As shown in our paper, the back-propogate-max-operation loss is

$$L_{\mathrm{BM}} = -\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c}|}{|\boldsymbol{\Sigma}^{\mathbf{z}_c}|} + \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\Sigma}^{\mathbf{z}_c})^{-1}\boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c}\right) - \frac{1}{2}(\boldsymbol{\mu}_j^{\hat{\mathbf{z}}_c} - \mu^{\mathbf{z}_c})^\top(\boldsymbol{\Sigma}^{\mathbf{z}_c})^{-1}(\boldsymbol{\mu}_j^{\hat{\mathbf{z}}_c} - \boldsymbol{\mu}^{\mathbf{z}_c}),$$

Note that we have $\boldsymbol{\mu}^{\mathbf{z}_c}, \boldsymbol{\mu}_j^{\hat{\mathbf{z}}_c} \in \mathbb{R}^{256}$ and $\boldsymbol{\Sigma}^{\mathbf{z}_c}, \boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c} \in \mathbb{R}^{256\times256}$. In our code, `z_c_mu, z_c_mu_hat_j` $\in \mathbb{R}^{256}$ and `z_c_logsigma, z_c_logsigma_hat_j` $\in \mathbb{R}^{256}$ which represent the diagonal elements of $\boldsymbol{\Sigma}^{\mathbf{z}_c}, \boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c}$ under our assumptions.

In this way, we can get this form of the code:

$$\begin{aligned}
-\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c}|}{|\boldsymbol{\Sigma}^{\mathbf{z}_c}|} &= -\frac{1}{2}\log\frac{\prod_{i=1}^{256}\mathrm{z_c_logsigma[i].exp()}}{\prod_{i=1}^{256}\mathrm{z_c_logsigma_hat_j[i].exp()}} \\
&= -\frac{1}{2}\times\left(\sum_{i=1}^{256}\mathrm{z_c_logsigma[i]} - \sum_{i=1}^{256}\mathrm{z_c_logsigma_hat_j[i]}\right) \\
&= -\frac{1}{2}\times(\mathrm{z_c_logsigma} - \mathrm{z_c_logsigma_hat_j})
\end{aligned}$$

In the second term,

$$\begin{aligned}
\frac{1}{2}\mathrm{tr}\left((\boldsymbol{\Sigma}^{\mathbf{z}_c})^{-1}\boldsymbol{\Sigma}_j^{\hat{\mathbf{z}}_c}\right) &= \frac{1}{2}\left(\sum_{i=1}^{256}(1/\mathrm{z_c_logsigma[i].exp()})\times\mathrm{z_c_logsigma_hat_j.exp()}\right) \\
&= \frac{1}{2}\times(1/\mathrm{z_c_logsigma.exp()})\times\mathrm{z_c_logsigma_hat_j[i].exp()}
\end{aligned}$$

In the third term,

$$-\frac{1}{2}(\boldsymbol{\mu}_j^{\hat{\mathbf{z}}_c} - \boldsymbol{\mu}^{\mathbf{z}_c})^\top(\boldsymbol{\Sigma}^{\mathbf{z}_c})^{-1}(\boldsymbol{\mu}_j^{\hat{\mathbf{z}}_c} - \boldsymbol{\mu}^{\mathbf{z}_c}) = -\frac{1}{2}(\mathrm{z\_c\_mu\_hat\_j} - \mathrm{z\_c\_mu})^2\times(1/\mathrm{z\_c\_logsigma.exp()})$$

We can now summarize our bm loss of this form in our PyTorch codes:

$$\begin{aligned}
\mathrm{self.loss_bm} = -(1/2) * ((&\mathrm{z_c_logsigma_hat_j} - \mathrm{z_c_logsigma}) \\
&- ((1/\mathrm{z_c_logsigma.exp()}) * \mathrm{z_c_logsigma_hat_j.exp()}) \\
&+ (\mathrm{z_c_mu_hat_j} - \mathrm{z_c_mu}) * *2 * (1/\mathrm{z_c_logsigma.exp()})).\mathrm{sum}(1).\mathrm{mean}()
\end{aligned}$$