**Motivation**

- Previous works usually propose both new architecture and regression losses;
- Few comparisons are made between loss functions using same architecture.

We propose:
- D3-Net: new architecture based on the reuse of previous feature maps that achieves top results on depth estimation (code on github);
- End-to-end conditional-GAN for depth estimation based on [1];
- Different experiments to compare the performances of the regression losses from the state-of-art.

**Network Architecture**

D3-Net is an encoder-decoder based on U-Net and DenseNet-121 for the encoder.

- The reuse of previous feature maps improve information flow during learning;
- Skip and dense connections easy gradient back-propagation to the bottom layers reducing vanishing gradient problems.

**Patch-GAN**

- We adapt the conditional patch GAN proposed in [1] to depth estimation using the LSGAN and a smaller discriminator;
- Instead of classifying if the entire image is True/False, the discriminative network classifies by patches.

**Regression Losses**

Our experiments make use of the following common losses for regression:

<table>
<thead>
<tr>
<th>Loss</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute</td>
<td>( \sum_{i}</td>
</tr>
<tr>
<td>Mean square</td>
<td>( \sum_{i} t_i^2 )</td>
</tr>
<tr>
<td>SI loss [2]</td>
<td>( \sum_{i} \left( t_i - \frac{1}{\sigma^2} \sum_{j} d_{ij}^2 \right</td>
</tr>
<tr>
<td>SI loss with gradients [3]</td>
<td>( \sum_{i} \left( t_i - \frac{1}{\sigma^2} \sum_{j} d_{ij}^2 \right</td>
</tr>
<tr>
<td>BerHu [4]</td>
<td>( \mathcal{L}<em>{\text{BerHu}} \left( \frac{\lambda</em>{\text{BerHu}}(t_i)}{\lambda_{\text{BerHu}}(t_i)} \right) )</td>
</tr>
<tr>
<td>Huber [4]</td>
<td>( \mathcal{L}<em>{\text{Huber}} \left( \frac{\lambda</em>{\text{Huber}}(t_i)}{\lambda_{\text{Huber}}(t_i)} \right) )</td>
</tr>
<tr>
<td>LSGAN [5, 1]</td>
<td>( \mathcal{L}<em>{\text{LSGAN}} \left( \frac{\lambda</em>{\text{LSGAN}}(t_i)}{\lambda_{\text{LSGAN}}(t_i)} \right) )</td>
</tr>
</tbody>
</table>

Let \( y_i \) and \( \hat{y}_i \) be the ground truth and the estimated distance in meters, \( i = y_i - \hat{y}_i \), \( d_i = \log(y_i) - \log(\hat{y}_i) \). \( G \) the generator network, \( D \) the discriminator network and \( x \) the input image.

**Experiments and Results**

- Different Front-End Comparison (ResNet vs. DenseNet);
- Convergence Speed Comparison;
- Influence of the Dataset Training Size on Performances;
- Results were adapted from original paper.

**Work**

In [9], we explore D3-Net to perform depth estimation with the presence of defocus blur.

**Deep-DFD in the wild with outdoor scenes**

We now observe the network generalization to outdoor challenging scenes.

**Conclusions**

- Combined information of geometrical structure and defocus blur avoids classical limitations of DFD techniques;
- Deep-DFD is a promising method to generalize learning depth estimation;
- Co-conception of a sensor and a deep depth estimation methods.

**References**


