Appendix for CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows

A. Relationship with the flow framework

The loss function for the *reverse* $D_{KL}\left[\hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta})\|p_Z(\boldsymbol{z})\right]$ objective [1], where $\hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta})$ is the model prediction and $p_Z(\boldsymbol{z})$ is a target density, is defined as

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta})}\left[\log \hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta}) - \log p_Z(\boldsymbol{z})\right]. \tag{5}$$

The first term in (5) can be written using (4) definition for a standard MVG prior ($\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})$) as

$$\log \hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta}) = \log(2\pi)^{-D/2} - E^2(\boldsymbol{u})/2 + \log\left|\det \boldsymbol{J}\right|, \tag{5.1}$$

where $E^2(\boldsymbol{u}) = \|\boldsymbol{u}\|_2^2$ is a squared Euclidean distance of $\boldsymbol{u}$.

Similarly, the second term in (5) can be written for MVG density (2) using a square of Mahalanobis distance as

$$\log p_Z(\boldsymbol{z}) = \log(2\pi)^{-D/2} + \log\det \boldsymbol{\Sigma}^{-1/2} - M^2(\boldsymbol{z})/2. \tag{5.2}$$

By substituting (5.1-5.2) into (5), the constants $\log(2\pi)^{-D/2}$ are eliminated and the loss is

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\hat{p}_Z(\boldsymbol{z},\boldsymbol{\theta})}\left[\frac{M^2(\boldsymbol{z}) - E^2(\boldsymbol{u})}{2} + \log\frac{|\det \boldsymbol{J}|}{\det \boldsymbol{\Sigma}^{-1/2}}\right]. \tag{6}$$

B. CFLOW decoders for likelihood estimation

We train CFLOW-AD using a maximum likelihood objective, which is equivalent to minimizing the *forward* D_{KL} objective [1] with the loss defined by

$$\mathcal{L}(\boldsymbol{\theta}) = D_{KL}\left[p_Z(\boldsymbol{z})\|\,\hat{p}_Z(\boldsymbol{z},\boldsymbol{c},\boldsymbol{\theta})\right], \tag{7}$$

where $\hat{p}_Z(\boldsymbol{z},\boldsymbol{c},\boldsymbol{\theta})$ is a conditional normalizing flow (CFLOW) model with a condition vector $\boldsymbol{c} \in \mathbb{R}^C$.

The target density $p_Z(\boldsymbol{z})$ is usually replaced by a constant because the parameters $\boldsymbol{\theta}$ do not depend on this density during gradient-based optimization. Then by analogy with unconditional flow (4), the loss (7) for $\hat{p}_Z(\boldsymbol{z},\boldsymbol{c},\boldsymbol{\theta})$ can be written as

$$\mathcal{L}(\boldsymbol{\theta}) = -\mathbb{E}_{p_Z(\boldsymbol{z})}\left[\log p_U(\boldsymbol{u}) + \log\left|\det \boldsymbol{J}\right|\right] + \text{const}. \tag{7.1}$$

In practice, the expectation operation in (7.1) is replaced by an empirical train dataset $\mathcal{D}_{\text{train}}$ of size N. Using the definition of base distribution with $p_U(\boldsymbol{u})$, the final form of (7) can be expressed as

$$\mathcal{L}(\boldsymbol{\theta}) \approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\|\boldsymbol{u}_i\|_2^2}{2} - \log\left|\det \boldsymbol{J}_i\right|\right] + \text{const}, \tag{7.2}$$

where the random variable $\boldsymbol{u}_i = g^{-1}(\boldsymbol{z}_i,\boldsymbol{c}_i,\boldsymbol{\theta})$ and the Jacobian $\boldsymbol{J}_i = \nabla_{\boldsymbol{z}} g^{-1}(\boldsymbol{z}_i,\boldsymbol{c}_i,\boldsymbol{\theta})$ depend both on input features $\boldsymbol{z}_i$ and conditional vector $\boldsymbol{c}_i$ for CFLOW model.

References

[1] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 2021.