StyleGAN3 Generated Images Detection

ISPL Lab, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano

POLITECNICO
MILANO 1863

Method Description: training stage

Fusion of 5 GAN-Generated Image detectors

Training Stage - Common Features among the 5 detectors:

- BackBone architecture: EfficientNet-B4
- Random Extraction of N RGB patches (128 x 128) from training images
- Max number of epochs: 500
- Initial Learning Rate: 0.001, reduced by a factor 10 if validation does not improve for 10 consecutive epochs

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: testing stage

Fusion of 5 GAN-Generated Image detectors

Testing Stage - Common Features among the 5 detectors:

- Extraction of M RGB patches (128 x 128) from each test image
- Two scores are obtained for each patch: the score related to class “real”, the score related to class “GAN-generated”
- Whether just 1 patch (among the M ones) is detected as GAN-generated, the image is assigned the best score related to the class “GAN-generated” among the M patches. Otherwise (i.e., all patches are detected as real), the image is assigned the best score related to the class “Real” among the M patches

Final Score of the image:

The final score of the image is the average among the scores obtained for the image by the 5 detectors

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: key insights

- Training a detector on a **specific image and/or processing operation** (e.g., only on animals, only on humans, only on JPEG compressed images, etc.) helps enhancing the detector performances on that specific category/operation
- **Fusion** among diverse detectors helps generalizing to multiple image categories (e.g., animals, humans, etc.), to diverse processing operations (e.g., compressions), and to unseen image generation methods (e.g., StyleGAN3)
- **Generic training augmentations** help obtaining a detector robust to common processing operations applied on images (i.e., noise addition, resizing, compressions, etc.) and not to overfit on image generation methods seen in training phase
- **Taking care of JPEG compression** (in training and testing phases) helps obtaining a detector robust to potential misalignment of testing images with respect to the 8 x 8 pixel grid typically introduced by JPEG compression (for details, please check [1])

[1] S. Mandelli, N. Bonettini, P. Bestagini and S. Tubaro, "Training CNNs in Presence of JPEG Compression: Multimedia Forensics vs Computer Vision," *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2020, pp. 1-6, doi: 10.1109/WIFS49906.2020.9360903.

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: Detector 1

Training stage:

- Real class: ~100K real images selected from unrestricted AFHQ-V2-TRAIN, METFACES-U, METFACES, FFHQ-U, FFHQ
- GAN-generated class: ~200K synthetic generated version of AFHQ-V2, METFACES, FFHQ by means of StyleGAN-2, StarGAN-v2, Taming Transformers, FaceVid2Vid, Score-based models.
- Training Augmentations applied on *Images* with probability 50%: Horizontal Flip, Vertical Flip, Random 90-deg-Rotation, Histogram Equalization, Random Blur, Random changes in Brightness/Contrast/Colors/Saturation, Random Downscale and Upscale, JPEG Compression (quality factors from 30 to 100, applied with probability 70%)
- Random Extraction of 1 RGB patch per image

Testing stage:

- 200 RGB patches (128 x 128) are randomly extracted per image

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: Detector 2

Training stage:

- Real class: ~100K real images selected from unrestricted AFHQ-V2-TRAIN, METFACES-U, METFACES, FFHQ-U, FFHQ
- GAN-generated class: ~200K synthetic generated version of AFHQ-V2, METFACES, FFHQ by means of StyleGAN-2, StarGAN-v2, Taming Transformers, FaceVid2Vid, Score-based models.
- Training Augmentations applied on *Patches* with probability 50%: Horizontal Flip, Vertical Flip, Random 90-deg-Rotation, Histogram Equalization, Random Blur, Random changes in Brightness/Contrast/Colors/Saturation, Random Downscale and Upscale, JPEG Compression (quality factors from 30 to 100, applied with probability 70%)
- Random Extraction of 1 RGB patch per image

Testing stage:

- ~ 180 RGB patches (128 x 128) are extracted per image, not in random positions. Patches are extracted such to be aligned with the 8 x 8 pixel grid starting from upper-left corner of the image

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: Detector 3

Training stage:

- Real class: ~ 14K real images selected from unrestricted AFHQ-V2-TRAIN
- GAN-generated class: ~14K synthetic generated version of AFHQ-V2 by means of StyleGAN-2, StarGAN-v2
- Training Augmentations applied on *Patches* with probability 50%: Horizontal Flip, Vertical Flip, Random 90-deg-Rotation, Histogram Equalization, Random Blur, Random changes in Brightness/Contrast/Colors/Saturation, Random Downscale and Upscale
- Random Extraction of 10 RGB patches per image

Testing stage:

- ~ 180 RGB patches (128 x 128) are extracted per image, not in random positions. Patches are extracted such to be aligned with the 8 x 8 pixel grid starting from upper-left corner of the image

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: Detector 4

Training stage:

- Real class: ~ 16K real images selected from unrestricted AFHQ-V2-TRAIN, METFACES-U, METFACES
- GAN-generated class: ~ 24K synthetic generated version of AFHQ-V2, METFACES by means of StyleGAN-2, StarGAN-v2
- Training Augmentations applied on *Patches*: Horizontal Flip, Vertical Flip, Random 90-deg-Rotation, Histogram Equalization, Random Blur, Random changes in Brightness/Contrast/Colors/Saturation, Random Downscale and Upscale
- Random Extraction of 10 RGB patches per image

Testing stage:

- ~ 180 RGB patches (128 x 128) are extracted per image, not in random positions. Patches are extracted such to be aligned with the 8 x 8 pixel grid starting from upper-left corner of the image

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Method Description: Detector 5

Training stage:

- Real class: 100K real images selected from unrestricted FFHQ-U, FFHQ
- GAN-generated class: ~ 170K synthetic generated version of FFHQ by means of StyleGAN-2, Taming Transformers, FaceVid2Vid, Score-based models.
- Training Augmentations applied on *Patches*: Horizontal Flip, Vertical Flip, Random 90-deg-Rotation, Histogram Equalization, Random Blur, Random changes in Brightness/Contrast/Colors/Saturation, Random Downscale and Upscale, JPEG Compression (quality factors from 30 to 100, applied with probability 70%)
- Random Extraction of 1 RGB patch per image

Testing stage:

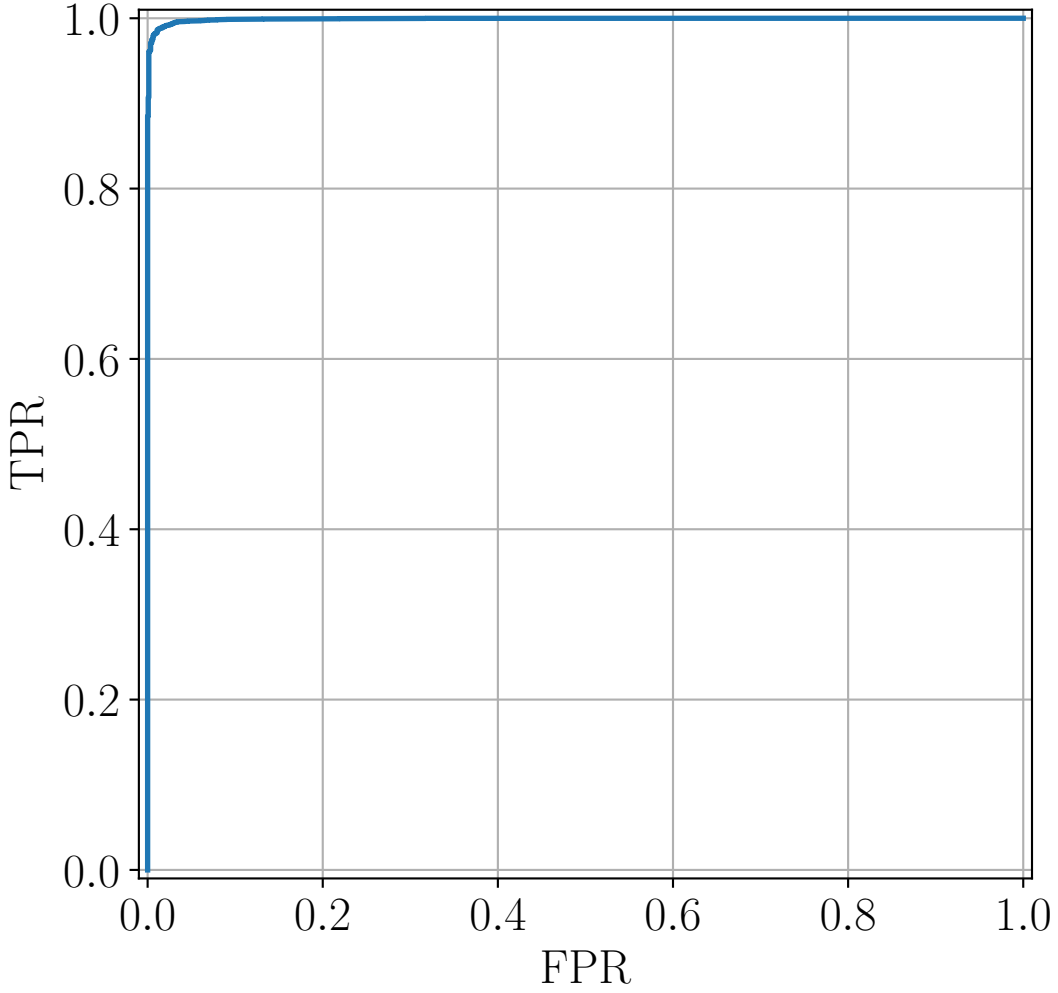
- ~ 180 RGB patches (128 x 128) are extracted per image, not in random positions. Patches are extracted such to be aligned with the 8 x 8 pixel grid starting from upper-left corner of the image

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 1. AFHQ-V2-TEST vs STYLEGAN3-R-AFHQV2

AUC: 0.999103



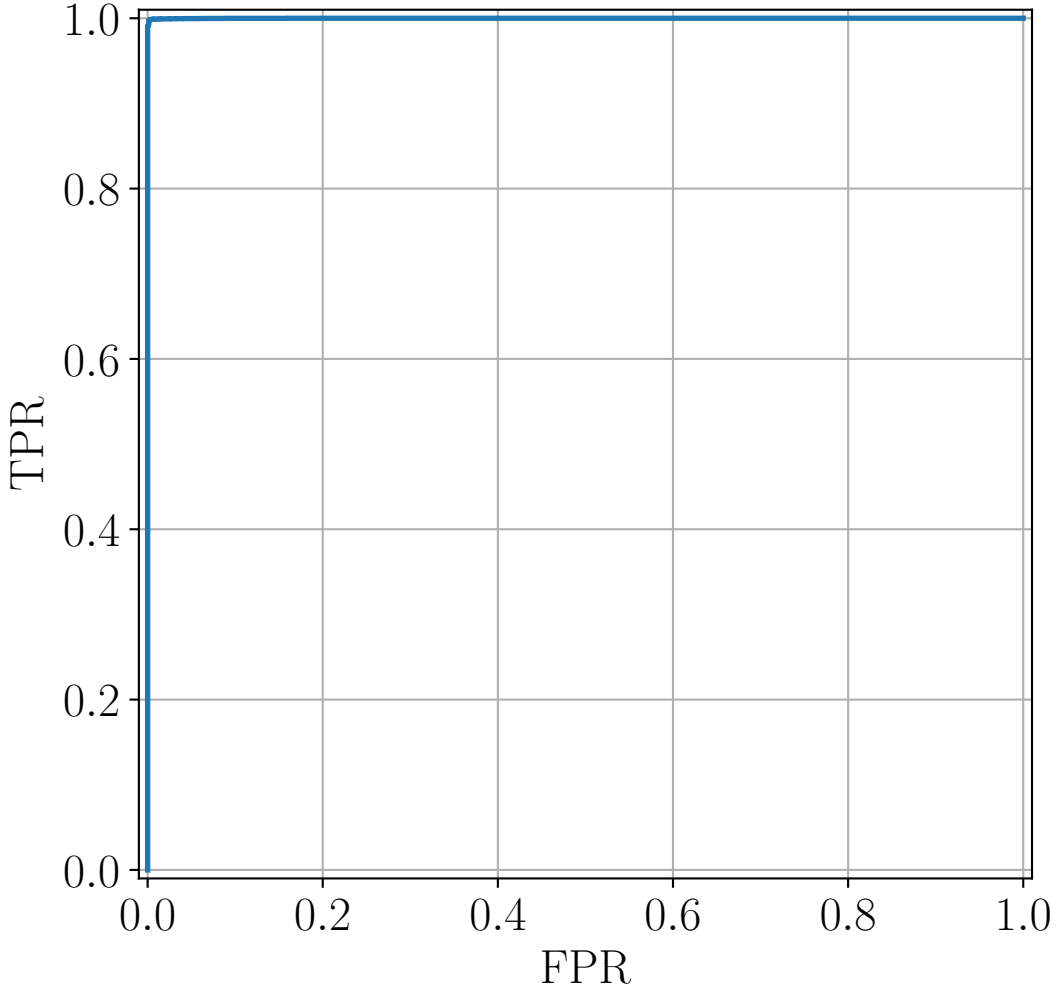
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 2. AFHQ-V2-TEST vs STYLEGAN3-T-AFHQV2

AUC: 0.999936



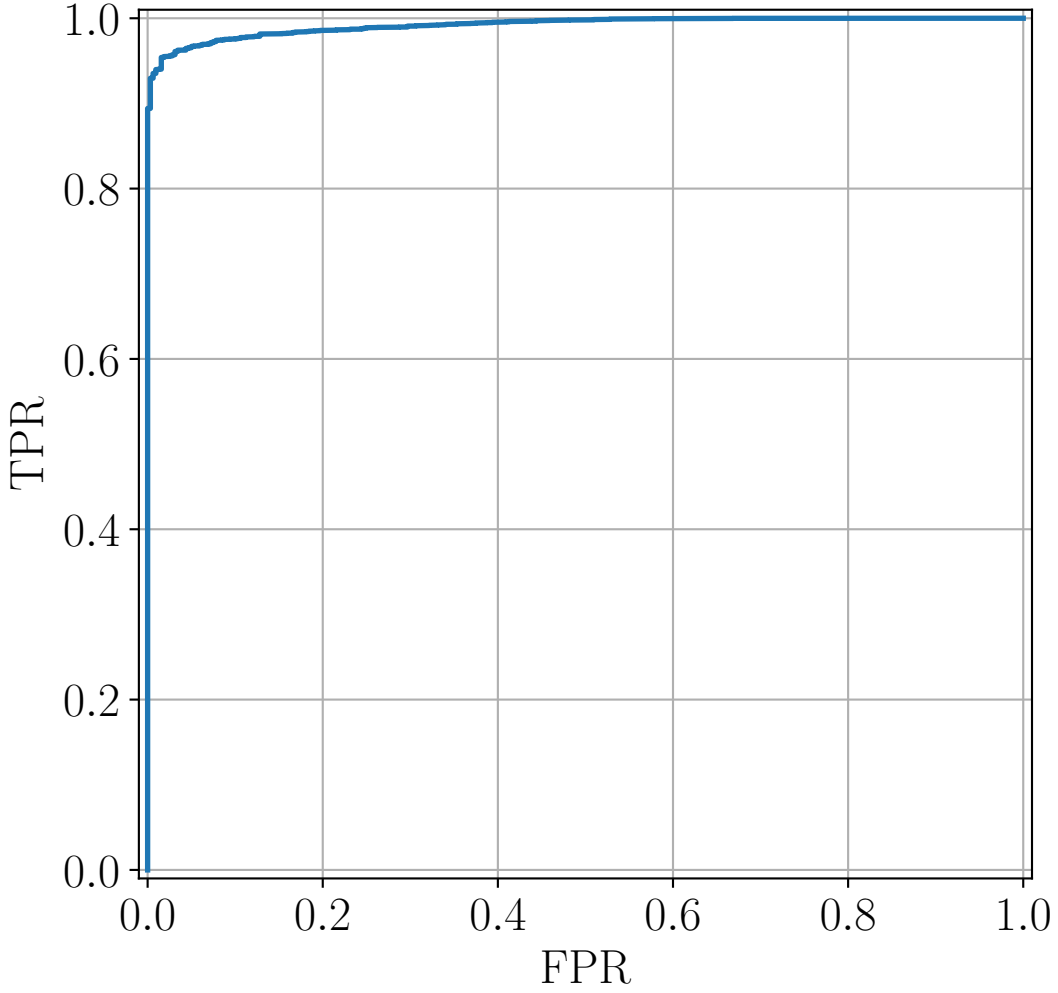
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 3. METFACES-IN-THE-WILD-TEST vs STYLEGAN3-R-METFACESU

AUC: 0.991906



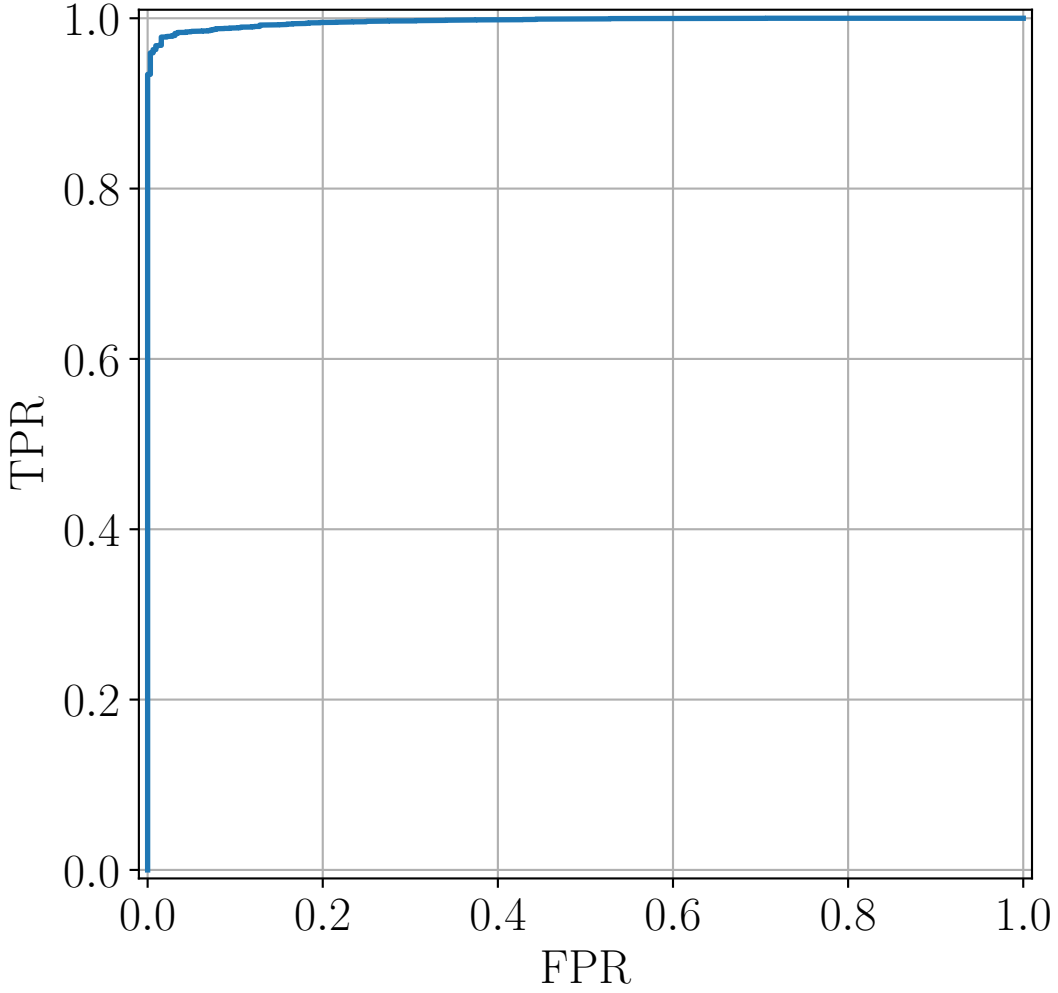
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 4. METFACES-IN-THE-WILD-TEST vs STYLEGAN3-T-METFACESU

AUC: 0.996449



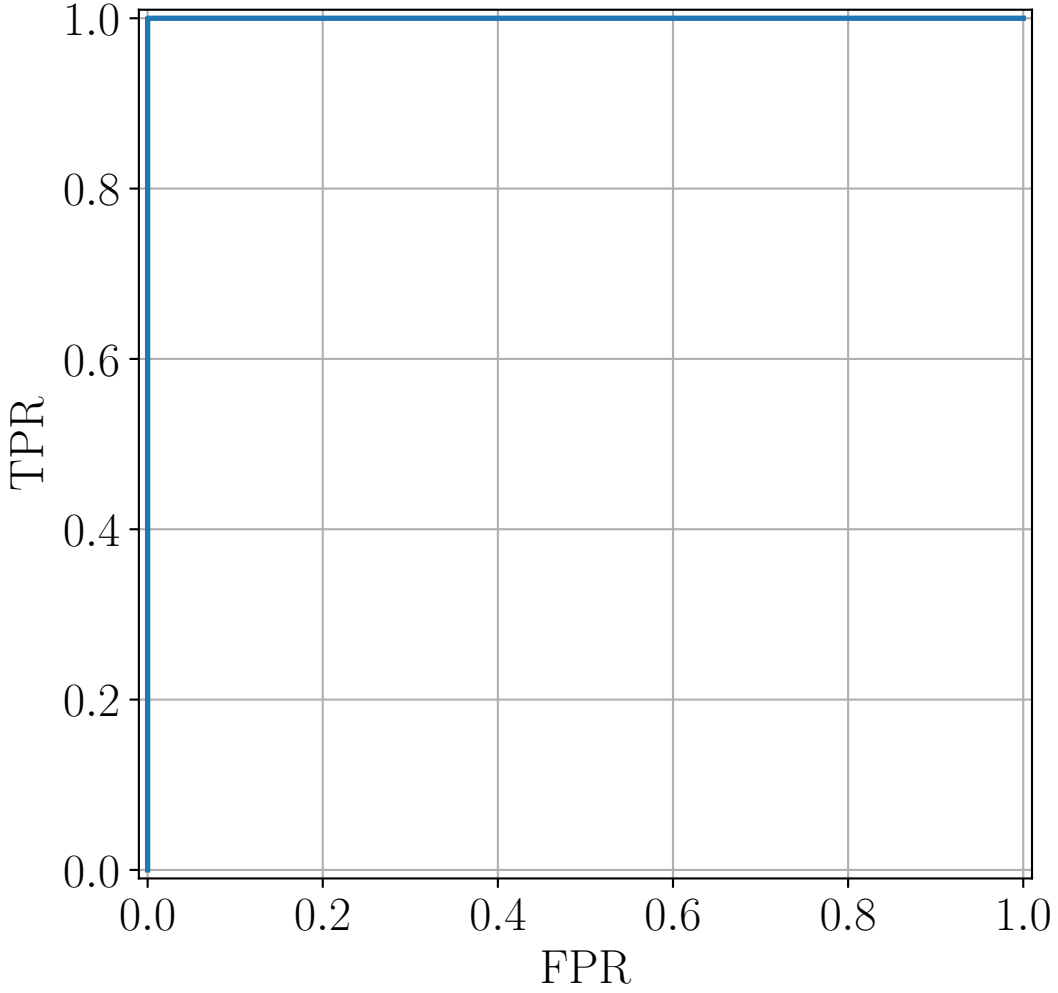
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 5. FFHQ-IN-THE-WILD-20K vs STYLEGAN3-R-NO-COMPRESSION

AUC: 0.999999



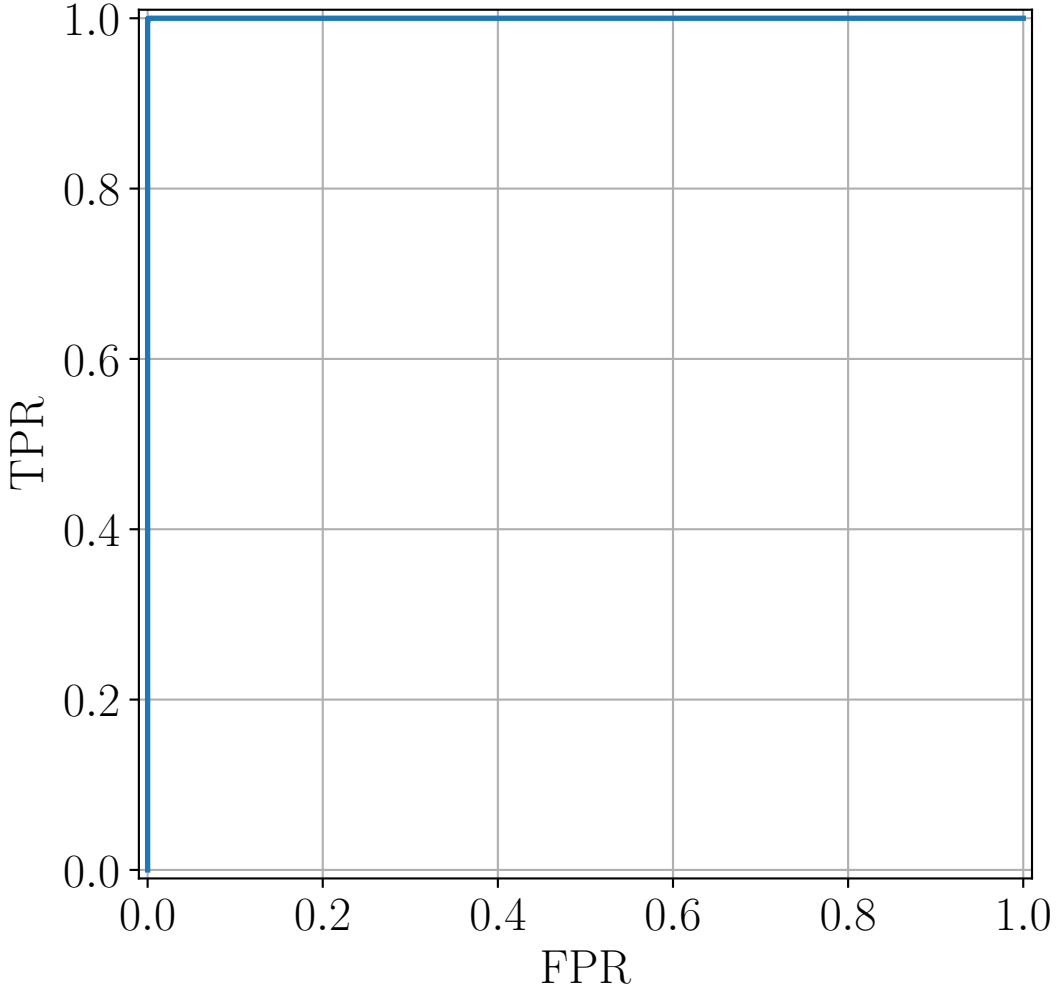
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 6. FFHQ-IN-THE-WILD-20K vs STYLEGAN3-T-NO-COMPRESSION

AUC: 0.999999



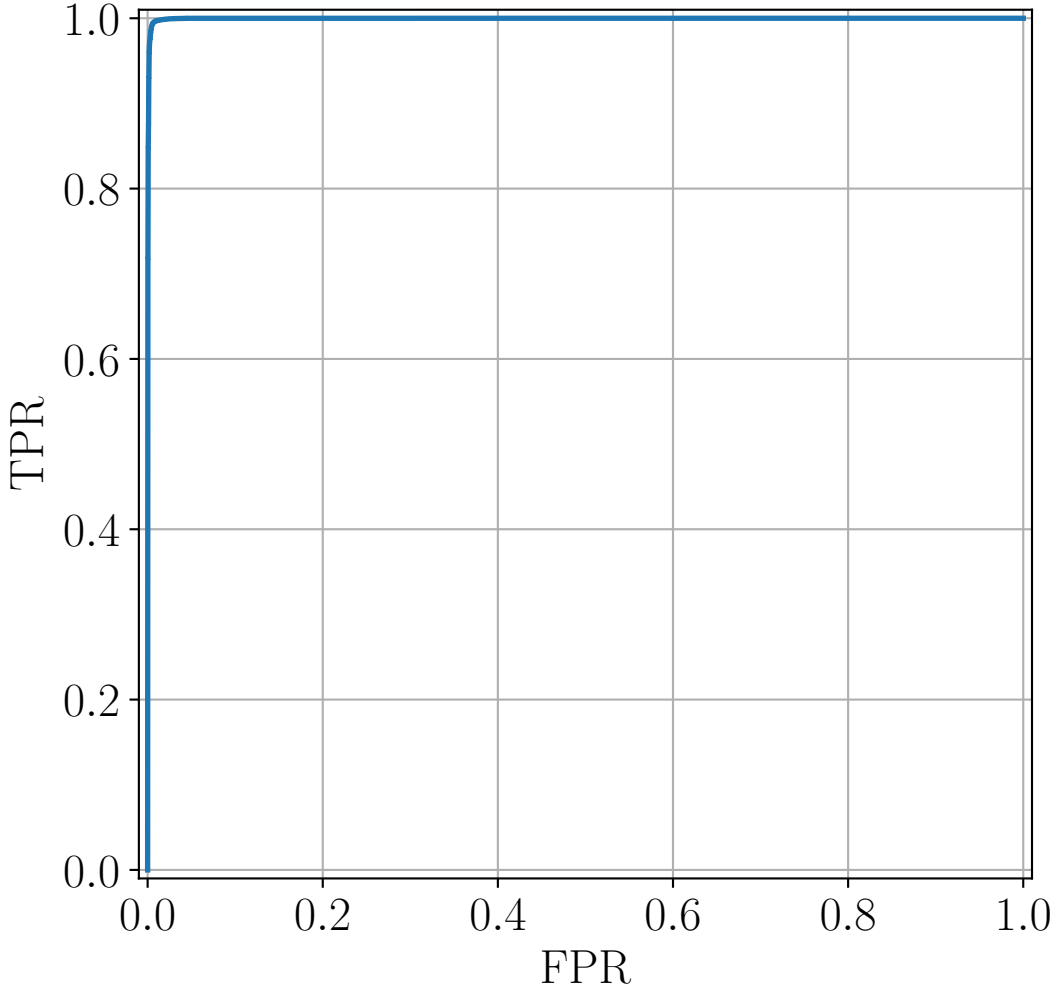
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 7. FFHQ-IN-THE-WILD-20K vs STYLEGAN3-R-COMPRESSION

AUC: 0.999582



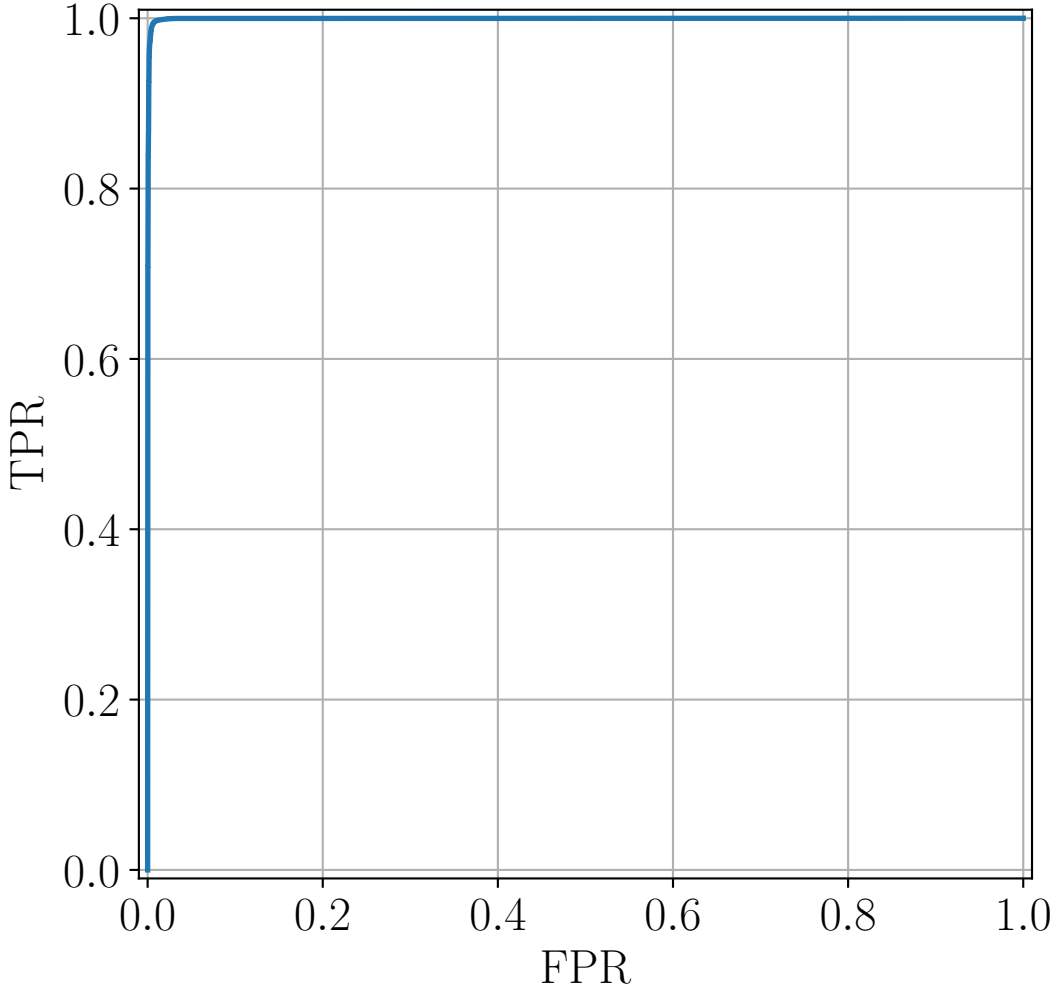
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: 8. FFHQ-IN-THE-WILD-20K vs STYLEGAN3-T-COMPRESSION

AUC: 0.999566



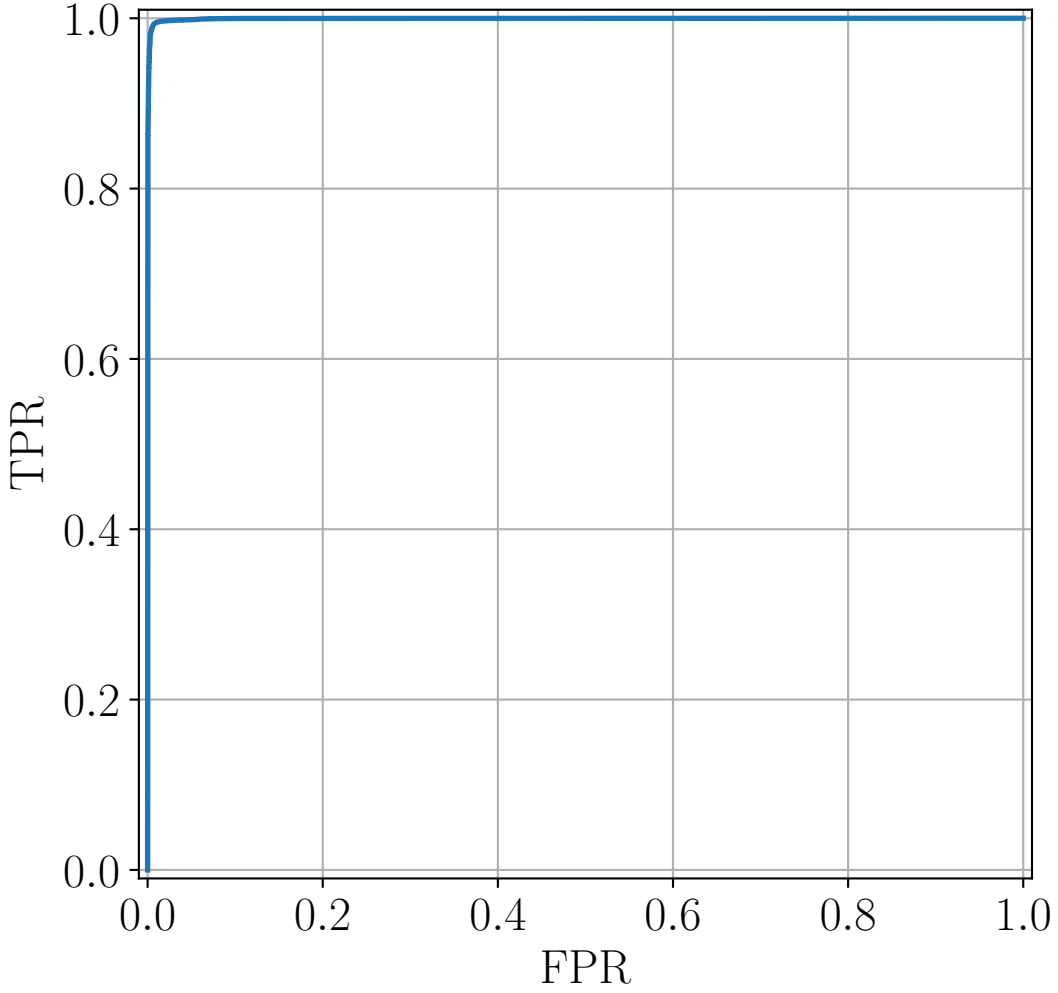
!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

ROC Curves: GLOBAL

AUC: 0.999561



!!! No StyleGAN3 images seen in training

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab

Project Page:

Try out our GAN-generated Image Detector on your query image:

https://github.com/polimi-ispl/GAN-image-detection

POLITECNICO MILANO 1863

ISPL Image and Sound Processing Lab