RAFT-3D: Supplementary Material

Figure 1. Network architecture. The components include (1) a feature encoder (2) a context encoder with a ResNet50 backbone, and (3) and GRU-based updated operator. The GRU uses a dilated convolution pattern as shown. In contrast to RAFT[3] and GRU-based updated operator. The GRU uses a dilated convolution pattern as shown. In contrast to RAFT[3] and GRU-based updated operator. The GRU uses a dilated convolution pattern as shown. In contrast to RAFT[3] and GRU-based updated operator.

1. Network Architecture

Details of the network architecture, including feature encoders and the GRU-based update operator are shown in Figure 1.

2. bi-Laplacian Optimization Layer Gradients

This layer minimizes an objective function in the form

\[ \|D_x u\|_{w_x}^2 + \|D_y u\|_{w_y}^2 + \|u - v\|^2 \]

where \(D_x\) and \(D_y\) are linear finite difference operators, and \(v\) is the flattened feature map.

First consider the case of single channel, \(v \in \mathbb{R}^{HW}\). Let \(W_x = \text{diag}(w_x), W_y = \text{diag}(w_y) \in \mathbb{R}^{HW \times HW}\). We can solve for \(u^*\)

\[ (I + D_x^T W_x D_x^T + D_y^T W_y D_y^T)u^* = v \]

We perform sparse Cholesky factorization and backsubstitution to solve for \(u^*\) using the Cholmod library[2].

Gradients: In the backward pass, given the gradient \(\frac{\partial L}{\partial u}\), we need to find the gradients with respect to the boundary weights \(\frac{\partial L}{\partial w_x}\) and \(\frac{\partial L}{\partial w_y}\).

Given the linear system \(H u = v\), the gradients with respect to \(H\) and \(v\) can be found by solving the system in the backward direction [1]

\[ \frac{\partial L}{\partial v} = H^{-T} \frac{\partial L}{\partial u^*} \]  \hspace{1cm} (3)
\[ \frac{\partial L}{\partial H} = u^* d_v^T \]  \hspace{1cm} (4)
\[ d_v = H^{-T} \frac{\partial L}{\partial u} \]  \hspace{1cm} (5)

Here the column vector \(d_v\) is defined for notational convenience. Since \(H\) is positive definite, \(H^{-T} = H^{-1}\) so we can reuse the factorization from the forward pass.

To compute the gradients with respect to \(w_x\) and \(w_y\)

\[ \frac{\partial L}{\partial w_x} = \text{diag} \left( \frac{\partial L}{\partial H} \frac{\partial H}{\partial W_x} \right) \]

\[ = \text{diag} \left( (D_x u^*) (D_x d_v)^T \right) \]  \hspace{1cm} (6)

where \(\circ\) is elementwise multiplication. Similarly

\[ \frac{\partial L}{\partial w_y} = (D_y u^*) (D_y d_v) \]  \hspace{1cm} (7)

Multiple Channels: We can easily extend Eqn. 2 to work with multiple channels. Since the matrix \(H\) does not depend on \(v\), it only needs to be factored once. We can solve Eqn. 2 for all channels by reusing the factorization, treating \(v\) as an \(HW \times C\) matrix. The gradient formulas can also be updated by summing the gradients over the channel dimensions.

References

