

# Supplementary Material for RayNet: Learning Volumetric 3D Reconstruction with Ray Potentials

Despoina Paschalidou<sup>1,5</sup> Ali Osman Ulusoy<sup>2</sup> Carolin Schmitt<sup>1</sup> Luc van Gool<sup>3</sup> Andreas Geiger<sup>1,4</sup>

<sup>1</sup>Autonomous Vision Group, MPI for Intelligent Systems and University of Tübingen

<sup>2</sup>Microsoft <sup>3</sup>Computer Vision Lab, ETH Zürich & KU Leuven

<sup>4</sup>Computer Vision and Geometry Group, ETH Zürich

<sup>5</sup>Max Planck ETH Center for Learning Systems

{firstname.lastname}@tue.mpg.de vangool@vision.ee.ethz.ch

## Abstract

*In this supplementary document, we present our inference algorithm as well as additional qualitative results. We start by deriving the message expressions for the sum-product belief propagation algorithm used for approximate inference in our Markov Random Field. Subsequently, we show additional results from the aerial dataset from different viewpoints and we compare RayNet with commonly used disparity methods like ZNCC, with a probabilistic method [5] and with a learning-based approach [2]. In all cases, our method outperforms the baseline methods, thus resulting in significantly smoother reconstructions, while retaining sharp object boundaries. Finally, we present visualizations of the completeness of the predicted reconstructions on the DTU dataset and we compare against SurfaceNet [3].*

## 1. Markov Random Field Inference

Our inference algorithm is based on the sum-product belief propagation [4]. The detailed derivations of the belief propagation equations is presented in Section 1.1.

### 1.1. Sum-product Message Derivations

The sum-product algorithm for factor graphs works by passing messages between factors and variables. The general form of messages is given by

$$\mu_{f \rightarrow x}(x) = \sum_{X_f \setminus x} \phi_f(X_f) \prod_{y \in X_f \setminus x} \mu_{y \rightarrow f}(y) \quad (1)$$

$$\mu_{x \rightarrow f}(x) = \prod_{g \in F_x \setminus f} \mu_{g \rightarrow x}(x) \quad (2)$$

where  $X_f$  denotes all variables associated with factor  $f$  and  $F_x$  is the set of factors to which variable  $x$  connects. Upon termination (namely either at convergence or until a maximum number of iterations has been reached), the approximate marginal distribution of each variable can be computed as the product of messages from all neighbouring factors according to Equation 3.

$$p(x) \propto \prod_{g \in F_x} \mu_{g \rightarrow x}(x) \quad (3)$$

In our case, the factor graph of the MRF consists of the occupancy variables  $\mathbf{o} = \{o_i | i \in \mathcal{X}\}$ , where  $\mathcal{X}$  is the set of voxels in the voxel grid, the unary factors  $\varphi_i(o_i)$ , defined for every voxel  $i \in \mathcal{X}$  and the ray factors  $\psi_r(\mathbf{o}_r)$ , defined for every ray  $r \in \mathcal{R}$ , where  $\mathcal{R}$  is the set of all rays and  $\mathbf{o}_r = (o_1^r, o_2^r, \dots, o_{N_r}^r)$  denotes the ordered set of occupancy variables associated with the voxels which intersect ray  $r$ .

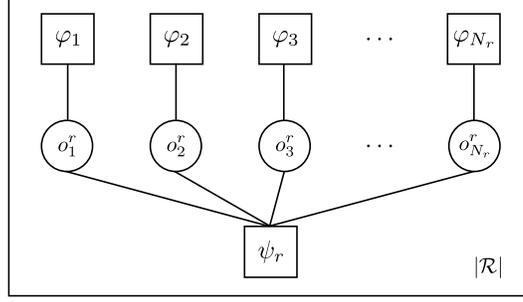


Figure 1: **Factor Graph of the MRF for a single ray.** Unary factors  $\varphi_i(o_i)$  are defined for every voxel  $i \in \mathcal{X}$  of the voxel grid. Ray factor  $\psi_r(\mathbf{o}_r)$  is defined for the query ray  $r \in \mathcal{R}$  and connects to the variables along the ray  $\mathbf{o}_r = (o_1^r, \dots, o_{N_r}^r)$ .

Consider a single ray  $r$  associated with a ray factor  $\psi_r(\mathbf{o}_r)$ . The ray factor  $\psi_r$  is associated with  $N_r$  occupancy variables, namely this ray passes through  $N_r$  voxels. The corresponding factor graph is illustrated in Fig. 1.

### 1.1.1 Ray factor to occupancy variables messages $\mu_{\psi_r \rightarrow o_i^r}$

The ray potentials encourage the predicted depth at pixel/ray  $r$  to coincide with the first occupied voxel along the ray, thus we can rewrite the expression for the ray potentials as follows

$$\psi_r(\mathbf{o}_r) = \sum_{i=1}^{N_r} o_i^r \prod_{j<i} (1 - o_j^r) s_i^r = \begin{cases} s_1^r, & \text{if } o_1^r = 1 \\ s_2^r, & \text{if } o_1^r = 0, o_2^r = 1 \\ \dots & \\ s_{N_r}^r, & \text{if } o_1^r = 0, o_2^r = 0, \dots, o_{N_r-1}^r = 0, o_{N_r}^r = 1 \end{cases} \quad (4)$$

where  $s_i^r$  is the probability that the surface intersects ray  $r$  at voxel  $i$  as predicted by the neural network described in the main submission.

We first derive the message equations from the ray factor to the  $i$ th occupancy variable, when it is occupied ( $o_i^r = 1$ ). Eq. 1 becomes

$$\mu_{\psi_r \rightarrow o_i^r}(o_i^r = 1) = \sum_{o_1^r} \dots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \dots \sum_{o_{N_r}^r} \psi_r(\mathbf{o}_r) \prod_{\substack{j=1 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \quad (5)$$

We carry out the summations over the occupancy variables one by one. We begin the summation over  $o_1^r$  according to Equation 5

$$\begin{aligned} \mu_{\psi_r \rightarrow o_i^r}(o_i^r = 1) &= \overbrace{\mu_{o_1^r \rightarrow \psi_r}(o_1^r = 1) \left[ \sum_{o_2^r} \dots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \dots \sum_{o_{N_r}^r} \psi_r(o_1^r = 1, \dots, o_{N_r}^r) \prod_{\substack{j=2 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right]}^{(\diamond)} + \\ &\quad \underbrace{\mu_{o_1^r \rightarrow \psi_r}(o_1^r = 0) \left[ \sum_{o_2^r} \dots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \dots \sum_{o_{N_r}^r} \psi_r(o_1^r = 0, \dots, o_{N_r}^r) \prod_{\substack{j=2 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right]}_{(\dagger)} \end{aligned} \quad (6)$$

In the first term of Eq. 6, marked with  $(\diamond)$  we can replace the ray potential  $\psi_r(o_1^r = 1, \dots, o_{N_r}^r = 0) = s_1^r$  from Eq. 4. Since  $s_1^r$  only depends on the  $o_1^r$ , it can be brought out of the summations over the other occupancy variables. Furthermore,

the remaining term  $\sum_{o_2^r} \dots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \dots \sum_{o_{N_r}^r} \prod_{\substack{j=2 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r)$  evaluates to 1 because we assume that all incoming messages are normalized such that they sum to 1. The last term, marked with  $(\dagger)$ , does not simplify.

Subsequently, we expand the summation over  $o_2^r$

$$\begin{aligned} \mu_{\psi_r \rightarrow o_i^r}(o_i^r = 1) &= \mu_{o_1^r \rightarrow \psi_r}(o_1^r = 1)s_1^r + \mu_{o_1^r \rightarrow \psi_r}(o_1^r = 0) \left[ \right. \\ &\quad \mu_{o_2^r \rightarrow \psi_r}(o_2^r = 1) \left[ \sum_{o_3^r} \cdots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi_r(o_1^r = 0, o_2^r = 1, \dots, o_{N_r}^r) \prod_{\substack{j=3 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right] + \\ &\quad \left. \mu_{o_2^r \rightarrow \psi_r}(o_2^r = 0) \left[ \sum_{o_3^r} \cdots \sum_{o_{i-1}^r} \sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi_r(o_1^r = 0, o_2^r = 0, \dots, o_{N_r}^r) \prod_{\substack{j=3 \\ j \neq i}}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right] \right] \end{aligned} \quad (7)$$

The second term of Eq. 7 can be further simplified by replacing the ray potential from Eq. 4 and by assuming normalized assuming messages with  $\mu_{o_1^r \rightarrow \psi_r}(o_1^r = 0)\mu_{o_2^r \rightarrow \psi_r}(o_2^r = 1)s_2^r$ .

We continue carrying out the summations until (including) the  $o_{i-1}^r$  occupancy variables and we get

$$\begin{aligned} \mu_{\psi_r \rightarrow o_i^r}(o_i^r = 1) &= \sum_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 1) \prod_{k=1}^{j-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0)s_j^r + \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) \times \\ &\quad \left[ \underbrace{\sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1, \dots, o_{N_r}^r)}_{(\ddagger)} \prod_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right] \end{aligned} \quad (8)$$

In the last term of Eq. 8, marked with  $(\ddagger)$ , we can substitute the ray potential  $\psi(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1, \dots, o_{N_r}^r)$  with  $s_i^r$  and we get the final form of the message when the  $i$ th voxel is occupied as follows

$$\mu_{\psi_r \rightarrow o_i^r}(o_i^r = 1) = \sum_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 1) \prod_{k=1}^{j-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0)s_j^r + \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0)s_i^r \quad (9)$$

Following a similar technique we derive the message expression when the  $i$ th voxel is free, namely  $(o_i^r = 0)$  as follows

$$\begin{aligned} \mu_{\psi_r \rightarrow o_i^r}(o_i^r = 0) &= \sum_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 1) \prod_{k=1}^{j-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0)s_j^r + \prod_{k=1}^{i-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0) \times \\ &\quad \left[ \underbrace{\sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 0, \dots, o_{N_r}^r)}_{(\text{II})} \prod_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r) \right] \end{aligned} \quad (10)$$

The last term of Eq. 10, marked with (II) can be simplified by summing out all occupancy variables one by one as follows

$$\mu_{o_{i+1}^r \rightarrow \psi_r}(o_{i+1}^r = 1)s_{i+1}^r + \cdots + \mu_{o_{i+1}^r \rightarrow \psi_r}(o_{i+1}^r = 0) \cdots \mu_{o_{N_r-1}^r \rightarrow \psi_r}(o_{N_r-1}^r = 0) \mu_{o_{N_r}^r \rightarrow \psi_r}(o_{N_r}^r = 1)s_{N_r}^r \quad (11)$$

The resulting expression when the  $i$ th voxel is free becomes

$$\mu_{\psi_r \rightarrow o_i^r}(o_i^r = 0) = \sum_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 1) \prod_{k=1}^{j-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0)s_j^r + \sum_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 1) \prod_{\substack{k=1 \\ k \neq i}}^{j-1} \mu_{o_k^r \rightarrow \psi_r}(o_k^r = 0)s_j^r \quad (12)$$

### 1.1.2 Occupancy variable to ray factor messages $\mu_{o_i^r \rightarrow \psi_r}$

Using Equation 2, we derive the message from the  $i$ th occupancy variable to the ray factor  $\psi_r$ . This message can be defined as the product of all messages from the set of factors  $F_{o_i^r}$  to which the occupancy variable  $o_i^r$  is connected (except for the  $\psi_r$  factor). The  $F_{o_i^r}$  set consists of the unary to occupancy variable message, namely  $\mu_{\varphi_i \rightarrow o_i^r}(o_i^r)$  as well as all the factor-to-occupancy messages for all the  $k \in \mathcal{L}$  rays, except for the  $r$  ray.

$$\mu_{o_i^r \rightarrow \psi_r}(o_i^r) = \varphi_i(o_i^r) \prod_{\psi_k \in \mathcal{L} \setminus \psi_r} \mu_{\psi_k \rightarrow o_i^r}(o_i^r) \quad (13)$$

### 1.1.3 Messages from and to the unary factors $\mu_{\varphi_i \rightarrow o_i^r}$ and $\mu_{o_i^r \rightarrow \varphi_i}$

The messages from and to the unary factors are straightforward to compute using Equations 1 and 2. The message from the occupancy variable  $o_i^r$  to the unary factor  $\varphi_i$  is defined as the product of all messages from the set of factors  $F_{o_i^r}$  to which the occupancy variable  $o_i^r$  is connected (except for the  $\varphi_i$  factor). The set of factors  $F_{o_i^r}$  includes all the  $\psi_r(\cdot)$  factors, one for each ray that might pass through the  $i^{\text{th}}$  voxel. Let  $\mathcal{L}$  be the set of rays that pass through the  $i$ th voxel, the  $\mu_{o_i^r \rightarrow \varphi_i}$  becomes

$$\mu_{o_i^r \rightarrow \varphi_i}(o_i^r) = \prod_{g \in F_{o_i^r} \setminus \varphi_i} \mu_{g \rightarrow o_i^r}(o_i^r) = \prod_{\psi_r \in \mathcal{L}} \mu_{\psi_r \rightarrow o_i^r}(o_i^r) \quad (14)$$

Regarding the message from unary factor  $\varphi_i$  to the occupancy variable  $o_i^r$ , each unary factor is associated with the occupancy variable of the  $i$ th voxel  $o_i^r$ , thus the unary factor to occupancy variable message is expressed as

$$\mu_{\varphi_i \rightarrow o_i^r}(o_i^r) = \varphi_i(o_i^r) = \gamma^{o_i^r} (1 - \gamma)^{1 - o_i^r} \quad (15)$$

## 1.2. Depth Estimation Derivation

In this section, we provide details for our depth estimation procedure. As described in Section 3.2 of the submission, we associate the occupancy and depth variables along a ray using the Eq. 16.

$$d_r = \sum_{i=1}^{N_r} o_i^r \prod_{j < i} (1 - o_j^r) d_i^r \quad (16)$$

Using the sum-product belief propagation and the message expressions derived in Section 1.1 we estimate a probability distribution  $p(d_r = d_i^r)$  for each depth variable  $d_r$ .

Eq. 16 encourages that the predicted depth at pixel/ray  $r$  coincide with the first occupied voxel along the ray, thus it can be rewritten as follows

$$d_r = \sum_{i=1}^{N_r} o_i^r \prod_{j < i} (1 - o_j^r) d_i^r = \begin{cases} d_1^r, & \text{if } o_1^r = 1 \\ d_2^r, & \text{if } o_1^r = 0, o_2^r = 1 \\ \dots & \\ d_{N_r}^r, & \text{if } o_1^r = 0, o_2^r = 0, \dots, o_{N_r-1}^r = 0, o_{N_r}^r = 1 \end{cases} \quad (17)$$

Taking into account the above expression for the depth formation process and the fact that we always assume that there exists at least one occupied voxel along the ray, the probability distribution  $p(d_r = d_i^r)$  becomes

$$p(d_r = d_i^r) = p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) \quad (18)$$

We can easily estimate the marginal distribution of the occupancy variables from Eq. 3 as follows

$$p(o_1^r, \dots, o_{N_r}^r) \propto \psi_r(\mathbf{o}) \prod_{i=1}^{N_r} \mu_{o_i^r \rightarrow \psi_r}(o_i^r) \quad (19)$$

However, we only want to estimate the  $p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) \forall i = 1, \dots, N_r$ . This can be done by simply summing out all the other occupancy variables from Eq. 19

$$\begin{aligned}
p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) &= \sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1, o_{i+1}^r, \dots, o_{N_r}^r) \\
&\propto \sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i = 1, o_{i+1}^r, \dots, o_{N_r}^r) \\
&\times \underbrace{\left[ \mu_{o_i^r \rightarrow \psi_r}(o_i^r = 1) \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) \right]}_{(\square)} \prod_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r)
\end{aligned} \tag{20}$$

Note that the term marked with  $(\square)$  does not depend on the occupancy variables summed, hence it can be taken out of the summations as follows

$$\begin{aligned}
p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) &\propto \mu_{o_i^r \rightarrow \psi_r}(o_i^r = 1) \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) \times \\
&\underbrace{\sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \psi(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1, o_{i+1}^r, \dots, o_{N_r}^r) \prod_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r)}_{(\dagger)}
\end{aligned} \tag{21}$$

The ray potential  $\psi(o_1^r = 0, \dots, o_i^r = 1, o_{i+1}^r, \dots, o_{N_r}^r)$  term in the expression marked with  $(\dagger)$  can be replaced with  $s_i^r$  from Eq. 4, hence it can also be taking out of the summations, resulting in

$$p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) \propto \mu_{o_i^r \rightarrow \psi_r}(o_i^r = 1) \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) s_i^r \times \underbrace{\sum_{o_{i+1}^r} \cdots \sum_{o_{N_r}^r} \prod_{j=i+1}^{N_r} \mu_{o_j^r \rightarrow \psi_r}(o_j^r)}_{\text{this term evaluates to 1}} \tag{22}$$

However, as we have already mentioned before all incoming messages are normalized such that the sum to 1, thus the joint probability simplifies to

$$p(o_1^r = 0, \dots, o_{i-1}^r = 0, o_i^r = 1) \propto \mu_{o_i^r \rightarrow \psi_r}(o_i^r = 1) \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) s_i^r \tag{23}$$

By substituting Eq. 23 in Eq. 18, we get the final expression for the probability distribution

$$p(d_r = d_i^r) = \frac{1}{Z} \mu_{o_i^r \rightarrow \psi_r}(o_i^r = 1) \prod_{j=1}^{i-1} \mu_{o_j^r \rightarrow \psi_r}(o_j^r = 0) s_i^r \tag{24}$$

where  $Z$  is a normalization constant such that the depth distribution sums to 1.

## 2. Aerial Dataset

In this Section, we provide additional qualitative results on the aerial dataset. We compare RayNet with the zero-mean normalized cross correlation (ZNCC), which is a commonly used disparity method, calculated over multiple input views. In addition, we also compare our method with the probabilistic 3D reconstruction method by Ulusoy et al. [5]. We further compare against the learning based approach of Hartmann et al. [2]. For more detailed analysis regarding the implementations of these baselines please refer to Section 4.1 of our submission.

Fig. 2 and Fig. 3 depict the predicted depth maps for all methods for a set of images from different viewpoints. In all images, both ZNCC and the Hartmann et al. [2] baselines result in smoother depth maps with artefacts around the image boundaries. In contrast, our CNN baseline as well as the probabilistic approach by Ulusoy et al [5] yield sharpened boundaries with significant noise. In general, facades of the buildings are considered to be problematic areas, because they lack any texture. It is important to point out, that also in this case (Fig. 3m), our method results in more accurate reconstructions compared to the other baselines.

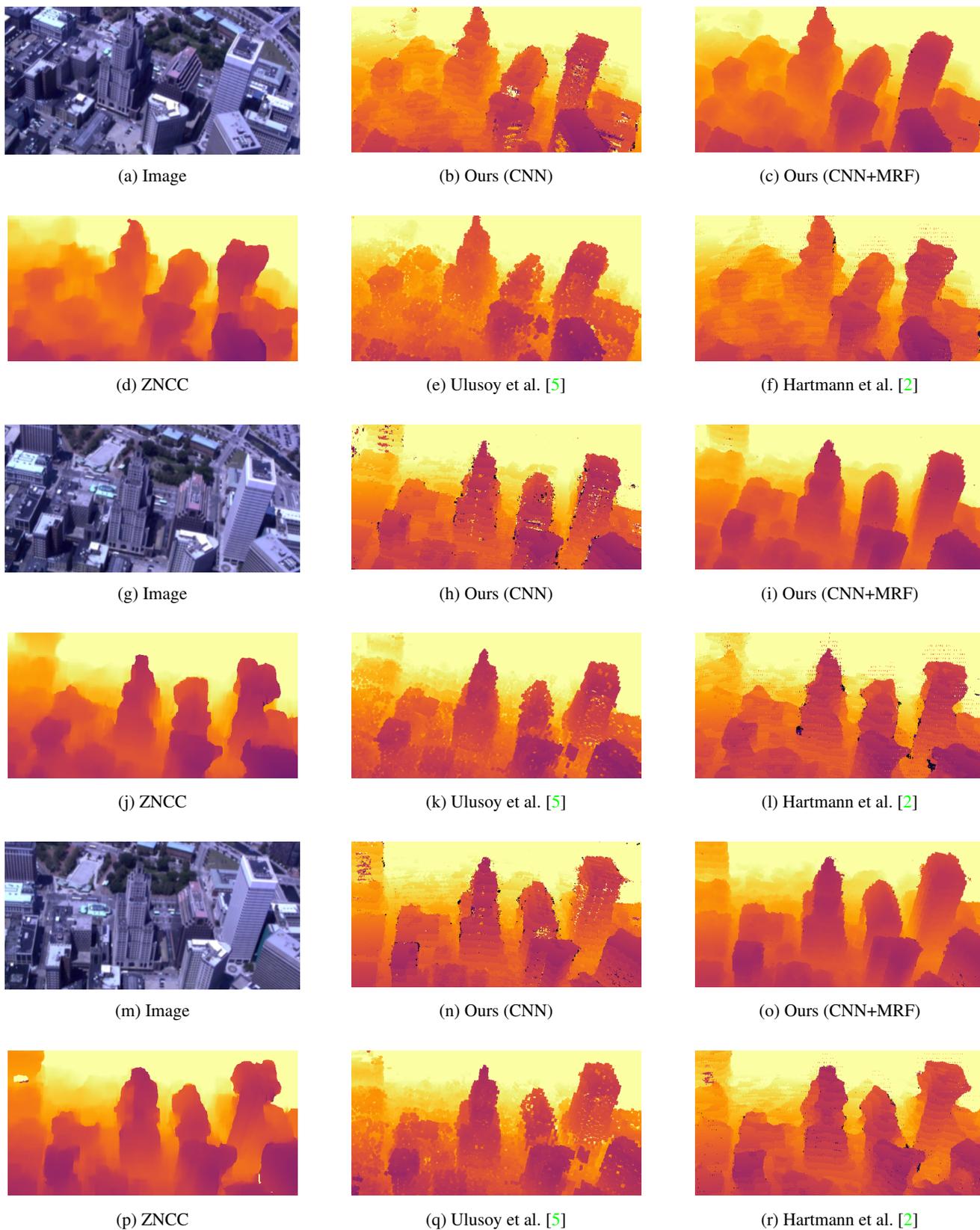


Figure 2: **Qualitative Results on Aerial Dataset** for images from different viewpoints. We show the depth maps predicted by our method (b+c,i+h, n+o) as well as three baselines (d-f,j-l, p-r) for three input images from different viewpoints (a,g, m). Darker colours correspond to closer regions.



Figure 3: **Qualitative Results on Aerial Dataset** for images from different viewpoints. We show the depth maps predicted by our method (b+c,i+h, n+o) as well as three baselines (d-f,j-l,p-r) for two input images from different viewpoints (a,g,m). Darker colours correspond to closer regions.

### 3. DTU Dataset

In this Section, we show additional qualitative results on the DTU dataset. We evaluate RayNet on six objects from this dataset: BUNNY, STONE EDGE, BIRD, SNOWMEN, BRICKS and BUDDHA and we compare the predicted reconstructions against SurfaceNet [3] on the same resolution, namely  $256^3$ . Fig. 4 shows the six objects.



(a) BUNNY



(b) STONEHENGE



(c) BIRD



(d) SNOWMEN



(e) BRICKS



(f) BUDDHA

Figure 4: **DTU dataset**. Sample images from DTU [1].

In all experiments, we observe that RayNet leads always to complete reconstructions, while SurfaceNet results in missing parts. It is also important to note that in the case of SurfaceNet we use the images in their original size, namely  $1400 \times 1600$ , because when we tried to use the downsampled images the reconstruction failed completely. However, this exposes SurfaceNet to more than 6 times more information/pixels compared to RayNet, which uses the downsampled images.

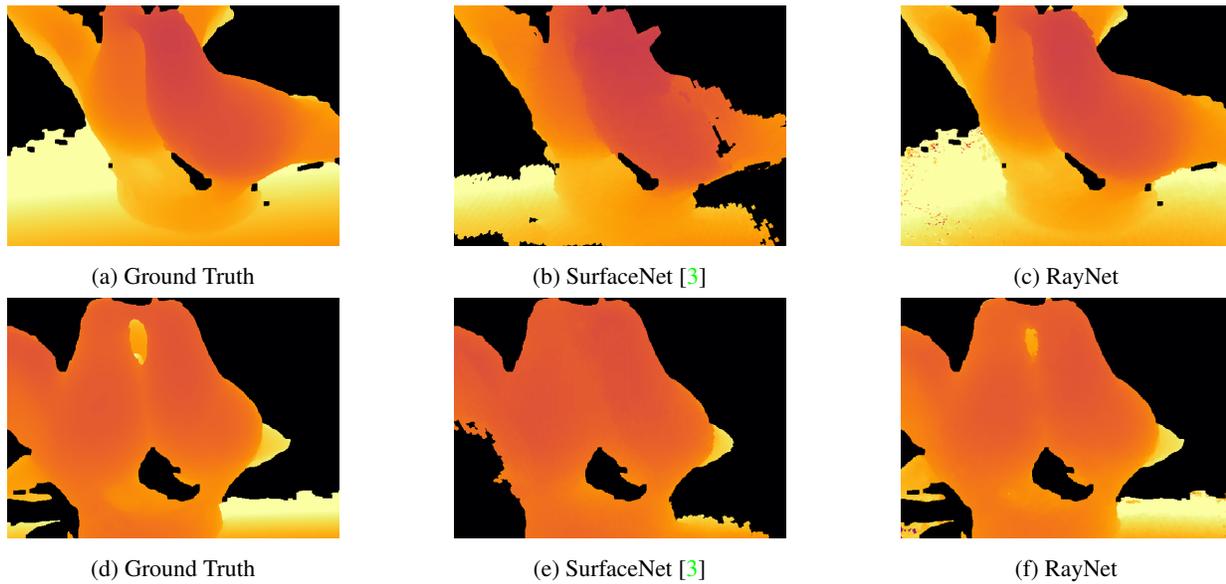


Figure 5: **Qualitative Results on the BIRD.** We visualize the ground-truth (a,d) and the depth reconstructions using SurfaceNet (LR) (b,e) and RayNet (c,f)

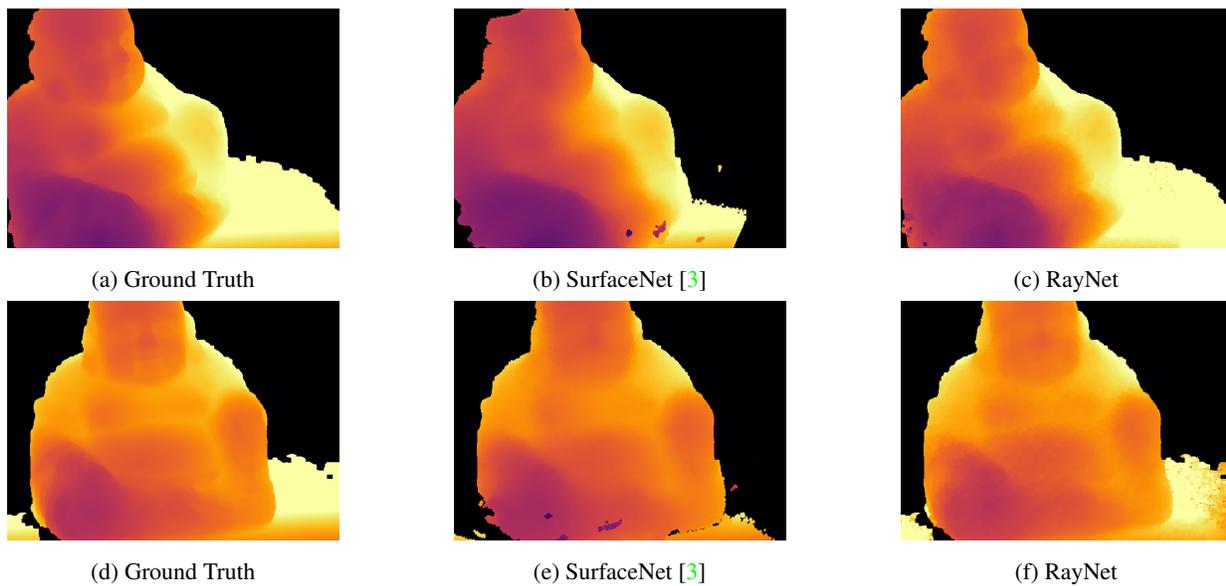


Figure 6: **Qualitative Results on the BUDDHA.** We visualize the ground-truth (a,d) and the depth reconstructions using SurfaceNet (LR) (b,e) and RayNet (c,f)

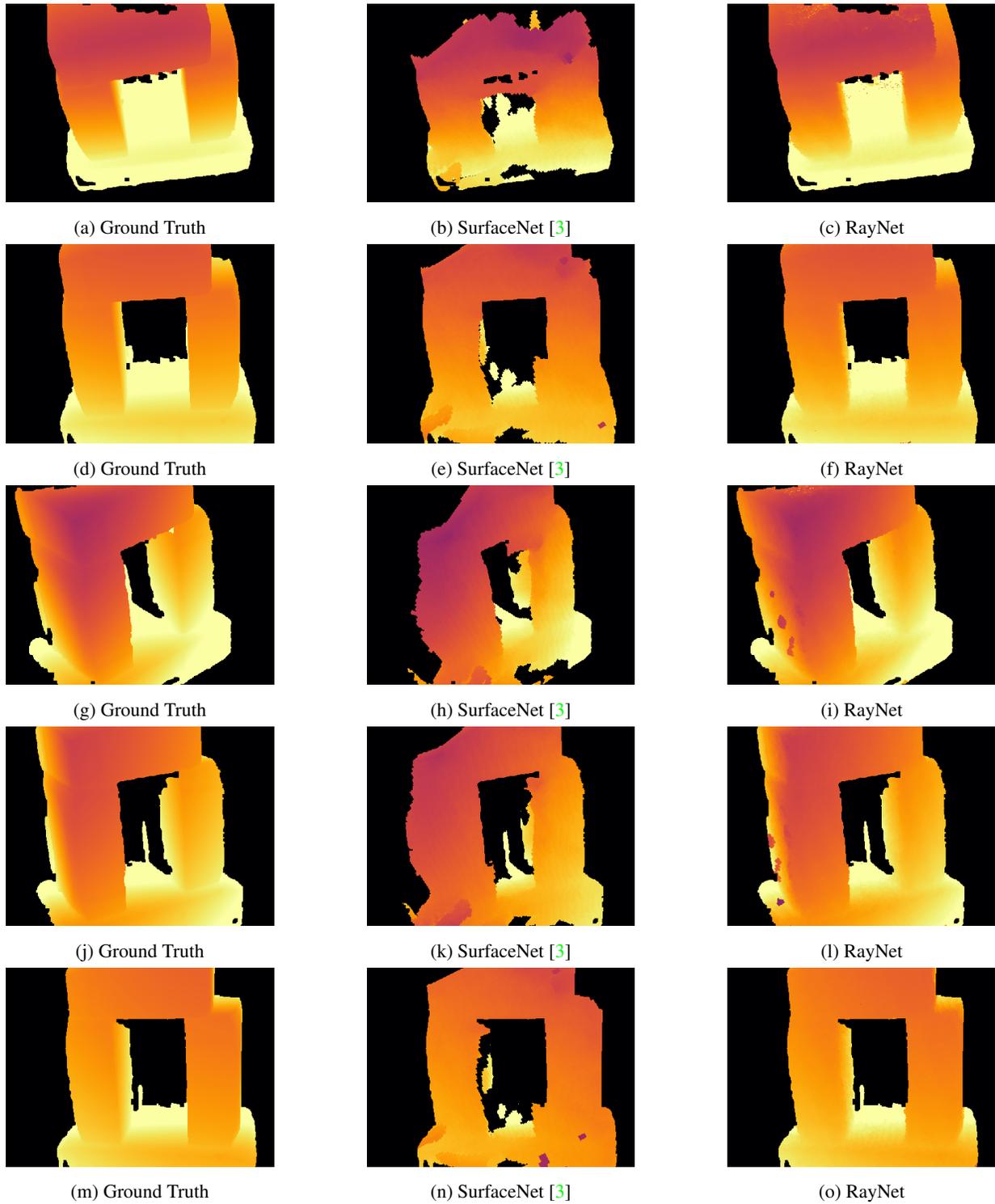


Figure 7: **Qualitative Results on the STONEHENGE.** We visualize the ground-truth (a,d,g,j,m) and the depth reconstructions using SurfaceNet (LR) (b,e,h,k,n) and RayNet (c,f,i,l,o)

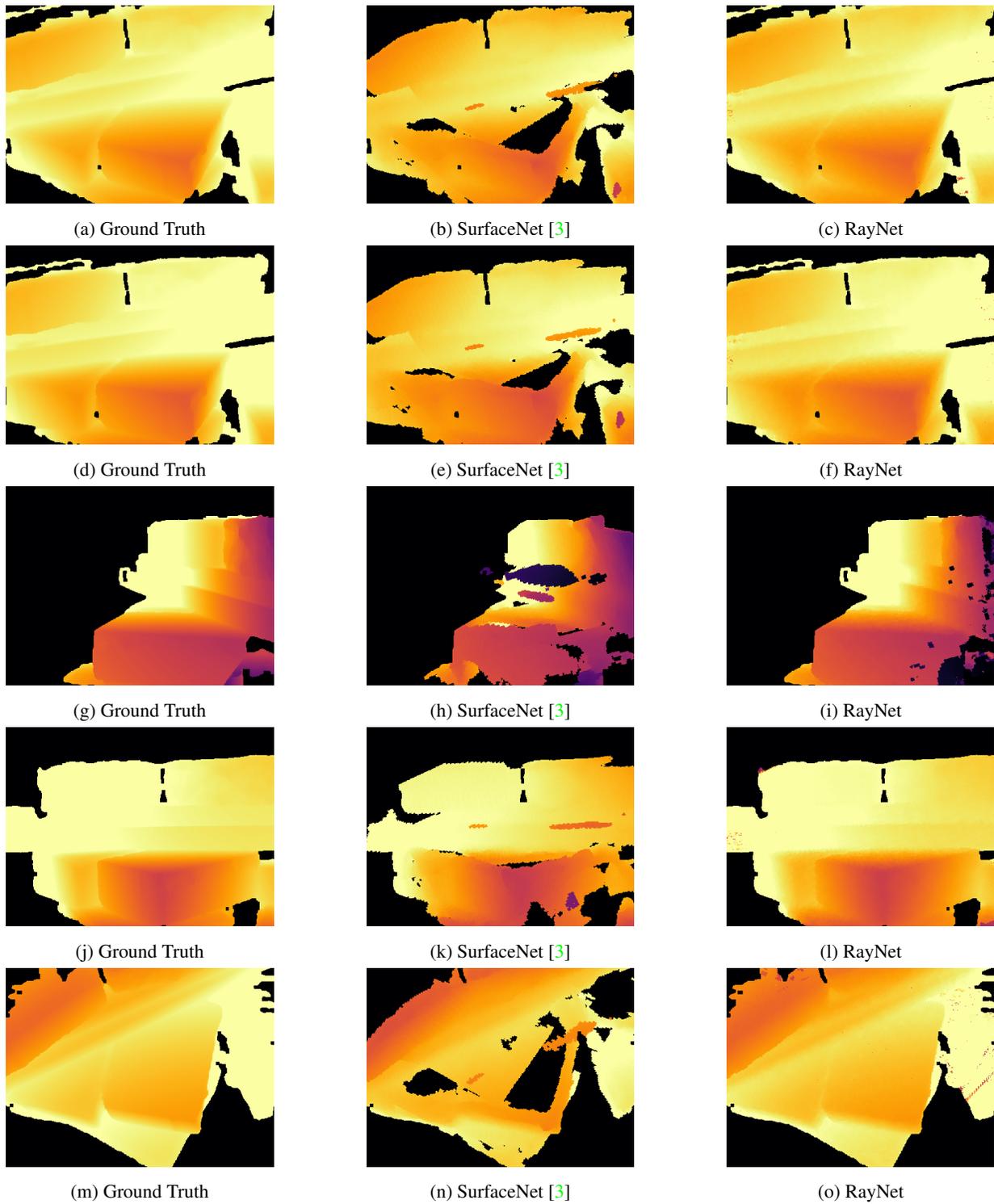


Figure 8: **Qualitative Results on the BRICKS.** We visualize the ground-truth (a,d,g,j,m) and the depth reconstructions using SurfaceNet (LR) (b,e,h,k,n) and RayNet (c,f,i,l,o)

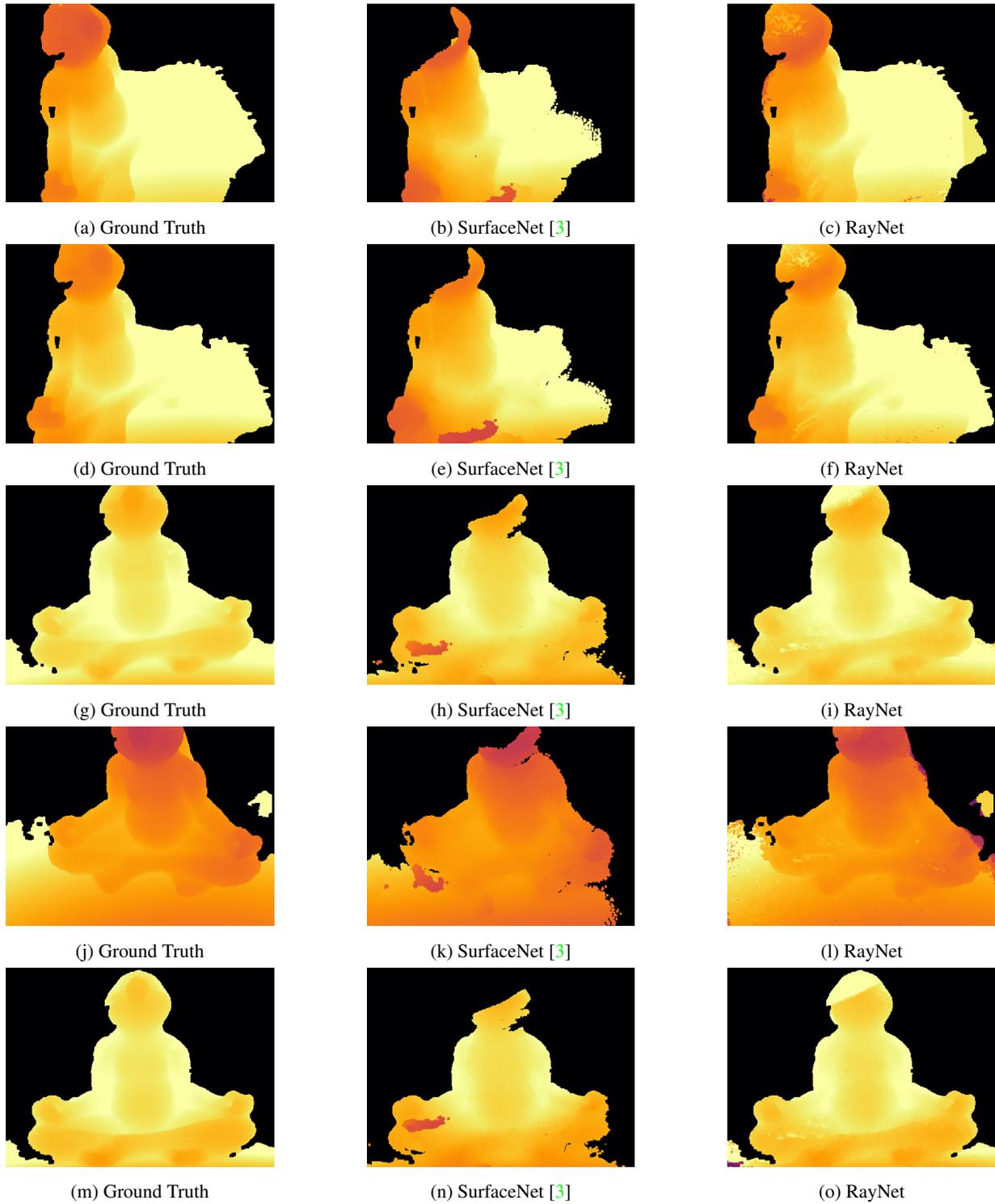


Figure 9: **Qualitative Results on the BUNNY.** We visualize the ground-truth (a,d) and the depth reconstructions using SurfaceNet (LR) (b,e) and RayNet (c,f)

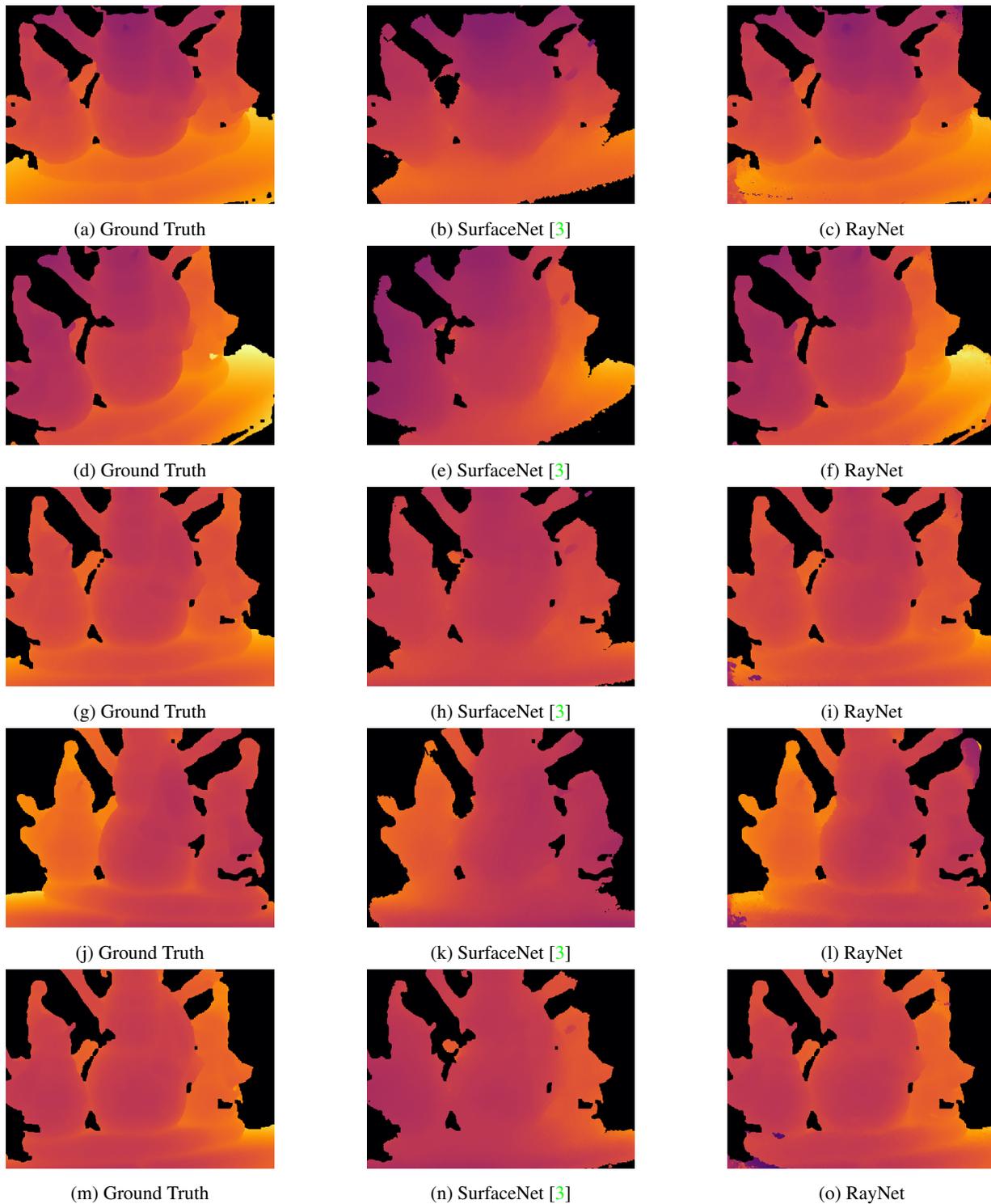


Figure 10: **Qualitative Results on the SNOWMEN.** We visualize the ground-truth (a,d) and the depth reconstructions using SurfaceNet (LR) (b,e) and RayNet (c,f)

## References

- [1] C. Häne, L. Heng, G. H. Lee, A. Sizov, and M. Pollefeys. Real-time direct dense matching on fisheye images using plane-sweeping stereo. In *Proc. of the International Conf. on 3D Vision (3DV)*, 2014. 8
- [2] W. Hartmann, S. Galliani, M. Havlena, L. Van Gool, and K. Schindler. Learned multi-patch similarity. In *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*, 2017. 1, 5, 6, 7
- [3] M. Ji, J. Gall, H. Zheng, Y. Liu, and L. Fang. SurfaceNet: an end-to-end 3d neural network for multiview stereopsis. In *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*, 2017. 1, 8, 9, 10, 11, 12, 13
- [4] S. Nowozin and C. H. Lampert. Structured learning and prediction in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 2011. 1
- [5] A. O. Ulusoy, A. Geiger, and M. J. Black. Towards probabilistic volumetric reconstruction using ray potentials. In *Proc. of the International Conf. on 3D Vision (3DV)*, 2015. 1, 5, 6, 7