From RealNVP to Glow: Anomaly Detection on Camelyon16 with MSFlow, SE-FPN, and Meta-Ensemble Fusion

Adi Inbar (025728999) & Noam Arian (311271829)

Abstract

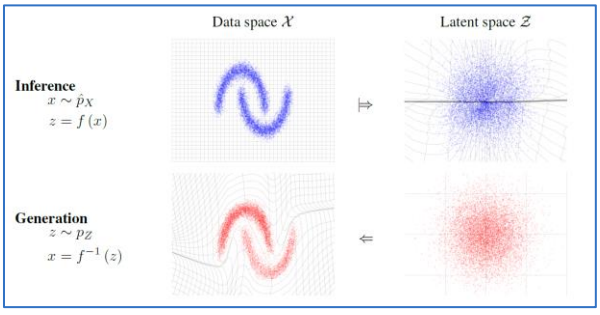
This paper explores a multi-stage, deep learning pipeline for medical anomaly detection on the Camelyon16 dataset, based on multi-scale normalizing flows (MSFlow). Starting with a RealNVP based MSFlow model, we iteratively improved the architecture through augmentations, channel calibration techniques, and advanced loss functions. After detecting structural weaknesses in RealNVP, we transitioned to a Glow based implementation enhanced with SE (Squeeze-and-Excitation) and FPN (Feature Pyramid Network) modules. We evaluated several Vision Transformer (ViT) backbones and ultimately rejected them due to architectural incompatibility with density-based models. Finally, we integrated two specialized models using a meta-ensemble strategy powered by XGBoost. Our approach achieved superior performance compared to existing benchmarks like BMAD and MedIAnomaly.

1. Introduction

Anomaly detection in histopathology is a critical task for early cancer detection. Automated tools are increasingly being adopted to assist pathologists in detecting abnormal tissue regions in whole-slide images (WSIs). The rarity and subtlety of anomalies in such images present a significant challenge, especially under limited labeled data. In this study, we investigate the use of Normalizing Flow, particularly MSFlow as a density-based method for medical anomaly detection. We explore different flow architectures, feature extractors, augmentation strategies, and ensemble techniques to maximize performance. Our results highlight how combining traditional convolutional backbones with modern flow models and intelligent meta learners can yield state of the art results on Camelyon16

2. Background: Normalizing Flows for Anomaly Detection

Normalizing Flows are invertible deep learning models that map complex data distributions into tractable, known priors (e.g., Gaussian), allowing exact likelihood computation. They are particularly suited for out of distribution detection because they model the distribution of normal data and flag low probability instances as anomalies. MSFlow extends this idea by applying flows at multiple scales, each modeling different levels of abstraction in the image, and combining them using learnable fusion weights. This architecture enables localized anomaly detection and improves sensitivity to subtle changes.



Bidirectional Mapping in a Normalizing Flow: Inference and Generation

Normalizing flows constitute a generative model whose density is explicitly defined: an observation x is transformed by a bijective, differentiable map f into a latent code z = f(x) that follows a simple prior (usually $\mathcal{N}$-($\mathbb{I}$)). Because the transformation is invertible and its Jacobian determinant is tractable, the model can evaluate the exact log-likelihood.

$$\log(p_x(x)) = \log(p_z\,(f(x))) + \log\left|det\left(\frac{\partial f(x)}{\partial x^T}\right)\right|$$

where $\frac{\partial f(x)}{\partial x^T}$ is the Jacobian of f at x

and it can *generate* data by drawing z~pZ and applying the inverse f^{-1}. These two abilities set flows apart: they behave like an implicit sampler when we ignore the likelihood term (akin to GANs), yet they remain fully *explicit* when quantitative density estimates are required, something GANs and most diffusion models cannot offer.

To keep both likelihood computation and sampling efficient, each layer in the chain must satisfy three constraints: it must be invertible and differentiable so that f^{-1} exists; it must be strictly bijective to conserve probability mass; and the determinant of its Jacobian must be computable in closed form or with O(D) work, guaranteeing tractability in high dimensions. Affine coupling and related flow layers meet these criteria by designing triangular Jacobians whose determinants reduce to a cheap product of diagonal entries, while their inverses are analytic and inexpensive. Because of this reversible architecture, normalizing flows can synthesize sharp images, reconstruct inputs exactly, and provide calibrated anomaly scores from log-likelihoods in a single unified framework. They therefore bridge the historical divide between models that are easy to sample from but hard to score (implicit models) and those that are easy to score but slow to sample, offering the best of both worlds in medical mage anomaly detection, density estimation, compression, and controllable image manipulation.

3. The Camelyon16 Dataset

Camelyon16 is a gold standard dataset for evaluating automated tumor detection systems in WSIs. It comprises annotated tissue slides from lymph nodes, with pixel level labels indicating metastases. For this study, we extracted 5,800 "normal" patches and 580 "anomalous" ones, ensuring non overlapping splits for training, validation, and ensemble fusion. The dataset presents a high class imbalance and subtle texture variations, making it ideal for testing density based anomaly detection methods.

4. Data Augmentations

To simulate real world noise and variability, soft augmentations were applied during training. These included:

- **Gaussian Blur** to simulate defocus
- **Color Jitter** with low amplitude to simulate staining variations
- **Additive Gaussian Noise** to simulate sensor imperfections
  These augmentations help the model generalize beyond the training distribution without introducing synthetic patterns that could mislead the flow models.

Additionally, simple geometric augmentations such as Random Rotation and Random Horizontal Flip were applied.

- **Random Rotation** introduces orientation variability by rotating images by a small random angle (e.g., ±15°), improving robustness to positional variations in tissue.
- **Random Horizontal Flip** mirrors images horizontally with a fixed probability, helping the model generalize across symmetric patterns and structural variations.

5. Evaluation Metrics

The models were evaluated using the following metrics:

- **Precision & Recall:** to assess detection capabilities
- **F1-Score:** to balance false positives and false negatives
- **AUROC (Area Under the Receiver Operating Characteristic Curve):** for threshold-independent evaluation
- **AP (Average Precision):** effective for imbalanced datasets
- **Confusion Matrix:** to provide insight into classification tendencies
  These metrics helped us benchmark each architecture and identify whether performance gains were skewed toward detecting normal or anomalous samples.

6. MSFlow with RealNVP: The Starting Point

The initial model was MSFlow equipped with RealNVP blocks. It successfully modeled the density of normal tissue and demonstrated reasonable performance on clean inputs. However, it had several limitations:

- Dead channels in deep layers
- Underperformance in detecting small, subtle anomalies
- Inefficient channel utilization
  We addressed these issues by adding skip connections and increasing flow depth. While this led to marginal improvements, RealNVP's architecture proved structurally limiting. A critical bottleneck was its affine coupling mechanism, which could not sufficiently capture fine texture granularity in histological images.

7. Vision Transformers: Promise and Pitfalls

Several Vision Transformer variants, including ViT-B/16, vit_base_patch16_224.mae, vit_base_patch16_224_dino, deit_small_patch16_224, facebook/dinov2-base, vit_small_patch16_224.dino pretrained with self-supervised DINO, were explored. Although the transformers effectively captured long-range dependencies and contextual patterns, they produced token sequences rather than spatial feature maps. Integrating these sequences with MSFlow required artificially reshaping tokens into pseudo feature maps, an operation that was nontrivial and resulted in information loss. Furthermore:

- ViT-only pipelines without MSFlow exhibited a significant performance drop
- The lack of spatial consistency undermined anomaly localization
  Despite their popularity, ViTs were eventually abandoned in favor of convolutional backbones due to better spatial resolution and compatibility with flow-based models.

8. Transition to Glow and Architectural Improvements

Glow replaced RealNVP as the flow mechanism. It provided:

- Invertible 1x1 convolutions
- ActNorm layers for stable training
- Greater expressiveness with fewer dead filters
  We further enhanced Glow-based MSFlow with two modules:

Squeeze-and-Excitation (SE)

SE modules dynamically recalibrate feature channels based on global context. This allowed the model to focus on more informative channels, improving its discrimination between normal and anomalous patterns.

Feature Pyramid Network (FPN)

FPN aggregates features from multiple scales, preserving both fine-grained texture and semantic information. This helped detect both micro level and macro level anomalies, leading to robust performance.

These additions significantly improved anomaly localization and generalization, especially on edge cases previously misclassified by RealNVP.

9. Specialized Loss Functions: Focal, Center, Margin

To further enhance learning, several loss functions were incorporated:

- **Focal Loss:** reduced the weight of easy examples, focusing training on hard-to-classify patches
- **Center Loss:** encouraged embeddings of the same class to cluster together
- **Margin Loss:** enforced separation between normal and abnormal feature distributions
  These loss functions were applied during training and improved both stability and discriminative power.

10. Meta-Ensemble with XGBoost

Instead of hardcoded fusion rules, a data driven approach was adopted:

- Two MSFlow models were trained separately: one optimized for normal detection, the other for anomaly recall.
- Their predictions and embedding-based features were passed to an **XGBoost** meta-learner.
- The meta-model learned to combine predictions optimally, adjusting thresholds per case.

This ensemble strategy corrected many of the Type I and II errors from the base models, yielding better balance across metrics.

11. Results and Comparison with SOTA

The best precision model, msflow_Glow_Focal_Center_margin_Loss_AugThreshold, achieved:

AUC: 0.8642

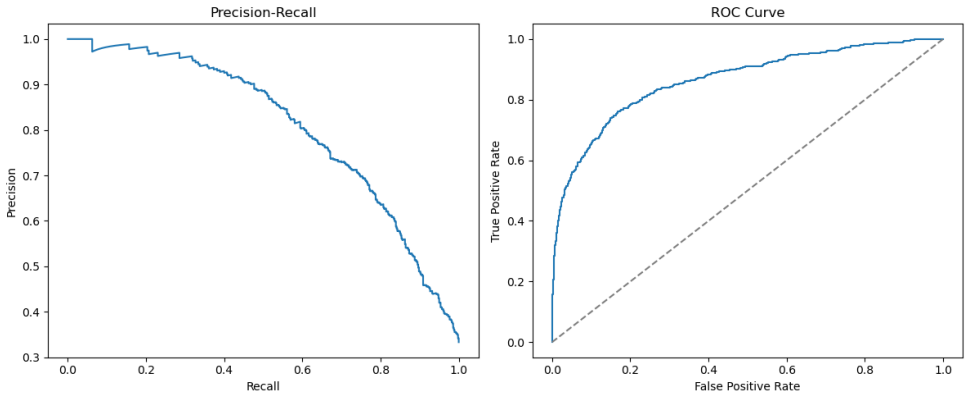
AP: 0.8088

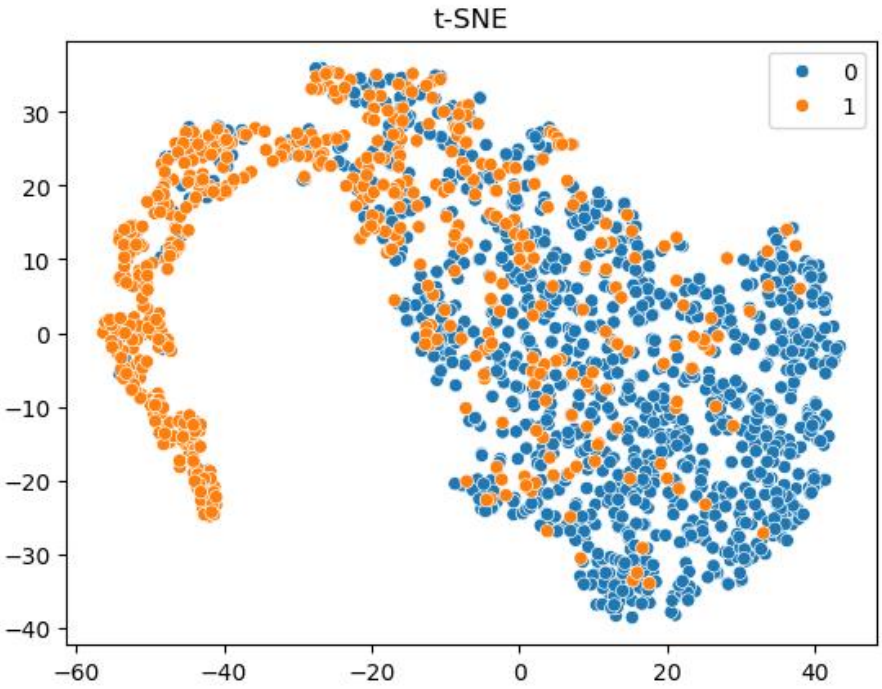
Confusion Matrix:

[[1104 16]

[346 214]]

Precision: 0.9304 , Recall: 0.3821





The best recall model, msflow_Glow_FocalLoss_Center_AugThreshold, achieved:

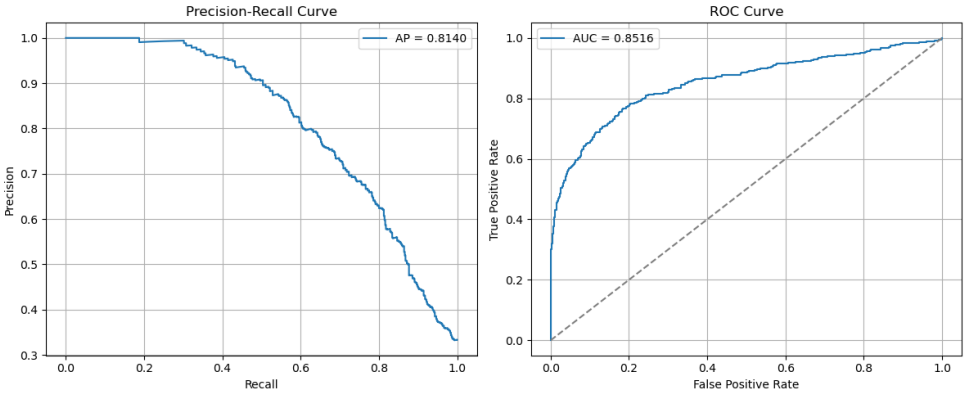
AUC: 0.8516 Confusion Matrix:

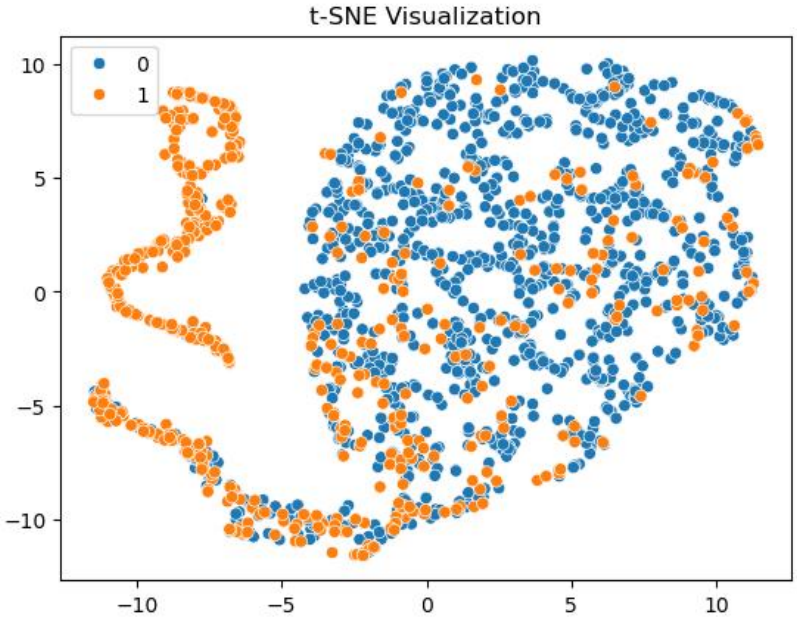
[[670 450]

[74 486]]

Average Precision (AP): 0.8140

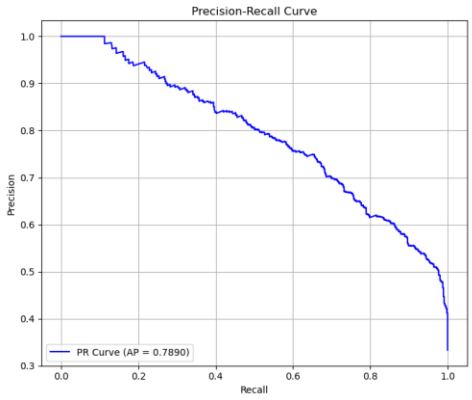
Precision: 0.5192 , Recall: 0.8679

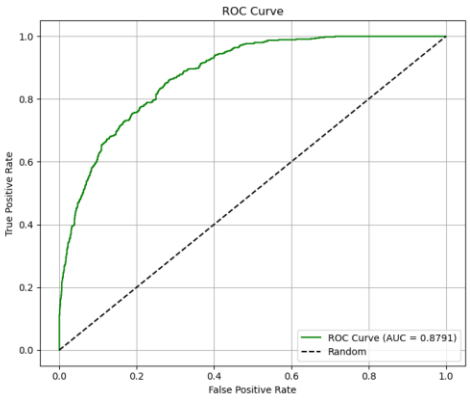


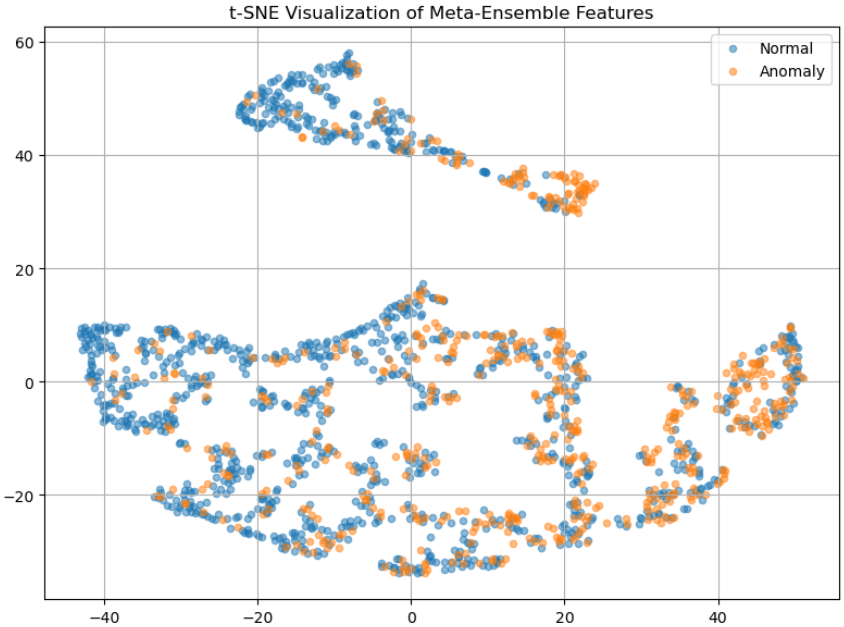


The meta-ensemble consisting of these two models achieved:"

- **AUROC**: 0.8791
- **AP**: 0.7890
- **F1-score**: 0.6642
- **Recall**: 0.9750
- **Precision**: 0.5037







Model	Visual Characteristics	Degree of Separation
Model 1	Large central cluster with class mixing	🔴 **Low to modearate**
Model 2	Two distinct elongated clusters; improved class separability	🟠 **Moderate to High**
Meta-Ensemble (XGBoost)	Multiple localized clusters with clearer class boundaries	🟢 **Relatively High**

Revised Interpretation Paragraph

The t-SNE plots provide an intuitive comparison of the internal feature representations across models. The first model (MSFlow 1) demonstrates substantial intermixing between normal and anomalous samples, reflected in the dense, overlapping central structure. In contrast, the second model (MSFlow 2) reveals a more structured and elongated separation, indicating improved but not perfect class distinction. The meta-ensemble model further refines this by forming more clearly delineated clusters in the embedded space. While none of the models achieve perfect separation, an expected outcome given the subtlety of anomalies in medical images, the ensemble model appears to leverage the combined strengths of the base models, leading to a more informative representation.

In comparison:

- **BMAD the 4 best results on Camelyon16**:

Method	Image AUROC
CutPaste	75.18 ± 0.41
MKD	77.54 ± 0.27
PatchCore	69.34 ± 0.2
UTRAD	69.96 ± 4.64

- **MedIAnomaly the 4 best results on Camelyon16**:

שיטה	AUC	AP
AutoDDPM	80.7±0.3	76.7±0.2
AE-PL	76.1±0.1	67.6±0.4
DAE	65.4±2.2	60.6±1.8
AE-U	60.6±2.7	55.5±1.0

12. Conclusion and Future Directions

Our study shows the strength of combining Normalizing Flows with modern convolutional enhancements for anomaly detection. Key takeaways include:

- ViTs are promising but not yet suited for flow-based modeling without adaptation.
- RealNVP is limited by dead channels; Glow offers superior density modeling.
- SE and FPN significantly boost anomaly localization and robustness.
- Meta-ensembles using XGBoost yield better overall performance than hard-coded rules.

Future Work:

Reinforcement Learning for Adaptive Thresholding

Another promising future direction is the use of **Reinforcement Learning (RL)** to dynamically determine the **optimal decision threshold** during inference. Instead of relying on static threshold tuning via validation or grid search, an RL agent can be trained to **adaptively select thresholds** based on real time score distributions, class imbalance, or feedback from performance metrics (e.g., F1-score, AUROC). This framing treats threshold selection as a sequential decision-making process, where the agent learns a thresholding policy that maximizes reward over multiple environments or data distributions.
Such an approach could allow the system to self-calibrate over time, enabling **autonomous optimization** even when data shifts or new medical domains are introduced

Data Expansion and Preprocessing

To enhance model performance and address the observed overlap in the feature space (as seen in the t-SNE plots), we propose incorporating additional data from the **Camelyon17** dataset. This dataset extends the Camelyon16 cohort and offers a broader variety of histopathological images, improving generalization and anomaly separation. Due to the extremely high resolution of WSI files (1–3 GB per image), each image must be downscaled to a uniform size (e.g., 256×256) using anti aliasing to preserve structural fidelity. Corresponding ground truth annotations should be resized accordingly to maintain spatial accuracy. All data should be stored in standardized RGB format and organized into good and ungood folders based on lesion masks.

This process will support a more balanced training set and potentially improve the clustering and separability of normal and anomalous samples, especially in low-data regimes where overfitting or undergeneralization are critical concerns.

References

1. **BCEWithLogitsLoss Documentation, PyTorch Official Docs.**
   https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html
2. **BMAD.**
   https://arxiv.org/abs/2209.07682
3. **Camelyon16.**
   https://camelyon16.grand-challenge.org/
4. **CHATGPT.**
   https://chat.openai.com
5. **FPN.**
   https://arxiv.org/abs/1612.03144
   https://arxiv.org/abs/1803.01534
   https://arxiv.org/abs/1904.07392
   https://arxiv.org/abs/1911.09070
6. **Jadon (2020).** A Survey of Loss Functions for Semantic Segmentation.
   https://arxiv.org/abs/2006.14822
7. **Kingma, D. P., & Dhariwal, P. (2018).** Glow: Generative Flow with Invertible 1x1 Convolutions. NeurIPS.
   https://papers.neurips.cc/paper/8224-glow-generative-flow-with-invertible-1x1-convolutions.pdf
8. **Lin et al. (2017).** Focal Loss for Dense Object Detection. ICCV.
   https://doi.org/10.1109/ICCV.2017.324
9. **MedIAnomaly.**
   https://arxiv.org/abs/2404.04518
10. **MSFlow.**
    https://arxiv.org/abs/2308.15300
11. **RealNVP.**
    https://arxiv.org/abs/1605.08803
12. **SE (Squeeze-and-Excitation).**
    https://arxiv.org/abs/1709.01507
    https://arxiv.org/abs/1807.06521
    https://arxiv.org/abs/1910.03151
    https://arxiv.org/abs/2107.13586
    https://github.com/hujie-frank/SENet
13. **t-SNE: van der Maaten & Hinton (2008).** Visualizing Data using t-SNE. JMLR.
    https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf
    https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
14. **Zhang et al. (2021).** Delving Deep into Loss Functions for Binary Classification.
    https://arxiv.org/abs/2106.07353