Batuhan Karaman and Cagla Deniz Bahadir
kbk46 and cdb232

Decoupling Image Features for Understanding where Diffusion Models Fail in Generating Mitotic Figures

1 Team Information

Team Members: Batuhan K. Karaman (kbk46) and Cagla Deniz Bahadir (cdb232)

Team Name: PathoGen

2 Description of the Problem

Mitotic figure detection is an important task used for tumor grading purposes in several cancers, such as meningiomas [1] and breast cancer [2]. Many applications have focused on the detection and classification of mitotic figures [3, 4]. Conditional diffusion probabilistic models have recently been utilized for realistic image generation [5]. However, in some cases, the diffusion model fails to generate an image that corresponds to the class that was inputted as the condition value to the model. This issue may be attributed to the falsely learned class embeddings by the diffusion model, as it focuses on unrelated image features.

In this project, we propose a retrospective training strategy that decouples general image features from mitotic features, which are trained on the generated paired images from the diffusion model. Our code is available at: `https://github.com/cagladbahadir/Cagla_Batuhan_ML6990`.

3 Dataset Description

In our experiments, we use a publicly available data set: Canine Cutaneous Mast Cell Tumor (CCMCT) [4]. The CCMCT dataset has 32 Whole Slide Images (WSI) in total. Every WSI has been exhaustively annotated for mitotic figures by two pathologists and the final labels are created via a consensus process. There are a total of 262,481 annotations, 44,880 of which are classified as mitotic figures. We use the same training and test split (21 slides, 11 slides) as described in the original paper. We use patches of images with the size of 64x64 that is centered with the cell nucleus.

4 Background

4.a Generating and Classifying Mitotic Figures

The diffusion probabilistic model (DPM) is a widely-used generative model that has various applications [5]. The model is capable of producing paired examples with different class conditions, utilizing the same random seed as input. This feature enables the generation of training data where the structural information from the image can be decoupled from the features that relate to the mitotic figure.

Figure 1 illustrates that each pair of cells generated with a different condition value generally resembles each other, with a distinct difference in the chromatin, resulting in a change in class from a negative example to a mitotic figure.

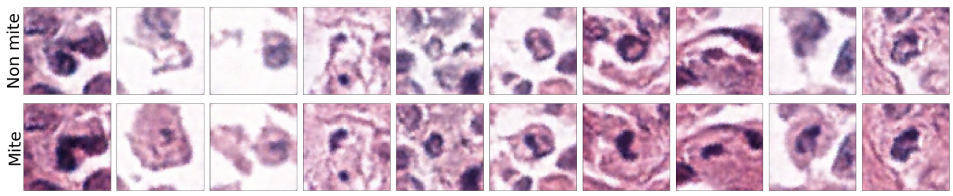

Figure 1: Paired cell images generated by the conditional diffusion model with condition values 0 and 1.

Recently ResNet and DenseNet style classification models have been used for the mitotic figure classification [4]. They reached accuracies around 90% and F1-Score around 81%. These classification models allow us to quickly

Batuhan Karaman and Cagla Deniz Bahadir
kbk46 and cdb232

test thousands of generated images without the need for an expert, and isolate the failed cases. In this work, we use a ResNet34 trained on the CCMCT dataset as the mitosis classifier.

4.b Understanding the Reasons why Diffusion Models Fail to Generate Images with the Conditional Class Values

In this study, we generated two classes of images using DPM: mitotic figures and negative examples. The conditional diffusion model takes in an integer as the mitosis probability, which can either be 0 or 1. Ideally, we aim for images generated with a condition value of 0 to be classified as 0 by the mitosis classifier, and images generated with a condition value of 1 to be classified as 1.

However, as discussed earlier and illustrated in Fig 2, we observed that when the conditional value is set to 1 for a set of 1000 images, the average classification score is only around 65%. This implies that the diffusion model fails to generate mitotic figures even when the conditional value is 1. As our mitosis classifier has been effective in identifying mitotic figures in real cell images, we suspect that the diffusion model struggles to isolate the features that are specific to mitotic figures. In other words, there is a possibility that the diffusion model is encoding a confounding feature in these images.

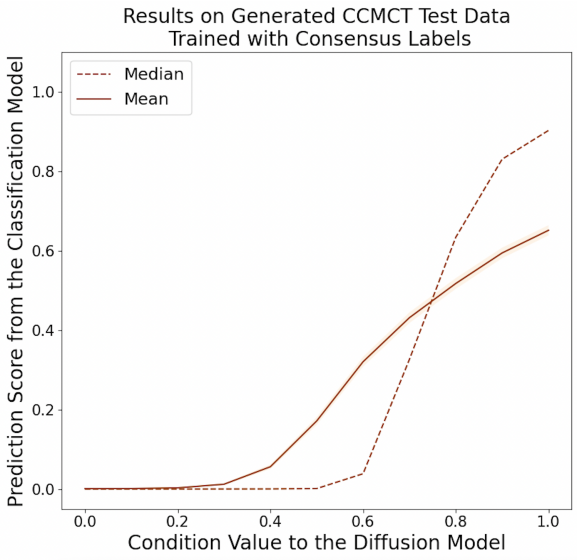

Figure 2: Generated images from DPM are tested on the mitosis classification model.

Fig 3 depicts the Class Activation Map (CAM) heatmaps generated using the mitosis classifier trained on real data, allowing us to comprehend the areas of focus for the model during the classification process. Since the diffusion model does not incorporate the mitosis condition in a convolutional model, the CAM approach cannot be applied to determine how the model encodes the mitotic features. Hence, in this work, we train a proxy model that contains a convolutional section, enabling us to identify the mitotic features proposed by the diffusion model and the reasons for its occasional failure.

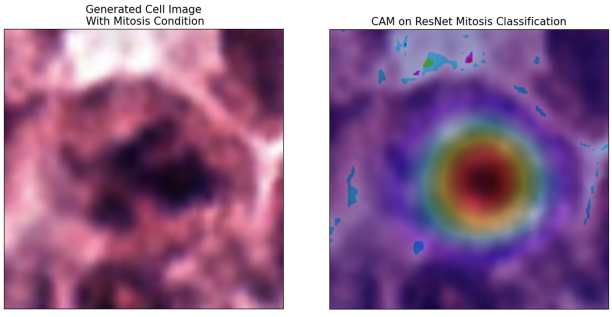

Figure 3: CAM results obtained for synthetic data using the mitosis classifier.

Batuhan Karaman and Cagla Deniz Bahadir
kbk46 and cdb232

5 Algorithmic Approach

It is possible to utilize the CAM (Class Activation Mapping) technique directly on a mitosis classifier in order to determine the specific area of a synthetic image that the classifier is focusing on for classification purposes. However, it is unclear which particular aspects of mitotic images are being considered by the DPM (Diffusion Probabilistic Model) as indicative of mitotic activity. This presents a challenging problem, as the DPM is composed of two distinct inputs: the probability of mitosis, and the noise (or seed). Therefore, in order to accurately perform CAM analysis, it is imperative that we need two proxy encoders that isolate the effects of each individual input on the overall generated mitotic image.

In order to build such encoders, we employ the methodology described in [6], with one modification. [6] addresses the issue of dataset bias in image classification models, which tend to rely on peripheral attributes of data items. The proposed feature-level data augmentation technique synthesizes diverse bias-conflicting samples by learning intrinsic and bias attributes and swapping their latent features, leading to improved debiasing and classification accuracy on both synthetic and real-world datasets.

In this work, we use the framework shown in Fig 4 during training. The mitotic encoder E_m learns the mitotic features, and the seed encoder E_s learns the non-mitotic features. The output size of both encoders is D. These encodings are swapped among different synthetic images and concatenated, as shown in Fig 4. A training mini-batch consists of two synthetic images per seed input (Seed 1 and Seed 2), one with 0 and the other one with 1 as the probability input (p_m).

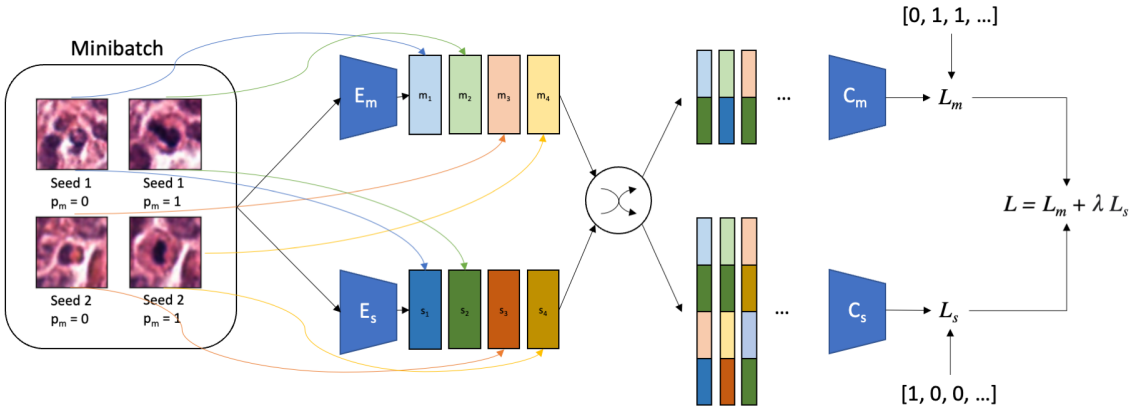

Figure 4: Overview of the training framework.

The mitosis classifier, C_m, and the bias classifier, C_b, both utilize fully connected layers, however, their inputs differ. C_m receives a single mitotic embedding and a randomly chosen seed embedding, which are concatenated to create a combined input of length $2D$. The true label for this input corresponds to the probability input of the image that the intrinsic part of the embedding relates to. On the other hand, C_b is a pair-wise classifier that requires the concatenation of two mitotic embeddings and two seed embeddings, resulting in a $4D$ dimensional input. The true label for this input is 1 if the seed embeddings are from images generated by the DPM using the same seed input, and 0 if they are from different seed inputs. For the mini-batch in Fig 4, we generate 16 (4×4) length $2D$ embeddings for C_m and 256 ($4 \times 4 \times 4 \times 4$) length $4D$ embeddings for the seed classifier.

Following completion of training, we consider E_m will function as an encoder that solely maps mitotic features of a cell into a latent space, while E_s will encode non-mitotic features.

In inference, as shown in Fig 5, we compute the D-dimensional mitotic embeddings of the images in our test set. Then, we compare the CAM visualizations of the mitotic encoder we train and the mitosis classifier that was trained on the real data.

Batuhan Karaman and Cagla Deniz Bahadir
kbk46 and cdb232

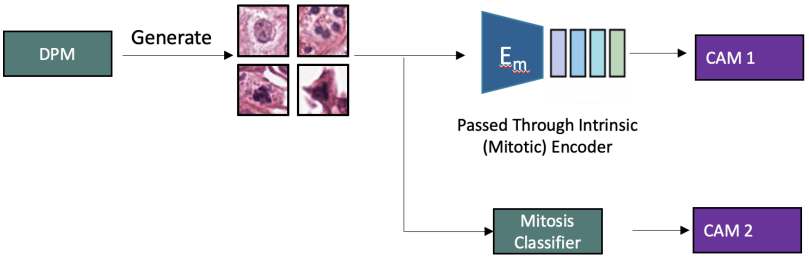

Figure 5: Overview of the inference.

6 Results

6.a Training

To train our model, we combined the loss functions of the mitosis and seed classifiers as shown in Eq 1, where $\mathcal{L}_m$, $\mathcal{L}_s$, λ are mitosis loss, seed loss, and balancing hyper-parameter, respectively. Each batch consisted of 16 x 4 images, providing the model with a total of 64 images. Our training set included 8000 images, while the validation set consisted of 1389 images. To ensure variation in our training, we shuffled the pairings in the training set at every epoch, while keeping the validation pairings constant throughout training. Furthermore, we generated a test set comprising 1300 images to assess the performance of our model.

$$\mathcal{L} = \lambda \mathcal{L}_m + (1 - \lambda)\mathcal{L}_s \tag{1}$$

To build the encoders E_m and E_s, we incorporated pre-trained Resnet-34s as the backbone architecture. This choice resulted in an embedding dimension of 512. As for the classifiers C_m and C_s, they comprised two fully connected layers with a ReLU activation function in between. We utilized the Adam optimizer with a learning rate of 1e-04 and weight decay of 1e-4. Throughout the training process, we kept track of the validation average accuracy, defined as the average of the accuracies of both mitosis and seed classifications. We then selected the model state that had the highest validation average accuracy for our testing purposes.

6.b Test Results

Table 1 shows the performance of the models with different lambda values. Three different models were compared, each with a different weighting of the mitosis $\mathcal{L}_m$ and seed $\mathcal{L}_s$ loss functions. We have generated three random pairings from the 1300 images and we share the average results. We would like to note that the Resnet mitosis classifier classified 70% of the 1300 images in the test set as mitotic figures.

	Mitosis Accuracy	Seed Accuracy	Average Accuracy
0.25 $\mathcal{L}_m$ + 0.75 $\mathcal{L}_s$	0.8611 [0.8593, 0.8629]	0.7957 [0.7938, 0.7981]	0.8284 [0.8266, 0.8305]
0.50 $\mathcal{L}_m$ + 0.50 $\mathcal{L}_s$	0.8326 [0.8316, 0.8331]	0.8194 [0.8055, 0.8331]	0.8245 [0.8193, 0.8285]
0.75 $\mathcal{L}_m$ + 0.25 $\mathcal{L}_s$	0.8766 [0.8754, 0.8783]	0.7556 [0.7498, 0.7622]	0.8161 [0.8126, 0.8202]

Table 1: Performance of the models trained with different λ values. The data format is mean [minimum, maximum]

The model that used a weighted combination of 0.25 for $\mathcal{L}_m$ and 0.75 for $\mathcal{L}_s$, achieved the highest accuracy for mitosis classification with a score of 0.8611, while the model that used an equal weighting of 0.5 for both $\mathcal{L}_m$ and $\mathcal{L}_s$, achieved the highest seed classification accuracy with a score of 0.8194. When looking at the average accuracy, which considers both mitosis and seed classifications, it is evident that the model with $\lambda = 0.25$ performed the best with an average accuracy of 0.8284, followed closely by the model with $\lambda = 0.50$ with an average accuracy of 0.8245. Based on these results, we picked the model with the highest average accuracy, which is the model trained with $\lambda = 0.25$.

6.c CAM Results

6.c.1 Acquiring CAM

We analyzed the CAM [7] results for the mitotic and seed encoders that were trained with synthetic images from the diffusion model and mitosis classifier that was trained with real images. All three networks had the ResNet34 architecture, where the mitotic and seed encoders had the final fully connected layer replaced with an identity function.

The synthesized images were resized to 256x256 pixels and passed through the convolutional blocks in ResNet34 architecture until the final average pooling layer. The final shape was 512x8x8 where 512 is the number of channels. To acquire the CAM results we averaged over the channel axis and the resultant low resolution activation map was interpolated onto the input images for visualization purposes.

6.c.2 Interpreting CAM Results

Figure 6 shows the CAM visualizations for the mitotic encoder and the mitosis classifier on a successfully generated synthetic mitotic figure image. The image on the left is the generated cell image, the second image is the CAM of the last convolutional layer of the mitotic encoder and third image depicts the CAM of the last convolutional layer of the mitosis classifier.

The mitotic encoder was trained with images from the diffusion model and labels were the conditional values. This means that the mitotic encoder captured what diffusion model "thinks" a mitotic figure is, rather than the image in actuality being a mitotic figure or not (either by mitosis classifier or expert verification). The mitosis classifier was trained on the real data that was trained on expert annotations and it achieved 90% accuracy on the real test data. This means the mitosis classifier learns what an expert deems a mitotic figure is.

In section 6.b it is mentioned that for the 1300 mitotic test cases generated by the diffusion model, 69% of them were classified as non-mitotic by the mitosis classifier. Therefore we expect to see differences in which features the diffusion model thinks are pertaining to mitotic figures and which features are pertaining to mitosis in actuality. The CAM results give insight into this problem.

In Figure 6 the mitotic encoder has a very concentrated activation in a small area on the chromatin whereas the mitosis classifier covers the entire nucleus. Here, the mitotic encoder is a proxy model to try to understand what features or regions diffusion model likely picked up the mitotic features from and ended up failing about 30 percent of the cases. This image shows that it is possible that the training mitosis images shown to the diffusion model may have had a strong signal, possibly darkness, in a small region in the chromatin. Therefore, the model, while learning how to embed the class features may have only picked up on the signal coming from this region.

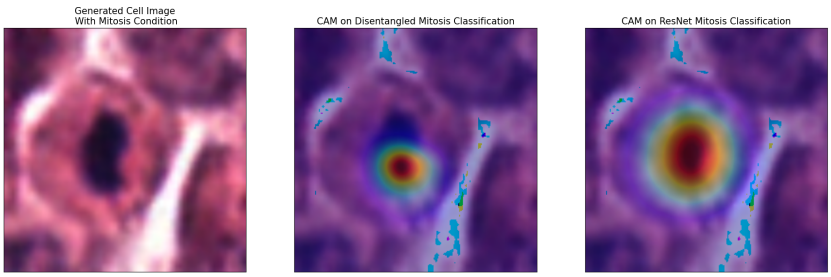

Figure 6: Generated cell image, CAM results on the mitotic encoder and CAM results on the mitosis classifier for a successful mitotic figure generation.

Furthermore in Figure 7 we are depicting the CAM results from mitotic and seed encoders for two images with class conditions non-mitosis and mitosis and with the same seed value. The images generated with the same seed and different conditional value often have similar features in areas besides the nucleus. The seed network was trained by showing pairs of images, which would eventually teach the model to look at areas in the images that are similar. In Figure 7 only a part of the chromatin is highlighted by the mitotic encoder for both images. Both highlights in the images are in the similar regions and are in similar sizes. Going from the non-mitotic example to the mitotic example the "bean" shape of the chromatin seems to remain similar, but the highlighted area in the mitotic example gets darker and the highlighted area in the bottom image transforms into a part of the cell border. The seed encoder encodes areas covering the top section of the nucleus and some regions around the nucleus where

we see the paired images have similar structures. Overall the proxy model approximately predicts which parts of the image is likely to change with the changing class condition and implicitly where the diffusion model "assumed" the mitotic features were located.

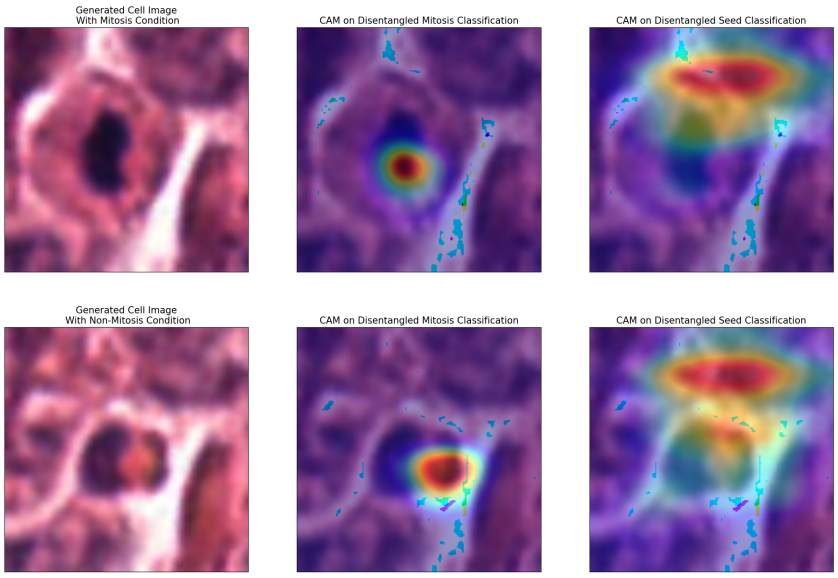

Figure 7: Top row from left to right: Generated mitotic figure image, CAM results on the mitotic encoder and CAM results on the mitosis classifier. Bottom row follows the same pattern for a non-mitotic cell image, that was created with the same seed as the mitotic image on the top row.

In addition to the successful generation of mitotic figures we also acquired CAM results for unsuccessful generation of mitotic figures. Figure 8 shows the results for a failed case. Similarly to the results in Figure 6 the mitotic encoder again only highlights a small region on the chromatin where the mitosis classifier highlights the entire nucleus.

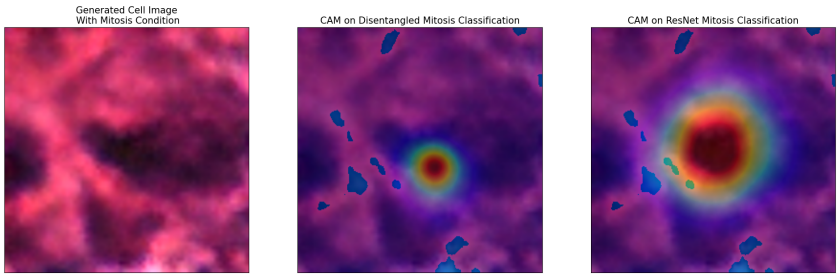

Figure 8: Generated cell image, CAM results on the mitotic encoder and CAM results on the mitosis classifier for a failed mitotic figure generation.

In Figure 9 the CAM results for the mitotic and seed encoders are shown for the mitotic and non-mitotic image pair similar to the Figure 7. However, in this case the mitotic example is a failed case. The mitotic encoder again highlights a small region in the nucleus where the mitotic version is darker and the non-mitotic version is lighter. The seed encoder covers the top part of the nucleus and extends further than the nucleus. The region highlighted by the seed encoder visually and structurally looks similar in the image pair. The results of the proxy models show that the diffusion model "thinks" that it's changing the class by darkening on a small region in the nucleus, where it should've made multitude of changes to the chromatin and cell borders.

7 Conclusion

The aim of this project was to understand why diffusion models fail to generate mitotic figures even when the condition is 1, for around 30% of the cases. In order to accomplish this we adapted a framework which separated

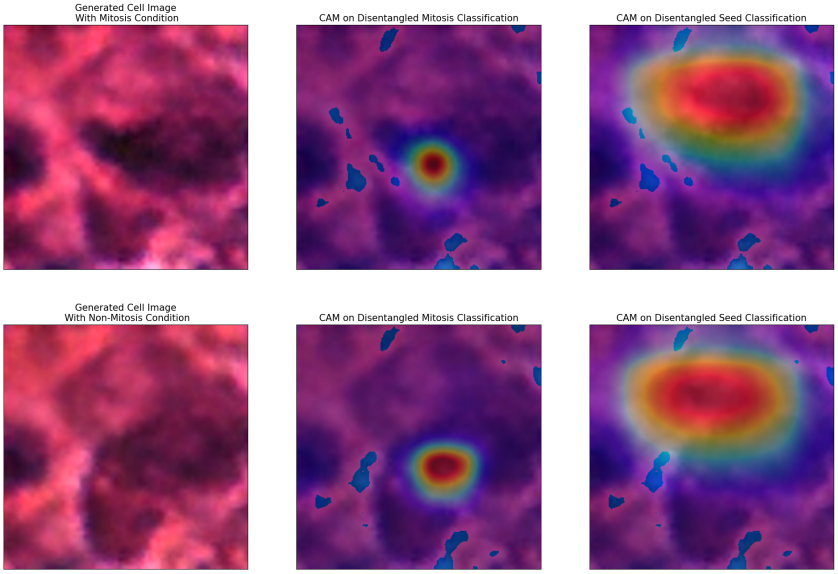

Figure 9: Top row from left to right: Generated failed mitotic figure image, CAM results on the mitotic encoder and CAM results on the mitosis classifier. Bottom row follows the same pattern for a non-mitotic cell image, that was created with the same seed as the mitotic image on the top row.

the encodings for the mitotic features and the seed features (the morphological features that are present in images generated with the same seed and different mitotic condition). This framework allowed us to isolate the features that the diffusion model "thinks" are pertaining to the mitotic figures.

Diffusion model encodes the class condition value in a fully connected network and in order to have a human intrepretable result we opted for a convolutional mitotic encoder going from an image to an embedding to generate class activation maps. In section 6.c.2 we presented multitude of CAM results for successful and unsuccessful generations. The main observation was that the CAM from the mitotic encoder showed a much smaller and concentrated highlighted region compared to the result of the mitosis classifier. The highlights from the mitotic encoder roughly corresponded with the structures that change with the class condition and the seed encoder roughly highlighted the structures that remain the same or similar with the changing class condition. These results overall showed that the diffusion model likely picked up on a strong signal in a small area in the nucleus while encoding the class condition rather than concentrating on the entire nucleus where many more structural changes should have occurred for the mitotic class. Overall, we believe that understanding the behaviour and failure of diffusion models with a proxy model paves the way to further guide and improve the diffusion models to accurately create images of certain classes.

8 Appendix

We generated CAM results for several successful and failed examples were we weren't able to share all of them due to space constraints. The appendix shows CAM results for another successful and failed example that shows the trend in the previous images holds for different cell generations.

In Figures 10 and 11 we see the mitotic and seed encoder and mitosis classifier CAM results for a successful generation. Similar to results shown before, the mitotic encoder only focuses on a small region on the chromatin where a significant change occurs going from non-mitotic to mitotic generation. In Figures 12 and 13 we see the mitotic and seed encoder and mitosis classifier CAM results for a failed generation. The mitotic encoder again only focuses on a small region on the chromatin where a the slight darkening in the highlighted area does not overall manage to generate a mitotic figure.

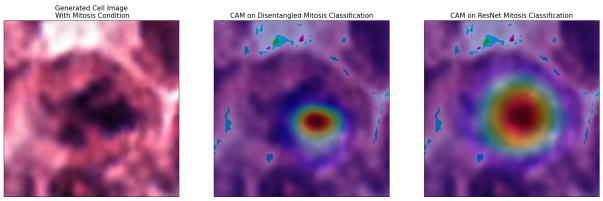

Figure 10: Generated cell image, CAM results on the mitotic encoder and CAM results on the mitosis classifier for a successful mitotic figure generation.

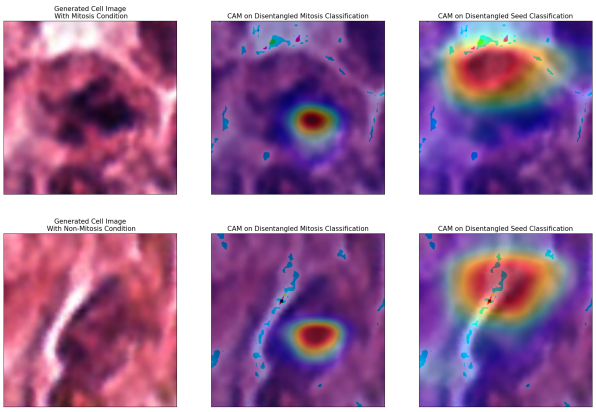

Figure 11: Top row from left to right: Generated mitotic figure image, CAM results on the mitotic encoder and CAM results on the mitosis classifier. Bottom row follows the same pattern for a non-mitotic cell image, that was created with the same seed as the mitotic image on the top row.

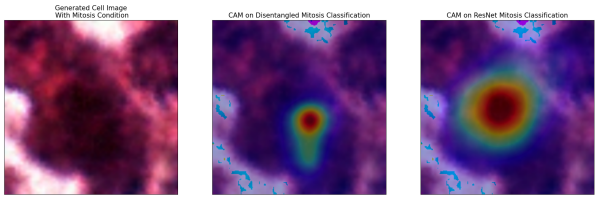

Figure 12: Generated cell image, CAM results on the mitotic encoder and CAM results on the mitosis classifier for a failed mitotic figure generation.

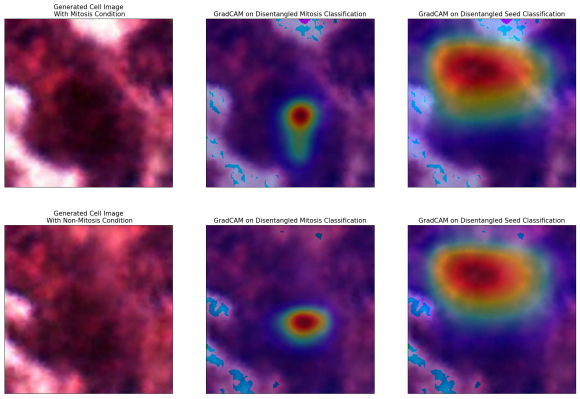

Figure 13: Top row from left to right: Generated failed mitotic figure image, CAM results on the mitotic encoder and CAM results on the mitosis classifier. Bottom row follows the same pattern for a non-mitotic cell image, that was created with the same seed as the mitotic image on the top row.

References

[1] J. Ganz, T. Kirsch, L. Hoffmann, C. A. Bertram, C. Hoffmann, A. Maier, K. Breininger, I. Blümcke, S. Jabari, and M. Aubreville, "Automatic and explainable grading of meningiomas from histopathology images," in *MICCAI Workshop on Computational Pathology*, pp. 69–80, PMLR, 2021.

[2] M. Veta, P. J. van Diest, and J. P. Pluim, "Detecting mitotic figures in breast cancer histopathology images," in *Medical Imaging 2013: Digital Pathology*, vol. 8676, pp. 70–76, SPIE, 2013.

[3] M. Aubreville, N. Stathonikos, C. A. Bertram, R. Klopfleisch, N. Ter Hoeve, F. Ciompi, F. Wilm, C. Marzahl, T. A. Donovan, A. Maier, *et al.*, "Mitosis domain generalization in histopathology images—the midog challenge," *Medical Image Analysis*, vol. 84, p. 102699, 2023.

[4] C. A. Bertram, M. Aubreville, C. Marzahl, A. Maier, and R. Klopfleisch, "A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor," *Scientific data*, vol. 6, no. 1, pp. 1–9, 2019.

[5] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[6] J. Lee, E. Kim, J. Lee, J. Lee, and J. Choo, "Learning debiased representation via disentangled feature augmentation," *arXiv:2107.01372 [cs]*, 10 2021.

[7] S. Poppi, M. Cornia, L. Baraldi, and R. Cucchiara, "Revisiting the evaluation of class activation mapping for explainability: A novel metric and experimental analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2299–2304, 2021.