Journal of Pathology Informatics 13 (2022) 100102

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Original Research Article

Multiscale generative model using regularized skip-connections and perceptual loss for anomaly detection in toxicologic histopathology

Philip Zehnder [1], Jeffrey Feng [1], Reina N. Fuji, Ruth Sullivan, Fangyao Hu *

Department of Safety Assessment, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA

ARTICLE INFO

Keywords:
Anomaly detection
Digital pathology
Deep learning
Toxicological pathology

ABSTRACT

Background: Automated anomaly detection is an important tool that has been developed for many real-world applications, including security systems, industrial inspection, and medical diagnostics. Despite extensive use of machine learning for anomaly detection in these varied contexts, it is challenging to generalize and apply these methods to complex tasks such as toxicologic histopathology (TOXPATH) assessment (i.e.,finding abnormalities in organ tissues). In this work, we introduce an anomaly detection method using deep learning that greatly improves model generalizability to TOXPATH data.

Methods: We evaluated a one-class classification approach that leverages novel regularization and perceptual techniques within generative adversarial network (GAN) and autoencoder architectures to accurately detect anomalous histopathological findings of varying degrees of complexity. We also utilized multiscale contextual data and conducted a thorough ablation study to demonstrate the efficacy of our method. We trained our models on data from normal whole slide images (WSIs) of rat liver sections and validated on WSIs from three anomalous classes. Anomaly scores are collated into heatmaps to localize anomalies within WSIs and provide human-interpretable results.

Results: Our method achieves 0.953 area under the receiver operating characteristic on a real-worldTOXPATH dataset. The model also shows good performance at detecting a wide variety of anomalies demonstrating our method's ability to generalize to TOXPATH data.

Conclusion: Anomalies in both TOXPATH histological and non-histological datasets were accurately identified with our method, which was only trained with normal data.

Background

The goal of anomaly detection is to identify abnormal data residing outside the distribution of normal data.[1–4] Anomaly detection is useful in many fields. For example, it is critical to identify potential weapons manifesting as common, hidden, or unfamiliar "abnormalities" in normal X-ray security passenger screening data.[5,6] Similarly, industrial production is rich with normal data, but detecting random anomalies is crucial for quality assurance.[7,8] Developments of anomaly detection in the medical domain include studies on chest X-ray,[9,10] brain MRI,[11–13] and retinal OCT[14–16] to screen for potential signs of disease.

Traditional, supervised methods use labeled normal and abnormal data to form decision boundaries between normal and abnormal class distributions. However, supervised methods have several drawbacks that limit their applicability to anomaly detection. One issue is that the feature space for abnormal data is infinitely large, and thus it is not possible to prepare labeled data for all possible abnormal scenarios. Another issue is the frequent imbalance between normal and abnormal data, as the sparsely spread abnormal classes may not occur in the quantities required for supervised approaches.

Toxicologic histopathology (TOXPATH) exhibits similar issues that render supervised anomaly detection approaches particularly challenging. The underlying distribution of abnormal samples is undefinable, as pathologists may encounter rare or previously unobserved anomalies. Moreover, the balance of data is extremely skewed in favor of normal data. Even a whole slide image (WSI) with an "abnormal"/lesion label typically comprises a few abnormal regions and abundant normal regions, which contributes to the data imbalance. Despite the paucity of abnormal samples, practical application demands selective attention towards anomalies.

To address the issues with supervised learning methods, modern anomaly detection utilizes one-class classification algorithms. One-class classification models are trained on abundant normal data but are applied to abnormal data which makes them an attractive solution to the TOXPATH anomaly detection problem.

Histopathological data poses unique challenges in the application of one-class classification. A primary challenge is that histopathological data

* Corresponding author.
E-mail address: hu.fangyao@gene.com (F. Hu).
[1] Equally contributed.

http://dx.doi.org/10.1016/j.jpi.2022.100102
Received 22 February 2022; Received in revised form 12 April 2022; Accepted 2 May 2022
Available online 26 May 2022

is extremely complex. Abnormalities can be subtle, resembling normal tissue, making it difficult for algorithms and non-experts to differentiate the data. In addition, abnormal data can manifest in infinite forms.[17] Moreover, identifying anomalous regions within WSIs can require assessment at multiple magnifications because of the importance of context in interpreting tissue changes in histologic samples.[17] Many current anomaly detection applications focus on problems involving fixed camera perspectives, whereas TOXPATH data must consider multiple scales of information.[6,8,18,19] In general, existing one-class classification methods require better generalizability for application on real-worldTOXPATHdata.

Many generative adversarial network (GAN)[16,20–22] and autoencoder[23,24] anomaly detection applications have been introduced recently. Although these methods have shown promise at distinguishing abnormal from normal data, the algorithms have often been evaluated on datasets that cannot address true anomaly detection semantics. For example, many anomaly detection algorithms are developed for artificial problems such as singling a CIFAR10 class as abnormal, or detecting anomalies from low variance distribution.[20,21] These methods are often reconstruction-based. In other words, these methods fundamentally rely on the model failing to reconstruct input data after mapping it to some sparse latent feature space.

GANs learn a loss function on reconstructed images by generating plausible fake images and comparing them to real, in-distribution images.[18] AnoGAN was the first to introduce the use of Bidirectional GANs to leverage latent space encodings for anomaly scoring.[22] Drawbacks of these approaches include computational complexity, the unaddressed objective of inverse mapping learning, and uninterpretable scores. EGBAD addresses the inefficiencies of AnoGAN, but still faces performance issues regarding anomaly scoring.[25] GANomaly and Skip-GANomaly reduce the architecture to standard GAN components to improve the speed of the learning process, and also develop a normalized score that denotes interpretable local anomalies.[20,21]

Autoencoders attempt to learn the identity function via an encoding function from the input image to a compressed latent space and a decoding function which maps from latent space back to an image.[26] Autoencoders have proven useful for anomaly detection. Since the latent space represents the input data, poor reconstructions indicate inputs that deviate from the distribution of normal samples. Skip-GANomaly builds on traditional autoencoders by adding UNet style skip-connections.[21] By transmitting data across these skip connections, the model is able to create higher quality reconstructed images.

We hypothesized that UNet style methods such as Skip-GANomaly lack perceptual context due to their reliance on the use of traditional per-pixel loss and convolutional neural networks. Thus, we adopted existing GAN architectures and introduced novel perceptual components. We also modified architectural components such as skip-connections to better facilitate anomaly detection. When evaluating model performance, we also utilized complex data to gauge realistic performance.

In this manuscript, we describe several techniques to improve UNet style anomaly detection algorithms. First, we introduced operational skip-connections to enhance model interpretability of normal data. Second, we added a Markovian discriminator (PatchGAN) to improve perceptual detectionof localized anomalies.[18] Next, we incorporated perceptual loss to account for local inter-dependencies in complex images. We adopted multi-scale information from the WSI to enhance regional interpretation of anomalies for TOXPATH data. Our proposed method demonstrated promising utility to detect anomalies in TOXPATHdata.

Methods

Our model concurrently trains a generator and discriminator as per the traditional GAN framework.[27] In addition to the adversarial training objective, two loss functions are designed to motivate anomaly detection: (1) perceptual reconstruction loss and (2) latent loss. An anomaly score is derived from these loss components using the outputs of both the generator and discriminator. As the model trains on normal data, the objectives drive the anomaly score for normal samples to a minimum. The model learns to isolate the feature space of normal data such that the generator understands the reconstruction parameters and the discriminator understands differentiating parameters. Consequently, exposing the model to unseen, abnormal data should result in the simultaneous failure to reconstruct and discriminate, yielding a higher loss and therefore higher anomaly score. Fig. 1 summarizes the anomaly detection process. The ability of the model to separate the normal and abnormal anomaly score distributions measures its success.

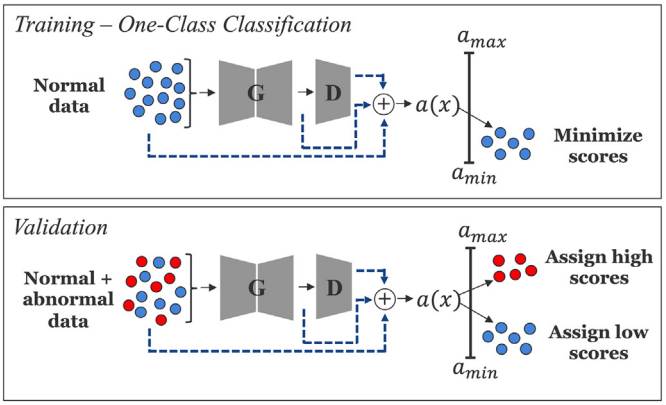

Fig. 1. Illustration of one-class classification. Training only uses normal data and minimizes the loss and the anomaly score metric. Trained models evaluate both normal and unseen abnormal data.

Generator

The generator G is based on the encoder–decoder architecture of UNet (See Fig. 2).[28] The layers of the generator network consist of symmetrical downsampling and upsampling convolutions. This architecture enables the model to learn the encoding of an input image into the latent feature space, as well as the decoding of latent vectors into the spatial dimension. The adversarial objective trains the model to output $G(x) \in \mathrm{R}^{w \times h \times c}$ given the input $x \in \mathrm{R}^{w \times h \times c}$, such that the discriminator fails to distinguish x and $G(x)$. The anomaly detection objective reduces the spatial difference between x and $G(x)$. The goal is to enhance the generator's ability to reconstruct normaldata.

Significant architectural components of the generator are the skip-connections between the downsampling and upsampling layers. Traditionally, UNet employs skip-connections to assist with the reconstruction of high-frequency features.[28] Leveraging skip-connections has proven to be a powerful technique that improves generator performance.[21] Since skip-connections supply high-level information during image reconstruction, as well as providing alternative pathways during back-propagation, they support the generator's success achieving the adversarial objective. However, the generator's ability to accurately reconstruct images is detrimental to the anomaly detection objectives if it is able to generalize to abnormal inputs. To address this issue, we introduce skip-connection processing, denoted as F_s, to enable the skip-connections to functionally support the anomaly detection objective.

A useful operation is applying dropout[29] to regularize the skip-connections and prevent the generator from overly depending on the connections during reconstruction. Other possibilities include using autoencoders within the skip-connections to transform the connection from a direct concatenation of feature maps to a more contextual source of information. Preliminary exploration of this idea with details and results are provided in the supplementary materials (Supplementary Figures1, 2 and Supplementary Table1). In general, employing functional skip-connections can greatly enable the generator towards the anomaly detection objective.

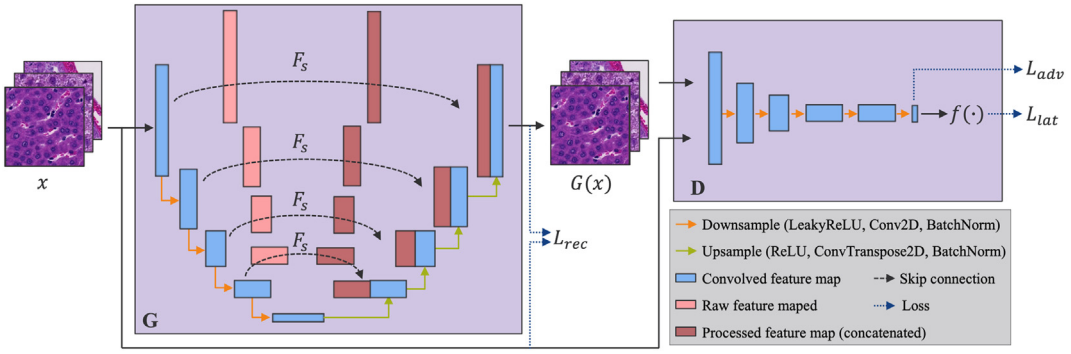

Fig. 2. Model architecture. The model architecture comprises a UNet generator G with operable skip-connections F_s, and a discriminator D with its corresponding feature extractor f. This general architecture is used for all models, with notable variations in F_s, D, L_{rec}, and dimension of input x. See text for further details.

Discriminator

The discriminator D is a convolutional neural network with the adversarial objective of distinguishing between real or generated images (see Fig. 2). In addition to the adversarial objective, the discriminator minimizes the difference between its real and fake features, $f(D(x))$ and $f(D(G(x)))$, which are typically vectors$z \in \mathrm{R}^{n\times 1}$. The purpose is for the generator and discriminator to learn the underlying distribution of normal data. We discovered that the large receptive field observed when using such an architecture, results in instability. We propose an alternative discriminator architecture based on PatchGAN that reduces the size of the receptive fields to small, overlapping patches.[30] As a result, each localized patch receives a decision from the discriminator as opposed to a uniform decision for the input image. Instead of a latent vector, $z^{\in} \mathrm{R}^{1\times m\times m}$ where each element z_{ij} refers to a 70 px by 70 px field on the original image. Anomaly detection performance improves because of the increase in perceptual precision, as the discriminator measures the per-patch normality of images.

Training

Model training involves a combination of adversarial, reconstruction, and latent losses. The adversarial loss is a GAN loss that optimizes the model for precise reconstruction of normal histology images.[27]

$$L_{adv} = \underset{x\sim p_x}{E}[\log\ D(x)] + \underset{x\sim p_x}{E}[\log\ (1 - \mathrm{D}(\mathrm{G}(\mathrm{x}))] \tag{1}$$

The reconstruction loss refines the generator performance in the spatial dimensions. Using a perceptual loss based on the structural similarity metric improves performance since it accounts for the contextual information of local image regions rather than per-pixel differences.[7,31]

$$L_{rec} = \underset{x\sim p_x}{E}[L_{SSIM}(x, G(x))] \tag{2}$$

$$L_{SSIM}(x,y) = \left\| \frac{-(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2\mu_y^2 + c_1\right)\left(\sigma_x^2\sigma_y^2 + c_2\right)} \right\| \tag{3}$$

The latent loss refines anomaly detection in the high-level latent dimensions. It optimizes the generator and discriminator to simultaneously learn the underlying distribution of normal data.

$$L_{lat} = \underset{x\sim p_x}{E}|f(D(x)) - f(D(G(x)))|_2 \tag{4}$$

The total loss is the weighted sum of the three losses, with $\lambda_{adv} = 1$, $\lambda_{rec} = 50$, and $\lambda_{lat} = 1$.

$$L_{total} = \lambda_{adv}L_{adv} + \lambda_{rec}L_{rec} + \lambda_{lat}L_{lat} \tag{5}$$

We implemented the model in PyTorch and trained with a batch size of 128 on a NVIDIA V100 GPU. Both the generator and discriminator are trained on Adam optimizers with a learning rate of 10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. All models are trained for a minimum of 20 epochs and a maximum of 200 epochs with early stopping.

Data pre-processing

A single WSI can measure over 100,000 pixels in one dimension and contain gigabytes of information. Current memory limitations demand the pre-processing of WSIs into tiles of a more manageable dimension (see Fig. 3). Consequently, the anomaly detection problem occurs on a per-tile basis. A shortcoming is the ratio of anomalous to normal tissue within the tile is inconsistent, since anomalies can vary greatly in size, combined with the loss of contextual information between tiles.[17] Abnormal regions can span over several tiles, and pathologists often inspect these regions at multiple magnifications to comprehensively understand the findings. Naively tiling a WSI neglects the spatial dependence between tiles. Therefore, we use multi-scaling techniques to incorporate high- and low-level information, resulting in a more comprehensive model that detects anomalies using global and local contexts. Implementation requires the concatenation of images of different scales that are symmetric about the same central pixel along the channel axis, resulting in a hyper-dimensional image containing multiple levels of information.

Anomaly scoring

The anomaly score is derived as the weighted sum of the reconstruction and latent losses. The parameter w_{rec} controls the relative importance of the losses. We use $w_{rec} = 0.9$ to emphasize the reconstruction loss.

$$a(x) = w_{rec}L_{rec} + (1 - w_{rec})L_{lat} \tag{6}$$

Since the model learns the distribution of normal data, it is expected that the anomaly scores for normal data are low at inference time. Conversely, a higher loss is indicative of poor image reconstruction, which implies that the input data is likely abnormal data. Given the separation of normal and abnormal anomaly scores into lower and higher categories respectively, the area under the receiver operator characteristic curve (AUROC) is an effective performance metric.[32,33] The AUROC metric provides a single scalar representing the ability of the model to properly segregate the abnormal and normal samples. Practical usage of the anomaly scores also requires normalization to the range [0,1] which is performed as:

$$\hat{a}(x) = \frac{a(x) - \min\ (a)}{\max\ (a) - \min\ (a)} \tag{7}$$

Fig.3 summarizes the pipeline of anomaly detection for TOXPATH. In practice, all tiles of individual WSIs feed into the model to obtain batches

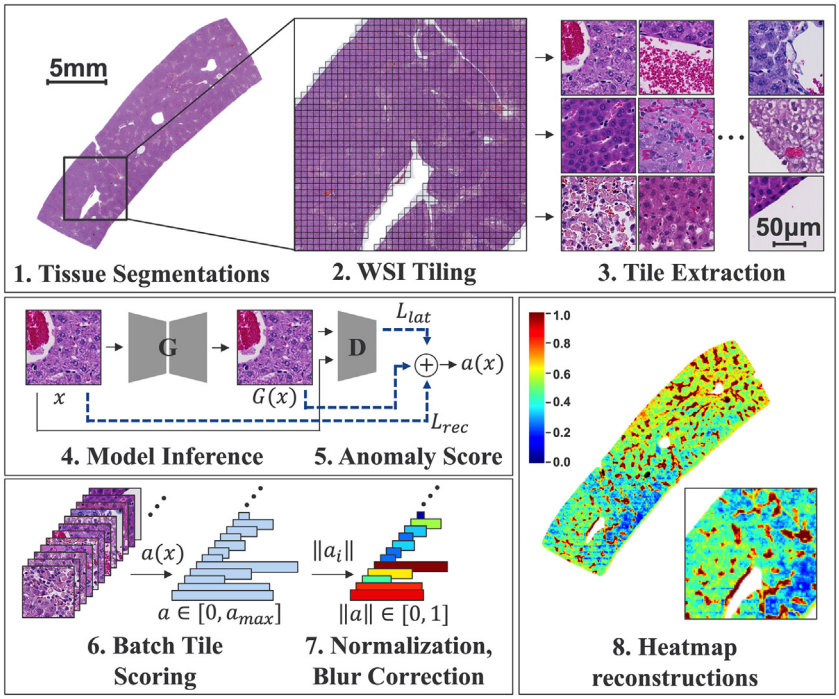

Fig. 3. Overview of anomaly detection for TOXPATH. The anomaly score is a function of the loss between the pre-processed WSI tiles (x) and generator output $G(x)$, as well as the loss between discriminator outputs $D(x)$ and $D(G(x))$. Batches of anomaly scores are normalized and processed into reconstructed heatmaps to visualize anomalous histology findings.

of anomaly scores. It is then possible to normalize the batch of scores of a WSI for visualization. It is typical to have a control slide in TOXPATH studies, in which case normalization more appropriately occurs with respect to the controlWSI.

Datasets

An in-houseWSI TOXPATH dataset of normal and abnormal histology from rat liver that were gathered from studies between 2012 and 2019 was utilized in this manuscript. Study pathologists provided global diagnoses and pixel-level annotations were generated for this project from 50 WSIs under the guidance of pathologists. Thirteen WSIs were allocated as normal data, contributing 33,940 256 px by 256 px tiles for training. The remaining 37 WSIs were split amongst 4 abnormal classes and used for validation. Of interest were 3 lesion types: necrosis, peritonitis, and infiltrate. Since many liver WSIs included a sample of spleen (a non-liver tissue), we defined spleen as a fourth anomaly class. Sample normal regions were annotated on WSIs with abnormalities to ensure that the validation datasets were balanced.

As seen in Table1, most abnormal samples were of necrosis and spleen, with fewer instances of infiltrate and peritonitis since the infiltrates were small, and peritonitis occurred infrequently. The scarcity of abnormal data further supports the value of one-class classification.

In contrast to existing histopathology datasets in the literature, the intention of our dataset was to represent a realistically challenging anomaly detection task. There is variance between the abnormal classes, requiring generalization for high performance. Additionally, some of the abnormal classes such as necrosis are similar to the normal data in color and shape, requiring a high degree of precision. Thus, we postulate that this dataset is more difficult for anomaly detection than other histopathology datasets for detection tasks.[34,35]

Publicly available anomaly detection datasets were also used to determine the generalizability of our method. The MVTec dataset[7] contains 15 object or texture classes with normal and abnormal samples for each respective class. The total dataset included 3629 training and 1725 validation images. This dataset was intended for industrial inspection, in which the anomaly detection task is generally focused on a single local abnormality in contrast to the diffuse and global anomalies in TOXPATH.

Table 1
Break down of in-house TOXPATH dataset.

	Whole slide image	Normal tiles	Abnormal tiles
Normal	13	33940	-
Necrosis	6	4277	4281
Peritonitis	11	470	470
Inflammation	15	53	53
Spleen	5	4745	4748
Total	50	43287	9552

Results

Ablations

We quantitatively and qualitatively evaluated our model against the baseline SkipGANomaly model.[21] To validate our contributions, we ablated single modifications and also combined the changes. The individual modifications include a higher dimensional input, dropout on the skip-connections, perceptual SSIM loss as the reconstruction loss, and a PatchGAN discriminator. We evaluated each model's ability to separate the anomaly scores for normal and abnormal tiles using four different abnormal classes. We then visualized the loss and reconstructed heatmaps to qualitatively assess performance. Table2 summarizes the AUROC results and shows that Dropout on the skip-connections results in the highest individual contribution, while aggregating the multiple changes yields the best overall performance. In Fig. 4, we present representative examples for the normal case, all abnormal classes, and an additional tissue tearing anomaly. The reconstruction loss images demonstrate that our model is more intensely failing to reconstruct abnormal data, which results in better anomaly detection performance. The inter-class variance of the abnormal classes reveals that our method is able to generalize to unseen classes. The abnormal tear shown in the last row of Fig. 4 is not considered in the quantitative evaluations; however, the model still detects the tears, which further

Table 2
Ablation studies on all abnormal histological findings.

	Necrosis	Infiltration	Peritonitis	Spleen	Average
SG, 3c	0.726 ± 0.017	0.936 ± 0.029	0.984 ± 0.013	0.952 ± 0.014	0.899 ± 0.012
SG, 6c	0.802 ± 0.019	0.924 ± 0.019	0.982 ± 0.010	0.980 ± 0.006	0.922 ± 0.007
SG + DP, 3c	0.809 ± 0.032	0.947 ± 0.021	0.985 ± 0.006	0.951 ± 0.010	0.923 ± 0.005
SG + SS, 3c	0.852 ± 0.030	0.836 ± 0.091	**0.994 ± 0.003**	0.886 ± 0.083	0.892 ± 0.008
SG + PG, 3c	0.777 ± 0.052	**0.959 ± 0.011**	0.983 ± 0.008	0.936 ± 0.011	0.914 ± 0.015
SG + All, 6c	**0.927 ± 0.009**	0.936 ± 0.023	0.968 ± 0.033	**0.981 ± 0.014**	**0.953 ± 0.010**

Performance is evaluated using AUROC. The baseline is shown in the first row. From the second to fifth rows inclusive, we add only one of the following: 6-channel multi-scale input, Dropout on skip-connections, SSIM loss, and PatchGAN discriminator. The sixth row combines all modifications.

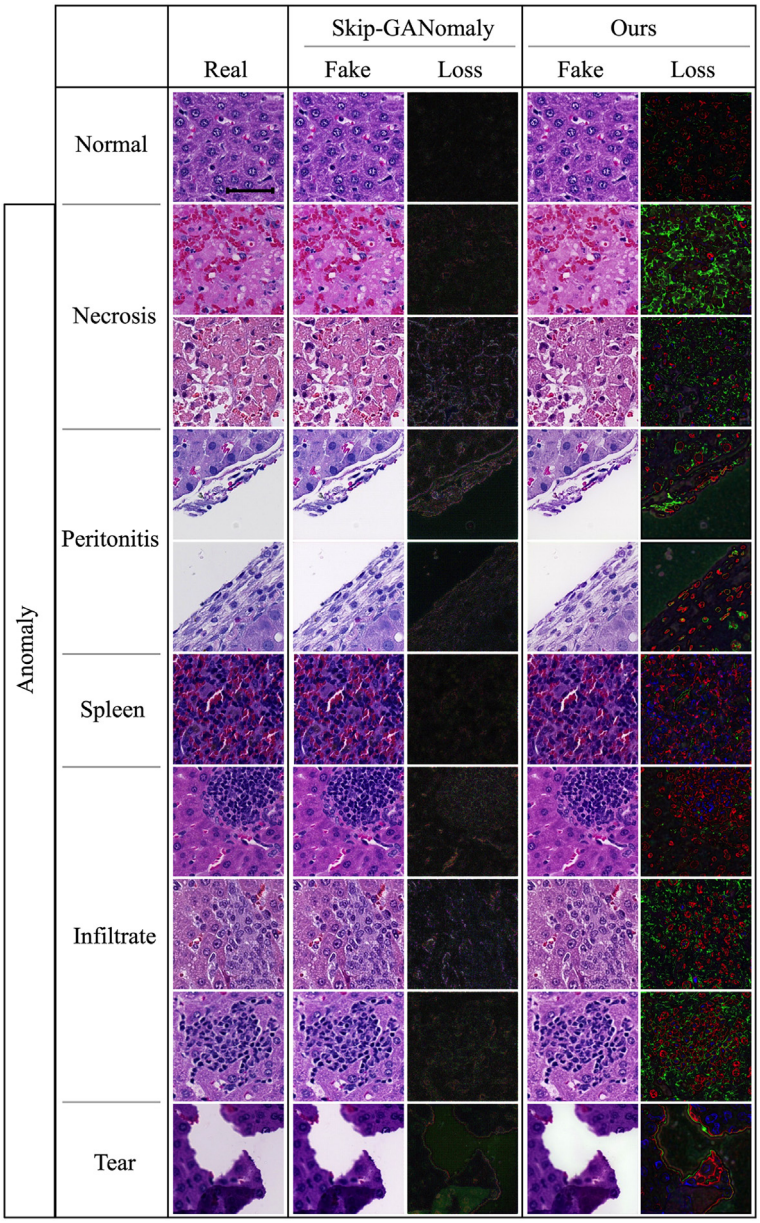

Fig. 4. Comparing the baseline against the proposed model on representative reconstructions. The high reconstruction loss when using our method on abnormal histology indicates better detection of anomalies. Further examples of MVTec reconstructions are shown in the supplementary materials. (Note the scale bar in the real, normal figure measures 50 μm. All figures are at the same scale).

demonstrates the ability of our model to generalize. The loss is also low for the normal image, which is important for quantitative performance.

From Table2, we see that dropout on the skip-connections significantly contributes towards performance on TOXPATH data. To verify the performance,we demonstrate that using the modification outperforms Skip-GANomaly on 12 of the 15 MVTec classes. Table3 shows that modifyingthe skip-connections has a significant impact on performance. Supplementary materials (Supplementary Figs 1, 2 and Supplementary

Table 3
Dropout on the skip-connections outperforms Skip-GANomaly on industrial detection datasets.

	Baseline	Dropout on skip-connections
Bottle	0.780 ± 0.025	**0.897 ± 0.011**
Cable	0.596 ± 0.016	**0.740 ± 0.067**
Capsule	0.695 ± 0.009	**0.741 ± 0.002**
Carpet	0.895 ± 0.006	**0.913 ± 0.005**
Grid	0.960 ± 0.015	**0.971 ± 0.008**
Hazelnut	0.934 ± 0.012	**0.999 ± 0.002**
Leather	0.972 ± 0.024	**0.994 ± 0.003**
Metal nut	0.558 ± 0.032	**0.690 ± 0.062**
Pill	**0.908 ± 0.010**	0.850 ± 0.011
Screw	**1.00 ± 0.000**	**1.00 ± 0.000**
Tile	**0.962 ± 0.003**	0.934 ± 0.024
Toothbrush	0.846 ± 0.017	**0.879 ± 0.059**
Transistor	0.550 ± 0.020	**0.786 ± 0.057**
Wood	**0.990 ± 0.002**	0.997 ± 0.001
Zipper	0.752 ± 0.026	**0.898 ± 0.021**

Table1) show additional preliminary results on functional skip-connections. Dropout provides regularization that mitigates the over-dependence of reconstructions on the skip-connections and optimizes the model for anomaly detection rather than naive reconstructions. The other modifications studied in the ablations did not result in significant changes in performance on the MVTec data, as the perceptual methods are more tailored for histological anomalies.

Toxicologic histopathology anomaly detection

Given the normalized anomaly scores, we can map the scores to colors and reconstruct heatmaps to allow pathologists to visualize anomaly detection performance over the entire tissue section. Post-processing using the variance of the Laplacian to deblur the images improves the quality of the reconstructions.[36]

Fig.5 shows that both Skip-GANomaly and our method perform as expected on the normal WSIs, with no significant signal except for highlighting the irrelevant spleen tissue. The necrotic samples show that our model yields higher anomaly scores within the abnormal regions. Similarly, our model outperforms the baseline when highlighting regions of peritonitis, which generally manifest at the edges of liver sections. The baseline model struggles to highlight the spleen, despite the marked and obvious visual differences between spleen and liver tissues. Both models struggle to identify multifocal infiltrates thus resulting in small differences in anomaly scores. Finally, our model is more sensitive to unwanted white space such as tears.

Conclusion

One-class classification achieves high performance on TOXPATH data and demonstrates high applicability towards real-world data. The performance enhancement we achieved is due to perceptual improvements, including SSIM loss, a PatchGAN discriminator, and multi-scale inputs. Functional skip-connections result in improvement on both TOXPATH and industrial inspection datasets. We experimented primarily on a rat liver TOXPATH dataset and show that our method greatly outperforms anomaly detection based on GAN reconstructions. Future plans are to expand our one-class classification algorithm to different organ tissues, and further study the impact of operable skip-connections on the general performance ofUNet.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

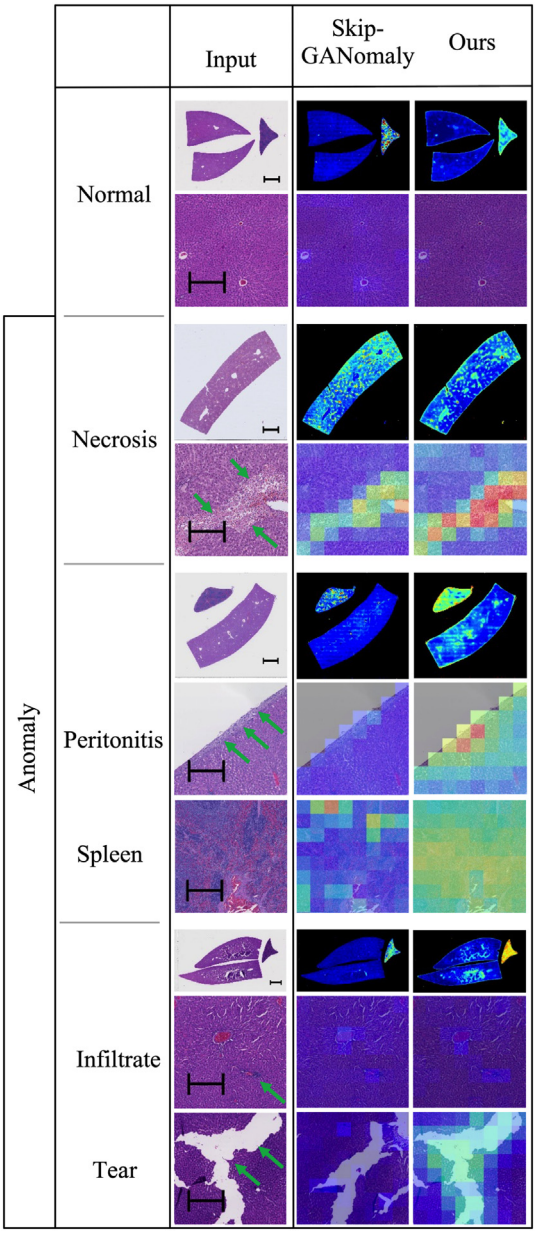

Fig. 5. Representative heatmaps and regions of interest. Our model demonstrates significantly higher contrast between normal and abnormal regions compared to SkipGANomaly. Arrows highlighted the abnormal regions. (Note the scale bar in the WSI measures 5 mm. In the tiles, scale bar measures 2 mm).

Fangyao Hu reports a relationship with Genentech Inc that includes: employment. Philip Zehnder reports a relationship with Genentech Inc that includes: employment. Jeffrey Feng reports a relationship with Genentech Inc that includes: employment. Ruth Sullivan reports a relationship with Genentech Inc that includes: employment. Reina Fuji reports a relationship with Genentech Inc that includes: employment. Fangyao Hu has patent #63/161864 pending to Genentech.

Acknowledgements

This work is funded by Genentech.

Appendix A. Supplementarydata

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2022.100102.

References

1. Chondola V, Banerjee A, Kumar V. Anomaly detection: a survey. Article no. 15. Comput Surv 2009;41:1-58.
2. Di Mattia F, Galeone P, De Simoni M, Ghelfi E. A Survey on GANs for Anomaly Detection. arXiv:1906.11632 2019.
3. Hodge VJ, Austin J. A survey of outlier detection methodologies. Artif Intel Rev 2004;22: 85-126.
4. Pimentel MAF, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Signal Process 2014;99:215–249.
5. Dumagpi J, Jung W-Y, Jeong Y-J. A new GAN-Based Anomaly Detection (GBAD) approach for multithreat object classification on large-scale X-Ray security images. IEICE Trans Inform Syst 2020;E103.D:454–458.
6. Miao C, Xie L, Wan F, et al. A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 2019. p. 2114–2123.
7. Bergmann P, Fauser M, Sattlegger D, Steger C. A comprehensive real-world dataset for unsupervised anomaly detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA; 2019. p. 9584–9592.
8. Tang T-W, Kuo W-H, Lan J-H, Ding CF, Hsu H, Young H-T. Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications. Sensors 2020;20:3336.
9. Tang Y-X, Tang YB, Han M, Xiao J, Summers RM. Deep adversarial one-class learning for normal and abnormal chest radiograph classification Medical Imaging 2019: Computer-Aided Diagnosis, 10950; 2019. p. 1095018.
10. Tuluptceva N, Bakker B, Fedulova I, Schulz H, Dylov DV. Anomaly Detection with Deep Perceptual Autoencoders. arXiv:2006.13265 October 2020.
11. Baur C, Denner S, Wiestler B, Navab N, Albarqouni S. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. Med Image Anal 2021;69, 101952.
12. Baur C, Wiestler B, Albarqouni S, Navab N. Deep Aut oencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images arXiv:1804.04488.
13. Chen X, Konukoglu E. Unsupervised Det ection of Lesions in Brain MRI Using Constrained Adversarial Auto-Encoders. arXiv:1806.04972 2018.
14. Ouardini K, Yang H, Unnikrishnan B, et al. Towards practical unsupervised anomaly detection on retinal images. In: *Wang Q, Milletari F, Nguyen HV, Albarqouni S, Cardoso MJ, Rieke N, et al, eds.* Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, Lecture Notes in Computer Science. Springer International Publishing; 2019. p. 225–234.
15. Seebock P, Waldstein SM, Klimscha S, et al. Unsupervised identification of disease marker candidates in retinal OCT imaging data. IEEE Trans Med Imag 2019;38:1037–1047.
16. Zhou K, Gao S, Cheng J, et al. Sparse-Gan: sparsity-constrained generative adversarial network for anomaly detection in retinal OCT image. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020. p. 1227–1231.
17. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Computat Struct Biotechnol J 2018;16:34–42.
18. Siddiquee MMR, Zhou Z, Tajbakhsh N, et al. Learning fixed points in generative adversarial networks: from image-to-image translation to disease detection and localization. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea; 2019. p. 191–200.
19. Spahr A, Bozorgtabar B, Thiran J-P. Self-Taught Semi-Sup ervised Anomaly Detection on Upper Limb X-rays. arXiv:2102.09895 February 2021.
20. Akcay S, Atapour-Abarghouei A, Breckon TP. Semi-supervised anomaly detection via adversarial training. In: *Jawahar CV, Li H, Mori G, Schindler K, eds.* Computer Vision – ACCV 2018. Lecture Notes in Computer Science. Spinger International Publishing; 2019. p. 622–637.
21. Akcay S, Atapour-Abarghouei A, Breckon TP. Skip connected and adversarially trained encoder-decoder anomaly detection. 2019 International Joint Conference on Neural Networks (IJCNN); 2019.
22. Schlegl T, Seebock P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: *Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P-T, et al, eds.* Information Processing in Medical Imaging, Lecture Notes in Computer Science. Springer International Publishing; 2017. p. 146–157.
23. Roy M, Kong J, Kashyap S, et al. Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images. Sci Rep 2021;11:139.
24. Tuluptceva N, Bakker B, Fedulova I, A. K.. Perceptual image anomaly detection. In: *Palaiahnakote S, Sanniti di Baja G, Wang L, Yan WQ, eds.* Pattern Recognition. Lecture Notes in Computer Science. Springer International Publishing; 2020. p. 164–178.
25. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR. Efficient GAN-Based Anomaly Detection. arXiv:1802.06222 May 2019.
26. Bank D, Koenigstein N, Giryes R. Autoencoders. arXiv:2003.05991 March 2020.
27. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Adv Neural Inform Process Syst 2014:2672–2680.
28. Ronneberger O, Fischer P, Brox T. Convolutional networks for biomedical image segmentation. In: *Navab N, Hornegger J, Wells WM, Frangi AF, eds.* Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. Springer International Publishing; 2015. p. 234–241.
29. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. arXiv:1207.0580 July 2012.
30. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 2017.Honolulu, HI.
31. Bergmann P, Lowe S, Fauser M, Sattlegger D, Steger C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, arXiv:1807.02011; p. 373-80.
32. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recog 1997;30:1145–1159.
33. Ling, Charles X. et al. "AUC: a Statistically Consistent and more Discriminating Measure than Accuracy." IJCAI (2003). https://www.ijcai.org/Proceedings/03/Papers/077.pdf
34. Bandi P, Geessink O, Manson Q, et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. IEEE Trans Med Imaging 2019;38:550–560.
35. Litjens G, Bandi P, Ehteshami Bejnordi B, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. Gigascience 2018:7.
36. Bansal R, Raj G, Choudhury T. Blur image detection using Laplacian operator and OpenCV. 2016 International Conference System Modeling Advancement in Research Trends (SMART); 2016. p. 63–67.