Learning from Synthetic Data

Thomas Brox
University of Freiburg, Germany

Research funded by the ERC Starting Grant VideoLearn, the German Research Foundation, and the Deutsche Telekom Stiftung
Deep learning wants data... input and output

UCF 101

ImageNet

Diving
Golf Swing
Running
Skateboarding

Microsoft COCO
Common Objects in Context
Typical strategy to set up a data set

1. Collect lots of images/videos
   - usually from the web

2. Manually annotate the desired output

Lock students in a cellar
Crowd sourcing
Workers at companies

Works well for recognition
• Can networks learn to find correspondences?

• New learning task!
  (very different from classification, etc.)

Dosovitskiy et al.
ICCV 2015
Getting ground truth optical flow for real videos is hard

Existing datasets are small:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Frames with ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middlebury</td>
<td>8</td>
</tr>
<tr>
<td>KITTI</td>
<td>194</td>
</tr>
<tr>
<td>Sintel</td>
<td>1041</td>
</tr>
<tr>
<td>Needed</td>
<td>&gt;10000</td>
</tr>
</tbody>
</table>

Solution: large scale synthetic data
Realism is overrated: the “flying chairs” dataset

Image pair

Optical flow
What about generalization?

Although the network has only seen flying chairs for training, it predicts good optical flow on Sintel.
Generalization to real data

FlowNet
P. Fischer,
A. Dosovitskiy,
E. Ilg,
P. Häusser,
C. Hazırbaş,
V. Golkov,
P. v.d. Smagt,
D. Cremers,
T. Brox
Can handle large displacements

Input images

Ground truth

EPE: 20.82

FlowNet

EPE: 32.56

EpicFlow (Revaud et al. CVPR 2015)
Weak spot (old FlowNet): small displacements

Example from UCF 101

FlowNet

EpicFlow

Identical images
DispNet: disparity estimation in 60ms

Mayer et al.
CVPR 2016
DispNet: disparity estimation in 60ms

DispNet result on KITTI 2015
Synthetic 3D datasets

Driving, Monkaa, FlyingThings3D datasets publicly available

Mayer et al. CVPR 2016
Powerful concept: render input and any output

Rendered input

Rendered depth image

Rendered segmentation
Indoor and outdoor scenes possible
Adding animated persons
Modeling plants swinging in the wind
Existing scenes vs. new scenes

• Open source movies (Sintel, Monkaa), 3D Games
  – Effort: hacking the rendering engine to produce ground truth
  – Limited to the scenes that exist

• Manually created scenes
  – Full control about the scene
  – Enormeous work

• Procedurally created scenes
  – Scales well, allows for diversity
  – Harder to create realistic scenes
Some dataset studies on FlowNet

<table>
<thead>
<tr>
<th>Model Configuration</th>
<th>Sintel Clean Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlowNet, 600k on FlyingChairs</td>
<td>4.45</td>
</tr>
<tr>
<td>FlowNet, 1.2M on FlyingChairs</td>
<td>4.24</td>
</tr>
<tr>
<td>FlowNet, 1.7M on FlyingChairs</td>
<td>4.21</td>
</tr>
<tr>
<td>FlowNet, 1.2M on FlyingThings3D</td>
<td>5.07</td>
</tr>
<tr>
<td>FlowNet, 1.7M on FlyingThings3D</td>
<td>4.51</td>
</tr>
<tr>
<td>FlowNet, 1.2M on FlyingChairs, 500k on FlyingThings3D</td>
<td>3.79</td>
</tr>
</tbody>
</table>

- Lessons:
  - Longer training helps a little
  - FlyingChairs works better than FlyingThings3D
  - Fine-tuning on FlyingThings3D improves results
How important are realistic scenes?

**Dynamic lighting**

**Static lighting**

**No lighting**
How important is realistic lighting?

- No lighting, 4.41
- Static lighting, 4.19
- Dynamic lighting, 4.08
Breaking news: FlowNet 2.0

- FlowNet 1.0
  - 12fps

- FlowNet 2.0
  - @20fps
  - @10fps
  - @15fps
  - @36fps

Presentation at ECCV VARVAI Workshop 16.10.2016
Good on small motion, too

- **EpicFlow**
  - FlowNet 1.0
    - 12fps

- **DeepFlow**
  - FlowNet 2.0
    - 20fps
  - FlowNet 2.0
    - @10fps
  - FlowNet 2.0
    - @15fps
  - FlowNet 2.0
    - @36fps

- **LDOF**
  - FlowNet 2.0
    - @10fps
Good on small motion, too

<table>
<thead>
<tr>
<th>EpicFlow</th>
<th>DeepFlow</th>
<th>LDOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>@20fps</td>
<td>Old @12fps</td>
<td>@15fps</td>
</tr>
<tr>
<td>@10fps</td>
<td>@36fps</td>
<td>Ground truth</td>
</tr>
</tbody>
</table>

Presentation at ECCV VARVAI Workshop 16.10.2016
Super precise at small details

<table>
<thead>
<tr>
<th>EpicFlow</th>
<th>DeepFlow</th>
<th>LDOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>@20fps</td>
<td>Old @12fps</td>
<td>@15fps</td>
</tr>
<tr>
<td>@10fps</td>
<td>@36fps</td>
<td>Ground truth</td>
</tr>
</tbody>
</table>

Presentation at ECCV VARVAI Workshop 16.10.2016
Super precise at small details

<table>
<thead>
<tr>
<th>EpicFlow</th>
<th>DeepFlow</th>
<th>LDOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>@20fps</td>
<td>Old @12fps</td>
<td>@15fps</td>
</tr>
<tr>
<td>@10fps</td>
<td>@36fps</td>
<td>Ground truth</td>
</tr>
</tbody>
</table>
## Preliminary numbers on Sintel.train.clean

<table>
<thead>
<tr>
<th>Method</th>
<th>EPE</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>EpicFlow (Revaud et al. 2015)</td>
<td>2.4</td>
<td>15-20s</td>
</tr>
<tr>
<td>FlowNet 1.0 (Dosovitskiy et al. 2015)</td>
<td>4.5</td>
<td>80ms (12 fps)</td>
</tr>
<tr>
<td>FlowNet 2.0</td>
<td>2.1</td>
<td>90ms (11 fps)</td>
</tr>
<tr>
<td>FlowNet 2.0</td>
<td>2.6</td>
<td>28ms (36 fps)</td>
</tr>
<tr>
<td>FlowNet 2.0</td>
<td>4.5</td>
<td>9ms (110 fps)</td>
</tr>
</tbody>
</table>

Sintel was not used for training
(Training on FlyingChairs and FlyingThings3D)
Realism is “only” important for learning good priors

There are many ways to create synthetic training data

Procedurally generated datasets are important for diversity

FlowNet 2.0 with up to 100fps and very precise flow to come