Output Recycling for Segmentation Tasks and Patch-Based Segmentation with CoordConv

Matthew Phillips
1 LifeOmic, Morrisville NC 27510 USA

Editors: Under Review for MIDL 2019

Abstract
We present a novel training protocol for segmentation networks, whereby we add input channels to an existing network containing segmentation masks generated by models trained on the same dataset ("output recycling"). We find that output recycling produces substantial improvements in multi-task scenarios (up to 0.057 Dice units), despite the fact that the network has no new information when recycling is used, as compared to baseline. We also combine CoordConv (Liu et al., 2018) with patch-based segmentation and find that CoordConv significantly improves performance on difficult segmentation tasks (up to 0.192 Dice units). Experiments run using the PALM Retinal Fundus image dataset (N=400).

Keywords: Image segmentation, Deep Learning, Data augmentation, Attention, U-Net, CoordConv

1. Introduction
There is a strong tendency in the Deep Learning community to operate on raw data, and avoid hand-crafted inputs or features of any kind. Here we report cases where ancillary input improves performance. Most significantly, in a multi-class segmentation task, when the segmentation maps generated by our network are fed back as additional inputs into the same network, the network achieves higher performance upon retraining. Critically, the network trained with the additional input has no information not available in the baseline case. Additionally we reason that in many biomedical imaging tasks there are regularities in global position of different class items, and CoordConv (Liu et al., 2018) could enable a network to learn these, particularly in a patch-based context. Our second finding is that indeed CoordConv improves performance significantly in difficult segmentation tasks.

Background We are not aware of any direct precedent for output recycling. Boosting algorithms (see e.g. (Hastie et al., 2009)) feed back error cases into the system in order to learn most effectively, and similar strategies have been employed with DNNs as "hard example mining" (Shrivastava et al., 2016). Output recycling by contrast does not modify the input distribution; rather, it exploits the fact that in segmentation tasks the output (a segmentation mask) has the same structure as the input (an image) and thus can be re-used as a prior. Thus there is a natural relation to region proposal networks, e.g. (He et al., 2017), although no separate architecture is employed in this case. Similarly, despite its simplicity and obvious applicability, we are aware of no precedent for CoordConv in the domain of biomedical image segmentation either.
2. Methods

We implemented the Attention U-Net of (Oktay et al., 2018), which adds additive attention gates to the skip-connections of the original U-Net (Ronneberger et al., 2015). In order to accommodate potentially overlapping segmentation regions we extended the output layer so that the network produced binary classifications for each predicted class separately. We used the Adam optimizer and early stopping, with an initial learning rate of 0.0001 which decreased by a factor of 5 after three successive epochs with no decrease in test loss; this was repeated twice then the experiment was ceased. Loss was binary cross entropy weighted by class size. Input size was 448x448 pixels.

Output recycling In the core case, the network is trained separately on individual categories, generating corresponding segmentation masks for each image in the dataset. Then these segmentation masks are provided as additional input channels to the network, and concatenated at every convolutional and deconvolutional layer, resampled appropriately. The network is then re-trained on the same dataset to predict both classes simultaneously, using these additional inputs (Figure 1).

![Output recycling](image)

**Figure 1:** Output recycling from two from single-task Attention U-Nets (Oktay et al., 2018), atrophy (top) and OD (bottom), to help solve the corresponding multi-task segmentation (right). Orange lines, skip connections; blue arrows, upsampling/downsampling; red arrows, attention gating.

Patch-based segmentation Only single-task versions of the experiment were run, and patch size (25%, 50%, 75%, 100%) and the presence or absence of CoordConv were varied. Horizontal, vertical, and radial gradients were broadcast to (concatenated with) all convolutional and deconvolutional layers. Critically, these gradients were given the identical crop used to create the input patch before being fed into the network.

3. Experiments

Data The dataset for these experiments is from the PALM Grand Challenge\footnote{https://palm.grand-challenge.org/} and consists 400 color Retinal Fundus (RF) images, mostly 2000x2000pix Figure 2. We used 340 for training and reserved 60 for validation. The data have expert-labeled segmentation masks.
for the optic disk (OD) and atrophied regions, as well as the x-y coordinates of the fovea. We created masks consisting of disks, 5\% of total image size in diameter, centered on the fovea location and thus were able to treat fovea detection as a segmentation task as well.

Figure 2: From left to right, a) RF image with OD and fovea labeled; b) RF image with atrophy labeled; c)-e) Segmentation masks produced for b) by single-task versions of the network (OD, fovea, atrophy), to serve as inputs to the multi-task version.

**General results** The three tasks defined for this dataset—optic disk (OD), atrophy, and fovea segmentation—differ considerably in difficulty with OD segmentation being the easiest followed by atrophy and then fovea segmentation; Dice scores of 0.894, 0.707, and 0.538 respectively.

**Output recycling results** Our key finding is that output recycling improves performance dramatically on the harder task when an easy and hard task are paired (Table 1). Baseline (a) corresponds to the right half of Figure 1 with no additional mask input. There is no performance gain when output is recycled in the single-task case—in fact performance appears to degrade on hard tasks (not shown). All reported values are averages of at least four experimental runs, with most SEMs $<0.01$.

Table 1: Dice scores on multi-task segmentations, (a) without and (b) with output recycling. Cases where output recycling leads to significant improved performance over both non-recycled and single-task versions are marked with *. "X+Y" indicates a multi-task experiment on classes X and Y; OD: Optic disk, A: Atrophy, F: Fovea.

<table>
<thead>
<tr>
<th></th>
<th>(a) OD+A</th>
<th>OD+F</th>
<th>A+F</th>
<th>(b) OD+A</th>
<th>OD+F</th>
<th>A+F</th>
</tr>
</thead>
<tbody>
<tr>
<td>OD</td>
<td>0.771</td>
<td>0.877</td>
<td>n/a</td>
<td>0.884</td>
<td>0.886</td>
<td>n/a</td>
</tr>
<tr>
<td>Atrophy</td>
<td>0.736</td>
<td>n/a</td>
<td>0.702</td>
<td>0.793*</td>
<td>n/a</td>
<td>0.727*</td>
</tr>
<tr>
<td>Fovea</td>
<td>n/a</td>
<td>0.532</td>
<td>0.492</td>
<td>n/a</td>
<td>0.566*</td>
<td>0.557*</td>
</tr>
</tbody>
</table>

**Patch-based segmentation results** In the fovea segmentation task (but not OD or atrophy segmentation) we found that CoordConv led to improved performance across experimental conditions, with improvement of 0.192, 0.034, 0.064, and 0.012 Dice units in the 25\%, 50\%, 75\%, and 100\% patch size cases, respectively. In particular, in the 25\% patch size case, training collapsed unless CoordConv was used.
References


