

# Appendix

```
import torch
import torch.nn as nn
import torch.nn.parameter as pm
class GaussianContextTransformer(nn.Module):
    def __init__(self, c=2, alpha=3, beta=1,
                 is_learnable=False):
        super(GaussianContextTransformer,
              self).__init__()
        #GCT-B0: is_learnable=False:
        #GCT-B1: is_learnable=True:
        #c: standard deviation
        #alpha and beta control the range of c
        if is_learnable:
            self.theta=pm.Parameter(torch.zeros(1))
            self.alpha = alpha
            self.beta = beta
            self.sig = nn.Sigmoid()
        self.c = c
        self.is_learnable = is_learnable
        self.avg_pool = nn.AdaptiveAvgPool2d(1)

    def forward(self, x):
        y = self.avg_pool(x)
        y = y - y.mean(dim=1, keepdim=True)
        std = y.std(dim=1, keepdim=True) + 1e-5
        y = y / std
        y = torch.pow(y, 2)
        if self.is_learnable:
            self.c = self.alpha *
                    self.sig(self.theta) + self.beta
        y = torch.exp(-y / (2 * self.c * self.c))
        y = y * x
        return y
```

Listing 1: PyTorch code of our GCT.

## 1. Appendix A: Code of GCT

Listing 1 gives PyTorch code of our GCT.

## 2. Appendix B: Other context transform functions

We investigate the performance of other transform functions in GCT-B0 based on ResNet50 on ImageNet validation set. Here we select two functions similar to Gaussian function: Witch of Agnesi  $\frac{5}{x^2+5}$  and hyperbolic function  $\frac{2}{e^{-0.5x}+e^{0.5x}}$ . The results are shown in Table 1. It is observed that Witch of Agnesi function achieves comparable Top-1 accuracy to Gaussian function. Hyperbolic function

Function	Top-1	Top-5
Gaussian	<b>77.51</b>	<b>93.86</b>
Witch of Agnesi	77.45	93.78
Hyperbolic	77.28	93.74

Table 1. Comparisons with other context transform functions. The results are based on the backbone ResNet50 on ImageNet validation set.

gets the worst performance, but also achieves a significant gain compared to the baseline ResNet50. Thus, we choose Gaussian function as the context transform function of GCT by default.