1. Abstract

TV-L1 is a classical diffusion-reaction model for low-level vision tasks, which can be solved by a duality based iterative algorithm. Considering the recent success of end-to-end learned representations, we propose a TV-LSTM network to unfold the duality based iterations into long short-term memory (LSTM) cells. To provide a trainable network, we relax the difference operators in the gate and cell update of TV-LSTM to trainable parameters. Then, the proposed end-end trainable TV-LSTMs can be naturally connected with various task-specific networks, e.g., optical flow estimation and image decomposition.

2. TV-L1 Model

We consider the following TV-L1 model:

\[
\hat{x} \in \arg \min_x TV_{\lambda}(x) + \lambda \| f(x) \|_{L^1(\Sigma)}.
\]

Here, \(\| \cdot \|_{L^1(\Sigma)}\) is the \(L^1\)-norm. An efficient way to solve the optimization problem defined in equation (1) is a duality based implementation:

\[
\begin{align*}
\hat{p}_{i,j}^{k+1} &= \frac{p_{i,j}^k + \tau (\nabla (\text{div} p^k - u/\theta))_{i,j}}{1 + \tau \| (\nabla (\text{div} p^k - u/\theta))_{i,j} \|}, \\
\hat{u}^{k+1} &= \hat{u}^k - \theta \nabla \hat{p}^k,
\end{align*}
\]

where \(\hat{p}^k\) and \(\hat{u}^k\) are the updated primal and dual variables, respectively.

3. Connection with LSTM-like Recurrent Networks

We present the qualitative results achieved by the TV-LSTMs and TVNet. From this table, it can be seen that the pose TV-LSTMs can achieve a speed of 7 fps TV-LSTMs based image decomposition algorithm.

4. The End-to-end Trainable Network

We mainly compare our method with TVNet [1] for optical flow estimation. We followed the same experimental settings as TVNet, performed on the Middlebury dataset. Table 1 presents the qualitative results achieved by TV-LSTMs and TVNet. From this table, it can be seen that the proposed method outperforms the original TVNet.

<table>
<thead>
<tr>
<th>Methods</th>
<th>No Training</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVNet(1-1-10)</td>
<td>3.47</td>
<td>1.24</td>
</tr>
<tr>
<td>TVNet(3-1-10)</td>
<td>2.00</td>
<td>0.52</td>
</tr>
<tr>
<td>TV-LSTMs(1-1-30)</td>
<td>2.55</td>
<td>1.30</td>
</tr>
<tr>
<td>TV-LSTMs(3-1-10)</td>
<td>1.77</td>
<td>0.51</td>
</tr>
<tr>
<td>TV-LSTMs(3-3-30)</td>
<td>1.67</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 1: The average EPEs on Middlebury

5. Results I

Figure 5: Examples of decomposition results achieved by different methods. (a) Input. (b) Structures [2]. (c) Structures (TV-LSTMs without training). (d) Structures (trained TV-LSTMs). (e) Textures (trained TV-LSTMs).

Figure 4: Optical flow results by trained TV-LSTMs

6. References


Future Research

This work proposes a neural network, namely TV-LSTMs, to achieve the TV-L1 model in an end-to-end manner. We show that the optimization of TV-L1 can be unfolded as a LSTM-like network with a novel computational unit. Furthermore, our TV-LSTMs can be naturally extended to a task-specific network by using a specific reaction term.

Contact Information

Email: fangyuqiang@nudt.edu.cn
Phone: +86 - 13521530926