

Deep Active Surface Models

Udaranga Wickramasinghe, Pascal Fua
Computer Vision Laboratory
EPFL

{udaranga.wickramasinghe, pascal.fua}@epfl.ch

Graham Knott
BioEM Laboratory
EPFL

graham.knott@epfl.ch

Abstract

Active Surface Models have a long history of being useful to model complex 3D surfaces. But only Active Contours have been used in conjunction with deep networks, and then only to produce the data term as well as meta-parameter maps controlling them. In this paper, we advocate a much tighter integration. We introduce layers that implement them that can be integrated seamlessly into Graph Convolutional Networks to enforce sophisticated smoothness priors at an acceptable computational cost.

We will show that the resulting Deep Active Surface Models outperform equivalent architectures that use traditional regularization loss terms to impose smoothness priors for 3D surface reconstruction from 2D images and for 3D volume segmentation.

1. Introduction

Triangulated meshes are one of the most popular and versatile kind of 3D surface representation. In recent years, one of the most popular approaches to inferring such representations from images has been to use deep networks to produce a volumetric representation and then running a marching-cube style algorithm to create the mesh. However, using marching-cubes tends to introduce artifacts and introduces additional complexities when trying to make the process end-to-end differentiable. Hence, deep-learning methods that go *directly* and without resorting to an intermediate stage from 2D images [40, 32, 41] and 3D image stacks [42] to 3D triangulated surfaces have recently been proposed.

Unfortunately, these direct methods are also prone to generating unwanted artifacts such as those shown at the top of Fig. 1. State-of-the-art methods handle them by introducing additional regularizing loss terms such as the edge length loss, the normal consistency loss, or the Laplacian loss during training [14]. To be effective without sacrificing reconstruction accuracy, these terms must be carefully weighted, which is typically difficult to achieve.

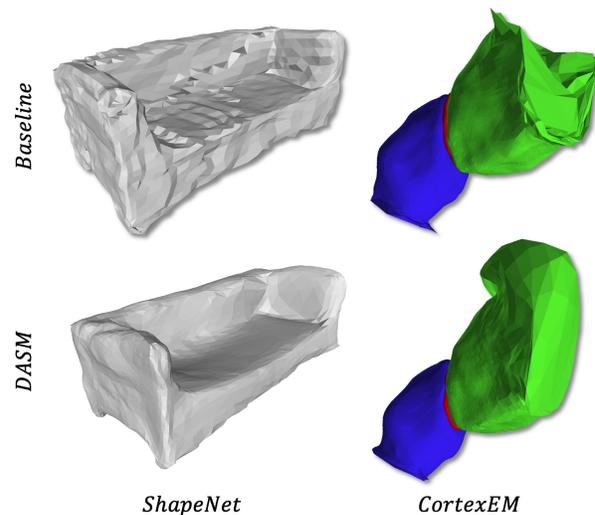


Figure 1: **Smoothness and Accuracy.** (Top) 3D surface meshes of a couch modeled from an RGB image and of a synaptic connection segmented from an electron microscopy stack by Mesh R-CNN [14] and by Vox2Mesh [42], two state-of-the-art mesh-generating methods. (Bottom) Results using the same backbones augmented by our DASM smoothing layers. The meshes have far fewer artifacts and we will show that they are also more accurate.

In this paper, we solve this problem by introducing into the surface generating architecture a special-purpose layer that regularize the meshes using a semi-implicit scheme that involves recursively solving sparse linear systems of linear equations. It propagates smoothness constraints much faster and more reliably than traditional gradient descent-based energy minimization without requiring much computational power and yields surface meshes that fit the data while remaining smooth, such as those shown at the bottom of Fig. 1. Furthermore, this scheme enables us to

- modulate locally the amount of regularization we impose so that we regularize only where it is needed and, hence, preserve accuracy;
- use meshes consisting of vertices with arbitrary

degrees which is not commonly seen in majority of Active Shape Models.

Both of these are important to model complex 3D objects that can be smooth in some places and very curvy elsewhere.

We took our inspiration from the Active Surface Models (ASMs) idea [38, 39], which were first introduced over 30 years ago and also used a semi-implicit optimization scheme to model complex 3D shapes from images. Today, they are mostly used in conjunction with deep networks that are used to compute the data term that is minimized when deforming the models and meta-parameter maps that controls its behavior [27, 10]. Even though these methods are end-to-end trainable, they do not embed the ASMs within the contour deforming graph convolution networks as we do. Furthermore, they are limited to 2D contours whereas we handle irregular 3D surface meshes, that is, meshes whose vertices can be of arbitrary degrees. To this end, we propose an original method to compute the derivatives required for back-propagation on such a mesh.

Our contribution therefore is *Deep Active Surface Models* (DASMs) that outperform equivalent architectures in which the smoothness constraints are imposed by minimizing a traditional loss function. We will demonstrate this for 3D surface reconstruction from 2D images and for 3D volume segmentation.

2. Related Work

2.1. Active Contour and Surface Models

Active contour models allow contours to be refined to account for local image properties while preserving global geometric primitives. They were first introduced in [19] for interactive delineation and then extended for many different purposes [12]. Active surface models operate on the same principle [38, 39] but replace the contours by triangulated meshes to model 3D surfaces. They have proved very successful for medical [28, 16] and cartographic applications [13], among others, and are still being improved [23, 35, 18].

Both active contours and surfaces operate by minimizing an objective function that is a weighted sum of data term derived from the images and a quadratic term that enforces global smoothness. They owe part of their success to the ability to perform the optimization using a semi-implicit scheme that propagates smoothness constraints much faster than gradient descent energy minimization would and gives them superior convergence properties. In our work, we integrate this scheme into our deep architecture.

2.2. Deep Surface Models

Before the advent of deep learning, mesh representations used to be dominant in the field of 3D surface

reconstruction [28, 12]. Since then, they have been eclipsed by methods that rely on continuous deep implicit-fields. They represent 3D shapes as level sets of deep networks that map 3D coordinates to a signed distance function [33, 43] or an occupancy field [29, 4]. This mapping yields a continuous shape representation that is lightweight but not limited in resolution. This representation has been successfully used for single-view reconstruction [29, 4, 43] and 3D shape-completion [6].

However, for applications requiring explicit surface parameterizations, the non-differentiability of standard approaches to iso-surface extraction, such as the many variants of the Marching Cubes algorithm [26, 31], as well as their tendency to produce artifacts, remain an obstacle to exploiting the advantages of implicit representations. The non-differentiability is addressed in [36] but the artifacts remain. Pixel2Mesh [40] and its newer variants [32, 41] represent attempts to overcome this difficulty by going *directly* from 2D images to 3D surface meshes without resorting to an intermediate stage. This approach has recently been extended to handle 3D image volumes [42]. These methods rely on graph-convolution layers to iteratively deform an initial mesh to match the target. While effective, they tend to produce large artifacts that detract both from their accuracy and their usability for further processing. This can be mitigated by introducing regularizing cost terms into the training loss function. However, it is difficult to weigh them properly to remove the artifacts without compromising the accuracy. Our method is designed to address this very issue.

2.3. Deep Contour Models

Similar to Pixel2Mesh and its variants for surface extraction, there exist its 2D counter parts. [24, 34] use graph-convolution networks to perform instance segmentation by deforming a contour. The same approach is used in [25] to perform interactive object annotation.

2.4. Active Contours/Surfaces and Neural Networks

Active contour models have been combined with deep networks by exploiting the differentiability of the active contour algorithm [27, 5, 15]. These approaches use deep networks to produce the data term as well as meta-parameter maps controlling the behavior of the active contour. This has also been exploited to correct errors in contour labels used in semantic segmentation [1]. However, we are not aware of any recent work that embeds active surfaces into deep networks. One potential reason is that most current approaches to generating 3D meshes in a deep learning context, such as those discussed above [40, 32, 41, 42], yield irregular meshes in which the vertices can have varying number of neighbors. This makes the computation of the derivatives required for back-propagation non-trivial.

This has been addressed in the ASM context by introducing finite element-based computations that notably complexify the approach [21] or non-easily differentiable elements such as quadric fitting in the neighborhood of some vertices [22]. In this paper, we propose an approach that is back-propagation friendly on potentially large meshes.

3. Method

We first introduce the general formulation of Active Surface Models (ASMs) and then show how we can apply it to meshes whose vertices can be of arbitrary degree. Finally, we discuss the integration of ASMs and mesh-deforming Graph-Convolutional Neural Networks (GCNNs), which produces our Deep Active Surface Models (DASMs).

3.1. Active Surface Model (ASM)

An ASM consists of a surface $\mathcal{S}(\Phi)$ whose shape is controlled by a vector of parameters Φ and can be deformed to minimize an objective function $E(\Phi)$, often referred to as an *energy*. We first introduce a continuous formulation and then its discretization, which is the one used in practice.

Continuous Formulation. \mathcal{S} is represented by the mapping from \mathbb{R}^2 to \mathbb{R}^3

$$v : (s, r; \Phi) \mapsto (v_x(s, r; \Phi), v_y(s, r; \Phi), v_z(s, r; \Phi)) , \quad (1)$$

where $(s, r) \in \Omega = [0, 1] \times [0, 1]$. Fig. 2 depicts this mapping and its derivatives. Φ is typically taken to be

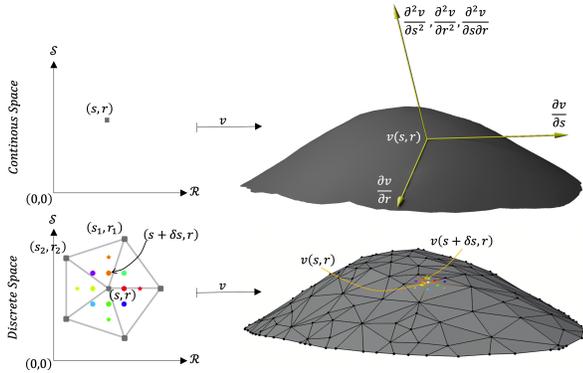


Figure 2: **Derivatives. Top.** The surface is represented by a differentiable mapping v from \mathbb{R}^2 to \mathbb{R}^3 . **Bottom.** After discretization, let $v(s, r)$ be a vertex. To approximate derivatives with respect to s using finite differences, we need to estimate quantities such as $v(s + \delta s, r)$ where δs is small. To this end, we estimate the barycentric coordinates λ , λ_1 , and λ_2 of $v(s + \delta s, r)$ in the facet it belongs to and take $v(s + \delta s, r)$ to be $\lambda v(s, r) + \lambda_1 v(s_1, r_1) + \lambda_2 v(s_2, r_2)$, where $v(s_1, r_1)$ and $v(s_2, r_2)$ are the other two vertices of the facet. The same operation can be performed for derivatives with respect to r .

$$\begin{aligned} \Phi^* &= \arg \min_{\Phi} E(\Phi) , \\ E(\Phi) &= E_{\text{dat}}(\Phi) + E_{\text{def}}(\Phi) , \end{aligned} \quad (2)$$

where E_{dat} is a data term that measures how well the surface matches the images and E_{def} is a deformation energy that is smallest when the surface is smooth. E_{def} is often written as

$$\begin{aligned} E_{\text{def}} &= \int_{\Omega} w_{10} \left\| \frac{\partial v}{\partial s} \right\|^2 + w_{01} \left\| \frac{\partial v}{\partial r} \right\|^2 + 2w_{11} \left\| \frac{\partial^2 v}{\partial s \partial r} \right\|^2 \\ &+ w_{20} \left\| \frac{\partial^2 v}{\partial s^2} \right\|^2 + w_{02} \left\| \frac{\partial^2 v}{\partial r^2} \right\|^2 dr ds . \end{aligned} \quad (3)$$

The surface Φ^* that minimizes the energy $E(\Phi)$ satisfies the associated Euler-Lagrange equation [9, 19]

$$\begin{aligned} F(v) &= -\frac{\partial}{\partial s} \left(w_{10} \frac{\partial v}{\partial s} \right) - \frac{\partial}{\partial r} \left(w_{01} \frac{\partial v}{\partial r} \right) \\ &+ 2\frac{\partial^2}{\partial s \partial r} \left(w_{11} \frac{\partial^2 v}{\partial s \partial r} \right) + \frac{\partial^2}{\partial s^2} \left(w_{20} \frac{\partial^2 v}{\partial s^2} \right) + \frac{\partial^2}{\partial r^2} \left(w_{02} \frac{\partial^2 v}{\partial r^2} \right) , \end{aligned} \quad (4)$$

where $F = -\nabla E_{\text{dat}}$.

Discrete Formulation. When $\mathcal{S}(\Phi)$ is discretized and represented by a triangulated mesh $\mathcal{M}(\Phi)$, Φ becomes the $3N$ -vector built by concatenating the 3D coordinates of its N vertices. Using the finite-difference approximation described in the appendix, Eq 4 can be written in matrix form as

$$\mathbf{A}\Phi^* = F(\Phi^*) , \quad (5)$$

where F is the negative gradient of E_{dat} with respect to Φ . Because \mathbf{A} is usually non-invertible, given an initial value Φ^0 , a solution to this equation can be found by iteratively solving

$$\begin{aligned} \alpha(\Phi^t - \Phi^{t-1}) + \mathbf{A}\Phi^t &= F(\Phi^{t-1}) , \\ \Rightarrow (\mathbf{A} + \alpha\mathbf{I})\Phi^t &= \alpha\Phi^{t-1} + F(\Phi^{t-1}) , \end{aligned} \quad (6)$$

where \mathbf{I} is the identity matrix. When the process stabilizes, $\Phi^t = \Phi^{t-1}$ and is a solution of Eq. 5.

The strength of this semi-implicit optimization scheme is that it propagates smoothness constraints much faster than traditional gradient descent that minimizes energy $E(\Phi)$ and at a low computational cost because \mathbf{A} is sparse, which means that the linear system of Eq. 6 can be solved efficiently. In this scheme α plays the role of the inverse of a step size: When α is large enough for the Froebinius norm of $\alpha\mathbf{I}$ to be much larger than that of \mathbf{A} , the optimizer performs a steepest gradient step given by $F(\Phi^{t-1})$ with learning rate $\frac{1}{\alpha}$ at each iteration. Conversely, when α is small, \mathbf{A} dominates and much larger steps can be taken.

In the original deformable contour models [19], the matrix $\mathbf{A} + \alpha\mathbf{I}$ was never inverted. Instead Eq. 6 was solved by LU decomposition. Instead, to implement this effectively on a GPU using sparse tensors and to speed up the computations of the losses and their derivatives, we approximate the inverse of $(\mathbf{A} + \alpha\mathbf{I})$ using the Neumann series

$$(\mathbf{A} + \alpha\mathbf{I})^{-1} \approx \sum_{n=0}^K (-1)^n \left(\frac{1}{\alpha}\right)^{n+1} \mathbf{A}^n. \quad (7)$$

and use it to solve Eq. 6. We use $K = 4$, which yields a sufficiently good approximation of actually solving Eq. 6.

Computing the Regularization Matrix \mathbf{A} . In most traditional ASMs, the meshes are either square or hexagonal and regular, which makes the computation of the derivatives of the mesh vertices possible using finite-differences and, hence, the regularization matrix \mathbf{A} of Eq. 5 easy to populate.

When the mesh is triangular and irregular, vertices can have any number of neighbors and the computation becomes more complex. Nevertheless the required derivatives, of order 2 and 4, can still be expressed as finite differences of weighted sums of vertex coordinates where the weights are barycentric coordinates of small perturbations of the original vertices. This is explained in more details in the appendix.

3.2. Deep Active Surface Model (DASM)

The update equation in a typical mesh-deforming graph-convolutional neural network (GCNN) that plays the same role as that of Eq. 6 is

$$\Phi^t = \Phi^{t-1} + \frac{1}{\alpha} F(\Phi^{t-1}, \mathbf{X}^{t-1}), \quad (8)$$

where F denotes the negative gradient of the loss function calculated using the feature vector \mathbf{X}^{t-1} associated with the mesh parameters Φ^{t-1} . In the case of our deep active surface models, it becomes

$$(\mathbf{A} + \alpha\mathbf{I})\Phi^t = \alpha\Phi^{t-1} + F(\Phi^{t-1}, \mathbf{X}^{t-1}), \quad (9)$$

as in Eq. 6. In Eq. 8, the loss function typically includes a regularization term to keep the mesh smooth, whereas in Eq. 9 our semi-implicit scheme enforces smoothness by solving the linear equation.

Uniform vs Adaptive DASMs Eq. 9 forms the basis of the simplest version of our DASMs, which we will refer to as Uniform DASMs because the same amount of smoothing is applied across the whole mesh. This may result in under- or over-smoothing because some parts of the objects require more smoothing while some parts do not.

To account for this, we also introduce Adaptive DASMs that are designed to smooth only where necessary,

as indicated by an auxiliary metric. Experimentally, adaptive smoothing is required when the GCNN produces particularly large deformations but only in a very specific part of the mesh or fails to smooth-out artifacts produced by mesh initialization algorithms. This could be eliminated by strongly smoothing everywhere but would degrade accuracy in high-curvature areas.

To solve this problem, we begin by using the approximation of $(\mathbf{A} + \alpha\mathbf{I})^{-1}$ from Eq. 7 to rewrite the evolution equation of Eq. 9 as

$$\begin{aligned} \Gamma^t &= \Phi^{t-1} + \frac{1}{\alpha} F(\Phi^{t-1}, \mathbf{X}^{t-1}), \\ \Phi^t &= (\mathbf{I} + \sum_{n=1}^K (-1)^n \left(\frac{1}{\alpha}\right)^n \mathbf{A}^n) \Gamma^t, \\ &= \Gamma^t + \mathbf{B}\Gamma^t \text{ with } \mathbf{B} = \sum_{n=1}^K (-1)^n \left(\frac{1}{\alpha}\right)^n \mathbf{A}^n. \end{aligned} \quad (10)$$

Γ^t represents Φ^{t-1} incremented by the negative gradient of the loss function $F(\Phi^{t-1})$ but not yet smoothed. In other words, we have rewritten the smoothing operation that transforms Γ^t into Φ^t as simply adding $\mathbf{B}\Gamma^t$ to Γ^t . This gives us the freedom to decide where we want to smooth and where we do not by introducing a diagonal matrix Λ and rewriting the update rule of Eq. 10 as

$$\Phi^t = \Gamma^t + \Lambda\mathbf{B}\Gamma^t. \quad (11)$$

This update rule is similar to the one of the Adagrad algorithm [11]. Here, each diagonal component $\lambda_{i,i}$ of Λ rescales the corresponding component $(\mathbf{B}\Gamma)_i$ of $\mathbf{B}\Gamma$. In Adagrad, adaptive re-scaling is a function of past gradients. Here we take it to be a function of current surface gradients because we have observed that $|\mathbf{B}\Gamma^t|_i$ tends to grow large when the facets increase in size and smoothing is required, and remains small otherwise. We therefore take the diagonal values of Λ to be

$$\lambda_{ii} = \sigma(|\mathbf{B}\Gamma^t|_i; \beta, \gamma), \quad (12)$$

where σ is the Sigmoid function and β, γ are its steepness and midpoint. In this way, for small values of $|\mathbf{B}\Gamma^t|_i$, there is almost no smoothing, but for larger ones there is. Fig. 3 illustrates this behavior.

Recursive smoothing Any single DASM step given by Eq. 11 can only rectify a finite amount of deformations. To mitigate this, we perform more than one adaptive-smoothing step in-between gradient updates. During these additional smoothing steps no gradient update is done and we use $F(\Phi^{t-1}, \mathbf{X}^{t-1}) = 0$. In practice we perform these steps until $\|\Phi^t - \Phi^{t-1}\| < \epsilon$, where ϵ is a preset constant. Fig. 4 illustrates this process.

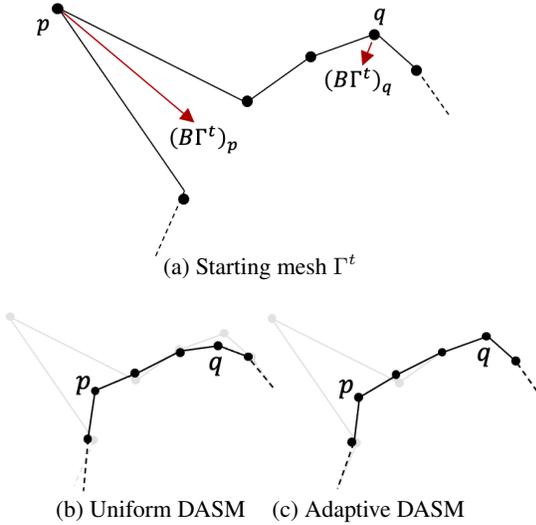


Figure 3: **Uniform vs Adaptive Smoothing.** (a) Mesh at time t . In general, $|\mathbf{B}\Gamma^t|_q < |\mathbf{B}\Gamma^t|_p$. (b) When using enough uniform smoothing to remove the irregularity at point p , the mesh will typically be oversmoothed at q . (c) When using adaptive smoothing, the mesh is smoothed around p but not oversmoothed around q .

Loss terms In architectures such as Mesh R-CNN [14] and Voxel2Mesh [42], a loss term is used to supervise the output of each mesh-refinement stage. We follow the same approach and add a loss term at the end of each DASM module. We write it as

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{data} + \lambda \mathcal{L}_{reg}, \\ \mathcal{L}_{data} &= \lambda_{cf} \mathcal{L}_{cf} + \lambda_{n.dist} \mathcal{L}_{n.dist}, \\ \mathcal{L}_{reg} &= \lambda_{edge} \mathcal{L}_{edge} + \lambda_{Lap} \mathcal{L}_{Lap} + \lambda_{n.cons} \mathcal{L}_{n.cons}. \end{aligned} \quad (13)$$

Here \mathcal{L}_{cf} , $\mathcal{L}_{n.dist}$ are Chamfer and Normal distances [14] and \mathcal{L}_{edge} , \mathcal{L}_{Lap} , $\mathcal{L}_{n.cons}$ are edge length loss, Laplacian loss and normal consistency loss, respectively [40]. All these loss terms are used in Voxel2Mesh [42] except \mathcal{L}_{Norm} . Similarly, they are all used in Mesh R-CNN [14] except \mathcal{L}_{Lap} and \mathcal{L}_{Norm} .

4. Experiments

In this section, we test DASM’s ability to predict 3D surfaces from 2D images on Shapenet [3] and to extract 3D surfaces from electron microscopy image stacks.

4.1. From 2D Images to 3D Surfaces

For prediction of 3D surfaces from 2D images, we benchmark our Adaptive DASM, which we will refer to as Ad.-DASM, on the ShapeNet dataset [3].

Baselines. We use Mesh R-CNN [14] both as a baseline and as the backbone of our network because, among

methods that use explicit surface representations, it is currently reported as yielding the best results on ShapeNet. We also compare against Pixel2Mesh [40].

Dataset. ShapeNet is a collection of 3D textured CAD models split into semantic categories. As in the Mesh R-CNN experiments, we use ShapeNetCore.v1 and corresponding rendered images from [7]. They are of size 137×137 and have been captured from 24 random viewpoints. We use the train / test splits of [14], that is, 35,011 models seen in 840,189 images for training and 8,757 models seen in 210,051 images for testing. We use 5% of the training data for validation purposes.

Metrics. We use the same metrics as in Mesh R-CNN. They are the Chamfer distance, Normal distance, and $F1^\tau$ at $\tau = 0.1, 0.3$ and 0.5 . For the Chamfer distance a lower value is better while a higher value is better for the others.

Implementation. We use the publicly available Pytorch implementation of Mesh R-CNN and incorporate Adaptive DASM layers after each mesh-refinement stage. We also add a Uniform DASM layer after the cubify operation to make the input to mesh refinement stages smooth. We train the networks for 12 epochs using Adam optimizer [20] with a learning rate 10^{-4} . We set $\alpha = 1, \beta = 6000$ and $\gamma = 15$.

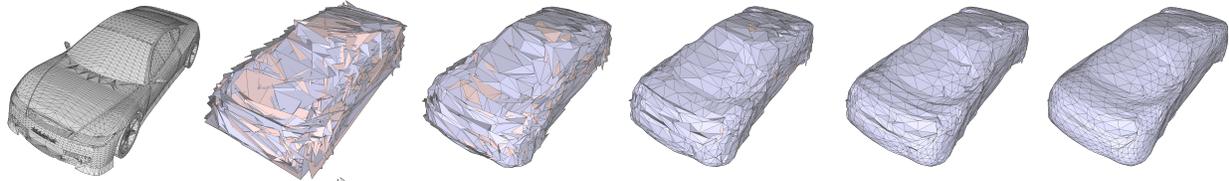
Mesh R-CNN only uses the \mathcal{L}_{edge} term of Eq. 13 for regularization purposes when training on ShapeNet and turns off the term \mathcal{L}_{Lap} because, according to remarks on Github by the authors, it has not helped to improve the results. For a fair comparison, we therefore do the same. In this setup, \mathcal{L}_{edge} , which penalizes increases in edge-length, is the only other source of geometric regularization besides the one we provide with our DASM layers. We will therefore experiment with different values of λ_{edge} , the weight parameter in Eq. 13 that controls how much influence it is given.

	Chf. (↓)	Normal	$F^{0.1}$	$F^{0.3}$	$F^{0.5}$
Pixel2Mesh	0.241	0.701	31.6	77.3	91.3
Mesh R-CNN	0.189	0.691	32.8	80.4	92.6
Ad.-DASM	0.183	0.727	33.5	81.0	92.6

Table 1: **Comparative results on ShapeNet.**

Results. We provide qualitative results in Fig. 5 and report quantitative results in Table 1 for $\lambda_{edge} = 0.2$. Ad.-DASM outperforms Pixel2Mesh and boosts the performance of Mesh R-CNN. Furthermore, the meshes it produces are of much a better visual quality.

In Table 2, we report similar results for different values of λ_{edge} , which are depicted qualitatively by Fig. 6. The trend is the same for $\lambda_{edge} = 0.6$ and 1.0 . However, for $\lambda_{edge} = 0.0$, the Chamfer distance for Mesh R-CNN



(a) Ground truth (b) Input (c) ASM 1-step (d) ASM 3-steps (e) ASM 5-steps (f) ASM 7-steps

Figure 4: **Performing multiple smoothing steps** We start with a very noisy mesh produced by the algorithm of [14] run without regularization. We have colored front-faces and back-faces of the meshes in blue and pink respectively for better visualization. After running one step of surface evolution given by Eq. 6, we obtain the mesh of Fig. 4 (c). Subsequent meshes shown in Fig. 4 (d,e,f) are obtained by continuing the surface evolution for 3, 5, and 7 steps respectively. In these subsequent steps we set $F(\Phi^{t-1})$ to zero.

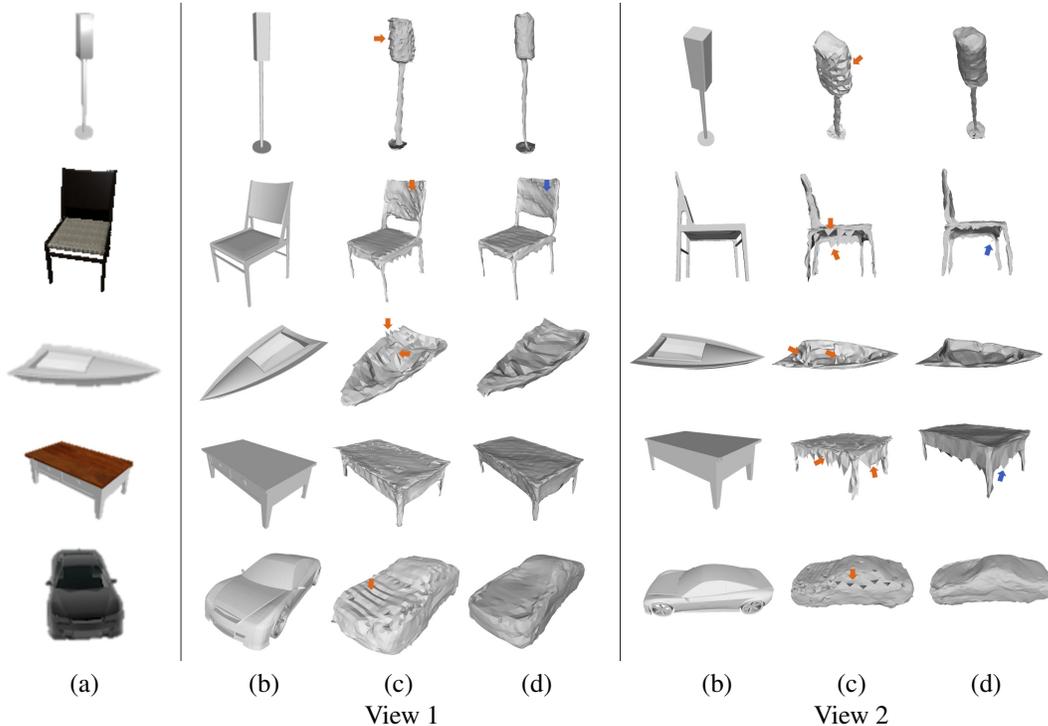


Figure 5: **ShapeNet Results.** (a) Input images (b) Mesh R-CNN results from two different viewpoints. The orange arrows highlight commonly seen Mesh R-CNN artifacts. (c) DASM results in the same two views. The meshes are much smoother and most artifacts have disappeared, except for a few highlighted by blue arrows.

λ_e		Chf. (\downarrow)	Norm.	$F^{0.1}$	$F^{0.3}$	$F^{0.5}$
1.0	Mesh R-CNN	0.232	0.684	29.7	76.6	89.4
	Ad.-DASM	0.231	0.691	29.3	76.6	89.3
0.6	Mesh R-CNN	0.212	0.681	30.6	79.2	91.4
	Ad.-DASM	0.206	0.693	31.4	79.5	91.4
0.2	Mesh R-CNN	0.189	0.691	32.8	80.4	92.6
	Ad.-DASM	0.183	0.727	33.5	81.0	92.6
0.0	Mesh R-CNN	0.144	0.713	35.8	85.1	94.2
	Ad.-DASM	0.167	0.718	34.3	84.3	93.9

Table 2: **Results on ShapeNet** as a function of λ_{edge} .

is lowest even though the resulting meshes are extremely noisy, as can be seen in the leftmost column of Fig. 6. Our

	Chf. (\downarrow)	Normal	$F^{0.1}$	$F^{0.3}$	$F^{0.5}$
PostProc-ASM	0.249	0.673	28.5	75.3	88.1
Uniform-DASM	0.201	0.699	31.4	78.3	91.3
Ad.-DASM	0.183	0.727	33.5	81.0	92.6

Table 3: **Ablation study on ShapeNet.**

interpretation for this somewhat surprising result is that, in this regime, the meshes produced by Mesh R-CNN are so noisy that DASM smoothing takes them away from the data they are trying to fit and degrades the Chamfer distance. In any event, even though the Chamfer distance is low, this can hardly be considered as a good results, hence confirming the observation made in [14, 40] that this metric might

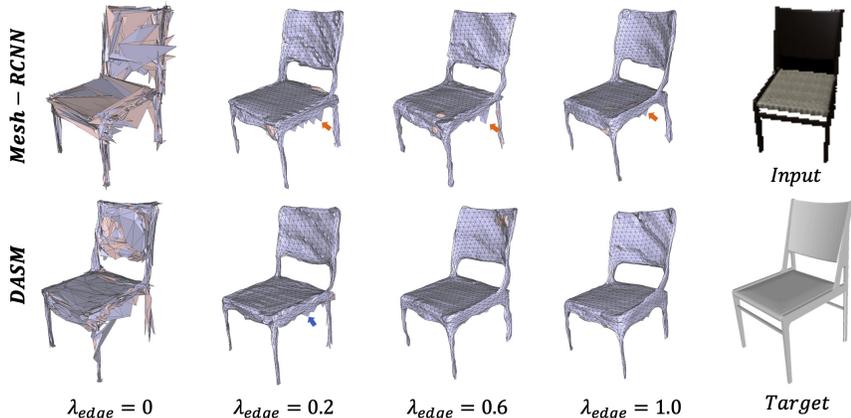


Figure 6: **Influence of the regularization term.** As λ_{reg} increases, the output of both methods becomes smoother but only DASM completely eliminates the artifacts. Note that in the case $\lambda_{reg} = 0$, the output of Mesh R-CNN is an extremely irregular mesh that nevertheless scores well on the Chamfer metric.

not be the best to evaluate the quality of a mesh. When $\lambda_{edge} = 1.0$, the difference between DASM and Mesh R-CNN performance is not statistically significant and this is because, when using higher λ_{edge} , there are not many anomalies for the DASM to fix.

Ablation Study It could be argued that we would have gotten similar results by simply smoothing our meshes as a post-processing step. To demonstrate this is not the case, we implemented PostProc-ASM that starts with Mesh R-CNN model trained with $\lambda_{edge} = 0.2$ that is then adaptively smoothed by running several times the surface evolution update of Eq. 11. In Table. 3, we compare PostProc-ASM against Ad.-DASM and the results are clearly worse. We also compare Ad.-DASM against Uniform-DASM, which clearly shows the benefit of the adaptive scheme of Section 3.2.

Failure modes The main source of failures are anomalies produced by the GCNN that are too large to be rectified. Remaining ones are denoted by blue arrows in Fig 5.

4.2. From 3D Image stacks to 3D Surfaces

In this section we benchmark our approach on CortexEM dataset and compare DASM against Voxel2Mesh [42] and several other baselines.

Baselines. We use Voxel2Mesh both as a baseline and as the backbone of our network. We also compare our performance against several architectures popular in the biomedical imaging community [8, 30, 17, 37, 2].

Dataset. CortexEM is a $500 \times 500 \times 200$ FIB-SEM image stack of a mouse cortex. From this 26 sub-volumes with dimension $96 \times 96 \times 96$ were extracted so that each one contains a synaptic junction that is roughly centered. 14 sub-volumes in the first 100 slices in the image stack are

used for training and the remaining 12 in the next 100 slice are used for testing. The task is to segment the pre-synaptic region, post-synaptic region, and synaptic cleft as shown in Fig. 7.

Metrics. As in [42] and many other papers, we use the intersection-over-union (IoU) as a measure of quality for volumetric segmentation. To compare the meshes, we use the Chamfer distance as in [42]. We repeat our experiments 3 times for each model and report the mean and the standard deviation.

	Pre-Syn.	Synapse	Post-Syn.
TernausNet [17]	73.5 \pm 1.3	64.4 \pm 0.5	78.4 \pm 1.3
LinkNet34 [37]	72.3 \pm 0.5	63.2 \pm 1.2	78.2 \pm 1.1
ResNet50 [2]	70.3 \pm 0.8	63.3 \pm 0.6	76.2 \pm 1.4
ResNet50-SE [2]	71.3 \pm 0.6	63.6 \pm 0.7	76.3 \pm 0.9
V-NET [30]	64.3 \pm 0.7	65.2 \pm 1.3	74.1 \pm 0.7
U-NET [8]	73.6 \pm 1.3	67.2 \pm 0.8	78.2 \pm 0.9
Voxel2Mesh _{0.25}	77.3 \pm 1.2	65.3 \pm 1.2	83.2 \pm 1.6
Voxel2Mesh _{0.025}	76.8 \pm 0.9	65.4 \pm 1.6	81.2 \pm 1.4
Ad.-DASM _{0.25}	77.7 \pm 0.8	65.3 \pm 0.7	83.3 \pm 0.8
Ad.-DASM _{0.025}	79.4 \pm 1.4	65.3 \pm 0.9	85.5 \pm 1.2

Table 4: **Comparative results on CortexEM.** We use the IoU metric to compare volumetric segmentations.

	Pre-Syn.	Synapse	Post-Syn.
Voxel2Mesh _{0.25}	1.62 \pm 0.4	0.19 \pm 0.2	2.41 \pm 0.7
Voxel2Mesh _{0.025}	1.53 \pm 0.6	0.18 \pm 0.4	2.35 \pm 0.4
Ad.-DASM _{0.25}	1.59 \pm 0.5	0.19 \pm 0.3	2.39 \pm 0.8
Ad.-DASM _{0.025}	1.47 \pm 0.6	0.18 \pm 0.5	2.31 \pm 0.6

Table 5: **Comparative results on CortexEM.** We use the Chamfer distance ($\times 10^{-2}$)

Implementation. As we did with Mesh R-CNN in Section 4.1, we incorporate Ad.-DASM layers into a

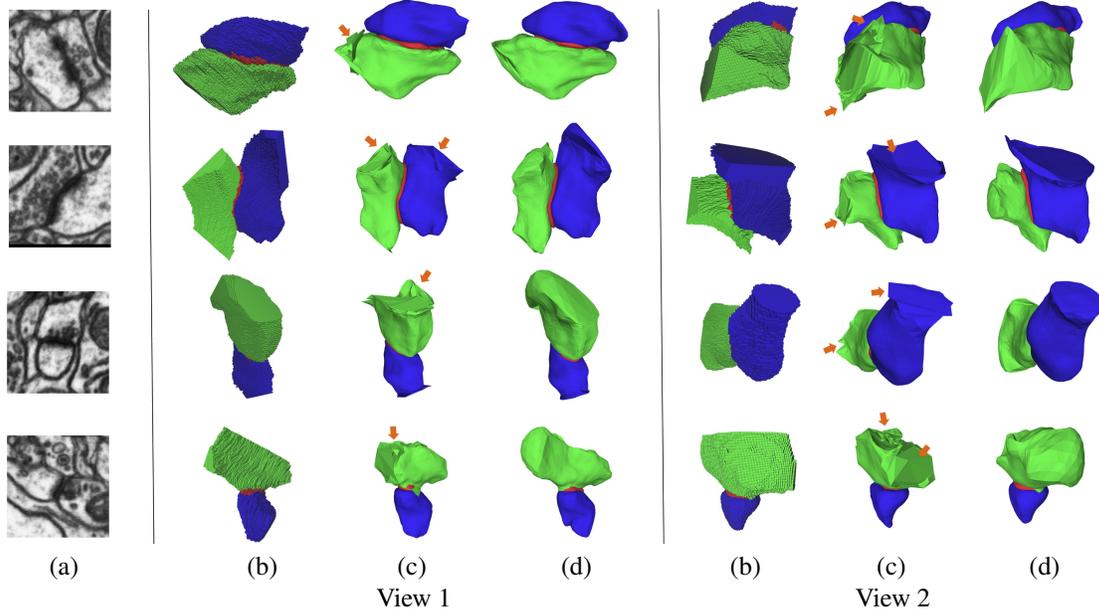


Figure 7: Modeling Synapses from Electron Microscopy Image Stacks (a) Representative slice from the input volume. (b) Voxel2Mesh results seen from two different views. (c) DASM results seen from the same two views. The pre-synaptic region, post-synaptic region, and synaptic cleft are shown in blue, green, and red, respectively. The DASM results are much smoother and without artifacts, while also being more accurate.

Voxel2Mesh backbone. We train the networks for 150000 iterations using Adam optimizer [20] with a learning rate of 10^{-4} . We set $\alpha = 1, \beta = 6000$ and $\gamma = 45$.

To match the conditions in [42], Ad.-DASM was trained using $\lambda_{edge} = \lambda_{Lap.} = \lambda_{n.cons.} = 0.25$ in Eq. 13. For comparison purposes, we also use $\lambda_{edge} = \lambda_{Lap.} = \lambda_{n.cons.} = 0.025$. To differentiate two versions, we add the regularization coefficients as a subscript to model names. This gives us Voxel2Mesh_{0.25}, Ad.-DASM_{0.25}, Voxel2Mesh_{0.025}, and Ad.-DASM_{0.025}.

Results We report quantitative results in Table 4 in IoU terms and in Table 5 in Chamfer distance terms. Fig. 7 depicts qualitative results for Ad.-DASM_{0.25}, and Voxel2Mesh_{0.025}. Ad.-DASM_{0.25} easily outperforms Voxel2Mesh when segmenting pre and post synaptic regions. For the smaller synaptic junction, Voxel2Mesh and DASM are statistically equivalent because, unlike for the other two regions, their shapes are simple and there is not much improvement for DASM to make. In fact, the best result is obtained by a vanilla U-Net.

	Time (sec.)	$ V $
Voxel2Mesh	1.06 ± 0.04	1534 ± 4
Ad.-DASM	1.47 ± 0.06	1535 ± 2

Table 6: **Run times.** We report the time required to perform a forward and backward pass. $|V|$ is the number of mesh vertices.

Computation time We report average execution time for a single forward and backward pass for Voxel2Mesh and DASM in Table 6. We run this test on a single Tesla V100 GPU. We have implemented the ASM module using custom CUDA kernels and uses sparse tensors. Assembling the regularization matrix and performing the update of Eq. 11 adds a 40% overhead, which is reasonable giving how large the matrices we deal with are.

5. Conclusion

We have developed an approach to incorporating Active Shape Models into layers that can be integrated seamlessly into Graph Convolutional Networks to enforce sophisticated smoothness priors at an acceptable computational cost. By embedding the smoothing directly into the update equations, we can deliver smoother and more accurate meshes than methods that rely solely on a regularization loss term.

In future work, we will further improve our adaptive DASM scheme. Currently, it relies on a hand-designed metric to detect where to smooth and where not to. We will replace it by an auxiliary network that predicts where smoothing is required, given the current state of the mesh and the input data.

6. Acknowledgment

This work was supported in part by the Swiss National Science Foundation.

References

- [1] D. Acuna, A. Kar, and S. Fidler. Devil is in the Edges: Learning Semantic Boundaries from Noisy Annotations. In *Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [2] A. Kavur and M. Selver. CHAOS Challenge - Combined (CT-MR) Healthy Abdominal Organ Segmentation. 2020. 7
- [3] A. Chang, T. Funkhouser, L. G., P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. Shapenet: An Information-Rich 3D Model Repository. In *arXiv Preprint*, 2015. 5
- [4] Z. Chen and H. Zhang. Learning Implicit Fields for Generative Shape Modeling. In *Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [5] D. Cheng, R. Liao, S. Fidler, and R. Urtasun. DARNet: Deep Active Ray Network for Building Segmentation. In *Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [6] J. Chibane, T. Alldieck, and G. Pons-Moll. Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion. In *Conference on Computer Vision and Pattern Recognition*, 2020. 2
- [7] C. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese. 3D-R2N2: A unified approach for single and multi-view 3d object reconstruction. In *European Conference on Computer Vision*, 2016. 5
- [8] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 424–432, 2016. 7
- [9] L.D. Cohen and I. Cohen. Finite-Element Methods for Active Contour Models and Balloons for 2D and 3D Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993. 3
- [10] S. Dong and H. Zhang. A Combined Fully Convolutional Networks and Deformable Model for Automatic Left Ventricle Segmentation Based on 3D Echocardiography. In *BioMed Research International*, 2018. 2
- [11] J. Duchi, E. Hazan, and Y. Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. In *Journal of Machine Learning Research*, 2011. 4
- [12] P. Fua. Model-Based Optimization: Accurate and Consistent Site Modeling. In *International Society for Photogrammetry and Remote Sensing*, July 1996. 2
- [13] P. Fua and Y. G. Leclerc. Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading. *International Journal of Computer Vision*, 16:35–56, September 1995. 2
- [14] G. Gkioxari, J. Malik, and J. Johnson. Mesh R-CNN. In *International Conference on Computer Vision*, 2019. 1, 5, 6
- [15] A. Hatamizadeh, D. Sengupta, and D. Terzopoulos. End-To-End Trainable Deep Active Contour Models for Automated Image Segmentation: Delineating Buildings in Aerial Imagery. In *arXiv Preprint*, 2020. 2
- [16] L. He, Z. Peng, B. E., X. Wang, C. Y. Han, K. L. Weiss, and W. G. Wee. A Comparative Study of Deformable Contour Methods on Medical Image Segmentation. *Image and Vision Computing*, 26(2):141–163, 2008. 2
- [17] V. Igloukov and A. Shvets. Terausnet: U-Net with VGG11 Encoder Pre-Trained on Imagenet for Image Segmentation. In *arXiv Preprint*, 2018. 7
- [18] A. Jorstad, B. Nigro, C. Cali, M. Wawrzyniak, P. Fua, and G.W. Knott. Neuromorph: A Toolset for the Morphometric Analysis and Visualization of 3D Models Derived from Electron Microscopy Image Stacks. *Neuroinformatics*, 13(1):83–92, 2014. 2
- [19] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1988. 2, 3, 4
- [20] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015. 5, 8
- [21] R. Lengagne, P. Fua, and O. Monga. Using Differential Constraints to Reconstruct Complex Surfaces from Stereo. In *Conference on Computer Vision and Pattern Recognition*, 1997. 3
- [22] R. Lengagne, P. Fua, and O. Monga. 3D Stereo Reconstruction of Human Faces Driven by Differential Constraints. *Image and Vision Computing*, 18(4):337–343, March 2000. 3
- [23] M. E. Leventon, W. E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *Conference on Computer Vision and Pattern Recognition*, pages 316–323, 2000. 2
- [24] J. Liang, N. Homyounfar, W. Ma, Y. Xiong, R. Hu, and R. Urtasun. Polytransform: Deep Polygon Transformer for Instance Segmentation. In *Conference on Computer Vision and Pattern Recognition*, 2020. 2
- [25] H. Ling, J. Gao, A. Kar, W. Chen, and S. Fidler. Fast Interactive Object Annotation with Curve-Gcn. In *Conference on Computer Vision and Pattern Recognition*, pages 5257–5266, 2019. 2
- [26] W.E. Lorensen and H.E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In *ACM SIGGRAPH*, pages 163–169, 1987. 2
- [27] D. Marcos, D. Tuia, B. Kellenberger, and R. Urtasun. Learning Deep Structured Active Contours End-To-End. In *Conference on Computer Vision and Pattern Recognition*, 2018. 2
- [28] T. Mcinerney and D. Terzopoulos. A Dynamic Finite Element Surface Model for Segmentation and Tracking in Multidimensional Medical Images with Application to Cardiac 4D Image Analysis. *Computerized Medical Imaging and Graphics*, 19(1):69–83, 1995. 2
- [29] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019. 2
- [30] F. Milletari, N. Navab, and S.-A. Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *arXiv Preprint*, June 2016. 7
- [31] T.S. Newman and H. Yi. A Survey of the Marching Cubes Algorithm. *Computers & Graphics*, 30(5):854–879, 2006. 2

- [32] J. Pan and K. Jia. Deep Mesh Reconstruction from Single RGB Images via Topology Modification Networks. In *International Conference on Computer Vision*, 2019. 1, 2
- [33] J. J. Park, P. Florence, J. Straub, R. A. Newcombe, and S. Lovegrove. Deepsdf: Learning Continuous Signed Distance Functions for Shape Representation. In *Conference on Computer Vision and Pattern Recognition*, 2019. 2
- [34] S. Peng, W. Jiang, H. Pi, X. Li, H. Bao, and X. Zhou. Deep Snake for Real-Time Instance Segmentation. In *Conference on Computer Vision and Pattern Recognition*, 2020. 2
- [35] R. Prevost, R. Cuingnet, B. Mory, D. L.D. C., and R. Ardon. Incorporating Shape Variability in Image Segmentation via Implicit Template Deformation. *Conference on Medical Image Computing and Computer Assisted Intervention*, pages 82–89, 2013. 2
- [36] E. Remelli, A. Lukoianov, S. Richter, B. Guillard, T. Bagautdinov, P. Baque, and P. Fua. Meshsdf: Differentiable Iso-Surface Extraction. In *Advances in Neural Information Processing Systems*, 2020. 2
- [37] A. Shvets, A. Rakhlin, A. Kalinin, and V. Iglovikov. Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning. In *arXiv Preprint*, 2018. 7
- [38] D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-Seeking Models and 3D Object Reconstruction. *International Journal of Computer Vision*, 1:211–221, 1987. 2
- [39] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion. *Artificial Intelligence*, 36(1):91–123, 1988. 2
- [40] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y. Jiang. Pixel2mesh: Generating 3D Mesh Models from Single RGB Images. In *European Conference on Computer Vision*, 2018. 1, 2, 5, 6
- [41] C. Wen, Y. Zhang, Z. Li, and Y. Fu. Pixel2mesh++: Multi-View 3D Mesh Generation via Deformation. In *International Conference on Computer Vision*, 2019. 1, 2
- [42] U. Wickramasinghe, E. Remelli, G. Knott, and P. Fua. Voxel2mesh: 3d mesh model generation from volumetric data. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2020. 1, 2, 5, 7, 8
- [43] Q. Xu, W. Wang, D. Ceylan, R. Mech, and U. Neumann. DISN: Deep Implicit Surface Network for High-Quality Single-View 3D Reconstruction. In *Advances in Neural Information Processing Systems*, 2019. 2