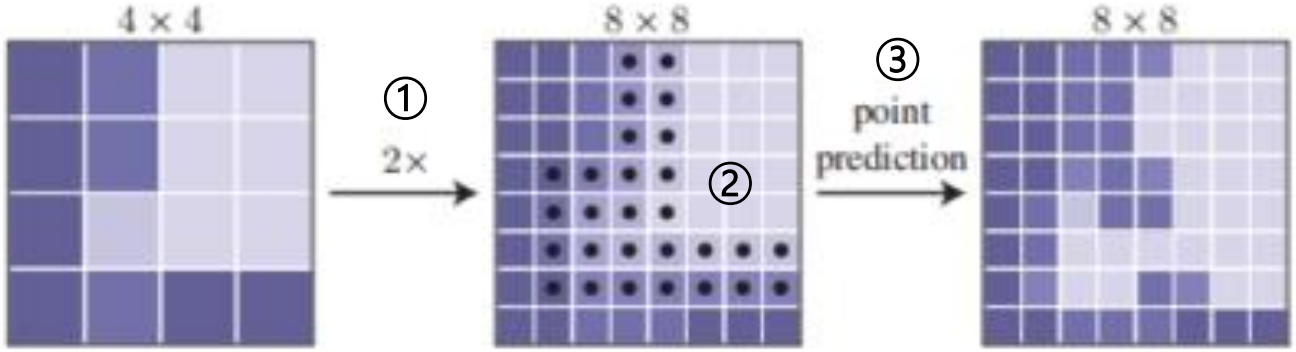
**Subdivision Rendering (C.G)**




**1 adaptive subdivision step**

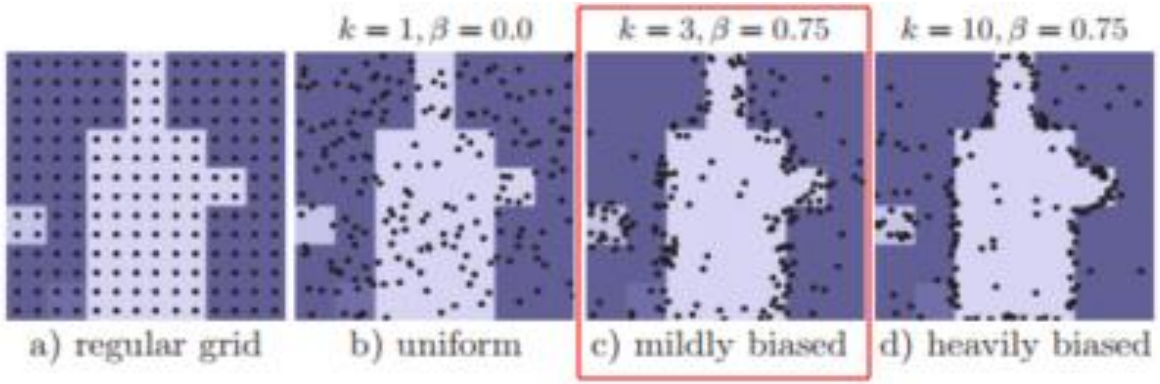
PointRend gradually increases resolution by making predictions for the most uncertain points.




① Low resolution prediction (coarse mask) is up-sampled with bilinear interpolation

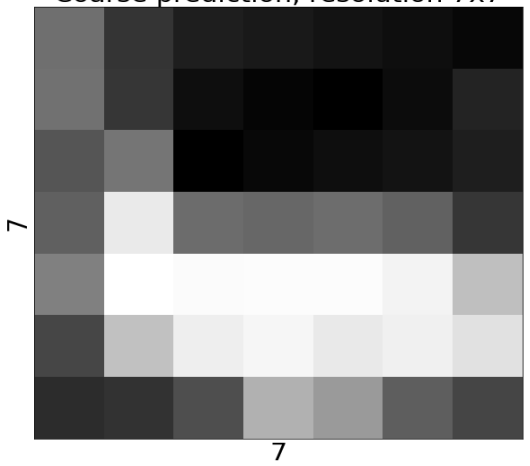
② The subset of the most uncertain points are adaptively selected.

③ Prediction for each selected point is refined using a lightweight MLP



Input

Coarse prediction, resolution 7x7

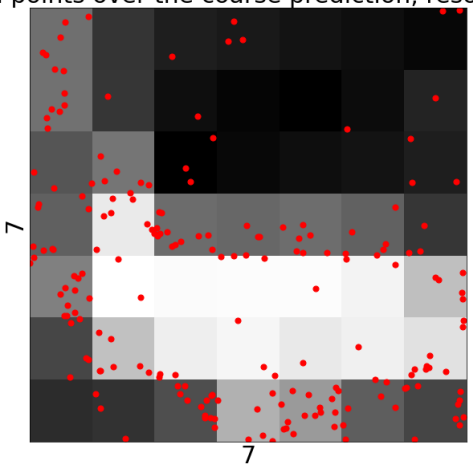



Coarse mask

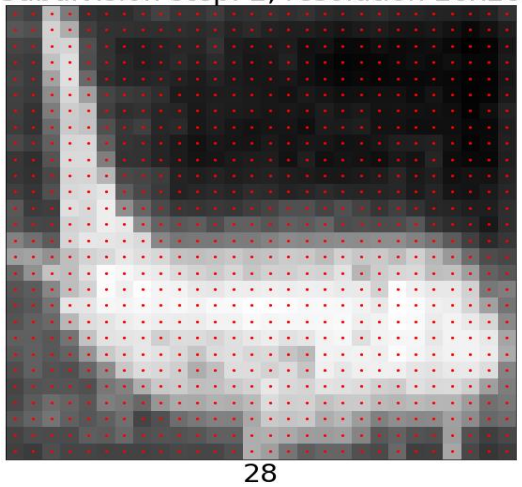
→

Point Sampling

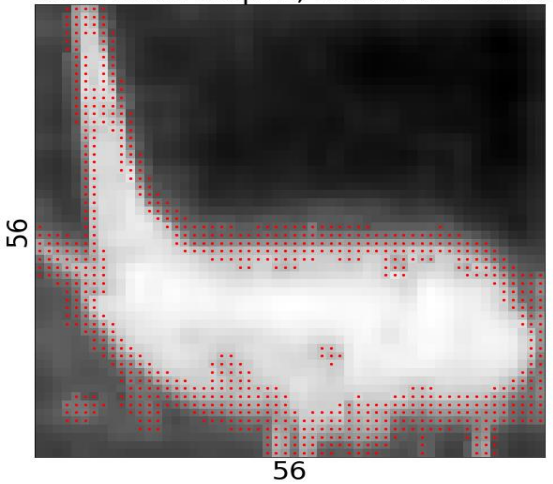
Sampled points over the coarse prediction, resolution 7x7



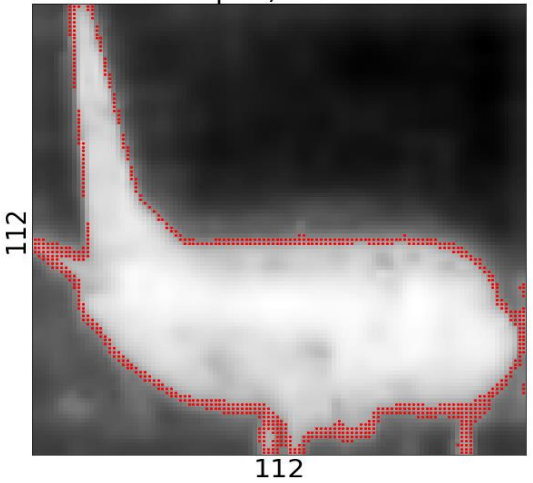

Subdivision step: 2, resolution 28x28



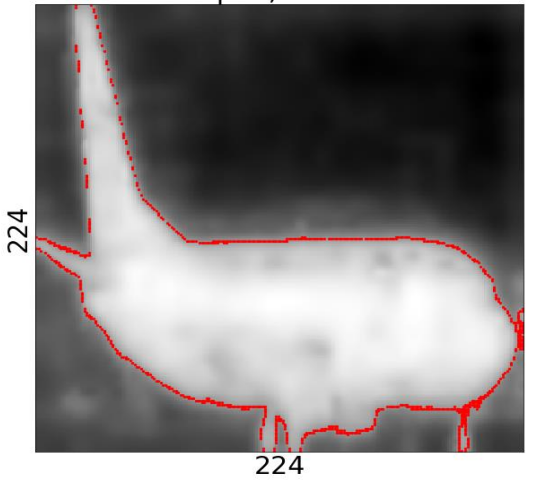

Subdivision step: 3, resolution 56x56




Subdivision step: 4, resolution 112x11




Subdivision step: 5, resolution 224x224




- **4 adaptive subdivision step → iteratively up-sampled & recover detail on the finer grid**