From TV-L^1 to Gated Recurrent Nets

{ Yuqiang Fang, Haiyan Fan, Yulan Guo, Lin Sun, Zhihao Ma}   Space Engineering University

1. Abstract

TV-L^1 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-to-end trainable TV-LSTMs can be naturally connected with various task-specific networks, e.g., optical flow estimation and image decomposition.

2. TV-L^1 model

We consider the following TV-L^1 model:

$$\text{Find} \quad \hat{x} \in \underset{x}{\operatorname{argmin}}\, \mathrm{TV}_i(x) + \lambda \|f(x)\|_{L^1(\Omega)}\,. \tag{1}$$

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

$$p^{k+1}_{i,j} = \frac{p^k_{i,j} + \tau\left(\nabla\left(\mathrm{div}p^k - v/\theta\right)\right)_{i,j}}{1+\tau\|\left(\nabla\left(\mathrm{div}p^k - v/\theta\right)\right)_{i,j}\|}, \tag{2}$$

$$x^{k+1} = v^{k+1} - \theta \mathrm{div}\left(p^{k+1}\right)\,. \tag{3}$$

3. Connection with LSTM-like Recurrent Networks

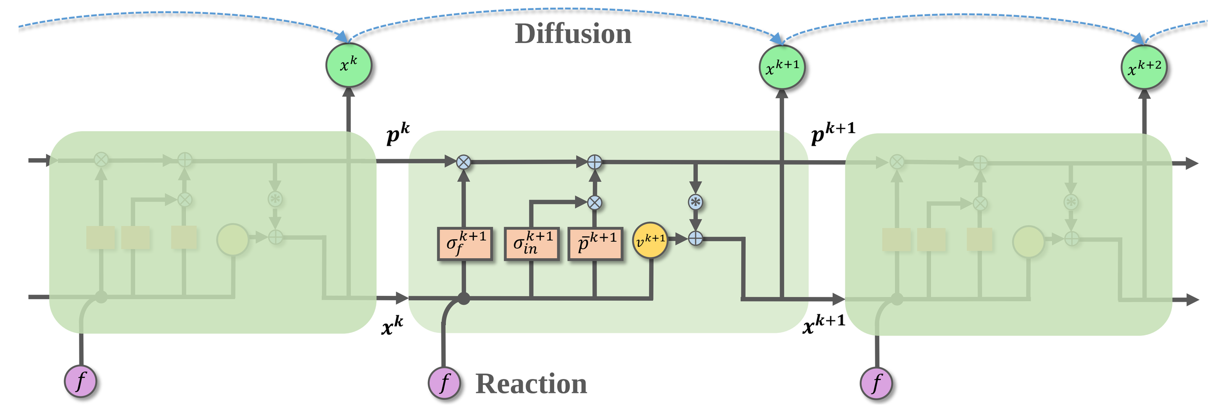

Figure 1: TV-L^1 iterations as LSTMs network

we can map specialized TV-L^1 iterations to a LSTM-like network. We partition the TV-L^1 iterations as gate updates:

$$\sigma_f^{k+1} = \sigma\left(\frac{1}{1+\frac{\tau}{\theta}\|\nabla x^k\|}\right), \tag{4}$$

$$\sigma_{\mathrm{in}}^{k+1} = \sigma\left(\frac{\frac{\tau}{\theta}}{1+\frac{\tau}{\theta}\|\nabla x^k\|}\right), \tag{5}$$

where $v^{k+1} = x^{k+1} + TH\left(x^{k+1}, \lambda\theta\right)$, and the cell updates

$$\bar{p}^{k+1} \leftarrow -\nabla x^k, \quad p^{k+1} \leftarrow \sigma_f^{k+1} \odot p^k + \sigma_{\mathrm{in}}^{k+1} \odot \bar{p}^{k+1}, \tag{6}$$

and output updates

$$x^{k+1} \leftarrow v^{k+1} - \theta \mathrm{div}\left(p^{k+1}\right)\,. \tag{7}$$

Starting from initial values of p^0 and x^0, the TV-L^1 implementation defined in equations (4)-(7) closely mirrors a canonical LSTM. The output of the network at time-step k is x^k, and p^k is considered as the internal LSTM memory cell, or the latent cell state. Thus, the iterative process in TV-L^1 model can be unfolded as a layer-to-layer LSTM-like network.

4. The End-to-end Trainable Network

To imitate the iterative process in TV-L^1, TV-LSTMs has two gates to protect and control the cell state step-by-step:

Forget Gate σ_f^{k+1}: the computation of gradient $\|\nabla x\|$ in (4) can be discretized as

$$\|\nabla x\| = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2}\left((x*w)^2[n_1,n_2] + (x*w^\top)^2[n_1,n_2]\right)^{\frac{1}{2}}, \tag{8}$$

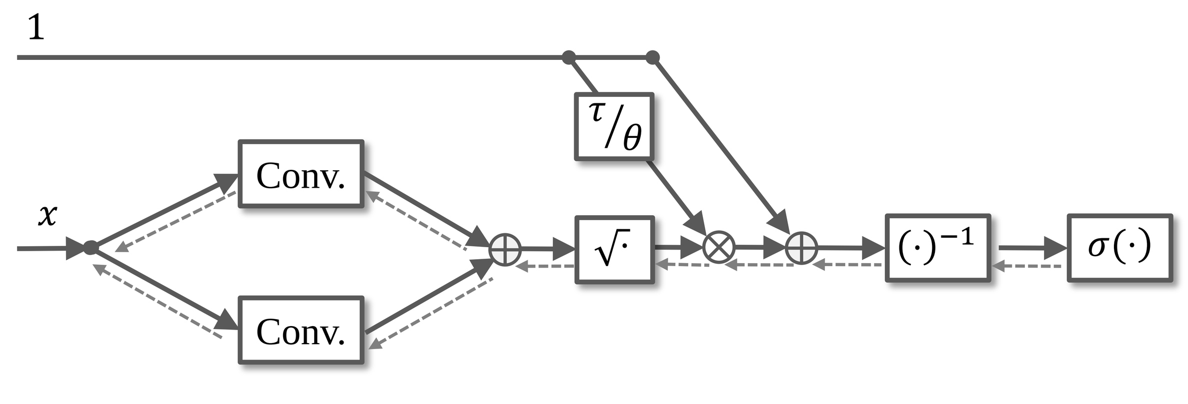

Figure 2: The computation graph of forget gate

Input Gate and Candidate Cell:

$$\bar{p}^{k+1} \leftarrow \left(-x^k * w_{in}, -x^k * w_{in}^\top\right). \tag{9}$$

Output Update:

$$\mathrm{div}\left(p\right) = \hat{p}_1 * w_o + \hat{p}_2 * w_o^\top, \tag{10}$$

Task-specific Networks: Our TV-LSTMs can be designed for different computer vision problems such as optical flow and image decomposition.

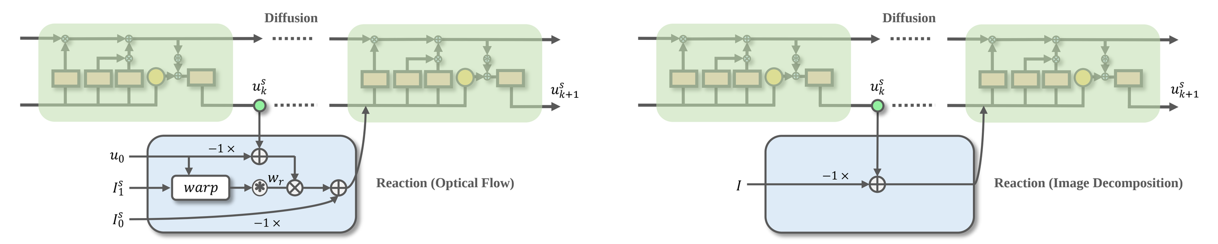

Figure 3: An illustration of feed forward networks in TV-LSTMs for optical flow (left) and image decomposition (right)

5. Results I

We mainly compare our method with TVNet [1] for optical flow estimation. We followed the same experimental settings as TVNet, performed on the MiddleBurry 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.

Methods*	No Training	Training
TVNet(1-1-10)	3.47	1.24
TVNet(3-1-10)	2.00	0.52
TV-LSTMs(1-1-30)	2.55	1.30
TV-LSTMs(3-1-10)	1.77	0.51
TV-LSTMs(3-1-30)	1.67	0.36

*TVNet / TV-LSTMs($N_{warps} - N_{scales} - N_{iters}$)

Table 1: The average EPEs on MiddleBurry

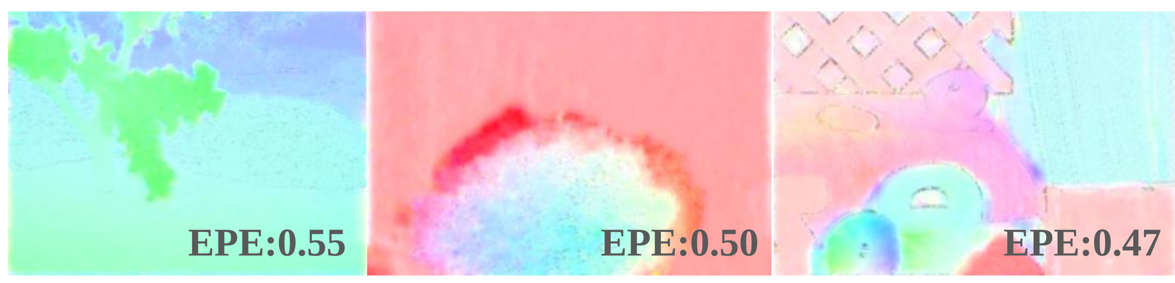

Figure 4: Optical flow results by trained TV-LSTMs

6. Result II

We present the qualitative results achieved by the TV-LSTMs based image decomposition algorithm. In figure 5, it can be seen that, our TV-LSTMs can produces well-decomposed textures. Moreover, our TV-LSTMs can achieve a speed of 7 fps on images with a resolution of 410×620 pixels.

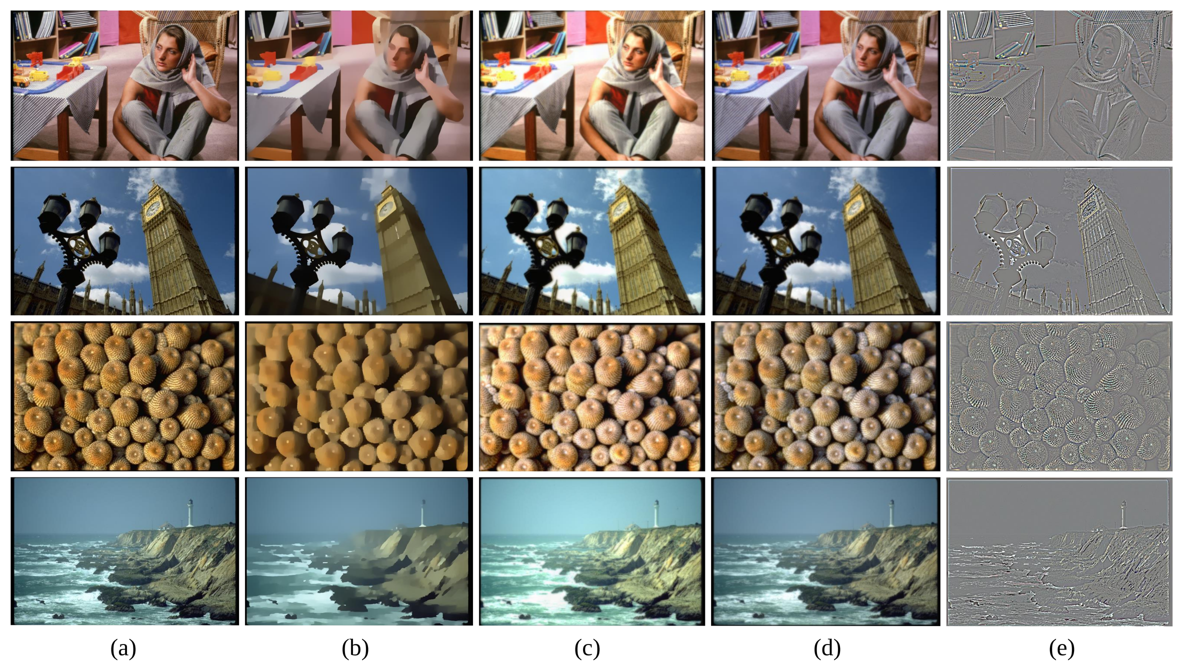

Figure 5: Examples of decomposition results achieved by different methods. (a) Input. (b) Structures [2]. (c) Structures (TV-LSTMs without training). (d) Structures (trained TV-LSTMs). (e) Textures (trained TV-LSTMs).

Future Research

This work proposes a neural network, namely TV-LSTMs, to achieve the TV-L^1 model in an end-to-end manner. We show that the optimization of TV-L^1 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.

References

[1] L. Fan. End to end learning of motion representation for video understanding. *CVPR*, June 2018.

[2] L. Xu. Structure extraction from texture via relative total variation. *SIGGRAPH*, 2012.

Contact Information

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