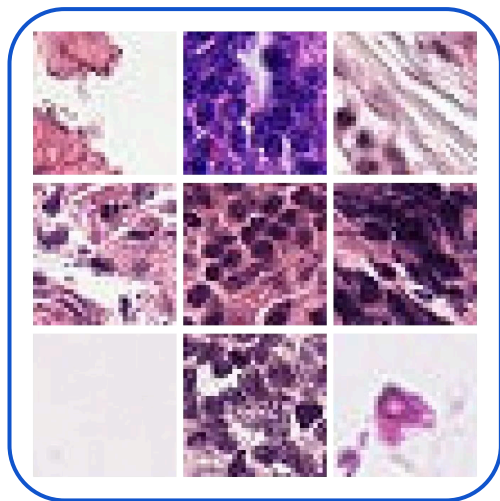
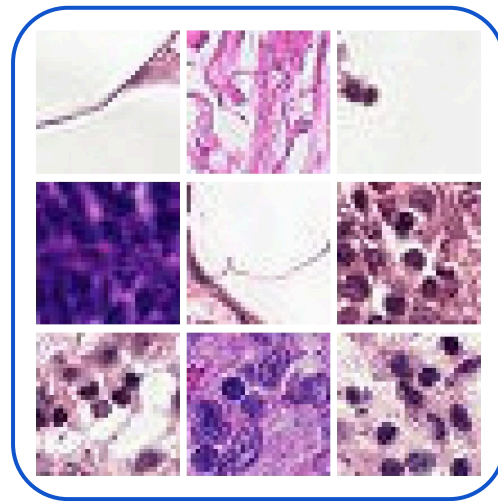
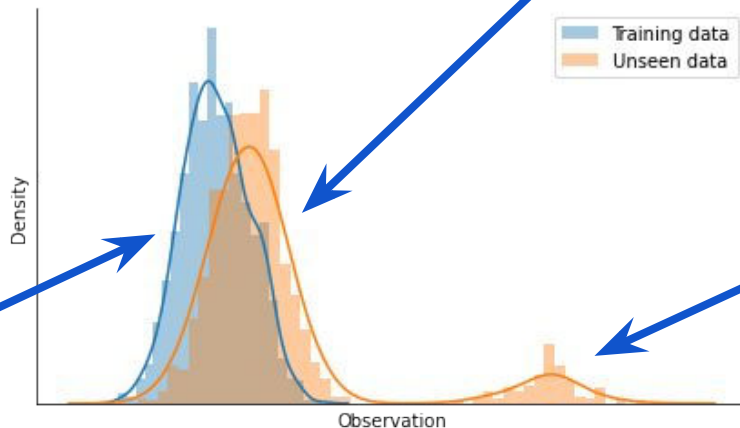


Abnormality Detection in Histopathology using Normalising Flows

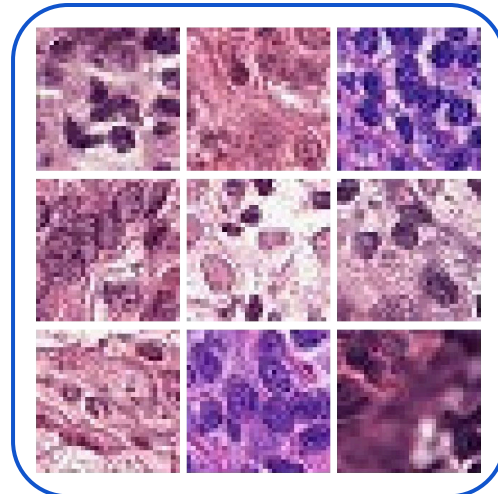
- 1) Train residual flow to estimate likelihood $p_{\phi}(x)$ of healthy samples from PatchCamelyon
- 2) Score unseen examples under trained model



Training data



Unseen healthy examples



Unseen unhealthy examples

Training Ensembling & Outlier Scores

The likelihood $p_{\phi}(x)$ is a bad outlier score

→ we test different metrics using multiple model checkpoints:

- Likelihood, mean likelihood and variance of the likelihood
- WAIC: $\mathbb{E}_i[\ln p_{\phi_i}(x)] - \text{Var}[\ln p_{\phi_i}(x)]$
- Adapted Typicality: $|\mathbb{E}_i[-\ln p_{\phi_i}(x) - \mathbb{E}_{x' \sim X_{train}}[-\ln p_{\phi_i}(x')]]|$

Results

Table 1: Comparison of AUROCs of correctly classified patches from the PCam test set using different methods. The first row compares different OOD scoring metrics using the trained Residual Flow. The single log-likelihood result is computed using the last model checkpoint. Typ. refers to our variation on the typicality test introduced by (Nalisnick et al.). GDensenet refers to the official supervised PatchCAM baseline (Veeling et al., 2018).

Method	$\log p_\phi$	$\mathbb{E}_i[\log p_{\phi_i}]$	$\text{Var}_i[\log p_{\phi_i}]$	WAIC	Typ.	Gaussian	GDensenet
AUROC [%]	53.4	81.6	92.4	25.3	61.8	31.8	96.3