MorphPool: Efficient Non-linear Pooling & Unpooling in CNNs
Rick Groenendijk, Leo Dorst & Theo Gevers

Goal
Down-sampling using max pooling can be formalised and improved using mathematical morphology, the mathematics of contact. In addition to down-sampling, encoder-decoder networks also require up-sampling. Seminal works [1, 2, 3] use a combination of unpooling and transposed convolutions to this end. In our article, both pooling and unpooling in CNNs are formalised as fully morphological operations without any linear interpolation scheme; the full procedure is called MorphPool.

Code is available at https://github.com/rickgroen/morphpool

Method
Max pool is a discrete local strided maximum operator

\[ f_v(x) = \max_{s \in Z} (f_s(x + z)) \]

where \( s \) is the stride and \( z \) denotes the indicator set over the local neighbourhood. By algebraic manipulation it can be shown that the max pool is in fact a special case of the morphological dilation, defined as

\[ f_v(x_v) = \bigvee_z f_s (x_v - z) + h(z). \]

Here \( h \) is the structuring element, which is flat in the case of max pooling, but can be freely parameterised.

Morphological Pool

<table>
<thead>
<tr>
<th>( f_s(x_v) )</th>
<th>( f_v(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 0.1 0.8 2.3 2.7</td>
<td>4.5 2.9</td>
</tr>
<tr>
<td>3.2 0.6 0.9 1.2 2.9</td>
<td>2.4 5.4</td>
</tr>
<tr>
<td>2.1 0.8 1.1 1.3 2.7</td>
<td>[ (-1, -1) ] ( (0.5, 1) )</td>
</tr>
<tr>
<td>2.2 2.4 2.3 3.6 3.4</td>
<td>[ (0, 0) ] ( (1, 1) )</td>
</tr>
<tr>
<td>1.8 2.1 2.3 5.4 5.1</td>
<td></td>
</tr>
</tbody>
</table>

Morphological unpooling is a two-stage process. First, place back values by means of provenance:

\[ f_s (x - z_v - w) \begin{cases} f_s (x_v) & \exists x_v \in x \left[ x_v = s x_{v'} \right] \\ -\infty & \text{otherwise.} \end{cases} \]

Second, fill \(-\infty\) using the upper bound of the sparse values by morphological interpolation:

\[ f_s (x) = \bigvee_w f_s (x - z_v - w) + h(w), \]

Results
MorphPool was evaluated on semantic segmentation and depth auto-encoding for 3 datasets: NYUv2, SUN-RGBD, and 2D-3D-S. It was compared to strided convolution and up-sampling by interpolation (Linear) and to pooling combined with transposed convolution (Standard Pool) as in [2, 3]. We use a general-purpose encoder-decoder network without any task-specific modules to isolate performance purely due to sampling.

For depth input, MorphPool outperforms the other sampling methods at no extra parameter cost. In the case of segmentation, semantic boundaries are more clearly delineated as measured by the Boundary F1 score (bf).

It is possible to parameterise the MM kernel \( h(\cdot) \) in MorphPool to make it more expressive.

RGB does not just express geometry but also illumination, which has aspects that are well described linearly, and then MorphPool has comparatively less effect. However, semantic boundaries are of significantly better quality for both RGB and depth input. In any case, the number of parameters required by MorphPool is significantly less than the Linear method.

Conclusions
We show that MorphPool can increase performance at much reduced parameter counts for semantic segmentation and depth auto-encoding. The beneficial effects are most pronounced on depth data and at semantic boundaries; mathematical morphology appears to be the natural language to express such nonlinear geometry.

References