

Normalizing Flows

Flow-based Deep Generative Models *Part 2: Anomaly & Rare Events Detection*

A.Demichev

March 2025

Основная задача:

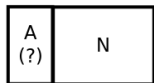
5. Выполнить вычислительные эксперименты и провести сравнение эффективности выделения **редких событий** с использованием нейросетевых моделей на базе автоэнкодеров и **нормализующих потоков**.

Последовательность постановки задачи:

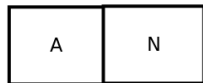
1. Аномалии и редкие события: разница в постановке задач
2. Анализ аномалий вне/до глубокого обучения
 - ▶ Transforming for normality \Leftrightarrow normalizing flows
3. Анализ аномалий с помощью глубокого обучения (Yu et al. FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows)
4. Использование размеченных (синтетических) аномальных событий для поиска аномалий на основе NF - как их учесть?
 - ▶ Chiu et al. Self-Supervised Normalizing Flows for Image Anomaly Detection and Localization
 - ▶ Смысл функций и метода правдоподобия
 - ▶ Минимизация функции правдоподобия аномалий (синтетических)
5. Упрощенный (предварительный) вариант постановки задачи
 - ▶ экстрактор \Rightarrow параметры Хилласа (сколько? какие?)
 - ▶ учебный NF (моя презентация)
 - ▶ анализ нормального распределения/аномалий

Anomaly vs. Rare events

Аномальные события



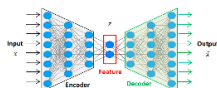
Обучающий набор данных: $A \ll N$



Тестовый набор данных: $A \sim N$



One-class learning

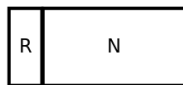


min "MSE"

Редкие события



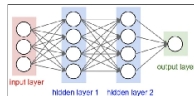
Обучающий набор данных: $R \sim N$



Тестовый набор данных: $R \ll N$

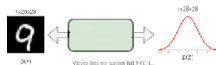


Two-class learning



Classifier

NF Classifier



Guess: max likelihood **normal** + min likelihood **rare** (???)

Общий алгоритм детектирования аномалий

1. Отбор/выделение существенных признаков (понижение размерности) данных
2. Преобразование распределения **неаномальных** данных в обучающей выборке в нормальное; *как - см. ниже*
 - ▶ **Важно:** Любая *непрерывная и биективная* функция является монотонной
 - ▶ \Rightarrow сэмпл с большой вероятностью в исходном распределении перейдет в **верхнюю часть** (многомерного) нормального распределения (**близко среднему**)
3. поскольку отображение оптимизируется под выборку **неаномальных** данных, аномалии имеют низкую вероятность
 - ▶ в исходном пространстве со сложным распределением (+имеем только выборку) сложно провести границу между нормальными и аномальными данными
 - ▶ для нормального распределения – легко, например, граница определяется 2σ (95%)

Общий алгоритм детектирования аномалий (2)

- 4 Многомерное (асимметричное) нормальное распределение + корреляции между измерениями (признаками) \Rightarrow подходящей мерой удаленности от среднего является *расстояние Махаланобиса* (Mahalanobis distance)

$$d_M(\vec{x}, Q) = \sqrt{(\vec{x} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{x} - \vec{\mu})}.$$

$$\Sigma_{ij} = \text{cov}(X_i, Y_j) \equiv \mathbb{E}[(X_i - \mathbb{E}X_i)(Y_j - \mathbb{E}Y_j)], \quad i, j = 1, \dots, n,$$

- 5 вероятность через d_M

$$\mathcal{N}[\vec{x}] = \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}} \exp\left(-\frac{d_M^2}{2}\right).$$

- 6 порог для аномалий: $\mathcal{N}[\vec{x}] < \epsilon \Rightarrow$ геометрически определяется d_M (многомерный овал)

Transforming Probability Distributions for Normality

"by hands"

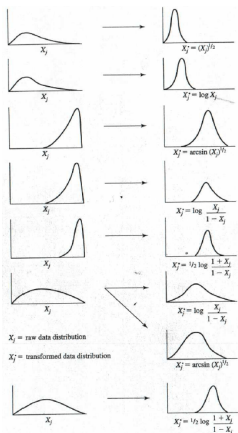


Рис.: Figure from Stevens (2002) Applied Multivariate Statistics for the Social Sciences 5th ed.

Anomaly Detection via DL+NF (Yu et al.): one-class

- ▶ to learn the parameter θ to map all x_i from the raw distribution $p_X(x)$ into the normal distribution $p_Z(z)$, with anomalous instances **mapped out** of the distribution.

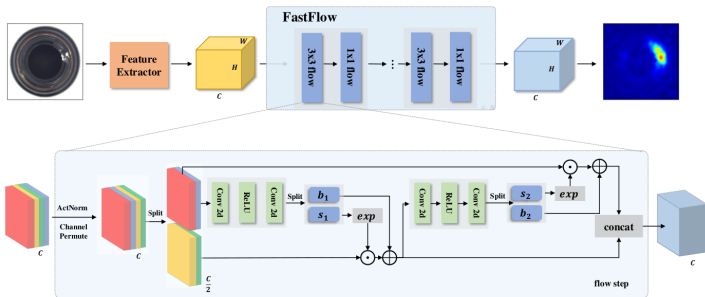


Рис.: FastFlow NN based on RealNVP. The *split*(\cdot) and *concat*(\cdot) functions perform splitting and concatenation operations along the **channel** dimension.

Code: <https://paperswithcode.com/paper/fastflow-unsupervised-anomaly-detection-and>

<https://anomalib.readthedocs.io/en/latest/markdown/guides/reference/models/image/fastflow.html>

Models with NF: RealNVP

affine coupling layer f: $\mathbf{x} \mapsto \mathbf{y}$:

$$\begin{aligned}\mathbf{y}_{1:d} &= \mathbf{x}_{1:d} \\ \mathbf{y}_{d+1:D} &= \mathbf{x}_{d+1:D} \odot \exp(\sigma(\mathbf{x}_{1:d})) + \mu(\mathbf{x}_{1:d})\end{aligned}\tag{1}$$

where $\sigma(\cdot)$ and $\mu(\cdot)$ are scale and translation functions and both map $\mathbb{R}^d \mapsto \mathbb{R}^{D-d}$. The operation \odot is the element-wise product.

► easily invertible

$$\begin{cases} \mathbf{y}_{1:d} &= \mathbf{x}_{1:d} \\ \mathbf{y}_{d+1:D} &= \mathbf{x}_{d+1:D} \odot \exp(\sigma(\mathbf{x}_{1:d})) + \mu(\mathbf{x}_{1:d}) \end{cases} \Leftrightarrow \tag{2}$$

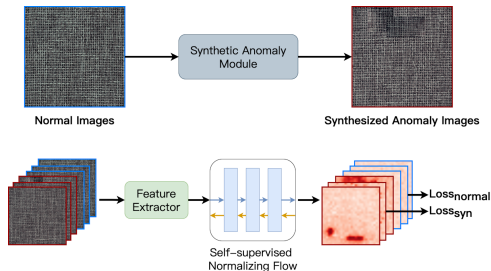
$$\Leftrightarrow \begin{cases} \mathbf{x}_{1:d} &= \mathbf{y}_{1:d} \\ \mathbf{x}_{d+1:D} &= (\mathbf{y}_{d+1:D} - \mu(\mathbf{y}_{1:d})) \odot \exp(-\sigma(\mathbf{y}_{1:d})) \end{cases} \tag{3}$$

NF for Anomalies \implies NF for Rare Events (RE)

Chiu, Li-Ling, et al. "Self-supervised normalizing flows...".

https://openaccess.thecvf.com/content/CVPR2023W/VAND/papers/Chiu_Self-Supervised_Normalizing_Flows_for_Image_Anomaly_Detection_and_Localization_CVPRW_2023_paper.pdf

- ▶ **two-class learning**
- ▶ using the **synthesized anomalous training samples**, a **novel** self-supervised NF (RealNVP)
- ▶ trained by **maximizing the likelihood** of normal images and **minimizing the likelihood** of anomalous images.



Anomaly/RE detection via DL+NF: two-class learning (1)

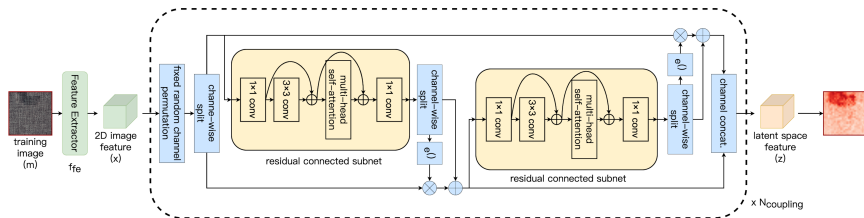


Рис.: The architecture of the model by Chiu et al.

- ▶ the architecture is very similar to the FastFlow
- ▶ but the training and loss function are essentially different
- ▶ Given a set of N normal/bulk features $B = \{x_b\}_{b=1}^N$ and a set of N (synthetic) anomalous/rare image features $R = \{x_r\}_{r=1}^N$
- ▶ the model is trained to maximize the likelihood of B and minimize the likelihood of R .

Anomaly/RE detection via DL+NF: two-class learning (2)

Loss function:

$$L_{total} = L_B - L_R$$

$$\begin{aligned} L_B &= \frac{1}{N} \sum_{x_b \in B} -\log p_B(x_b) \\ &= \frac{1}{N} \sum_{x_b \in B} \left[-\log p_Z(f_\theta(x_b)) - \log \left| \det \frac{\partial f_\theta(x_b)}{\partial x_b} \right| \right], \quad (4) \end{aligned}$$

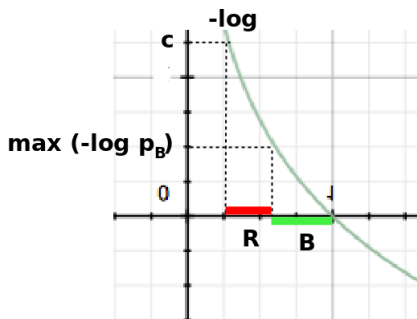
$$\begin{aligned} L_R &= \sum_{x_r \in R} -\log p_R(x_r) \cdot \mathbb{I} \left[\log p_R(x_r) > c \right. \\ &\quad \left. \wedge \log p_R(x_r) > \min \left(\log p_B(x_b) \right) \right] \quad (5) \end{aligned}$$

\mathbb{I} = indicator function

Anomaly/RE detection via DL+NF: two-class learning (3)

Meaning of the conditions in the last eq.

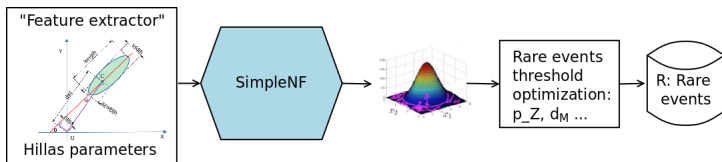
- ▶ In order to prevent the log-likelihood of synthesized samples $-\log p_R(x_r)$ from reaching to $+\infty$, [Kirichenko et al.]
- ▶ bulk images and rare may be slightly different \Rightarrow the likelihood of all B -samples should be larger than of R [Dohi et al.]



- ▶ Problem/question: the meaning of L_{total} ? **Likelihood???**
- ▶ порог для аномалий: $\mathcal{N}[\vec{x}] < \epsilon \Rightarrow$ геометрически определяется d_M

Возможная постановка **предварительной** задачи

1. feature extractor \rightarrow Hillas parameters
2. simpleNF \equiv simple NF, e.g.,
 - ▶ PyTorch:
 - ▶ https://github.com/gebob19/introduction_to_normalizing_flows/blob/master/normalizing_flows.ipynb
previous pres., page 16 – 24
 - ▶ <https://github.com/AxelNathanson/pytorch-normalizing-flows>
 - ▶ TensorFlow: previous pres., page 26, etc.
3. threshold - according to training set



Materials used:

- ▶ <https://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1/https://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1/>
- ▶ <https://math.stackexchange.com/questions/2082015/are-monotonic-and-bijective-functions-the-same>
- ▶ Yu, Jiawei, et al. "Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows." arXiv preprint arXiv:2111.07677 (2021).
- ▶ Chiu, Li-Ling, and Shang-Hong Lai. "Self-supervised normalizing flows for image anomaly detection and localization." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2927-2936. 2023.
https://openaccess.thecvf.com/content/CVPR2023W/VAND/papers/Chiu_Self-Supervised_Normalizing_Flows_for_Image_Anomaly_Detection_and_Localization_CVPRW_2023_paper.pdf
- ▶ Kirichenko, Polina, Pavel Izmailov, and Andrew G. Wilson. "Why normalizing flows fail to detect out-of-distribution data." Advances in neural information processing systems 33 (2020): 20578-20589.
<https://proceedings.neurips.cc/paper/2020/file/ecb9fe2fbb99c31f567e9823e884dbec-Paper.pdf>
- ▶ Dohi, Kota, Takashi Endo, Harsh Purohit, Ryo Tanabe, and Yohei Kawaguchi. "Flow-based self-supervised density estimation for anomalous sound detection." In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 336-340. IEEE, 2021. <https://arxiv.org/pdf/2103.08801>