

■ Nenadzorovana detekcija rakavih regij v histoloških slikah s pomočjo preslikav slika-v-sliko

Dejan Štepec^{1,2*}, Danijel Skočaj²

¹ XLAB d.o.o., Pot za Brdom 100, 1000, Ljubljana, Slovenia

² University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia
dejan.stepec@xlab.si

Izveček

V delu predstavimo nenadzorovan pristop za detekcijo vizualnih anomalij v medicinskih slikah, kjer je cilj odkriti vizualne značilnosti, ki pomembno odstopajo od pričakovane normalne predstavitve. Glede na naravo pojava anomalij in kompleksnih procesov, ki jih tvorijo, je težko pridobiti ekspertno označene podatke. Pridobitev tako označenih podatkov je še posebej zahtevna na biomedicinskem področju, kjer nam take podatke lahko zagotovijo zgolj domenski eksperti. Poleg tega so podatki na biomedicinski domeni veliko bolj kompleksni tako z vidika njihove predstavitve kot dimenzionalnosti. V tem delu ta problem naslovimo na nenadzorovan način z uporabo metode za preslikavo slika-v-sliko, ki pomembno izboljša trenutne nenadzorovane pristope in deluje z zmogljivostjo, ki je podobna nadzorovanemu pristopu.

Ključne besede: detekcija anomalij, nenadzorovano učenje, globoko učenje, generativne nasprotniške mreže, preslikave slika-v-sliko, digitalna patologija

Abstract

Detection of visual anomalies refers to the problem of finding patterns in different imaging data that do not conform to the expected visual appearance, and is a widely studied problem in different domains. Due to the nature of anomaly occurrences and underlying generating processes, it is hard to characterize them and obtain labelled data. Obtaining labelled data is especially difficult in biomedical applications, where only trained domain experts can provide labels, which are often diverse and complex to a large degree. The recently presented approaches for unsupervised detection of visual anomalies omit the need for labelled data and demonstrate promising results in domains where anomalous samples significantly deviate from the normal appearance. Despite promising results, the performance of such approaches still lags behind supervised approaches and does not provide a universal solution. In this work, we present an image-to-image translation-based framework that significantly surpasses the performance of existing unsupervised methods and approaches the performance of supervised methods in a challenging domain of cancerous region detection in histology imagery.

Keywords: Anomaly detection, unsupervised learning, deep-learning, generative-adversarial-networks, image-to-image translation, digital pathology

1 INTRODUCTION

Anomaly detection represents an important process of determining instances that stand out from the rest of the data. Detecting such occurrences in different data modalities has wide applications in different domains such as fraud detection, cyber-intrusion, industrial in-

spection, and medical imaging [Chandola et al., 2009]. Detecting anomalies in high-dimensional data (e.g. images) is a particularly challenging problem that has recently seen a significant rise of interest, due to the prevalence of deep-learning-based methods.

* Partially supported by the EU Horizon 2020 research project iPC (826121).

The success of current deep-learning-based methods has mostly relied on the abundance of available data. Anomalies generally occur rarely, in different shapes and forms, and are thus extremely hard or even impossible to label.

Supervised deep-learning-based anomaly detection approaches have seen great success in different industrial and medical application domains [Ehteshami Bejnordi et al., 2017, Tabernik et al., 2019]. The success of such methods is the most evident in the domains with well-known characterization (and possibly a finite set) of the anomalies and abundance of labeled data. Specific to the detection of visual anomalies, we usually also want to localize the actual anomalous region in the image. Obtaining such detailed labels to learn supervised models is a costly process and in many cases also impossible. There is an abundance of data available in the biomedical domain, but it is usually of much higher complexity and diversity. Domain complexity prevents large-scale crowd annotation efforts and only trained biomedical experts can usually annotate such data.

Weakly supervised approaches address such problems by requiring only image-level labels (e.g. disease present or not) and are able to detect and delineate anomalous regions solely from such weakly labeled data, without the need for detailed pixel or patch-level labels [Campanella et al., 2019]. On the contrary, few-shot approaches reduce the number of required labeled samples to the least possible amount [Tian et al., 2020].

In an unsupervised setting, only normal appearance samples are available (e.g. healthy, defect-free), which are usually available in larger quantities and are easier to obtain. Deep generative methods, in a form of autoencoders (AE) or generative adversarial networks (GAN), have been recently applied to the problem of unsupervised detection of visual anomalies and have shown promising results in different industrial and medical application domains [Schlegl et al., 2019, Baur et al., 2020b, Baur et al., 2020a, Bergmann et al., 2020]. Current approaches require normal appearance samples for training, in order to detect and segment deviations from that normal appearance, without the need for labeled data. They usually model normal appearance with low-resolution AE or GAN models and the overall performance still lags significantly behind supervised approaches.

In this work, we present a novel high-resolution image-to-image translation-based method for unsupervised detection of visual anomalies that significantly surpasses the performance of existing unsupervised approaches and closes the gap towards the supervised counterparts. We particularly focus on a challenging problem of cancerous region detection from gigapixel histology imagery, which has been already addressed in a supervised [Ehteshami Bejnordi et al., 2017], as well as in a weakly supervised setting [Campanella et al., 2019]. Extremely large histology imagery (patch-based processing) and the highly variable appearance of the different tissue regions represent a unique challenge for existing unsupervised approaches.

2 IMAGE-TO-IMAGE TRANSLATION AS A PRETEXT FOR ANOMALY DETECTION

Inspired by the multimodal image-to-image translation methods [Huang et al., 2018, Lee et al., 2020], we propose an example guided image translation method (Figure 1) [Štepec and Skočaj, 2021], which in comparison with SteGANomaly [Baur et al., 2020a] enables anomaly detection without cycle-reconstruction during the inference, specially crafted intermediate domain distribution, and Gaussian filtering. Similar to MUNIT [Huang et al., 2018], we assume that the latent space of images can be decomposed into content and style spaces. We also assume that images in both domains share a common content space C , as well as style space S (i.e. they both come from the same **healthy** domain). This differs from MUNIT [Huang et al., 2018], where style space is not shared, due to semantically different domains X and Y . Similar to MUNIT [Huang et al., 2018], our translation model consists out of encoder E_{ij} and decoder G_j networks for each space $i \in \{C, S\}$ and domains $j \in \{X, Y\}$. Those subnetworks are used for autoencoding, as well as cross-domain translation, by interchanging encoders and decoders from different domains. Style latent codes s_x and s_y are randomly drawn and additionally transformed by a multilayer perceptron (MLP) network f for a cross-domain translation. Randomness addresses the memorization effect, large present in autoencoder-based anomaly detection approaches.

During anomaly detection (Figure 1), an input image x is encoded with E_{cx} to produce content vector c_x , which is then joined with the style code s_y

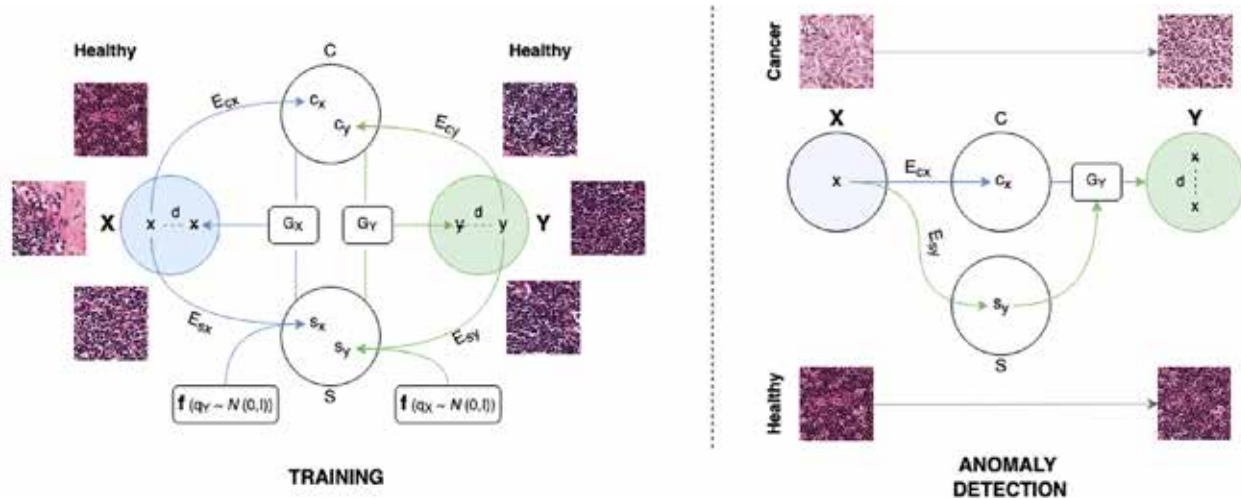


Figure 1: Our proposed unsupervised anomaly detection method, based on the image-to-image translation. We disentangle a latent space into shared content and style spaces, implemented via domain-specific (blue and green colors) encoders E and decoders G. Anomaly detection is performed with an example-guided image translation. Best viewed in digital version with zoom.

extracted from the original image x , with the style encoder E_{s_y} of the target domain Y . This presents an input to decoder G_y which generates y . This is basically an example guided image translation, used also in MUNIT [Huang et al., 2018] and DRIT++ [Lee et al., 2020] methods. Content-style space decomposition is especially well suited for histopathological analysis due to different staining procedures, which causes the samples to significantly deviate in their visual appearance. Style-guided translation ensures that the closest looking normal appearance is found, taking into account also the staining appearance. We then measure an anomaly score using distance metric d (e.g. perceptual LPIPS distance [Zhang et al., 2018] or Structure Similarity Index (SSIM) [Wang et al., 2004]), between the original image x and its reconstruction*.

3 EXPERIMENTS AND RESULTS

3.1 Histology Imagery Dataset.

We address the aforementioned problems of anomaly detection pipeline on a challenging domain of digital pathology, where whole-slide histology images (WSI) are used for diagnostic assessment of the spread of cancer. This particular problem was already addressed in a supervised setting [Ehteshami Bejnordi et al., 2017], as a competition², with provided

clinical histology imagery and ground truth data. A training dataset with ($n=110$) and without ($n=160$) labeled cancerous regions (used as anomalies) is provided, as well as a test set of 129 images (49 with and 80 without labeled cancerous regions). Raw histology imagery, presented in Figure 2a, is first preprocessed, in order to extract the tissue region (Figure 2b). We used the approach from IBM³, which utilizes a combination of morphological and color space filtering operations. Patches of 512×512 are then extracted from the filtered image and filtered according to the tissue (Figure 2c) and cancer (Figure 2d) coverage. We only use patches with tissue and cancerous region coverage over 90 % (i.e. green patches). With this procedure, we produce a dataset of healthy (i.e. no overlap with cancerous label) and cancerous patches (i.e. > 90% overlap with cancerous label).

We train the models on random 80,000 healthy tissue patches extracted from a training set of healthy and cancerous (coverage=0% - cancerous samples also contain healthy tissue) WSIs ($n=270$). The baseline supervised approach is trained on randomly extracted healthy ($n=25,000$) and cancerous patches ($n=25,000$). The methods (i.e. supervised baseline and proposed ones) are evaluated on healthy ($n=7673$) and cancerous ($n=16,538$) patches extracted from a cancerous test set of WSIs ($n=49$). We mix healthy training patches of both cohorts (i.e. healthy patches

² <https://camelyon16.grand-challenge.org/>
³ <https://github.com/CODAIT/deep-histopath>

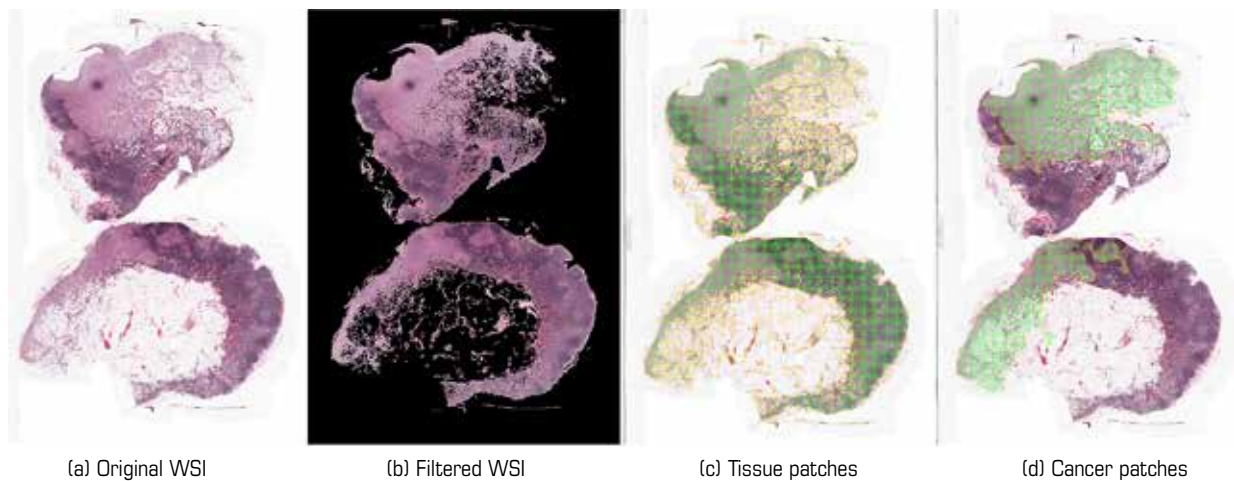


Figure 2: Preprocessing of the original WSI presented in a) consists of b) filtering tissue sections and c) extracting tissue patches, based on the tissue (green $\geq 90\%$, orange $\leq 10\%$ and yellow in-between) and d) cancerous region coverage (green $\geq 90\%$, orange $\leq 30\%$ and yellow in-between). Best viewed in a digital version with zoom.

from cancerous WSIs) in order to demonstrate the robustness of the proposed approach against a small percentage of possibly contaminated healthy appearance data (e.g. non-labeled isolated tumor cells in cancerous samples).

3.2 Unsupervised Anomaly Detection.

We compare the proposed method against GAN-based f-AnoGAN [Schlegl et al., 2019] and StyleGAN2 [Karras et al., 2020] methods. Both methods separately model normal appearance and perform latent space mapping for anomaly detection. The f-AnoGAN method models normal appearance using Wasserstein GANs (WGAN) [Arjovsky et al., 2017], which is limited to a resolution of 64^2 and uses an encoder-based fast latent space mapping approach. The StyleGAN2 method enables high-resolution image synthesis (up to 1024^2) and also implements an iterative optimization procedure, based on Learned Perceptual Image Patch Similarity (LPIPS) [Zhang et al., 2018] distance metric. We evaluate the performance

of the proposed and StyleGAN2 methods on patches of 512^2 , while center-cropped 64^2 patches are used for the f-AnoGAN method. Additionally, we compare the performance against the supervised DenseNet-121 [Huang et al., 2017] baseline model, trained and evaluated on 512^2 patches. We evaluate the proposed method using Structural Similarity Index Measure (SSIM) [Wang et al., 2004] and LPIPS reconstruction error metrics as an anomaly score. We use the same metrics (i.e. SSIM and LPIPS) as also as an alternative to the original f-AnoGAN anomaly score implementation, as well as to measure StyleGAN2 reconstruction errors.

We first evaluate the methods by inspecting the distribution of anomaly scores across healthy and cancerous patches, as presented in Figure 3. We compare our proposed approach (Figures 3a and 3b) against f-AnoGAN (Figure 3c) and StyleGAN2 (Figure 3d) methods and report significantly better distribution disentanglement between healthy and cancerous patches.

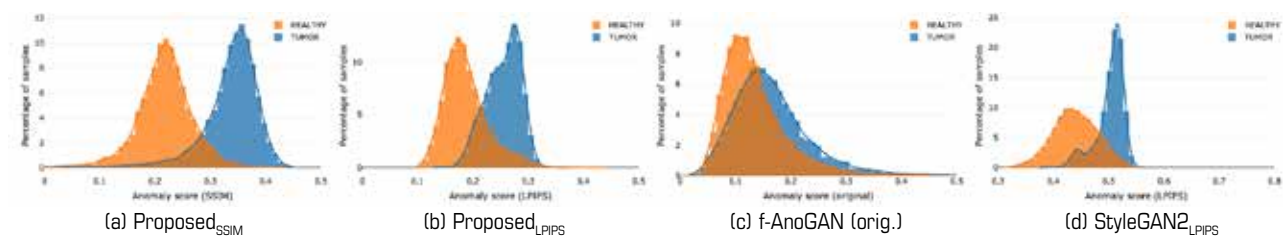


Figure 3: Distribution of anomaly scores on healthy and cancerous histology imagery patches (a) for the proposed method (SSIM metric), (b) proposed method (LPIPS metric), (c) f-AnoGAN (original metric) and (d) StyleGAN2 (LPIPS metric). Results for the proposed and StyleGAN2 methods are reported for 5122 patches, while 642 patches are used for f-AnoGAN.

The area under the ROC curve (AUC) and Average Precision (AP) scores are reported in Table 1 for all the methods and different anomaly scores. We also report F1 and classification accuracy measures, calculated at the Youden index of the ROC curve. We notice that the performance of the proposed method approaches the performance of the supervised baseline in terms of both reconstruction error metrics (i.e. LPIPS and SSIM). The performance of the f-AnoGAN significantly improves using SSIM and LPIPS metrics, in comparison with the originally proposed anomaly score. This shows the importance of the selection of the appropriate reconstruction error metric. The StyleGAN2 method shows good distribution disentanglement using the LPIPS distance metric, while the SSIM metric fails to capture any significant differences between the two different classes (i.e. healthy and anomalous). The proposed method demonstrates consistent performance across both anomaly score metrics, as well as different evaluation measures.

Table 1: **Performance statistics (F1, Classification Accuracy - CA) calculated at Youden index of Receiver Operating Characteristic (ROC) curve and the corresponding area under the ROC curve (AUC) and Average Precision (AP) scores summarizing ROC and Precision-Recall (PR) curves.**

	AUC	AP	F1	CA
Supervised	0.954	0.974	0.925	0.901
Proposed (SSIM)	0.947	0.976	0.920	0.895
Proposed (LPIPS)	0.900	0.914	0.886	0.847
StyleGAN2 (LPIPS)	0.908	0.940	0.872	0.836
StyleGAN2 (SSIM)	0.580	0.711	0.674	0.588
f-AnoGAN (original)	0.650	0.443	0.502	0.637
f-AnoGAN (SSIM)	0.887	0.916	0.886	0.846
f-AnoGAN (LPIPS)	0.865	0.902	0.875	0.830

4 CONCLUSION

In this work, we presented an image-to-image translation-based unsupervised approach that significantly surpasses the performance of existing GAN-based unsupervised approaches for the detection of visual anomalies in histology imagery and also approaches the performance of supervised methods. The method is capable of closely reconstructing presented healthy histology tissue samples, while unable to reconstruct cancerous ones and is thus able to detect such samples with an appropriate visual distance measure. The image-to-image translation-

-based framework offers a promising multi-task platform for a wide range of problems in the medical domain and can be now further extended with the capabilities for anomaly detection and applied to the completely new set of domains where labeled data is hard to obtain. Additional research is needed to investigate effectiveness in other biomedical modalities, as well as to exploit the benefits of using such a framework in a multi-task learning setting.

REFERENCES

- [1] [Arjovsky et al., 2017] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *ICML*, pages 214–223. PMLR.
- [2] [Baur et al., 2020a] Baur, C., Graf, R., Wiestler, B., Albarqouni, S., and Navab, N. (2020a). Steganomaly: Inhibiting cycle-gan steganography for unsupervised anomaly detection in brain mri. In *MICCAI*, pages 718–727. Springer.
- [3] [Baur et al., 2020b] Baur, C., Wiestler, B., Albarqouni, S., and Navab, N. (2020b). Scale-space autoencoders for unsupervised anomaly segmentation in brain mri. In *MICCAI*, pages 552–561. Springer.
- [4] [Bergmann et al., 2020] Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *CVPR*, pages 4183–4192.
- [5] [Campanella et al., 2019] Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Silva, V. W. K., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., and Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309.
- [6] [Chandola et al., 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Comput. Surv.*, 41(3):15:1–15:58.
- [7] [Ehteshami Bejnordi et al., 2017] Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J. A. W. M., and the CAMELYON16 Consortium (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, 318(22):2199–2210.
- [8] [Huang et al., 2017] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *CVPR*, pages 4700–4708.
- [9] [Huang et al., 2018] Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). Multimodal unsupervised image-to-image translation. In *ECCV*, pages 172–189.
- [10] [Karras et al., 2020] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *CVPR*, pages 8110–8119.
- [11] [Lee et al., 2020] Lee, H.-Y., Tseng, H.-Y., Mao, Q., Huang, J.-B., Lu, Y.-D., Singh, M., and Yang, M.-H. (2020). Dri++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, pages 1–16.
- [12] [Schlegl et al., 2019] Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., and Schmidt-Erfurth, U. (2019). f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, 54:30–44.

- [13] [Štepec and Skočaj, 2021] Štepec, D. and Skočaj, D. (2021). Unsupervised detection of cancerous regions in histology imagery using image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3785–3792.
- [14] [Tabernik et al., 2019] Tabernik, D., Šela, S., Skvarč, J., and Skočaj, D. (2019). Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *Journal of Intelligent Manufacturing*.
- [15] [Tian et al., 2020] Tian, Y., Maicas, G., Pu, L. Z. C. T., Singh, R., Verjans, J. W., and Carneiro, G. (2020). Few-shot anomaly detection for polyp frames from colonoscopy. In *MICCAI*, pages 274–284. Springer.
- [16] [Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.
- [17] [Zhang et al., 2018] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595.

■

Dejan Štepec is a Lead Data Scientist at XLAB d.o.o. and a Ph.D. student at the Faculty of Computer and Information Science at the University of Ljubljana. He completed his master's studies at the University of Ljubljana in 2017. His main research interests lie in the fields of computer vision and machine learning. He is currently mostly focusing on advancing the field of digital pathology with approaches that require as little labeled data as possible.

■

Danijel Skočaj is an associate professor at the University of Ljubljana, Faculty of Computer and Information Science. He is the head of the Visual Cognitive Systems Laboratory. He obtained a Ph.D. in computer and information science from the University of Ljubljana in 2003. His main research interests lie in the fields of computer vision, machine learning, and cognitive robotics.