# Building Scene Models by Completing and Hallucinating Depth and Semantics —Supplementary Material—

Miaomiao Liu[1], Xuming He[1], Mathieu Salzmann[2]

[1]Data61, CSIRO, and ANU, Australia,
CVLab, EPFL, Switzerland
[1]miaomiao.liu@data61.csiro.au, xuming.he@anu.edu.au,
[2]mathieu.salzmann@epfl.ch

In this document, we first provide mathematical details regarding our optimization of the visible and hidden layers. In particular, we discuss the derivation of $\mathbf{K}^{-1}$ in Section 1.3 and provide the proximal operators for the hidden layer in Section 2. We further give the detailed derivation of the proximal operators for the hidden layer as an example in Section 2.1. A similar derivation can be used for the other proximal operators. We finally provide some additional details about our experiments.

## 1 Details on optimizing the visible layer $\mathbf{u}^v$, $\mathbf{s}^v$

First, we review the optimization problem corresponding to our alternating procedure and then discuss the derivation of $\mathbf{K}^{-1}$.

### 1.1 Optimizing $\mathbf{u}^v$ with fixed $\mathbf{s}^v$

Recall that, when optimizing the visible layer in Section 4.2, we first fix the semantic variable $\mathbf{s}^v$ and optimize $\mathbf{u}^v$. This results in an objective in the standard Mumford-Shah form of Eq. (10) in the paper, with

$$D_{\mathbf{u}^v}(\mathbf{u}^v) = \sum_{\mathbf{x}} \|d \cdot (\mathbf{p}^T \mathbf{u}^v - y^o)\|^2 \,, \tag{1}$$

$$R_{\mathbf{u}^v}(\mathbf{K}\mathbf{u}^v) = \eta_{rv} \sum_{\mathbf{x}\in\Omega} \min(\alpha_1 \|\mathbf{K}\mathbf{u}^v\|^2 + e_{uv}, \lambda_1) \,, \tag{2}$$

where $e_{uv} = \|\mathbf{K}\mathbf{s}^v\|^2$. Here, $\|\mathbf{K}\mathbf{u}\|^2 := \sum_j \|\mathbf{K}\mathbf{u}_j\|^2$ denotes the Euclidean norm, where $\mathbf{u}_j$ is the $j$-th channel in the multi-channel variable $\mathbf{u}$.

The corresponding proximal operators can be written as

$$prox_{\tau, D_{\mathbf{u}^v}}(\tilde{\mathbf{u}}^v) = (\mathbf{I} + 2d\tau \mathbf{p}\mathbf{p}^T)^{-1}(\tilde{\mathbf{u}}^v + 2d\tau \mathbf{p} y^o) \,, \tag{3}$$

$$prox_{\sigma, R^*_{\mathbf{u}^v}}(\tilde{\mathbf{q}}_u) = \begin{cases} \frac{2\eta_{rv}\alpha_1}{\sigma + 2\eta_{rv}\alpha_1} \tilde{\mathbf{q}}_u & \text{if } |\tilde{\mathbf{q}}_u| \leq \sqrt{\frac{(\lambda_1 - e_{uv})\sigma(\sigma + 2\eta_{rv}\alpha_1)}{\alpha_1}} \\ 0 & \text{else,} \end{cases} \,, \tag{4}$$

where $\mathbf{I}$ denotes the identity matrix, $\tilde{\mathbf{u}}^v = (\mathbf{u}^v)^n - \tau_n \mathbf{K}^{-1}\mathbf{q}_u^{n+1}$ and $\tilde{\mathbf{q}}_u = \mathbf{q}_u^n + \sigma_n \mathbf{K}(\bar{\mathbf{u}}^v)^n$ as defined in Eqs. (13) and (12) in the paper, respectively. Note that $\mathbf{A}^{-1}$ becomes $\mathbf{K}^{-1}$, which equals $-div \cdot T_I$. In section 1.3, we provide the details for the derivation of $\mathbf{K}^{-1}$.

### 1.2 Optimizing $\mathbf{s}^v$ with fixed $\mathbf{u}^v$.

We fix $\mathbf{u}^v$ and optimize $\mathbf{s}^v$. This results in an objective in the standard Mumford-Shah form of Eq. (10) in the paper, with

$$D_{\mathbf{s}^v}(\mathbf{s}^v) = \sum_{\mathbf{x}} \eta_d \|(\mathbf{s}^v - \mathbf{s}^o)\|^2 + \eta_c \sum_{\mathbf{x}} (\mathbf{f}^T \mathbf{s}^v - m + b)^2 \,, \tag{5}$$

$$R_{\mathbf{s}^v}(\mathbf{K}\mathbf{s}^v) = \eta_{rv} \sum_{\mathbf{x}\in\Omega} \min(\alpha_1 e_{sv} + \|\mathbf{K}\mathbf{s}^v\|^2, \lambda_1) \,, \tag{6}$$

where $e_{sv} = \|\mathbf{K}\mathbf{u}^v\|^2$, and $\mathbf{f}$ is a binary vector with 1s in the position corresponding to the foreground classes and 0 everywhere else. The corresponding proximal operators are then given by

$$prox_{\tau, D_{\mathbf{s}^v}}(\tilde{\mathbf{s}}^v) = ((1 + 2\tau\eta_d)\mathbf{I} + 2\tau\eta_c \mathbf{f}\mathbf{f}^T)^{-1}(\tilde{\mathbf{s}}^v + 2\tau\eta_d \mathbf{s}^o + 2(m-b)\eta_c \tau \mathbf{f}) \,, \tag{7}$$

$$prox_{\sigma, R^\star_{\mathbf{s}^v}}(\tilde{\mathbf{q}}_s) = \begin{cases} \frac{2\eta_{rv}}{\sigma + 2\eta_{rv}} \tilde{\mathbf{q}}_s & \text{if } |\tilde{\mathbf{q}}_s| \leq \sqrt{(\lambda_1 - e_{sv}\alpha_1)\sigma(\sigma + 2\eta_{rv})} \,, \\ 0 & \text{else} \,, \end{cases} \tag{8}$$

where, again, $\tilde{\mathbf{s}}^v = (\mathbf{s}^v)^n - \tau_n \mathbf{K}^{-1}\mathbf{q}_s^{n+1}$ and $\tilde{\mathbf{q}}_s = \mathbf{q}_s^n + \sigma_n \mathbf{K}(\bar{\mathbf{s}}^v)^n$ as in in Eqs. (13) and (12) in the paper, respectively. In the following section, we discuss the detailed derivation of $\mathbf{K}^{-1}$.

### 1.3 Derivation of $\mathbf{K}^{-1}$

Recall that $\mathbf{K} = T_I \cdot \nabla$ denotes an image adaptive gradient operator. In particular,

$$T_I = \exp(-\beta |\nabla I|^\gamma)\mathbf{n}\mathbf{n}^T + \mathbf{n}^\perp \mathbf{n}^{\perp T}, \tag{9}$$

where $\mathbf{n} = \frac{\nabla I}{|\nabla I|}$ and $\mathbf{n}^\perp$ is the normal vector to the image gradient. Minimizing Eq. (16) in the paper is equivalent to solving the saddle-point problem

$$\min_{\mathbf{u}^v} \max_{\mathbf{q}} \;\; D_{\mathbf{u}^v}(\mathbf{u}^v) + < \mathbf{q}, \mathbf{K}\mathbf{u}^v > - R^\star_{\mathbf{u}^v}(\mathbf{q}), \tag{10}$$

where $\mathbf{u}^v$ denotes the primal variable and $\mathbf{q}$ the dual variable. To apply the primal-dual procedure, we need to perform gradient descent for the primal variable and gradient ascent for the dual variable. For the primal gradient, we use the equivalent form of the inner product $< \mathbf{q},\ \mathbf{K}\mathbf{u}^v > = < \mathbf{K}^{-1}\mathbf{q},\ \mathbf{u}^v >$. This enables us to compute the gradient efficiently in a pixel-wise manner.

Here we provide the details for the derivation of $\mathbf{K}^{-1}$. Without loss of generality, we consider a single channel function $u$ in the following. For multi-channel

functions, we have $\mathbf{K}\mathbf{u}^v = vec(\mathbf{K}u_k^v)_{k\in\{1,2,\cdots,K\}}$, where $u_k^v$ denotes the $k$-th channel of $\mathbf{u}^v$, and $vec(\cdot)$ denotes the vectorization operation.

Note that $T_I$ is a symmetric matrix and we can represent it as $T_I = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$. As typically done in practice, we compute $\nabla u$ by finite difference. This can be written as

$$\nabla u = \begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_{i+1,j} \\ u_{i,j} \\ u_{i,j+1} \end{bmatrix} = \begin{bmatrix} u_{i+1,j} - u_{i,j} \\ u_{i,j+1} - u_{i,j} \end{bmatrix} , \tag{11}$$

where $u_x$ and $u_y$ represent the gradient of $u$ w.r.t. $x$ and $y$ respectively, and where we used $(i,j) = (x,y)$ to denote the pixel coordinate. This lets us compute

$$T_I \nabla u = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_{i+1,j} \\ u_{i,j} \\ u_{i,j+1} \end{bmatrix} = \begin{bmatrix} a(u_{i+1,j} - u_{i,j}) + c(u_{i,j+1} - u_{i,j}) \\ c(u_{i+1,j} - u_{i,j}) + b(u_{i,j+1} - u_{i,j}) \end{bmatrix} . \tag{12}$$

Let us now write $\mathbf{q} = (q^1, q^2)$. The terms involving $u_{i,j}$ are then

$$\begin{aligned} \langle \mathbf{K}u_{i,j}, q_{i,j} \rangle &= \langle \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_{i+1,j} \\ u_{i,j} \\ u_{i,j+1} \end{bmatrix}, q_{i,j} \rangle \\ &= \langle \begin{bmatrix} a(u_{i+1,j} - u_{i,j}) + c(u_{i,j+1} - u_{i,j}) \\ c(u_{i+1,j} - u_{i,j}) + b(u_{i,j+1} - u_{i,j}) \end{bmatrix}, \begin{bmatrix} q_{i,j}^1 \\ q_{i,j}^2 \end{bmatrix} \rangle , \end{aligned} \tag{13}$$

$$\begin{aligned} \langle \mathbf{K}u_{i-1,j}, q_{i-1,j} \rangle &= \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_{i,j} \\ u_{i-1,j} \\ u_{i-1,j+1} \end{bmatrix} \\ &= \langle \begin{bmatrix} a(u_{i,j} - u_{i-1,j}) + c(u_{i-1,j+1} - u_{i-1,j}) \\ c(u_{i,j} - u_{i-1,j}) + b(u_{i-1,j+1} - u_{i-1,j}) \end{bmatrix}, \begin{bmatrix} q_{i-1,j}^1 \\ q_{i-1,j}^2 \end{bmatrix} \rangle , \end{aligned} \tag{14}$$

$$\begin{aligned} \langle \mathbf{K}u_{i,j-1}, q_{i,j-1} \rangle &= \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_{i+1,j-1} \\ u_{i,j-1} \\ u_{i,j} \end{bmatrix} \\ &= \langle \begin{bmatrix} a(u_{i+1,j-1} - u_{i,j-1}) + c(u_{i,j} - u_{i,j-1}) \\ c(u_{i+1,j-1} - u_{i,j-1}) + b(u_{i,j} - u_{i,j-1}) \end{bmatrix}, \begin{bmatrix} q_{i,j-1}^1 \\ q_{i,j-1}^2 \end{bmatrix} \rangle . \end{aligned} \tag{15}$$

Collecting all the terms involving $u_{i,j}$ in the three inner products above yields

$$\begin{aligned} &\langle -(a+c)q_{i,j}^1 - (c+b)q_{i,j}^2 + aq_{i-1,j}^1 + cq_{i-1,j}^2 + cq_{i,j-1}^1 + bq_{i,j-1}^2, \ \ u_{i,j} \rangle \\ &= \langle -a(q_{i,j}^1 - q_{i-1,j}^1) - c(q_{i,j}^1 - q_{i,j-1}^1) - c(q_{i,j}^2 - q_{i-1,j}^2) - b(q_{i,j}^2 - q_{i,j-1}^2), \ \ u_{i,j} \rangle \\ &= \langle -div \left( \begin{bmatrix} aq_{i,j}^1 + cq_{i,j}^2 \\ cq_{i,j}^1 + bq_{i,j}^2 \end{bmatrix} \right), \ \ u_{i,j} \rangle = \langle -div \left( \begin{bmatrix} a & c \\ c & b \end{bmatrix} \cdot \begin{bmatrix} q_{i,j}^1 \\ q_{i,j}^2 \end{bmatrix} \right), u_{i,j} \rangle \\ &= \langle -div(T_I q_{i,j}), u_{i,j} \rangle , \end{aligned} \tag{16}$$

where *div* is defined as

$$div\ p_{i,j} := p^1_{i,j} - p^1_{i-1,j} + p^2_{i,j} - p^2_{i,j-1}\ . \tag{17}$$

Since $< \mathbf{q},\ \mathbf{K}\mathbf{u}^v > = < \mathbf{K}^{-1}\mathbf{q},\ \mathbf{u}^v >$, we have $\mathbf{K}^{-1} = -div \cdot T_I$. Recall that $\mathbf{K} = T_I\nabla$. If $T_I$ is the identity matrix, then $\mathbf{K} = \nabla$. We therefore have $\nabla^{-1} = -div$.

## 2 Details on optimizing the hidden layer $\mathbf{u}^h$, $\mathbf{s}^h$

Let us now provide the details of optimizing the hidden layer, as discussed in Section 4.3 of the main paper.

### 2.1 Optimizing $\mathbf{u}^h$ with fixed $\mathbf{s}^h$

By fixing the semantic variable $\mathbf{s}^h$, we can write the objective in Eq. (25) of the paper in the standard Mumford-Shah form with

$$D_{\mathbf{u}^h}(\mathbf{u}^h) = \gamma_{uh}\sum_{\mathbf{x}}(1-m)(\mathbf{p}^T\mathbf{u}^h - \mathbf{p}^T\mathbf{u}^v)^2 + m\ \eta_s \sum_j s^h_j(\mathbf{p}^T\mathbf{u}^h - y^s_j)^2\ , \tag{18}$$

$$R_{\mathbf{u}^h}(\nabla\mathbf{u}^h) = \eta_{rh} m \min(\alpha_2\|\nabla\mathbf{u}^h\|^2 + e_{uh}, \lambda_2)\ , \tag{19}$$

where $e_{uh} = \|\nabla\mathbf{s}^h\|^2$.

The corresponding proximal operators can be computed as

$$prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h) = (\mathbf{I} + 2\tau\eta_s\mathbf{p}\mathbf{p}^T)^{-1}(\tilde{\mathbf{u}}^h + 2\tau\eta_s\mathbf{p}\sum_j s_j y^s_j)\cdot m$$
$$+ (1-m)\cdot(\mathbf{I} + 2\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T)^{-1}(\tilde{\mathbf{u}}^h + 2\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T\mathbf{u}^h)\ , \tag{20}$$

$$prox_{\sigma,R^\star_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u) = \begin{cases} \frac{2\eta_{rh}\alpha_2 m}{\sigma + 2\eta_{rh}\alpha_2}\tilde{\mathbf{q}}_u & \text{if } |\tilde{\mathbf{q}}_u| \le \sqrt{\frac{(\lambda_2 - e_{uh})\sigma(\sigma + 2\eta_{rh}\alpha_2)}{\alpha_2}}\ , \\ 0 & \text{else}\ , \end{cases} \tag{21}$$

where $\tilde{\mathbf{u}}^h = (\mathbf{u}^h)^n - \tau_n\nabla^{-1}\mathbf{q}_u^{n+1}$ and $\tilde{\mathbf{q}}_u = \mathbf{q}_u^n + \sigma_n\nabla(\bar{\mathbf{u}}^h)^n$, following Eqs. (13) and (12) in the paper, respectively. Note that, as derived above, $\nabla^{-1} = -div$ where *div* denotes the divergence operator.

**Derivation of the proximal operators.**
Here, we provide the detailed derivation of $prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h)$ and $prox_{\sigma,R^\star_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u)$.

$$prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h) = \operatorname*{argmin}_{\mathbf{u}_h} \frac{(\mathbf{u}_h - \tilde{\mathbf{u}}^h)^2}{2\tau} + D_{\mathbf{u}^h}$$
$$= \operatorname*{argmin}_{\mathbf{u}_h} \underbrace{\frac{(\mathbf{u}_h - \tilde{\mathbf{u}}^h)^2}{2\tau} + \gamma_{uh}\sum_{\mathbf{x}}(1-m)(\mathbf{p}^T\mathbf{u}^h - \mathbf{p}^T\mathbf{u}^v)^2 + m\ \eta_s\sum_j s^h_j(\mathbf{p}^T\mathbf{u}^h - y^s_j)^2}_{\tilde{D}(\mathbf{u}^h)}\ .$$

We can obtain $prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h)$ as the stationary point of the minimization problem above by solving

$$\begin{aligned}\frac{\partial\tilde{\mathbf{D}}(\mathbf{u}^h)}{\partial\mathbf{u}^h} &= \frac{\mathbf{u}^h-\tilde{\mathbf{u}}^h}{2\tau} + 2\gamma_{uh}(1-m)\mathbf{p}(\mathbf{p}^T\mathbf{u}^h-\mathbf{p}^T\mathbf{u}^v)\\ &\quad + 2m\eta_s\mathbf{p}(\mathbf{p}^T\mathbf{u}^h-\sum_j s_j y_j^s)\\ &= 0. \end{aligned} \tag{22}$$

This yields

$$\begin{aligned}prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h) = \Big(\mathbf{I}+2(1-m)\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T+2m\tau\eta_s\mathbf{p}\mathbf{p}^T\Big)^{-1}\Big(\tilde{\mathbf{u}}^h\\ + 2(1-m)\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T\mathbf{u}^v + 2m\tau\eta_s\sum_j s_j y_j^s\mathbf{p}\Big)\end{aligned}$$

By making use of the fact that $m\in\{0,1\}$, we finally have

$$\begin{aligned}prox_{\tau,D_{\mathbf{u}^h}}(\tilde{\mathbf{u}}^h) &= (\mathbf{I}+2\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T)^{-1}(\tilde{\mathbf{u}}^h+2\tau\gamma_{uh}\mathbf{p}\mathbf{p}^T\mathbf{u}^h)(1-m)\\ &\quad + (\mathbf{I}+2\tau\eta_s\mathbf{p}\mathbf{p}^T)^{-1}(\tilde{\mathbf{u}}^h+2\tau\eta_s\sum_j s_j y_j^s\mathbf{u}^h)m. \end{aligned} \tag{23}$$

Let us now turn to the computation of the second proximal operator $prox_{\sigma,R^\star_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u)$. This operator is defined as

$$prox_{\tau,R^\star_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u) = \underset{\mathbf{q}_u}{\operatorname{argmin}}\ \frac{(\mathbf{q}_u-\tilde{\mathbf{q}}_u)}{2\tau} + R^\star_{\mathbf{u}^h}(\mathbf{q}_u). \tag{24}$$

To derive it, we follow the same idea in [1]. We first derive the proximal operator $prox_{\tau,R_{\mathbf{u}^h}}$, which can be expressed as

$$prox_{\tau,R_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u) = \begin{cases} \frac{1}{1+2\eta_{rh}\tau\alpha_2}\tilde{\mathbf{q}}_u & \text{if } |\tilde{\mathbf{q}}_u| \le \sqrt{\frac{(\lambda_2-e_{uh})(1+2\eta_{rh}\tau\alpha_2)}{\alpha_2}} \\ \tilde{\mathbf{q}}_u & \text{else.}\end{cases} \tag{25}$$

Based on Moreau's identity [2], $prox_{\sigma,R^\star}(\tilde{\mathbf{q}}_u) = \tilde{\mathbf{q}}_u - \sigma prox_{\frac{1}{\sigma},R_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u/\sigma)$. We therefore have

$$prox_{\sigma,R^\star_{\mathbf{u}^h}}(\tilde{\mathbf{q}}_u) = \begin{cases} \frac{2\eta_{rh}\alpha_2}{\sigma+2\eta_{rh}\alpha_2}\tilde{\mathbf{q}}_u & \text{if } |\tilde{\mathbf{q}}_u| \le \sqrt{\frac{(\lambda_2-e_{uh})\sigma(\sigma+2\eta_{rh}\alpha_2)}{\alpha_2}} \\ 0 & \text{else.}\end{cases} \tag{26}$$

### 2.2 Optimizing $\mathbf{s}^h$ with fixed $\mathbf{u}^h$

We then fix the disparity variable $\mathbf{u}^h$, which lets us write the objective in Eq. (25) in the paper in the standard Mumford-Shah form, with

$$D_{\mathbf{s}^h}(\mathbf{s}^h) = \gamma_{sh}\sum_{\mathbf{x}}(1-m)(\mathbf{s}^h-\mathbf{s}^v)^2 + m\,\eta_s\sum_j s_j^h(\mathbf{p}^T\mathbf{u}^h-y_j^s)^2\,, \tag{27}$$

$$R_{\mathbf{s}^h}(\nabla\mathbf{s}^h) = \eta_{rh}m\sum_{\mathbf{x}\in\Omega}\min(\alpha_2 e_{sh}+\|\nabla\mathbf{s}^v\|^2,\lambda_2)\,, \tag{28}$$

where $e_{sh} = \|\nabla \mathbf{u}^h\|^2$. We derive the corresponding proximal operators in a similar way as described in the previous section. This yields

$$prox_{\tau, D_{\mathbf{s}^h}}(\tilde{\mathbf{s}}^h) = (1-m)\frac{\tilde{\mathbf{s}}^h + 2\tau\gamma_{sh}\mathbf{s}^h}{1+2\tau\gamma_{sh}} + m \cdot (\tilde{\mathbf{s}}^h + 2\tau\eta_s(\mathbf{p}^T\mathbf{u}^h - y_j^s)^2) \,, \quad (29)$$

$$prox_{\sigma, R^\star_{\mathbf{s}^h}}(\tilde{\mathbf{q}}_s) = \begin{cases} \frac{2\eta_{rh}m}{\sigma+2\eta_{rh}}\tilde{\mathbf{q}}_s & \text{if } |\tilde{\mathbf{q}}_s| \leq \sqrt{(\lambda_2 - e_{sh}\alpha_2)\sigma(\sigma+2\eta_{rh})} \,, \\ 0 & \text{else} \,, \end{cases} \quad (30)$$

where, again, $\tilde{\mathbf{s}}^h = (\mathbf{s}^h)^n - \tau_n \nabla^{-1}\mathbf{q}_s^{n+1}$ and $\tilde{\mathbf{q}}_s = \mathbf{q}_s^n + \sigma_n \nabla(\bar{\mathbf{s}}^h)^n$, following Eqs. (13) and (12) in the paper, respectively.

## 3 Additional details about the experiments

In this section, we provide more details about the experiments on the KITTI and Stixel datasets.

### 3.1 Model Parameters

We estimated our parameters using the 10 validation images with annotated ground truth. This resulted in $\alpha_1 = 1, \lambda_1 = 100, \eta_d = 0.1$, for visible layer and $\alpha_2 = 0.01, \lambda_2 = 1$ $\eta_s = 0.004, \eta_c = 1, b = 0.1$ $\eta_c = 1$, $b = 0.1$ for hidden layer. Recall that the plane parameters are initialized to zero in large holes. We therefore start with a high regularizer weight $\eta_{rv} = 10000$ and reduce its value iteratively to $\eta_{rv} = 0.1$ by a fixed step size. The same strategy is applied to $\eta_{rh}$.

### 3.2 KITTI Dataset

To study the sensitivity of our approach to the initialisation of the semantic segmentation, we further use [3] to predict initial semantics. Our method improved per-pixel and per-class accuracies from 81.24% and 58.07% ([3]) to 82.68% and 58.92%. The Visible and Hidden RMSE are 5.00 and 15.97. We were not able to obtain a better semantic segmentation results using [3] compared with the FCN-32s model [4], which is most likely due to limited training data. However, even with a lower-quality semantics initialization, our approach can still improve the visible semantic segmentation as well as the visible RMSE results.

### 3.3 Results on Stixel

The images in this dataset correspond to the frames of video sequences. Although only every $10^{\text{th}}$ frame of each sequence was sampled to create the dataset, neighboring images still show important similarities. In our experiment, we therefore chose one in every 4 of such frames to constitute our own test set, which thus consists of 50 images. Note that the Stixel dataset provides ground-truth disparity with 72.05% image coverage on average. To be able to evaluate our depth completion results, we artificially created a random mask to preserve at most

| | ours | [5] | [6] |
|---|---|---|---|
| **visible-rmse** | **1.0253** | 1.1107 | 1.1319 |

**Table 1. Depth estimation.** Quantitative comparison with several baselines for the visible depth.

| road | building | sky | class-avg. | pixel-avg |
|---|---|---|---|---|
| 98.89 | 95.81 | 80.31 | 91.67 | 97.44 |

**Table 2. Estimating hidden semantics.** Per-class and overall accuracy of our approach.

20% of the observations in each image as input to our method. The value of 20% corresponds to the observation ratio of KITTI. We then provide a quantitative evaluation for the visible layer using the remaining data. These quantitative results are shown in Table 1 and qualitative results in Fig. 4, respectively. Due to the incomplete ground-truth disparity and coarse annotation for the semantics, we cannot generate high quality ground truth disparity for hidden layer. We thus only provide the quantitative evaluation for disparity in visible layer. Note that we outperform the baselines both quantitatively and qualitatively. The latter is particularly visible on the hidden layer predictions, where our approach is able to better remove the pedestrians and cars than the baselines.

In Table 2, we show the results of our semantics labeling estimates for the hidden regions. Here, again, since no baseline is available for this task, we only report the results of our approach. These results show that our model yields good accuracy on the more common and larger classes. We observed that the semantic labeling accuracy in the visible layer did not significantly change compared to our initialization. In particular, our method achieves 95.09% per-pixel accuracy and 88.38% per-class accuracy on average for the 5 semantic classes of the visible layer. In Fig. 5, we show the semantics estimated with our approach for the visible and hidden layer, respectively.

## References

1. Strekalovskiy, E., Cremers, D.: Real-time minimization of the piecewise smooth mumford-shah functional. In: ECCV. (2014)
2. Rockafellar, R.T.: Convex analysis. Princeton university press (2015)
3. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.: Conditional random fields as recurrent neural networks. In: ICCV. (2015)
4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR. (2015)
5. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM Transactions on Graphics (TOG). (2004)
6. Ferstl, D., Reinbacher, C., Ranftl, R., Rüther, M., Bischof, H.: Image guided depth upsampling using anisotropic total generalized variation. In: ICCV. (2013)

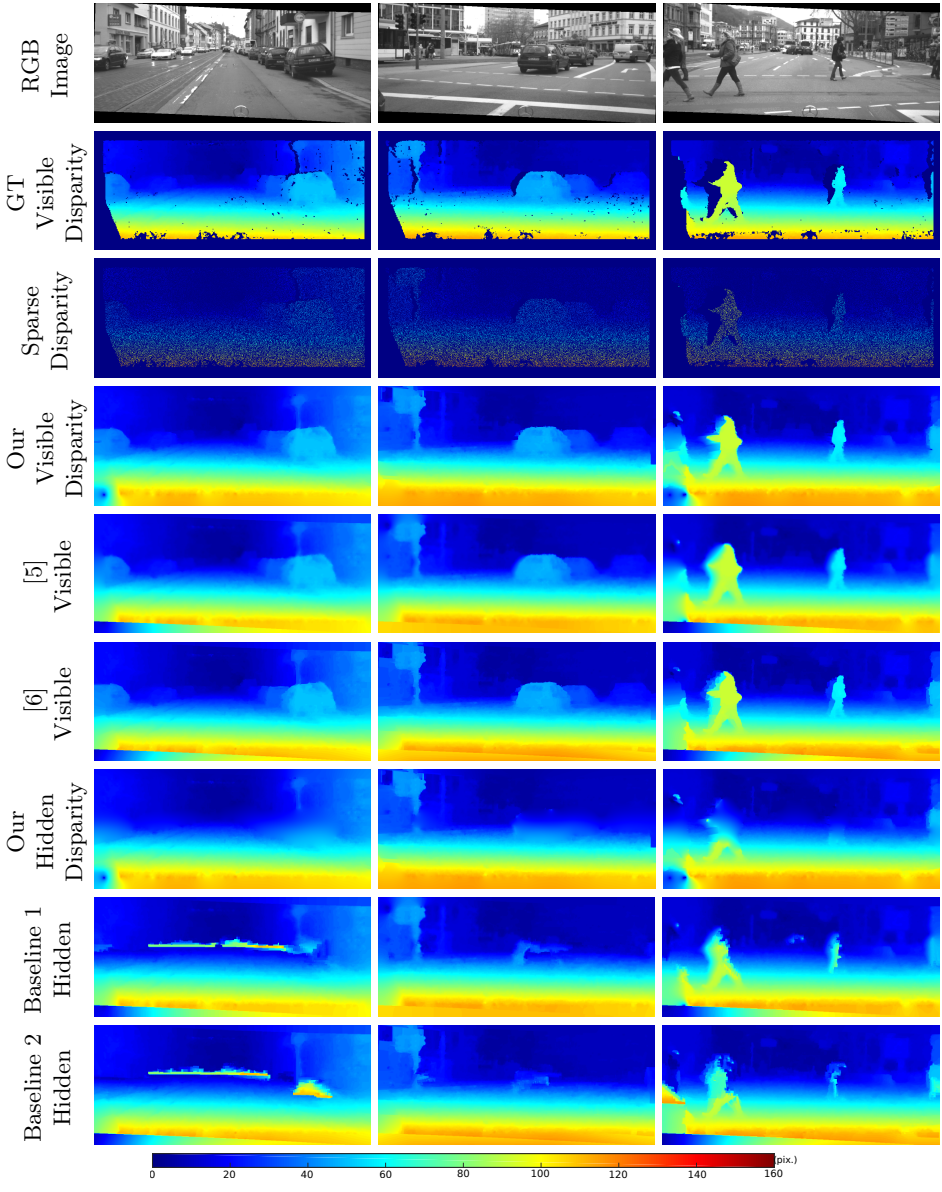


**Fig. 4. Qualitative results on Stixel.** For the disparity values, red denotes large values, *i.e.*, points close to the camera, and blue denotes small disparity values, *i.e.*, points far from the camera. **From top to bottom:** RGB image, ground-truth visible disparity map, sparse observations with large holes, our completed disparity map, two baselines for the visible layer, our disparity for the hidden layer, and two baselines for the hidden layer. Note that our method can remove the foreground objects as well as accurately fill in the background disparity behind them. Compared to the baselines, our approach can better complete the disparity for the visible and hidden layers.

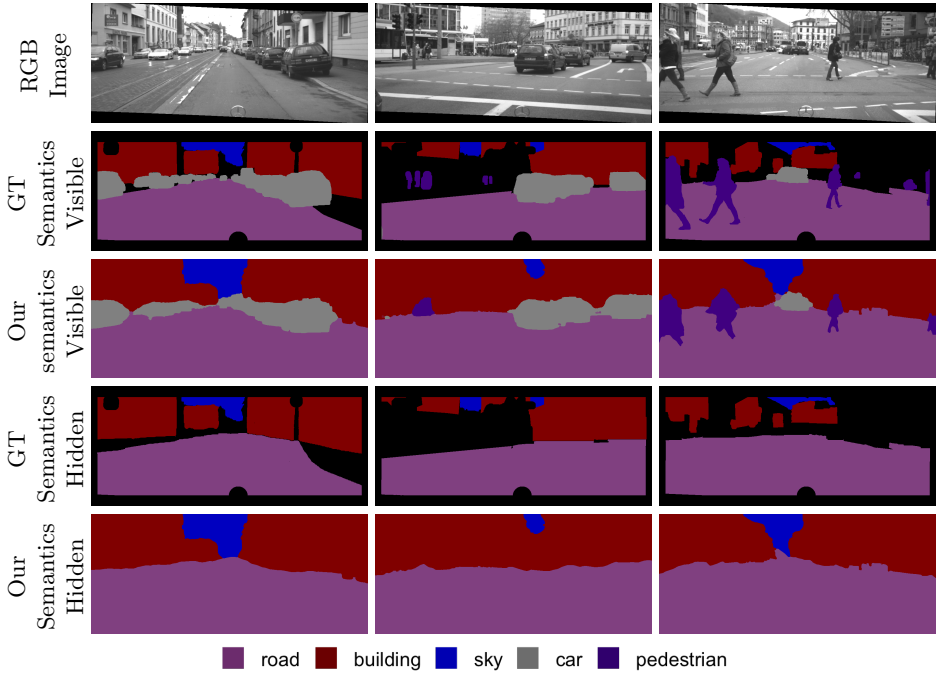


**Fig. 5. Qualitative results for semantic segmentation on the Stixel dataset. From top to bottom:** RGB image, ground-truth semantics for the visible layer, and our estimated semantics for the visible, ground-truth semantics for the hidden layer, and our estimated semantics for hidden layer, respectively.