Surface Snapping Optimization Layer for Single Image Object Shape Reconstruction

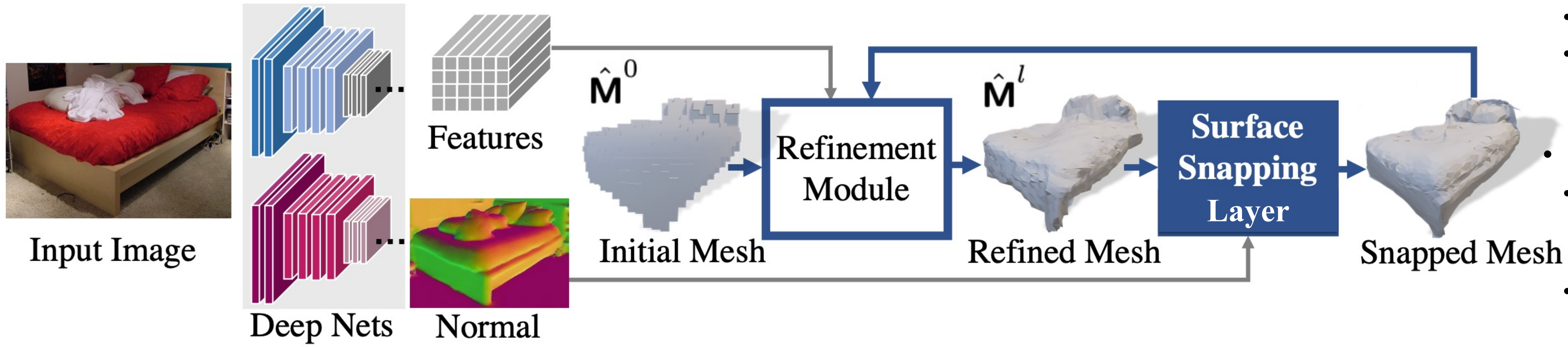
Yuan-Ting Hu[1], Alexander G. Schwing[1] and Raymond A. Yeh[2]

[1] University of Illinois Urbana-Champaign, [2] Purdue University

1. Introduction

- **Single Image 3D Shape Reconstruction**
  - Reason about the 3D geometry of objects from a monocular image
- **Motivation**
  - To better utilize the monocular cue, i.e., surface normal
- **Contributions**
  - A novel optimization layer **surface snapping Layer**
  - Improve 3D reconstruction quality

2. Overview



- **Problem definition**
  - Given: an image
  - Goal: predict the object shape characterized by a triangular mesh $\mathbf{M} = (V, F)$
- **Approach overview**
  - Iteratively update the initial mesh using the refinement module and the proposed surface snapping layer
  - Surface snapping optimizes the vertices of the mesh to match to the estimated normal

3. Surface Snapping Layer

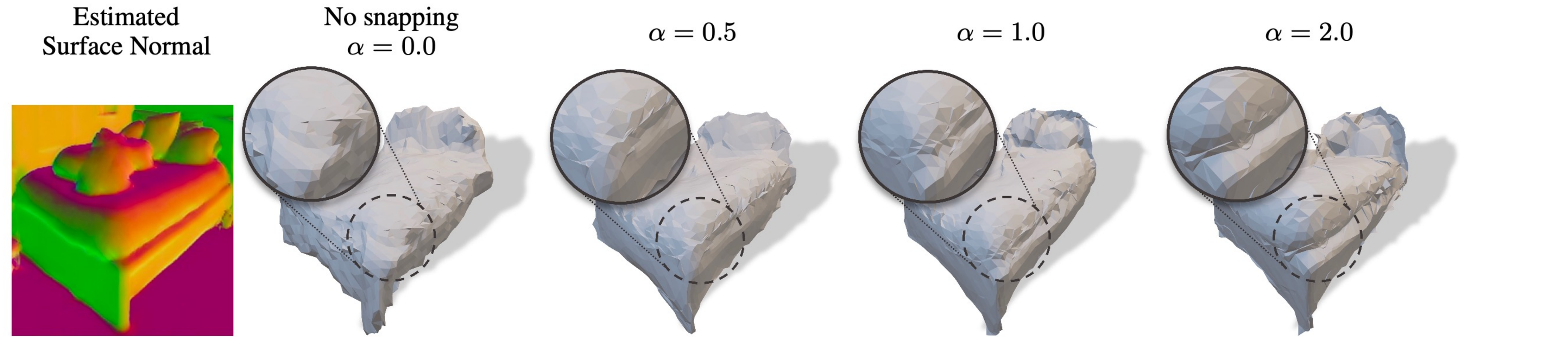
- **Input:** shape $\widehat{\mathbf{M}}^l = (\widehat{V}^l, \widehat{F})$ and normal estimate $\widehat{N}$
- **Output**: the refined shape $\widehat{\mathbf{M}}^{l+1} = (\widehat{V}^{l+1}, \widehat{F})$ with vertices updated by the following objective

$$\widehat{V}^{l+1} \triangleq X^* = \arg\min_X C_V(X, \widehat{V}^l) + \alpha C_N(X, \widehat{N})$$

- **Vertex cost** $C_V(X, \widehat{V}^l) = \|X - \widehat{V}^l\|_2^2$
- **Normal cost** $C_N(X, \widehat{N}) = \sum_{i=1}^{N_F} \sum_{j,k \in \widehat{F}_i} \langle \widehat{N}_i, X_j - X_k \rangle^2$
  - N_F: the number of faces
  - $\widehat{F}_i$: the i^{th} face
  - $\widehat{N}_i$: normal of the i^{th} face
  - X: optimization variable
- Solve by using an efficient conjugate gradient solver, use the sparsity pattern
- End-to-end trainable
  - The weighting term α is learnable
  - Learnt with the other parameters in the deep net

4. Experimental Results

- **Ablation Study**
  - Qualitative effect of surface snapping. Larger values of α result in stronger snapping



  - Quantitative effect of surface snapping

	Pix3D S_1			Pix3D S_2		
	AP[mesh]	Normal	Normal[Vis]	AP[mesh]	Normal	Normal[Vis]
$\alpha = 0.0$	53.4	21.5	45.4	29.1	21.4	46.5
$\alpha = 1.0$	53.4	22.2	47.1	28.7	20.8	44.7
$\alpha = 2.0$	52.7	21.4	44.8	28.5	21.6	44.6
α learned	**54.1**	**23.0**	**48.8**	**29.9**	**23.0**	**49.7**

- **Quantitative Results** Pix3D Split 1

	Method / %	chair	sofa	table	bed	desk	bkcs	wrdrb	tool	misc	AP[box]	AP[mask]	AP[mesh]	Normal	Normal[Vis]
Baseline	Pixel2Mesh+ (Wang et al., 2018)	30.9	59.1	40.2	40.5	30.2	50.8	62.4	18.2	26.7	93.5	**88.4**	39.9	18.0	39.8
	Sphere-Init	40.9	75.2	44.2	50.3	28.4	48.6	42.5	26.9	7.0	**94.1**	87.5	40.5	15.3	39.0
	Mesh R-CNN (Gkioxari et al., 2019)	48.2	71.7	60.9	53.7	42.9	70.2	63.4	21.6	27.8	94.0	**88.4**	51.2	21.6	46.5
	GCN Transformer	**49.9**	74.3	**67.3**	50.5	42.8	**75.4**	68.9	**37.4**	33.3	**94.1**	88.3	55.5	23.3	49.6
Snap+	Pixel2Mesh+	32.3	61.8	43.9	42.8	33.5	46.0	74.3	4.5	26.7	93.5	**88.4**	40.7	19.4	43.3
	Sphere-Init	35.1	61.7	44.1	40.0	31.3	55.7	46.1	23.6	13.5	**94.1**	87.6	39.0	17.1	39.0
	Mesh R-CNN	49.0	74.8	65.3	55.7	**46.1**	73.8	**70.9**	21.6	27.8	**94.1**	88.3	54.1	23.0	48.8
	GCN Transformer	49.5	**76.5**	64.3	**56.0**	44.3	73.8	**70.9**	31.8	**33.4**	**94.1**	88.3	**55.6**	**23.8**	**50.4**

- **Quantitative Results** Pix3D Split 2

	Method / %	chair	sofa	table	bed	desk	bkcs	wrdrb	tool	misc	AP[box]	AP[mask]	AP[mesh]	Normal	Normal[Vis]
Baseline	Pixel2Mesh+ (Wang et al., 2018)	26.7	58.5	10.9	38.5	7.8	34.1	**3.4**	10.0	0.0	71.1	63.4	21.1	19.5	42.2
	Sphere-Init	32.9	**75.3**	15.8	40.1	10.1	45.0	1.5	0.8	0.0	**72.6**	**64.5**	24.6	15.7	40.0
	Mesh R-CNN (Gkioxari et al., 2019)	**42.7**	70.8	27.2	40.9	18.2	**51.1**	2.9	5.2	0.0	72.2	63.9	28.8	21.4	46.5
	GCN Transformer	42.9	68.8	26.5	40.7	22.9	44.6	1.2	0.5	0.0	72.4	64.0	27.5	22.1	48.8
Snap+	Pixel2Mesh+	26.9	60.9	10.0	38.7	10.0	25.3	4.2	**10.1**	0.0	71.1	63.4	20.7	20.5	44.7
	Sphere-Init	31.2	72.7	11.7	42.1	7.8	38.2	1.0	1.2	0.0	**72.6**	**64.5**	22.9	16.0	40.2
	Mesh R-CNN	41.4	74.7	**28.1**	**42.6**	**20.1**	50.4	2.9	3.7	0.0	72.4	64.0	**29.9**	23.0	**49.7**
	GCN Transformer	42.2	75.3	28.6	41.3	21.1	50.0	2.8	6.1	0.0	72.4	64.0	29.7	**23.3**	**49.7**

- **Qualitative results**

- **Failure cases of our approach**
  - Surface normal estimation is not accurate