# Global Texture Enhancement for Fake Face Detection In the Wild: Supplementary File

## 1. Compare with existing works on natural images

To compare with existing works and evaluate the Gram-Net when training and testing images are in different semantic classes, we follow the setting in [4], which is also the "leave one out" setting in [5]. We evaluate on CycleGAN [6] dataset in which images of one category are set aside for testing while the remaining for training. There are a total number of 14 categories: Horse (H), Zebra (Z), Yosemite Summer (S), Yosemite Winter (W), Apple (A), Orange (O), Facades (F), CityScape Photo (City), Satellite Image (Map), Ukiyoe (U), Van Gogh (V), Cezanne (C), Monet (M) and Photo (P). Following [5], we also exclude the sketch and pixel-level semantic map from the dataset.

We train Gram-Net with the same training strategy as in the main paper. Table 1 shows that Gram-Net achieves 98.49% mean accuracy of all the settings, which outperforms existing works. To be noted, we use ResNet-50 as our backbone compared to DesnetNet-121[3] in [5] and Xception-71 [1] in [4]. We expect that deeper backbone networks will further benefit our performance.

More importantly, [5] fails when GANs adopted in training and testing are with different upsampling structures. However, as shown in Table 3 cross-GAN setting ( StyleGAN: nearest-upsampling and PGGAN: deconvolution upsampling ) in the main paper, our approach works almost perfectly in this setting.

## References


[1] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[4] Francesco Marra, Diego Gragnaniello, Davide Cozzolino, and Luisa Verdoliva. Detection of gan-generated fake images over social networks. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 384–389. IEEE, 2018.

[5] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images. *arXiv preprint arXiv:1907.06515*, 2019.

[6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.


| Method | backbone | H-Z | S-W | A-O | F | City | Map | U | V | C | M | average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xception [4] | Xception [1] | 99.16% | 76.76% | 95.91% | 98.56% | 100.0% | 76.79% | 100% | 99.93% | 100.00% | 95.10% | 94.49% |
| AutoGAN-Spec [5] | DenseNet-121 [3] | 98.4% | 99.9% | 98.3% | 100% | 100% | 78.6% | 99.9% | 97.5% | 99.2% | 99.7% | 97.2% |
| Gram-Net | RsNet-50 [2] | 96.46% | 99.54% | 94.73% | 98.0% | 99.53% | 98.86% | 98.62% | 99.47% | 99.90% | 99.82% | 98.49% |

Table 1. Compare with existing works on natural images.