We would like to thank the reviewers for their constructive feedback and comments. In this work, we looked at analyzing the high-frequency Fourier modes of real and deep network generated images and showed that deep network generated images fail to replicate the attributes of these high-frequency modes. Using this, we demonstrated that simple classifiers based on the frequency characteristics of the images can perform effectively and generalize well even when trained with minimal data. Furthermore, we observed the effects of image transformations on this classification task, showing that compression has a large impact on the frequency spectra and, consequently, the classifier, whereas the classifier was less affected by cropping and resolution reduction. To successfully disguise deep network generated images to classifiers based on the frequency spectra, we proposed a synthesis method for spoofing these images which was able to directly and more effectively homogenize the spectra in comparison to compression. We are pleased to see that the reviewers found the observations interesting and insightful to the NeurIPS community (**R1**, **R3**, **R4**), that the problem requires consideration (**R2**), and the experiment design to be systematic and comprehensive (**R1**, **R4**). Additionally, we are encouraged to see that the reviewers found the classification method effective even when using simple classifiers (**R1**, **R3**) and the 'spectrum synthesis' method to show potential in alleviating these spectrum discrepancies (**R1**, **R2**, **R4**).

Some concerns raised by **R1** and **R3** were with regard to the effectiveness of 'more complex' classifiers with respect to the simple low-dimensional classifier we proposed: whether they are 'using the same spectral cues just in a more opaque way (**R1**).' **R1** mentioned that this could perhaps be demonstrated by seeing if those detectors also suffered the same degradation in detection accuracy under compression. For this concern, we refer to the concurrent work of Wang *et al.* that was briefly referred to in the introduction and by **R2**, where they trained a DNN with a large training set to classify real and fake images generated by various generative models. Although their work focused on low-resolution images (224^2 pixels), their classification accuracy also severely deteriorated with compressed images, but this was partially alleviated by including compressed images in their training examples. However, we observed similar detrimental effects on classification at low-resolutions as in their work, and we also showed that compression more severely affected the spectra of high-resolution images. Therefore, it can be surmised that 'complex classifiers' would suffer even larger degradation under compression at higher resolution which would indicate that they are in fact using the same spectral cues in a more opaque way. Hence, we believe it is likely that the proposed 'spectrum synthesis' method would also detrimentally affect their accuracy. In further work, we could directly compare those classification methods.

In relation to the aforementioned Wang *et al.* paper, **R2** raised concerns that more generative models are needed to achieve a more reliable conclusion. However, to systematically analyze the effects of higher-resolution and compression on images, we required datasets or pretrained models that could provide *uncompressed, high-resolution* images, which limited the number of models we could evaluate relative to Wang *et al.* whose work was focused towards low-resolution images. To our knowledge at the time, of the six models which were capable of generating high-resolution (1024^2) images, we considered five for the manuscript. For the sixth model (IntroVAE), the images were only available in a heavily compressed form without an available pretrained model and hence were not suitable for our investigation.

Furthermore, we would like to clarify some of the results. **R3** mentioned the difference in *decay* between StyleGAN and StyleGAN2, indicating that next generation GAN architectures would solve this discrepancy problem. However, we note that the differences (in Fig. 5a) between these two models are actually along the b_1 axis, which indicates the level of high-frequency content. The decay rate b_2, the primary discrepancy between real and fake images, is largely unchanged between the two models. Therefore, we don't believe that there is indication that this problem is being implicitly solved in the current development path of these models. **R2** noted the poor performance (65.9%) of the classifier on highly-compressed StyleGAN2 images at high-resolutions. We reiterate that this is a level of compression where visually noticeable artifacts are observed, and it is not feasible to use it in cases where high-quality synthetic data is needed. In the literature on fake image detection, classification is usually performed using the original output of the generative models, for which we obtain a 97.4-99.9% classification accuracy. **R2** also raised the question of the sensitivity of the decay function fitting method. For this, we've looked at various decay functions (power law, exponential, etc.) and classifier and fitting parameters. As long as the fitting window (threshold frequency) was sufficiently high such that it was primarily fitting noise, the results showed very little sensitivity (on the order of $\pm 1\%$).

There was interest from **R1**, **R3**, and **R4** in having further discussion on explaining the behavior of the generative models shown in the paper. We posit that it is related to the checkerboard artifacts shown by Odena *et al.* as mentioned by **R3** and the AutoGAN method by Zhang *et al.*, which can be seen as an inverse of our 'spectrum synthesis' method. Additionally, we will note the formatting advice of **R3** and the request for more related work in Fourier space image forensics from **R2**. Although **R2** has pointed out that using the Fourier space for image forensics has been done before, this work performs it in relation to deep network generated images for which the mechanisms and behavior of the high-frequency content is quite different than in biometric spoofing attacks. **R3** also pointed out the classification can easily be defeated using the proposed spoofing method, but we believe that is a positive aspect of the work as it shows that these generative models need some post-processing of their original images. **R4** mentioned the need for a more explicit way to improve the spectra of fake images instead of compression, but we believe that our 'spectrum synthesis' method addresses this. We acknowledge the feedback from the reviewers and hope that these points address their concerns.