Straight-Through Estimator (STE) explanation

What is an STE?

Let's suppose we want to binarize the activations of a layer using the following function:

$$f(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

The problem with this function is that its gradient is zero and so it is not possible to perform the backpropagation. To overcome this issue we will use a Straight-Through Estimator in the backward pass.

A Straight-Through Estimator estimates the gradients of a function ignoring the derivative of the threshold function and passing on the incoming gradient as if the function was an identity function. The following diagram will help explain it better.

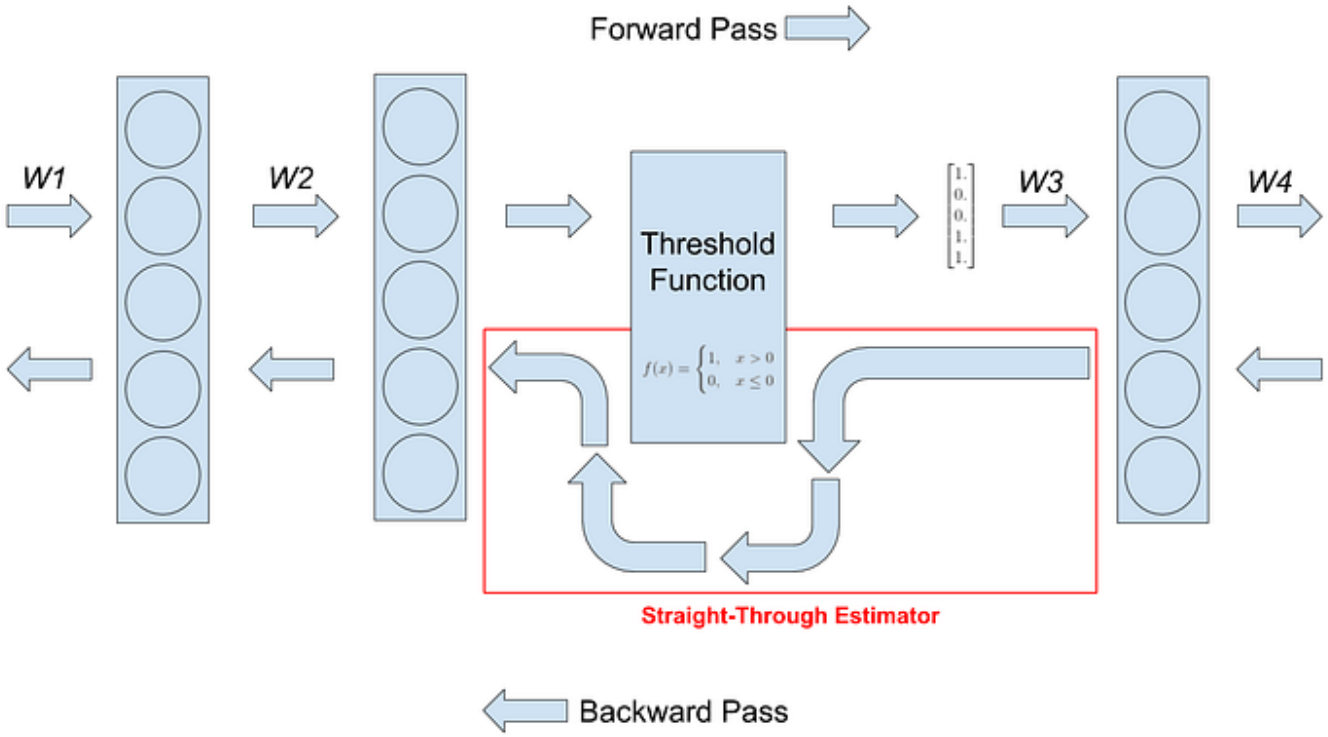

Figure 1: STE

Summary

An STE essentially bypasses the threshold function in the backward pass and makes the gradient of the threshold function look like the gradient of the identity function.

Specific implementation

In the specific case of our implementation, there are two nondifferentiable steps, both required to compute the loss based on the IoU ($loss_{IoU} = 1 - IoU(depth_{predicted}, depth_{GT})$):

1. cleaning the iToF predicted data from noise

$$iToF_{pred} = \begin{cases} 0, & |iToF_{pred}| < 0.05 \\ iToF_{pred}, & otherwise \end{cases} ;$$

2. performing the hard threshold on the depth data in order to obtain a binary mask

$$depth_{pred} = \begin{cases} 0, & depth_{pred} = 0 \\ 1, & depth_{pred} > 0 \end{cases} .$$

To overcome this issue we will use two different STEs, one for each step.

References

- Straight-Through Estimator original paper
- Straight-Through Estimator intuitive explanation