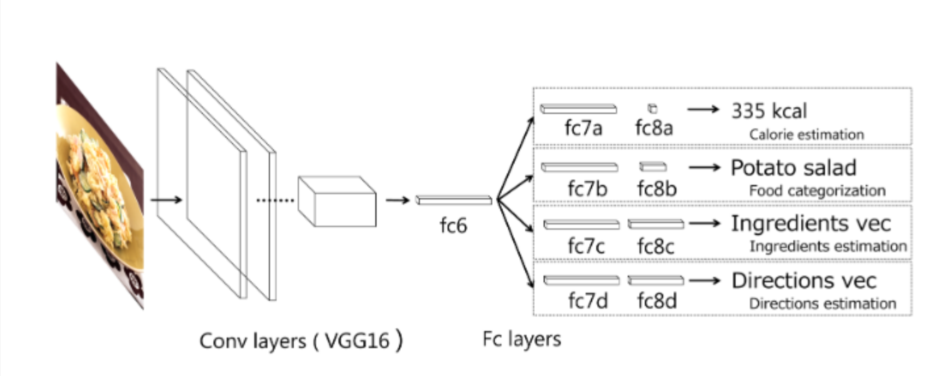
Hungry Networks: 3D Mesh Reconstruction of a Dish and a Plate from a Single Dish Image for Estimating Food Volume

Shu Naritomi, Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

Introduction

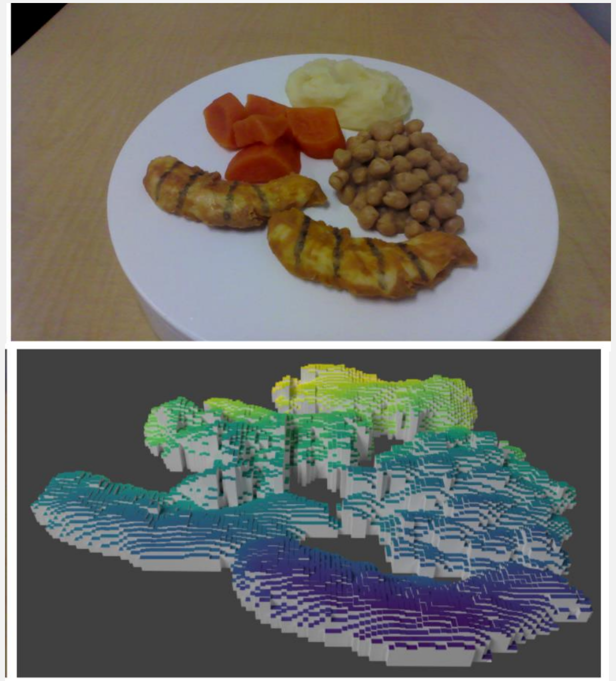
- Dietary calorie management has been an important topic.
- There is a lot of research on calorie estimation in the multimedia community.

2D based



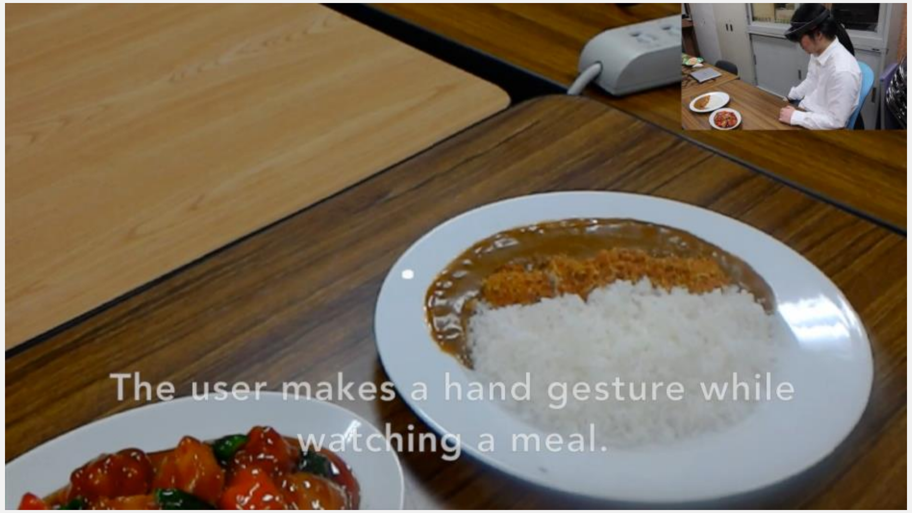
[Ege et al., IEICE2018]

Depth based



[Im2Calories, ICCV 2015]

Sensor based

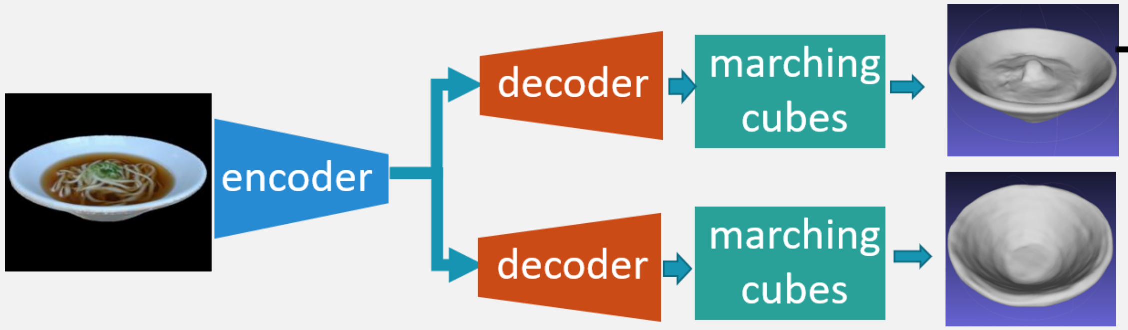


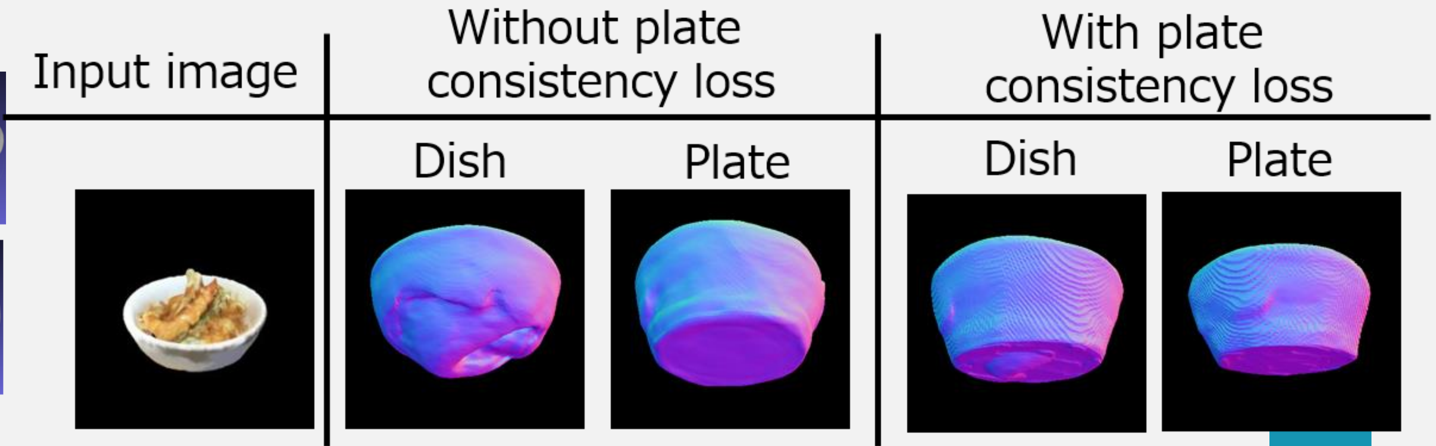
[CalorieCaptorGlass, IEEE VR 2020]

Introduction

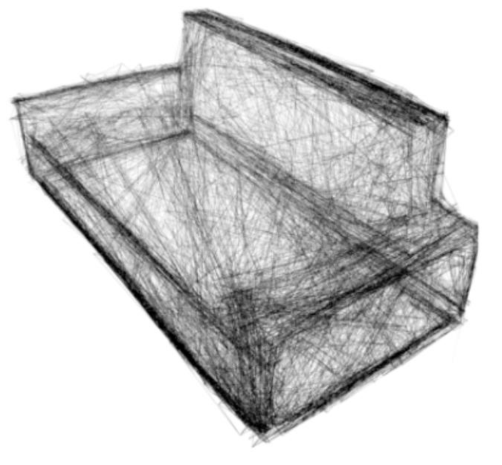
- Reconstruct 3D dish (food + plate) volume and 3D plate volume from a single dish image

- Achieve consistency between the plate part of the two reconstructed volumes introducing plate consistency loss.



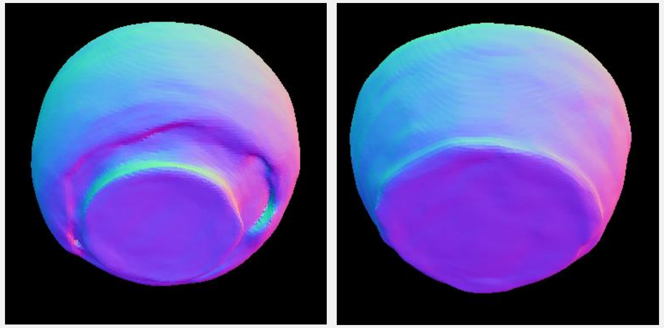


Appropriate 3D representation



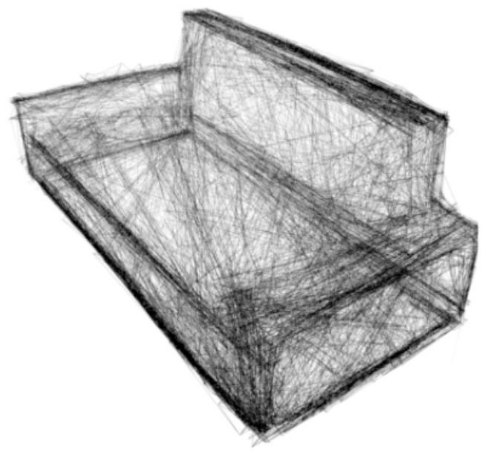
Self-intersection [Mesh R-CNN, ICCV2019]

- Purpose: estimate the food volume.
- Desired features.
  - The volume can be easily obtained.
  - Matching plate part shape of dish and plate.



The shapes of the dishes do not match.

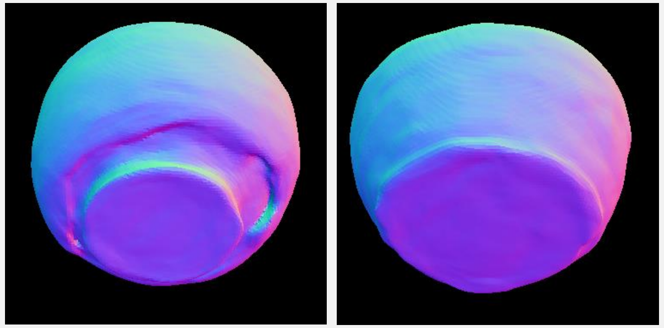
Appropriate 3D representation



Self-intersection [Mesh R-CNN, ICCV2019]

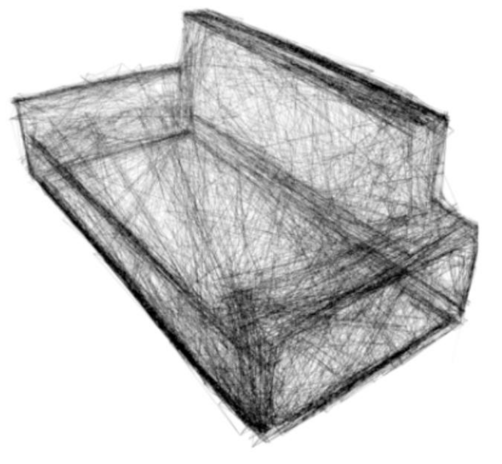
- Purpose: estimate the food volume.
- Desired features.
  - The volume can be easily obtained.
  - Matching plate part shape of dish and plate.

No Self-intersection
&
Watertight Mesh



The shapes of the dishes do not match.

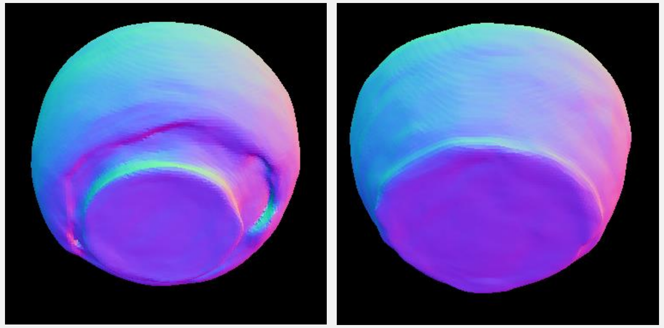
Appropriate 3D representation



Self-intersection [Mesh R-CNN, ICCV2019]

- Purpose: estimate the food volume.
- Desired features.
  - The volume can be easily obtained.
  - Matching plate part shape of dish and plate.

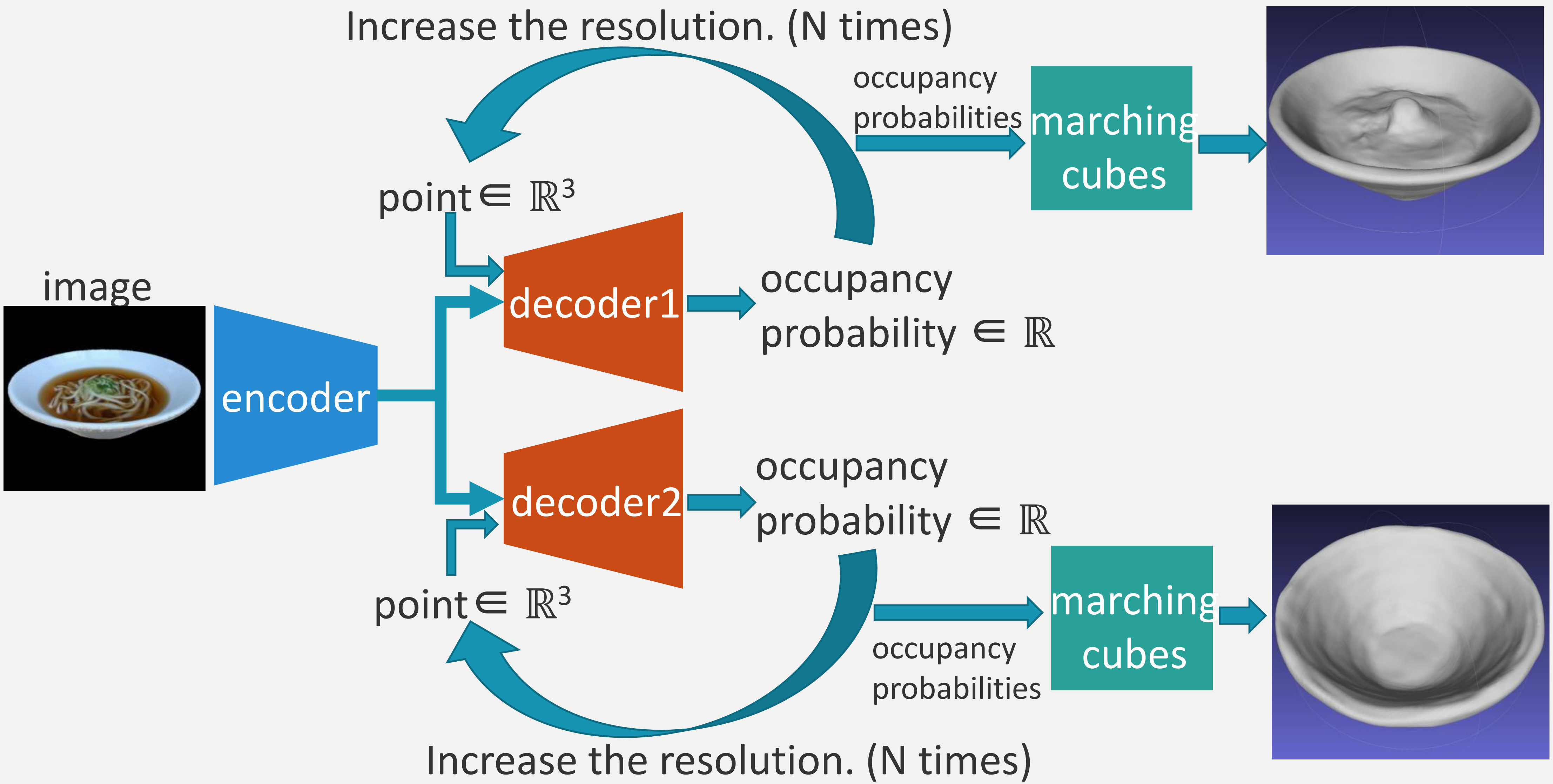
No Self-intersection
&
Watertight Mesh



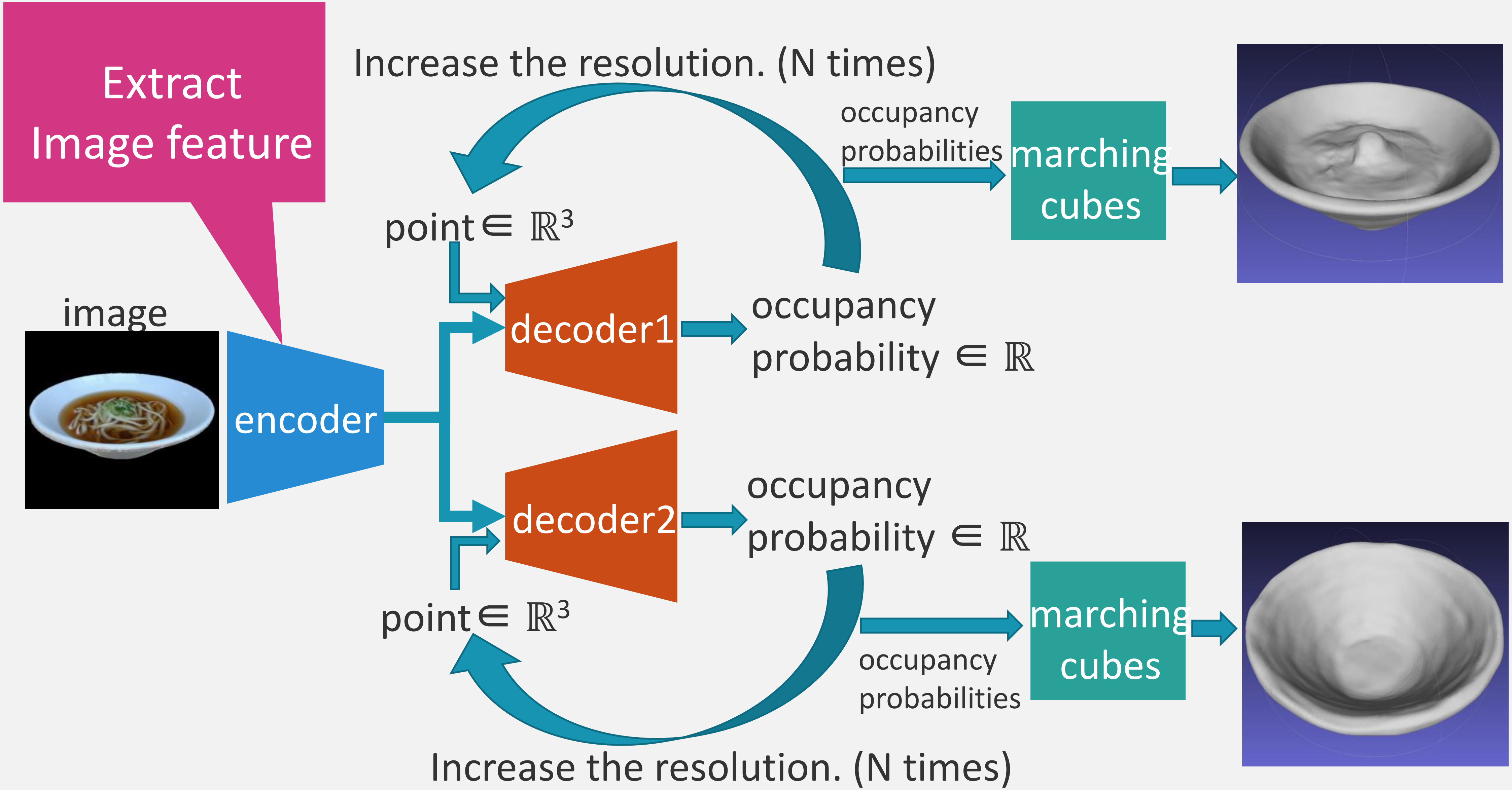
The shapes of the dishes
do not match.

Occupancy representation
is reasonable

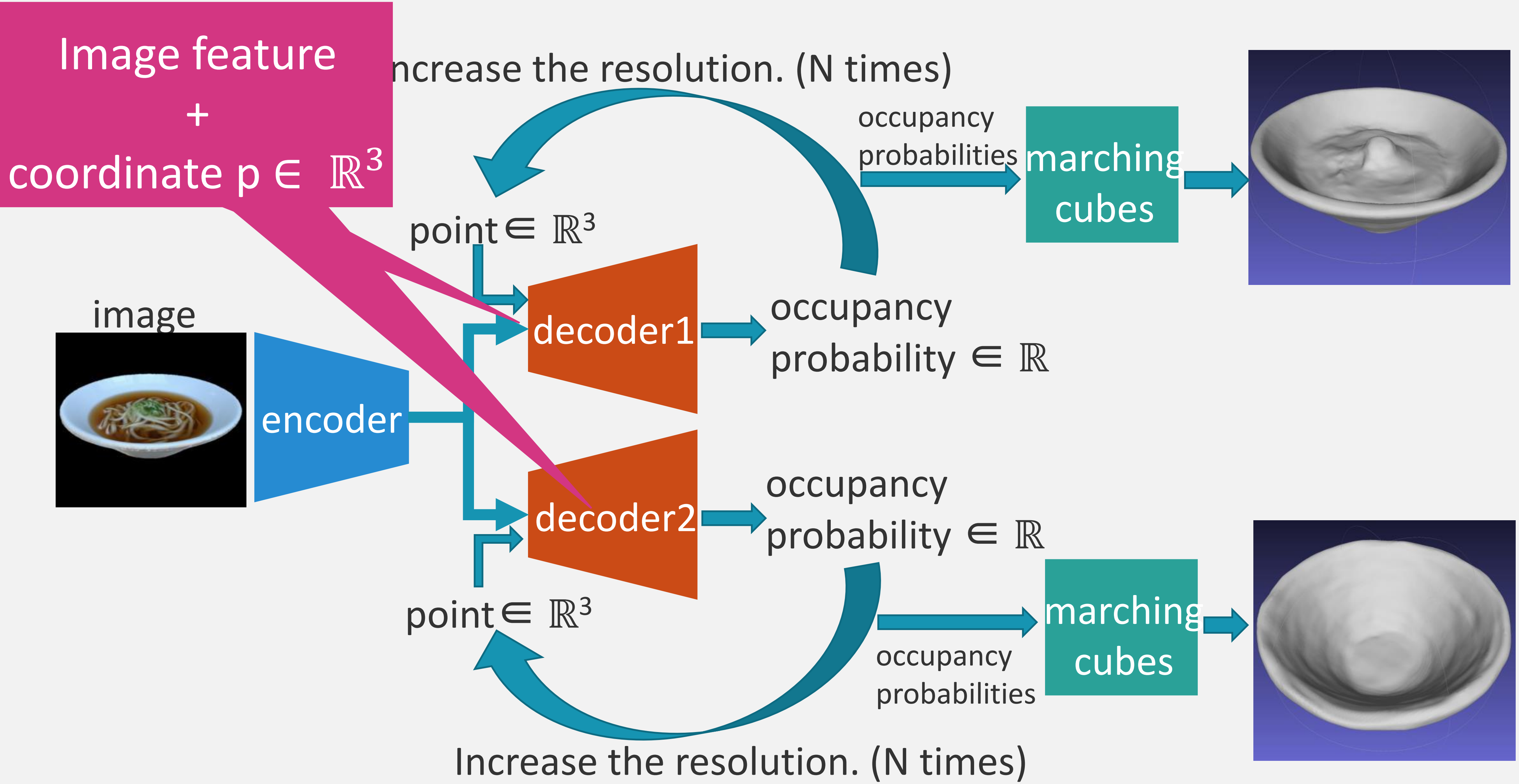
Hungry Networks : inference



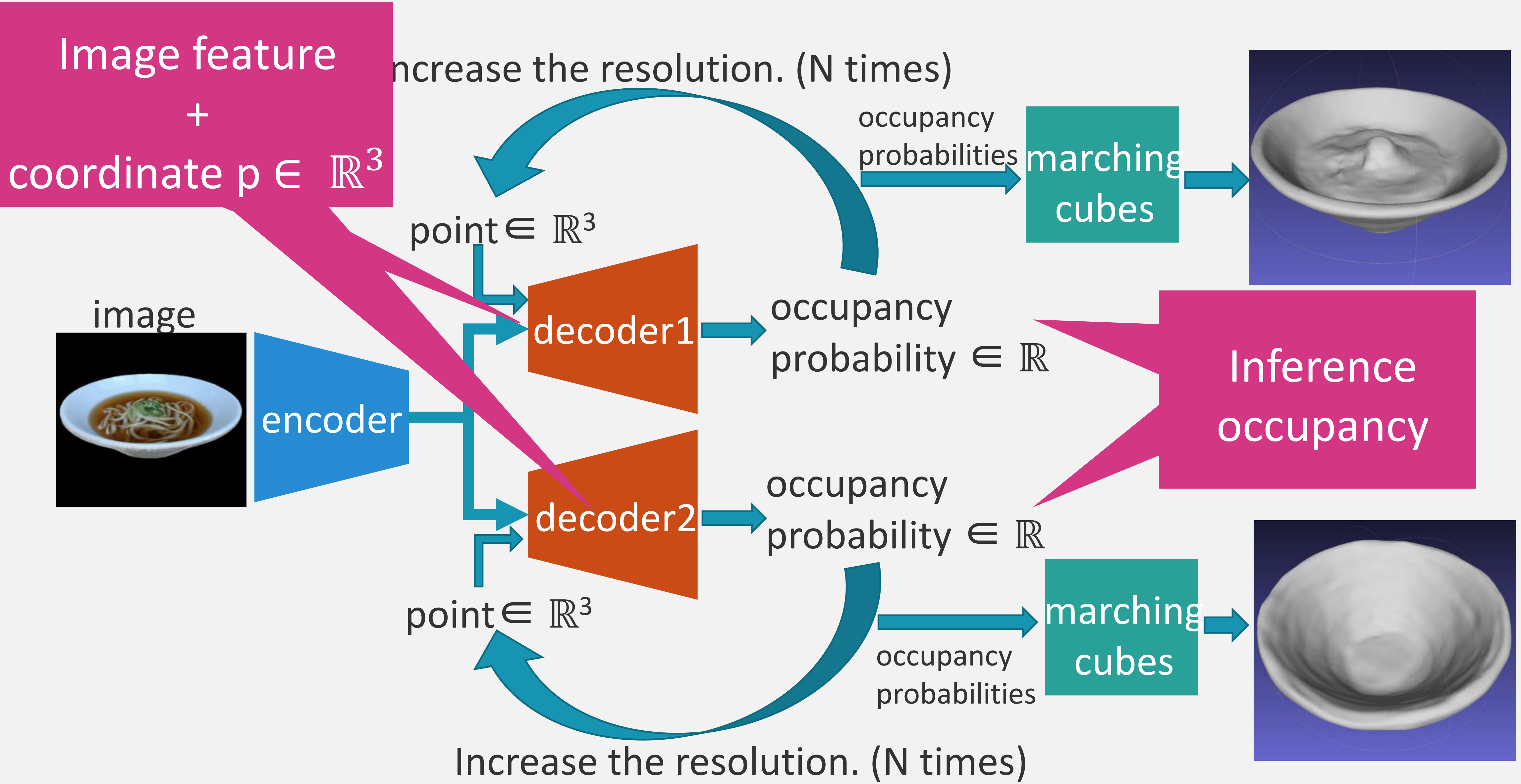
Hungry Networks : inference



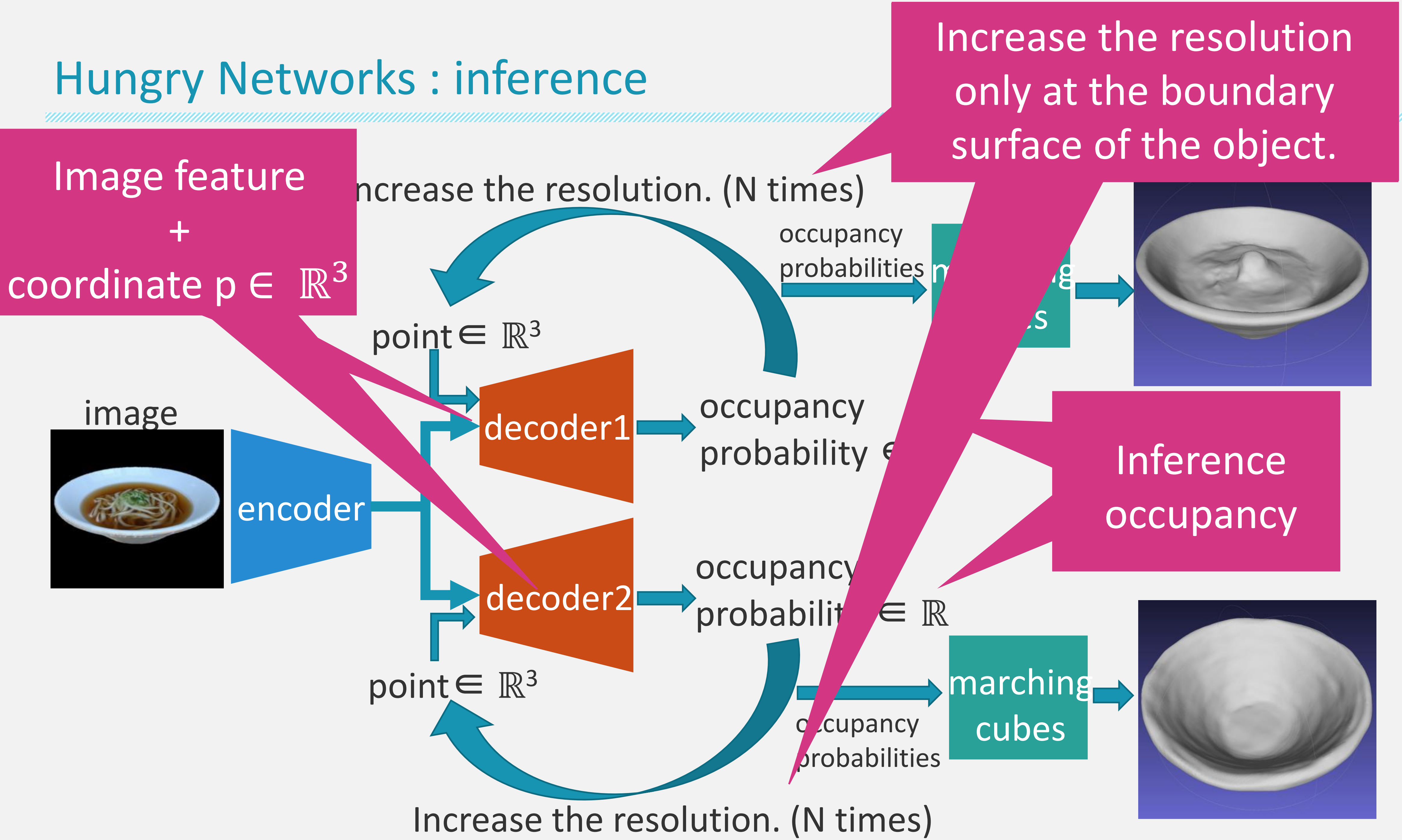
Hungry Networks : inference



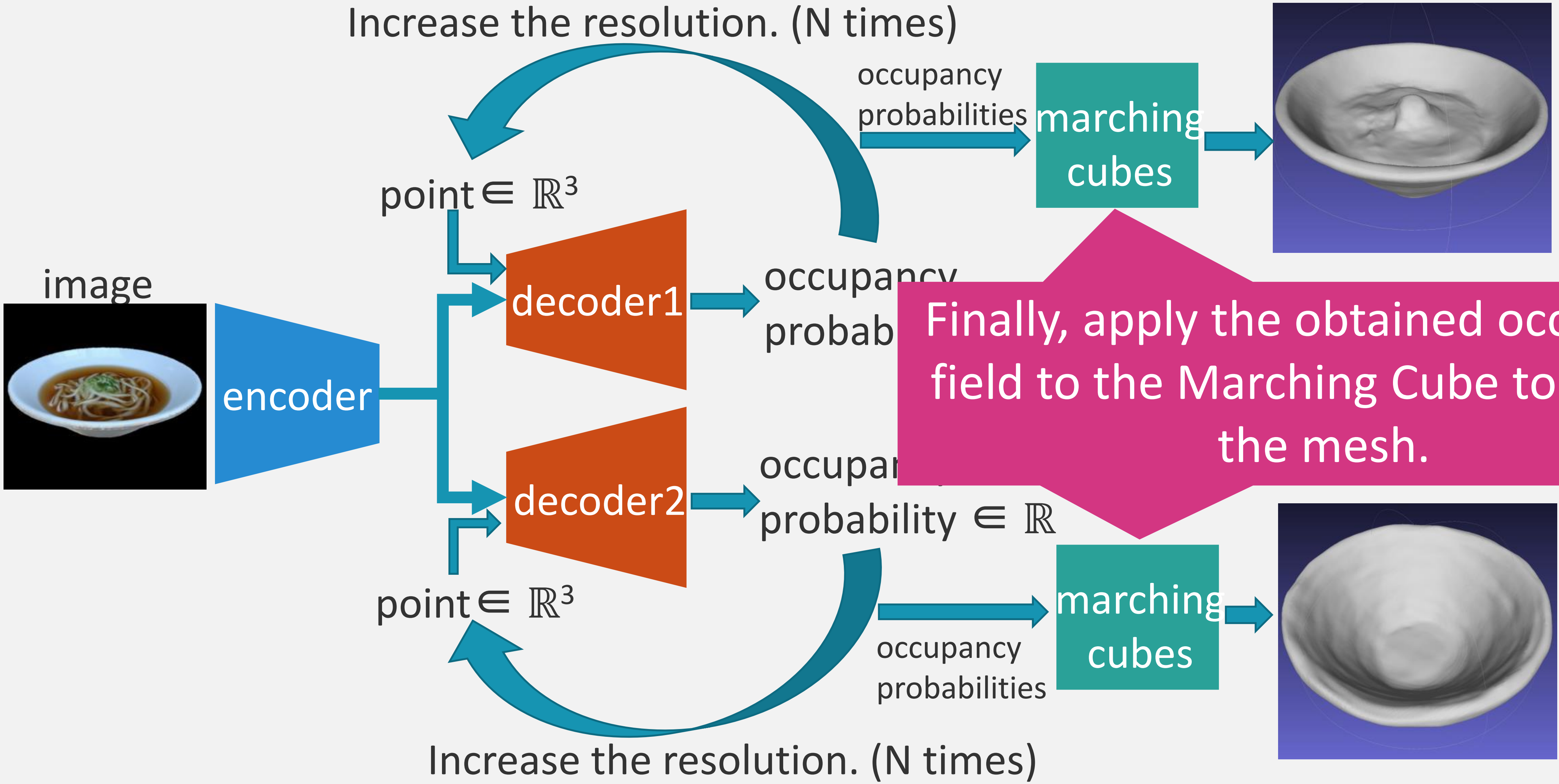
Hungry Networks : inference



Hungry Networks : inference



Hungry Networks : inference



Finally, apply the obtained occupancy field to the Marching Cube to extract the mesh.

Hungry Networks : training

- Learning the occupancy is actually a binary classification. (inside or outside)
    - Binary cross entropy loss

$$\mathcal{L}_{\mathcal{O}}(f_d(x,p), o(p)) = \mathcal{L}_{bce}(f_d(x,p), o(p))$$

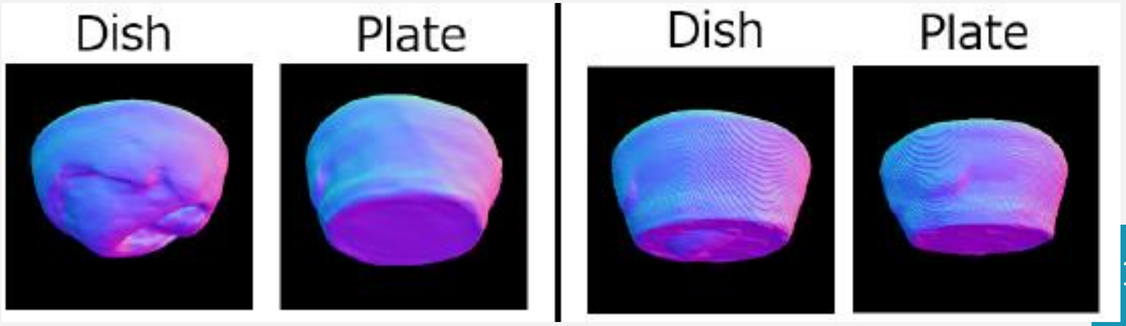
$p \in R^3$	: input point coordinate
x	: image feature vector
$o(p) \in R$	: occupancy of point p
$f_d(x,p) \in R$	: decoder that outputs occupancy

Hungry Networks : training

- **Plate consistency loss (proposal method)**
  - Loss function for matching plate parts of the 3D shape of dish and plat

Dish occupancy $f_{d1}(x,p)$	Plate occupancy $f_{d2}(x,p)$	$f_{d2}(x,p) - f_{d1}(x,p)$
0	**0**	**0**
1	**0**	**-1**
0	**1**	**1**
1	**1**	**0**

$$\mathcal{L}_C(f_{d1}(p), f_{d2}(p)) = \max(f_{d2}(p) - f_{d1}(p)\ ,0)$$



Hungry Networks : training

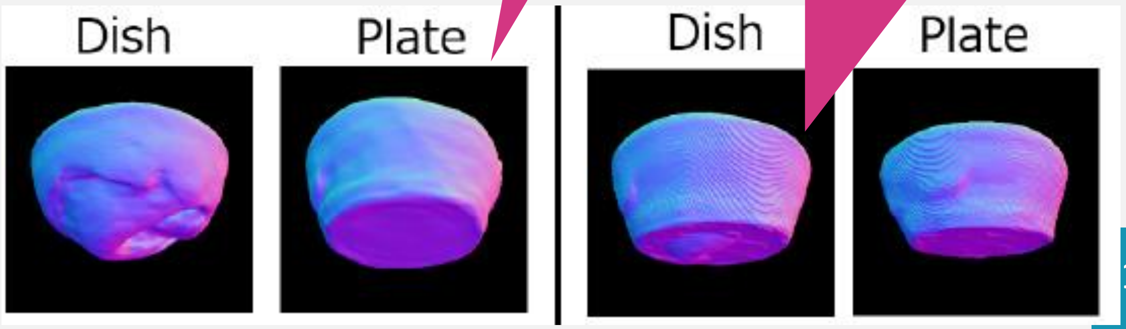
- **Plate consistency loss (proposal method)**
  - Loss function for matching plate parts of the 3D shape of dish and plat

Dish occupancy $f_{d1}(x,p)$	Plate occupancy $f_{d2}(x,p)$	$f_{d2}(x,p) - f_{d1}(x,p)$
0	**0**	**0**
1	**0**	**-1**
0	**1**	**1**
1	**1**	**0**

$$\mathcal{L}_C(f_{d1}(p), f_{d2}(p)) = \max(f_{d2}(p) - f_{d1}(p)\ , 0)$$

Without Plate consistency loss

With Plate consistency loss



Hungry Networks : training

- **Plate consistency loss (proposal method)**
  - Loss function for matching plate parts of the 3D shape of dish at

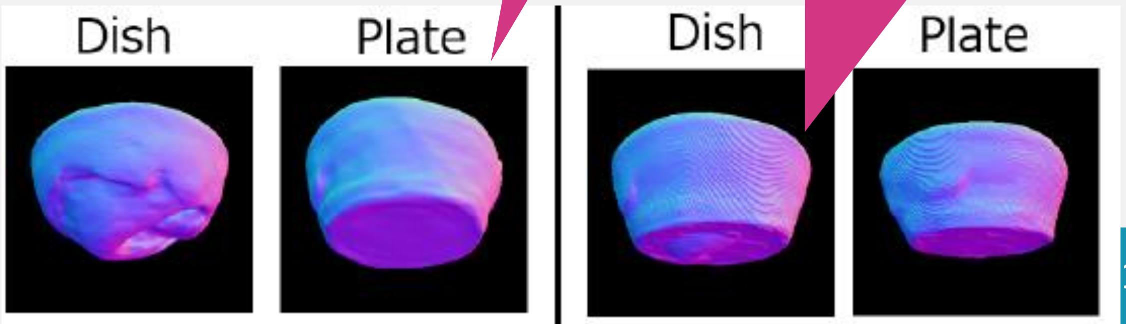
Dish occupancy $f_{d1}(x,p)$	Plate occupancy $f_{d2}(x,p)$	$f_{d2}(x,p) - f_{d1}(x,p)$
0	**0**	**0**
1	**0**	**-1**
0	**1**	**1**
1	**1**	**0**

There is a problem if the difference is **1**.

Without Plate consistency loss

With Plate consistency loss

$$\mathcal{L}_C(f_{d1}(p), f_{d2}(p)) = \max(f_{d2}(p) - f_{d1}(p)\ ,0)$$



Hungry Networks : training

- Mini batch loss

$$
\begin{aligned}
x_i &= f_e(I_i) \\
y1_{i,j} &= f_{d1}(x_i, p_{i,j}) \\
y2_{i,j} &= f_{d2}(x_i, p_{i,j})
\end{aligned}
$$

$f_e(I_i)$ Encoder that outputs image feature

I_i i-th image

$\mathcal{B}$ mini batch

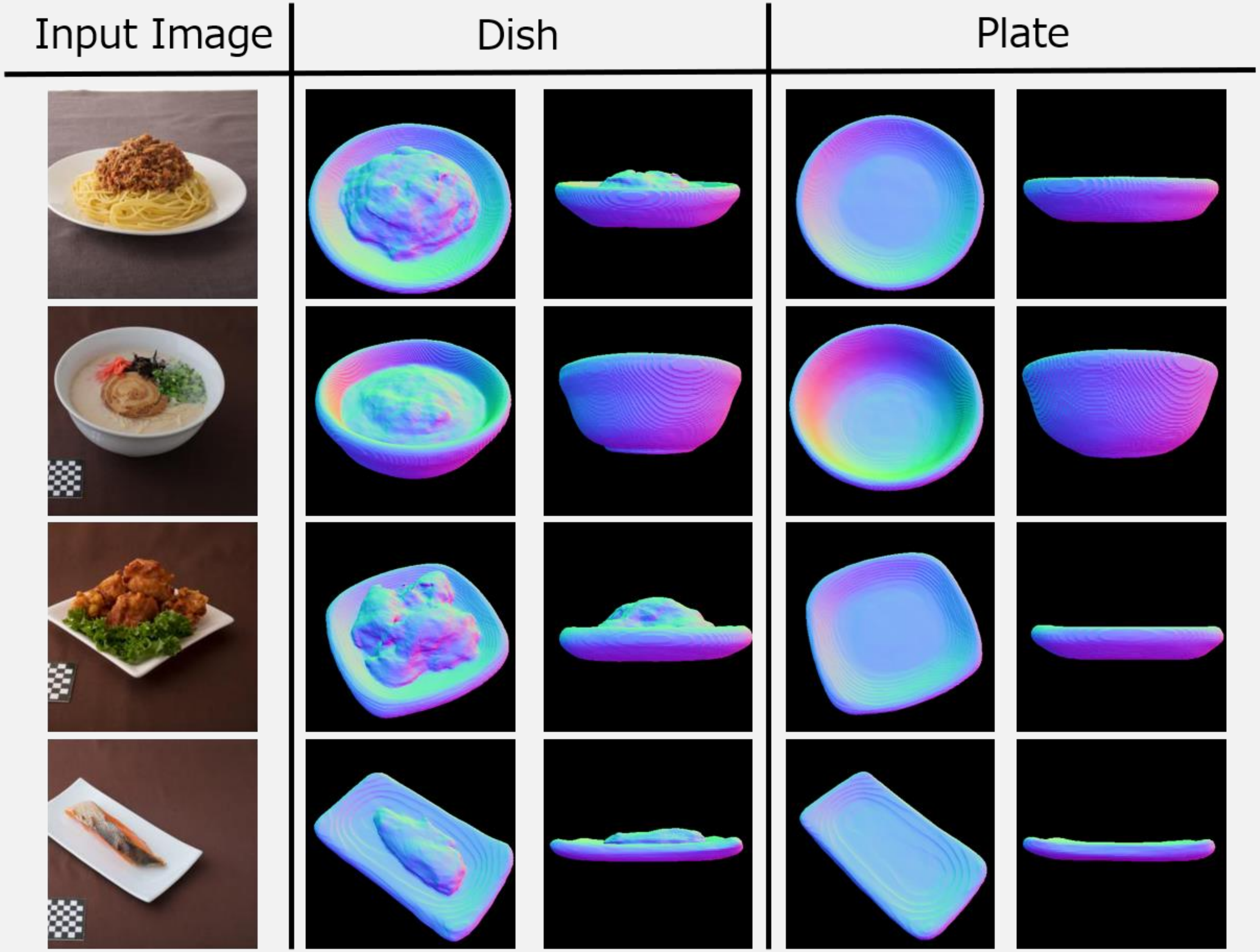
$$
\begin{aligned}
\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \sum_{j=1}^{K} \Bigg(& \lambda_1 \mathcal{L}_{\mathcal{O}}(y1_{i,j}, o1_i(p_{i,j})) \\
& + \lambda_2 \mathcal{L}_{\mathcal{O}}(y2_{i,j}, o2_i(p_{i,j})) \\
& + \lambda_3 \mathcal{L}_{\mathcal{C}}(y1_{i,j}, y2_{i,j}) \Bigg)
\end{aligned}
$$

Training dataset

- There is no dataset containing a 3D mesh of dish.
  - Build a new dataset
- 240 Dish 3D models、38 plate 3D models.
  - Using a commercially available 3D scanner.




Experiment : Qualitative evaluation



Experiment : Quantitative evaluation

- weighting plate consistency loss

λ_3	IoU (dish)	IoU (plate)	Chamfer L1 (dish)	Chamfer L1 (plate)	**plate consistency**	**Volume error**
0	**0.624**	**0.621**	**0.0189**	0.0186	0.0256	0.0252
20	0.550	0.607	0.0262	**0.0182**	0.0168	**0.0155**
50	0.542	0.610	0.0260	0.0209	**0.0152**	0.0161

$$
\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|}\sum_{i=1}^{|\mathcal{B}|}\sum_{j=1}^{K}\Big(\lambda_1\mathcal{L}_{\mathcal{O}}(y1_{i,j}, o1_i(p_{i,j}))
+ \lambda_2\mathcal{L}_{\mathcal{O}}(y2_{i,j}, o2_i(p_{i,j}))
+ \underline{\lambda_3}\mathcal{L}_{\mathcal{C}}(y1_{i,j}, y2_{i,j})\Big)
$$

Experiment : Quantitative evaluation

- weighting plate consistency loss

plate consistency loss contributes to reducing volume error.

λ_3	IoU (dish)	IoU (plate)	Chamfer L1 (dish)	Chamfer L1 (plate)	plate consistency	Volume error
0	**0.624**	**0.621**	**0.0189**	0.0186	0.0256	0.0252
20	0.550	0.607	0.0262	**0.0182**	0.0168	**0.0155**
50	0.542	0.610	0.0260	0.0209	**0.0152**	0.0161

$$\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \sum_{j=1}^{K} \Big(\lambda_1 \mathcal{L}_{\mathcal{O}}(y1_{i,j}, o1_i(p_{i,j})) + \lambda_2 \mathcal{L}_{\mathcal{O}}(y2_{i,j}, o2_i(p_{i,j})) + \lambda_3 \mathcal{L}_{\mathcal{C}}(y1_{i,j}, y2_{i,j}) \Big)$$

Conclusion

- Hungry Networks
  - Reconstruct 3D dish (food + plate) volume and 3D plate volume from a single dish image

- Introducing plate consistency loss
  - Matching plate parts of the 3D shape of dish and plate
  - Contributes to the accuracy of volume estimation

- Creating a 3D food dataset
  - We showed that it can correspond to the real dish image.