6. Supplementary Section

In this section we present our early approach for the ISIC-2018 data set that rely on methods proposed in the main text. Section 6.1 describes deep-learning-based approach for data purification, Section 6.2 and Section 6.3 address the problem of data imbalancedness for ISIC-2018 and propose a solution that rely on coupled DCGAN model proposed in the main text, Section 6.4 incorporates both algorithms into a classification network.

6.1. Data Purification

Hair removal algorithm described in main section is not scalable to massive datasets as it requires manual tuning of parameters. These parameters vary with type of images (dermoscopic or clinical), size of the image, color of the hair (black or blonde) and the amount of hair present in the image. Thus, a deep learning solution was desired to fully incorporate it into our classification pipeline for ISIC 2018 task. We utilized a Unet based encoder-decoder model architecture with convolution operations replaced by partial convolutions [22] to improve the performance of the model (Partial convolution is often referred as segmentation-aware convolution. The intuition behind using partial convolutions arises from the fact that given an input image and the corresponding binary mask, the output of the convolution operation should only depend on the regions of the image that are not zeroed-out by the mask and should not take into account regions where pixels values are zero.). We refer to this network as data purification network. The input of the network is the original image and the output is the same image after occlusion removal. Since the training data set for this model requires coupled pairs of images before and after occlusion removal and those, to the best of our knowledge, are not available in any public data sets, we created such training data using the algorithm described in the main text. The obtained training data-set consisted of $16,270$ images and was augmented through random masking. The data purification network was trained using the loss proposed for training networks with partial convolutions [22] and that targets both per-pixel reconstruction accuracy ans well as composition i.e how smoothly the predicted hole values transition into their surrounding context. The data purification network was trained using an Adam optimizer with beta values set to $(0.5, 0.99)$ and constant learning rate set to $2e^{-4}$ in the beginning of training and SGD optimizer with momentum 0.9 and learning rate $1e^{-4}$ in the later stages of learning. The model was trained over a week on 4 GTX 1080 12Gb GPU cards. Figure 10 show the results of data purification obtained by our model. Figure 10 show the results of data purification obtained by our model. We processed all the images in ISIC-2018 data-set to remove occluded objects. The preprocessed images were then added to the training data set of the lesion classification model to make it more robust to the presence of occlusions and prevent overfitting.

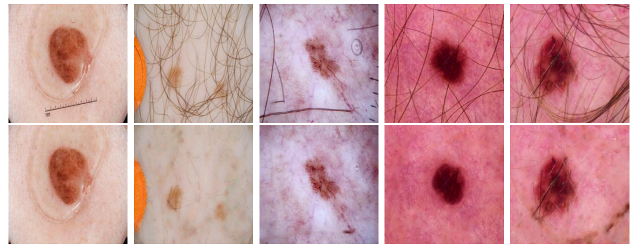

Figure 10. **(Top)**: Original images. **(Bottom)**: Images after removing occlusions, i.e. hairs and rulers, using Data Purification Network.

6.2. Data Imbalancedness

Figure 11 highlights the class imbalancedness problem for ISIC-2018 challenge dataset. The dataset contains 452 cases of acitinic keratosis (AK), 825 cases of basal cell carcinoma (BCC), 1833 samples of benign keratosis (BK), 187 cases of dermatofibroma (DF), 2285 cases of melanoma (M), 9786 cases of nevus (N), and 142 cases of vascular lesion (VL)). It is clearly observed Dermatofibroma constitute less than 2% of the entire dataset.

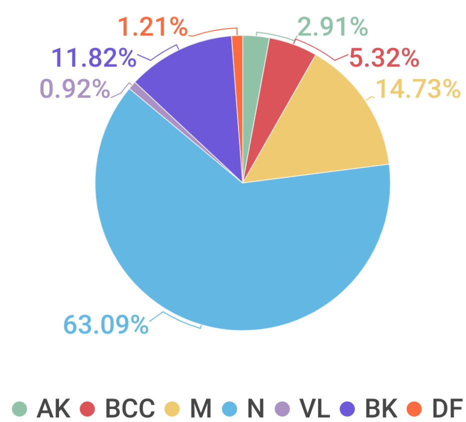

Figure 11. Sizes of the training data sets for the ISIC 2018 task. Data set is heavily imbalanced.

6.3. Data Generation with coupled DCGANs

As the number of classes in the ISIC 2018 task is larger than in case of ISIC 2017, using multiple separate DCGANs for the former becomes inefficient. Instead, for the ISIC 2018 we coupled seven DCGAN architectures. They share parametrization of their initial 4 layers with each other and the final 3 layers are class-specific. Figure 12 and Table 5 shows the idea behind the coupled DCGAN models. The same figure shows exemplary images generated using this approach. The coupled DCGAN models were trained using Adam optimizer with learning rate of $2e^{-4}$ and beta values

	Layer	Output Size	Kernel	Stride	Padding
Generator - Weight sharing block	TransConv	256×4×4	4×4	1	0
	TransConv	128×8×8	4×4	2	1
	TransConv	64×16×16	4×4	2	1
	TransConv	32×32×32	4×4	2	1
Generator- Class specific block	TransConv	16×64×64	4×4	2	1
	TransConv	8×128×128	4×4	2	1
	TransConv	3×256×256	4×4	2	1
Discriminator - Class specific bloc	Conv	16×128×128	4×4	2	1
	Conv	32×64×64	4×4	2	1
	Conv	64×32×32	4×4	2	1
Discriminator - Weight sharing block	Conv	128×16×16	4×4	2	1
	Conv	256×8×8	4×4	2	1
	Conv	512×4×4	4×4	1	1
	Conv	1×1×1	4×4	1	0

Table 5. Details of the architecture of the generative model. Let n be the number of features maps, h be the height and w be the width. Size of the output feature maps is represented as $n \times h \times w$. Each convolution layer in the generator, except for the last one, is followed by a batch normalization and a ReLU nonlinearity. The last convolution layer is followed by a hyperbolic tangent. Similarly, each layer in the discriminator, except for the last convolution layer, is followed by a batch normalization and a leaky ReLU nonlinearity with the leakage coefficient of 0.2. The last convolution layer is followed by a sigmoid. The class specific blocks are repeated for 7 different classes.

Method	Accuracy	Sensitivity	Specificity
Our Classification Model but without performing data purification at testing	0.675	0.561	**0.954**
Our Classification Model	**0.717**	**0.754**	0.837

Table 6. The effect of inducing data purification at testing on the performance of Classification Network.

of 0.5 and 0.999. The latent vector of length 100 that inputs the generator is obtained from standard Gaussian distribution with mean 0 and standard deviation 1. Binary cross entropy loss was used to train both discriminator and generator. This time we balanced the data online i.e the data was augmented at each mini-batch. Thus the model processes a balanced mini-batch before updating the model parameters. We additionally used standard online data augmentation techniques.

6.4. Classification

In Table 6 we demonstrate the advantage of using data purification using ISIC 2018. We report the results when at testing we either perform or not perform data purification. Note that ISIC 2018 does not publish labels for the test data and utilizes black box system to output relevant metrics.

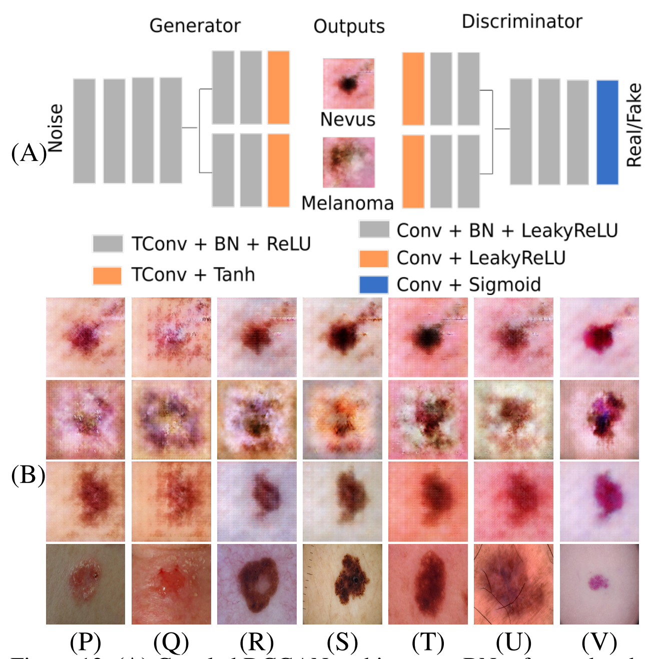

Figure 12. **(A)** Coupled DCGAN architecture. BN refers to batch normalization, Conv and TConv refer to convolution and transposed convolution, respectively. **(B) First three rows:** Generated image of size 256 × 256 (we show 3 exemplary images per class; images in the same row are generated from the same random latent vector) for **(P)**: Actinic Keratosis **(Q)** Basal Cell Carcinoma **(R)** Benign Keratosis **(S)** Melanoma **(T)** Nevus **(U)** Dermatofibroma **(V)** Vascular Lesion. **Fourth row:** Images of real lesion similar (in terms of the MSE) to the generated ones from the third row.