Boosting FFPE-to-HE Virtual Staining with Cell Semantics from Pretrained Segmentation Model

Yihuang Hu[1], Qiong Peng[1], Zhicheng Du[2], Guojun Zhang[3], Huisi Wu[4], Jingxin Liu[5], Hao Chen[6], and Liansheng Wang[1](✉)

[1] Department of Computer Science at School of Informatics, Xiamen University, Xiamen, China
{huyihuang, qpeng}@stu.xmu.edu.cn, lswang@xmu.edu.cn

[2] Fujian Key Laboratory of Precision Diagnosis and Treatment in Breast Cancer, Xiang'an Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China
chengmed2018@163.com

[3] Yunnan Cancer Hospital The Third Affiliated Hospital of Kunming MedicalUniversity Peking University Cancer Hospital Yunnan, Kunming, China
zhangguojun@kmmu.edu.cn

[4] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
hswu@szu.edu.cn

[5] School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China
jingxin.liu@xjtlu.edu.cn

[6] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, China
jhc@cse.ust.hk

Abstract. Histopathological samples are typically processed by formalin fixation and paraffin embedding (FFPE) for long-term preservation. To visualize the blurry structures of cells and tissue in FFPE slides, hematoxylin and eosin (HE) staining is commonly utilized, a process that involves sophisticated laboratory facilities and complicated procedures. Recently, virtual staining realized by generative models has been widely utilized. The blurry cell structure in FFPE slides poses challenges to well-studied FFPE-to-HE virtual staining. However, most existing researches overlook this issue. In this paper, we propose a framework for boosting FFPE-to-HE virtual staining with cell semantics from pretrained cell segmentation models (PCSM) as the well-trained PCSM has learned effective representation for cell structure, which contains richer cell semantics than that from a generative model. Thus, we learn from PCSM by utilizing the high-level and low-level semantics of real and virtual images. Specifically, We propose to utilize PCSM to extract multiple-scale latent representations from real and virtual images and align them. Moreover, we introduce the low-level cell location guidance for generative models, informed by PCSM. We conduct extensive experiments on our collected dataset. The results demonstrate a significant improvement of our method over the existing network qualitatively and quantitatively. Code is available at https://github.com/huyihuang/FFPE-to-HE.

Keywords: FFPE-to-HE virtual staining · Cell semantics · Generative adversarial network.

1 Introduction

Histopathological samples are typically available in both unstained and stained forms. Among unstained forms, formalin fixation and paraffin embedding (FFPE) is the most common due to the advantage of long-term preservation, while the structures of cells and tissue in FFPE slides are blurry. Different chemical dyes are commonly employed to visualize the structures, or to label the specific molecules, thereby assisting in diagnosis or research. Among them, hematoxylin and eosin (HE) is recognized as the standard staining method, constituting approximately 80% of histopathology slides worldwide [1]. Hence, it is obvious that FFPE-to-HE staining is widely applied in the medical field. However, the chemical staining process from FFPE to HE is typically labor-intensive and requires specific laboratory facilities [2].

With the rapid development of digital pathology and deep learning technology, virtual staining has shown enormous potential as an alternative to chemical staining [2]. Due to its wide application, FFPE-to-HE virtual staining has attracted the attention of many researchers. For example, Asaf *et al.*, [1] applied DCLGAN [4] on unstained skin tissue and compared its performance with CycleGAN [14] and CUT [9]; Khan *et al.* investigated the virtual staining performances of pix2pix and its variants (double convolution and dense convolution) on preclinical prostate tissue [6]; Koivukoski *et al.* utilized CycleGAN and prostate tissue to investigate the influence of section thickness [7]; Randa *et al.* proposed a novel loss function based on Pearson's correlation coefficient (PCC) to reduce the high-level tiling artifacts in the images generated by pix2pix [11].

It can be concluded that most researchers simply apply the Conditional Generative Adversarial Networks (cGANs) to achieve FFPE-to-HE virtual staining on different tissues, while few studies introduce the cell semantics. Additionally, the blurry cell structure in FFPE slides poses challenges to FFPE-to-HE virtual staining. Hence, the insufficient consideration of cell semantics will likely result in suboptimal performances of FFPE-to-HE virtual staining. Motivated by the analysis above, we take the cell semantics as the critical information for mitigating the challenge of blurry cell structure, thereby leading to superior outcomes. Specifically, cell semantics includes comprehensible low-level semantic information such as position, shape, and powerful high-level semantic representation for downstream tasks like classification and segmentation. However, cell semantics is typically inaccessible, and usually low-level semantic information can be obtained even through labor-intensive and time-consuming expert annotation. We notice that there are several well-pretrained cell segmentation models (PCSM) based on a vast number of medical images [3,12], which can simultaneously obtain low-level and high-level semantics. To leverage the powerful cell semantics, we propose a framework to boost the challenging virtual staining tasks like FFPE-to-HE with PCSM. Specifically, we feed the input image into PCSM to acquire low-level and high-level cell semantics that are provided by cell segmentation masks and latent representations, respectively. Then, we introduce the acquired low-level and high-level cell semantics into the cGAN in different ways. In summary, The main contributions of our work are listed as follows:

(1) The importance of cell semantics for boosting the challenging virtual staining like FFPE-to-HE is first noticed and investigated.

(2) We propose a framework to simultaneously introduce the low-level semantics and the high-level semantics of cells into the cGAN without expert annotation.

(3) We demonstrate the significant improvement of our framework on boosting the challenging FFPE-to-HE virtual staining through quantitative and qualitative evaluations on an internal test dataset and an external test dataset.

2 Methodology

A typical cGAN is composed of a generator G and a discriminator D. The overview of our framework to boost virtual staining of the cGAN is shown in Fig. 1. During training, the real FFPE image (denoted as x) is firstly input into G to generate a virtual HE image (denoted as $\hat{y}$). Both $\hat{y}$ and the aligned real HE image (denoted as y) are sent to PCSM to acquire their high-level semantics and the alignment between them is conducted. Additionally, the low-level semantic (denoted as m) of the real HE image is obtained from the output of PCSM and serves as the guidance to D. The details of high-level semantic alignment and low-level semantic guidance are elaborated below.

2.1 High-level Semantic Alignment

PCSM is typically based on an encoder-decoder architecture [3,12], which forms the basis of the following description. It is believed that the intermediate features of the multiple encoder blocks in PCSM contain high-level semantics about cells. To align the high-level semantics of the real and the virtual images, we design a CSLoss based on the intermediate features. We denote the i-th feature extractor as EC^i (from input to the i-th encoder block of PCSM). The corresponding i-th intermediate feature loss l^i_{cs} is then:

$$l^i_{\text{cs}} = ||EC^i(y) - EC^i(\hat{y}))||_1 \quad (1)$$

The CSLoss L_{CS} is then:

$$L_{\text{CS}} = \sum_{i=1}^{N} w^i_{\text{cs}} l^i_{\text{cs}} \quad (2)$$

where N is the total number of the pretrained CS model's encoder blocks, w^i_{cs} is the weight corresponding to l^i_{cs}.

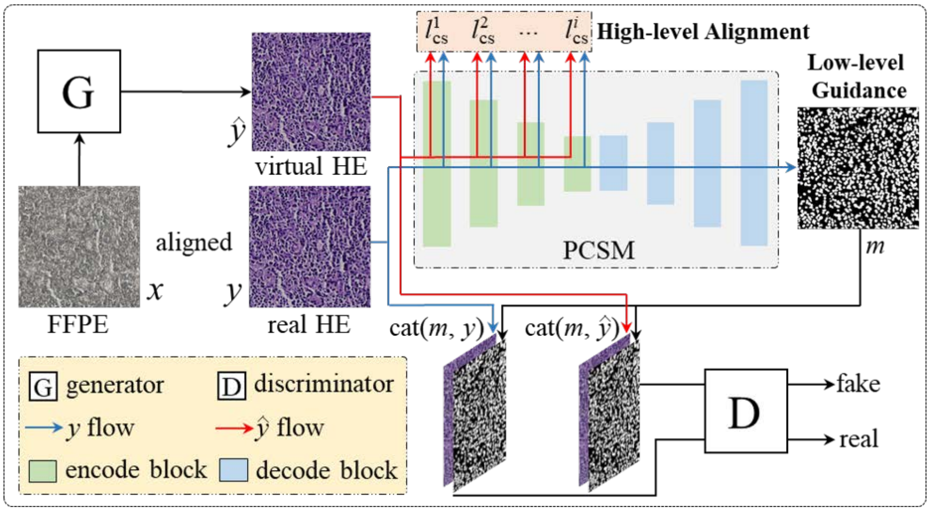

Fig. 1. Overview of the proposed framework to boost virtual staining of the cGAN.

2.2 Low-level Semantic Guidance

The GAN loss of typical cGANs like pix2pix and pix2pixHD is given by

$$E_{(x,y)}[\log D(x,y)] + E_x[\log(1 - D(x,\hat{y}))] \tag{3}$$

where x is the semantic label map and y is the input image [5,13]. It is worth noting that the input to D is a channel-wise concatenation of x and y, denoted as cat(x, y). That means it would be the channel-wise concatenation of a real FFPE image and a corresponding HE image when it comes to FFPE-to-HE virtual staining. This situation leads us to formulate a hypothesis that the blurry cell structure in the FFPE image may misguide D to determine the authenticity of the HE image.

The simplest way to tackle this potential issue is to remove the blurry FFPE image from the input to D, denoted as cat(~~FFPE~~, HE), which has been proved effective in our experiment (see Table 1). Differently, our method is replacing the misguidance of the FFPE image with the guidance of low-level semantics from y as part of the input to D (see Fig. 1), hoping to guide D to recognize the difference between the virtual and real images. We name the low-level semantics as CSMask to simplify the description. In this case, the GAN loss of our method $L_{\text{GAN}}(G, D)$ is given by

$$L_{\text{GAN}}(G,D) = E_{(x,y)}[\log D(m,y)] + E_x[\log(1 - D(m,\hat{y}))] \tag{4}$$

2.3 Objective Function

To establish an overall objective function that integrates CSLoss and CSMask, we employ CellPose [12] as the CS model and pix2pixHD [13] as the cGAN for

FFPE-to-HE virtual staining in this paper. In this case, Our objective function G^* is given by

$$G^* = \arg\min_G \max_D L_{\mathrm{GAN}}(G, D) + \lambda_{\mathrm{fm}} L_{\mathrm{FM}} + \lambda_{\mathrm{cs}} L_{\mathrm{CS}} \tag{5}$$

where L_{FM} is the feature matching loss based on the discriminator [13], λ_{fm} and λ_{cs} are employed to adjust the importance of their respective losses. In addition, the values of w^i_{cs} in L_{CS} are determined based on VGGLoss [13], which is calculated on five intermediate features in pretrained VGG19. Specially, we take the first four weights ($[\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}]$) in VGGLoss as w^i_{cs} in L_{CS}, considering that CellPose has only four encoder blocks. It should be emphasized that these weights may not be optimal for CSLoss. We encourage the community to contribute to refining them.

3 Experiment

3.1 Datasets

The dataset for the internal train-test is from the Affiliated Cancer Hospital of Shantou University Medical College in China. We collected 22 pairs of 40x magnification FFPE-HE whole slide images (WSI) of the axillary lymph node tissue samples with breast cancer. We first discarded the HE WSIs with unclear staining or tissue shedding and registered on the remaining pairs of aligned FFPE-HE WSIs. Then, we checked the registration results and obtained 13 well-registered pairs of FFPE-HE WSIs. Finally, we obtained 5098 pairs of aligned FFPE-HE patches with a size of 1152 x 1152 pixels, which were randomly split into train and internal test datasets with a ratio of 8:2.

Similarly, the dataset for the external test is from Yunnan Cancer Hospital in China, which contains 1399 pairs of aligned FFPE-HE patches with a size of 1024 x 1024 pixels.

3.2 Implementation Details

We train the cGAN with a combined data augmentation of random cropping (from 1152 x 1152 to 1024 x 1024 pixels) and random flipping. In particular, the 1075 pairs of internal test data are center-cropped into 1024 x 1024 pixels for dimensional consistency. The total number of training epochs is set to 100: the learning rate remains unchanged for the first 50 epochs and gradually decreases to 0 for the remaining 50 epochs. λ_{cs} of CSLoss is equal to λ_{fm}. The other hyper-parameters are consistent with the default pix2pixHD. The experiment is conducted on GeForce RTX 3080.

3.3 Evaluation Metrics

We use Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), multi-scale SSIM (MSSSIM), and Pearson's correlation coefficient (PCC) as the evaluation metrics for the quality of the generated image, which are widely used in virtual staining research [8,10]. Between a virtual image and a real image, PSNR measures the peak error; SSIM evaluates similarity based on luminance, contrast, and structure; MSSSIM extends SSIM by considering multiple scales; and PCC measures the linear correlation.

3.4 Quantitative Evaluation

Benchmark Results. pix2pixHD is designed with improved objective function and network of pix2pix for generating high-resolution images. Hence, it is reasonable that PSNR, SSIM, MSSSIM, and PCC metrics of pix2pixHD are all higher than those of pix2pix, as shown in Table 1. It is noteworthy that metrics of ours simultaneously achieve significant improvements over pix2pixHD, both on the internal and external test datasets. Additionally, our method does not increase the network capacity, and the training time is almost equal to that of pix2pixHD (about 1900 s/epoch). These results demonstrate the significant effect and strong generalization capability of our method on boosting the cGAN's learning of FFPE-to-HE virtual staining.

Table 1. Experimental results of different methods. The **best** values are highlighted. The second best values are underlined.

Method	Internal Test				External Test			
	PSNR	SSIM	MSSSIM	PCC	PSNR	SSIM	MSSSIM	PCC
Benchmark Results								
pix2pix	18.314	0.548	0.690	0.795	16.979	0.458	0.612	0.751
pix2pixHD	18.682	0.575	0.716	0.814	17.727	0.472	0.626	0.768
Ablation Study								
HD+VGGLos	19.087	0.606	0.739	0.829	18.076	0.518	0.666	0.790
HD+CSLoss	19.168	0.619	0.748	0.833	18.203	0.516	0.664	0.791
HD+cat(~~FFPE~~,HE)	19.238	0.605	0.736	0.831	18.833	0.541	0.675	0.813
HD+cat(CSMask,HE)	19.760	0.636	0.763	0.849	18.910	0.557	0.694	0.814
Ours	**19.861**	**0.651**	**0.771**	**0.853**	**18.973**	**0.569**	**0.701**	**0.817**

Ablation Study. We further explore the effectiveness of CSLoss and CSMask within our method through the ablation study. Firstly, we focus on the ablation experiments of CSLoss to evaluate its contribution to our method. Then, we turn to CSMask and discuss its impact on our approach. Through these analyses, we can gain a more comprehensive understanding of their roles and significance in our method.

As shown in Table 1, the addition of CSLoss simultaneously improves the PSNR, SSIM, MSSSIM, and PCC metrics of pix2pixHD on both the internal and external test datasets, which indicates its effectiveness in boosting the aligned cGAN's learning of FFPE-to-HE virtual staining. Additionally, the addition of

VGGLoss also results in comparable enhancements in these metrics, but lower than those achieved by CSLoss on the internal test dataset or nearly equivalent to those attained by CSLoss on the external test dataset. Given what is emphasized in the methodology regarding the potential optimal values of w_{cs}^{i} in L_{CS}, it is evident that CSLoss outperforms VGGLoss despite this caveat. In fact, this is not difficult to understand, as the pretrained model of VGGLoss is based on natural images rather than pathological images. Therefore, while VGGLoss may capture certain features relevant to image quality, it might not be as effective in capturing the specific characteristics of pathological images, such as those present in the staining process. On the other hand, CSLoss is specifically designed to consider these pathological image characteristics, leading to more accurate and effective evaluations.

As shown in Table 1, replacing the blurry FFPE image with CSMask as part of the input to D, denoted as cat(CSMask, HE), simultaneously and significantly improves the PSNR, SSIM, MSSSIM and PCC metrics of pix2pixHD on both the TRA and EXT datasets, which indicates its significant effectiveness on boosting the aligned cGAN's learning of FFPE-to-HE virtual staining. Additionally, simply removing the FFPE image from the original input to D, denoted as cat(~~FFPE~~, HE), also improves the performance a lot, which validates our speculation in the methodology that the concatenated blurry FFPE image would disturb D from determining the authenticity of the HE image. This also underscores the rationale behind replacing the blurry FFPE image with CSMask. It is reasonable to concatenate CSMask with HE to assist, rather than disturb the determination of D.

3.5 Qualitative Evaluation

To more intuitively reflect the difference between virtual and real images, we define a concept of pixel difference map, whose pixel value at position $(w,\, h)$ is given by

$$\frac{1}{C}\sum_{i=1}^{C}\frac{|y^{i}(w,h)-\hat{y}^{i}(w,h)|}{scale} \tag{6}$$

where C represents the number of channels of images. y^i and $\hat{y}^i$ represent one channel of the real image and the virtual image, respectively. *scale* is the maximum of the pixel value range of images.

As shown in Eq. 6, a smaller pixel value on the pixel difference map means a smaller difference between the virtual image and the real image. Fig. 2 shows one map result of the internal test dataset and one map result of the external test dataset, where a darker blue color on the pixel difference map indicates a smaller value. It can be observed from Fig. 2 that, the virtual image of ours is closer to the real image than the others on both the internal and external test datasets.

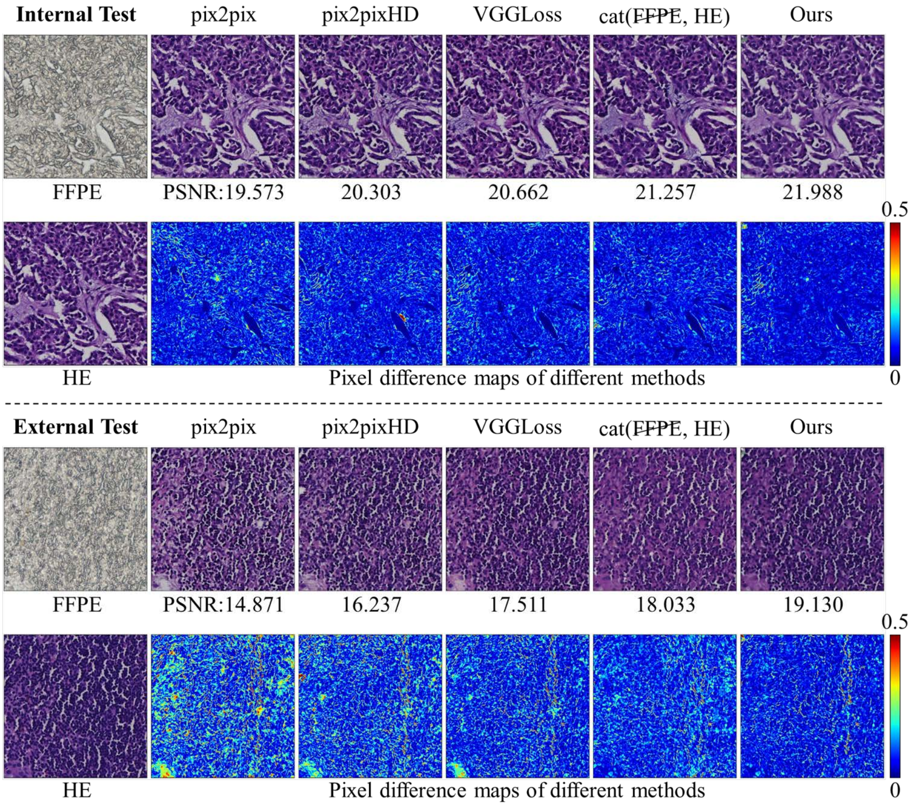

Fig. 2. The pixel difference maps between the real HE image and the corresponding virtual HE images of different methods on the internal and external test datasets.

4 Conclusion

Based on the intuition that cell semantics should be integrated into FFPE-to-HE virtual staining learning, we propose a framework to boost the learning of the cGAN. Our method integrates low-level and high-level semantics into the cGAN model in different ways (CSLoss and CSMask) and achieves a significant improvement over the typical cGAN without the cost of network capacity or training time. Furthermore, we demonstrate the improvements of CSLoss and CSMask within our method, respectively. This indicates that, even under insufficient conditions, employing CSLoss or CSMask individually can still enhance FFPE-to-HE virtual staining learning. The significant improvement and strong generalization capability of our method are also demonstrated on an external test dataset.

Acknowledgments. This work was supported by National Natural Science Foundation of China (Grant No. 62371409), National Natural Science Foundation of China

(62201474), Suzhou Science and Technology Development Planning Programme (Grant No.ZXL2023171).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Asaf, M.Z., Rao, B., Akram, M.U., Khawaja, S.G., Khan, S., Truong, T.M., Sekhon, P., Khan, I.J., Abbasi, M.S.: Dual contrastive learning based image-to-image translation of unstained skin tissue into virtually stained h&e images. Scientific Reports **14**(1), 2335 (2024)
2. Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., Ozcan, A.: Deep learning-enabled virtual histological staining of biological samples. Light: Science & Applications **12**(1), 57 (2023)
3. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical image analysis **58**, 101563 (2019)
4. Han, J., Shoeiby, M., Petersson, L., Armin, M.A.: Dual contrastive learning for unsupervised image-to-image translation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 746–755 (2021)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
6. Khan, U., Koivukoski, S., Valkonen, M., Latonen, L., Ruusuvuori, P.: The effect of neural network architecture on virtual h&e staining: Systematic assessment of histological feasibility. Patterns **4**(5) (2023)
7. Koivukoski, S., Khan, U., Ruusuvuori, P., Latonen, L.: Unstained tissue imaging and virtual hematoxylin and eosin staining of histologic whole slide images. Laboratory Investigation **103**(5), 100070 (2023)
8. Liu, S., Zhang, B., Liu, Y., Han, A., Shi, H., Guan, T., He, Y.: Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. IEEE Transactions on Medical Imaging **40**(8), 1977–1989 (2021)
9. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. pp. 319–345. Springer (2020)
10. Rana, A., Lowe, A., Lithgow, M., Horback, K., Janovitz, T., Da Silva, A., Tsai, H., Shan-mugam, V., Bayat, A., Shah, P.: Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. jama netw. open. 2020; 3. Publisher: American Medical Association.[Europe PMC free article][Abstract][Google Scholar] (2020)
11. Rana, A., Yauney, G., Lowe, A., Shah, P.: Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 828–834. IEEE (2018)
12. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature methods **18**(1), 100–106 (2021)

13. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8798–8807 (2018)
14. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)