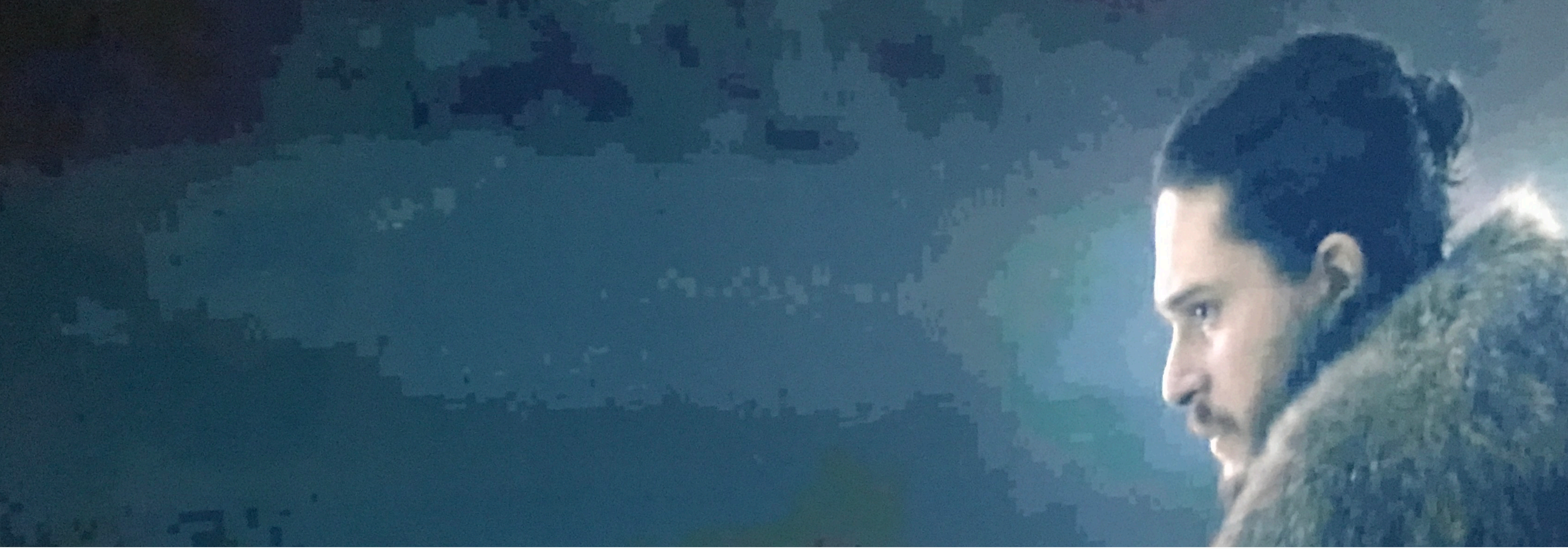
Improving Quality of Compressed Video Using GAN

Lorenzo Seidenari - Assistant Professor
University of Florence, Italy



Google

why the long night

why the long night **was bad** Remove

why the long night **was good**

why the long night **was so dark**

why the long night **was terrible**

why the long night **was a disappointment**

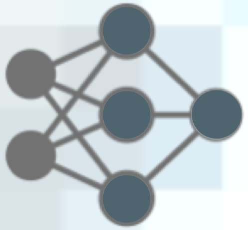
Why does this happens? (in 2019…)

- First Episode of Season 2 had 15M viewers
- Stream this at reasonable quality at a cost of 0,020 $/GB*

-1,125,000 $

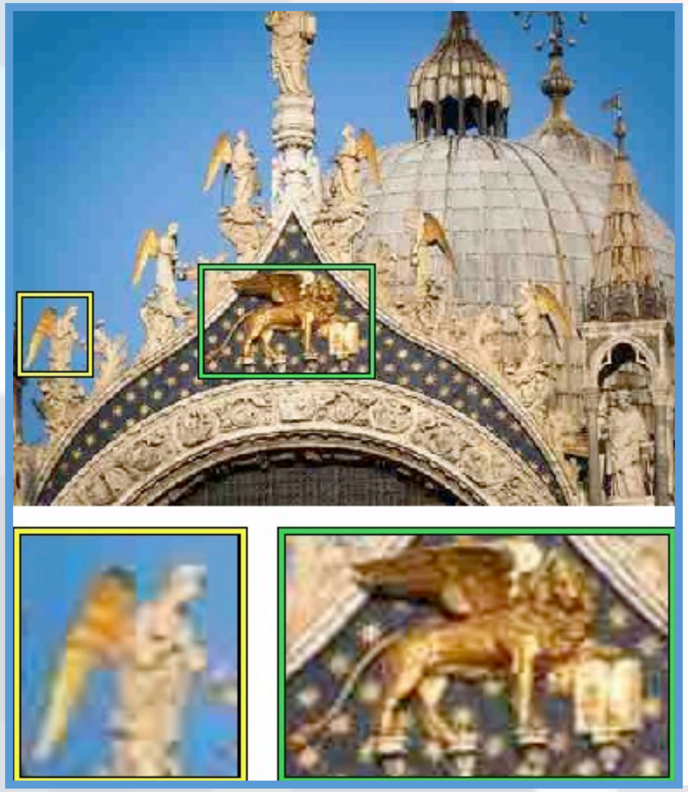
*Amazon Cloud Outbound Traffic

How can we fix it



Deep CNN

Improving Compressed Images with GANs



x_{LQ}

$G(x_{LQ})$

Given an uncompressed frame x_{HQ}

$$x_{LQ} = \mathcal{C}(x_{HQ}; \theta)$$

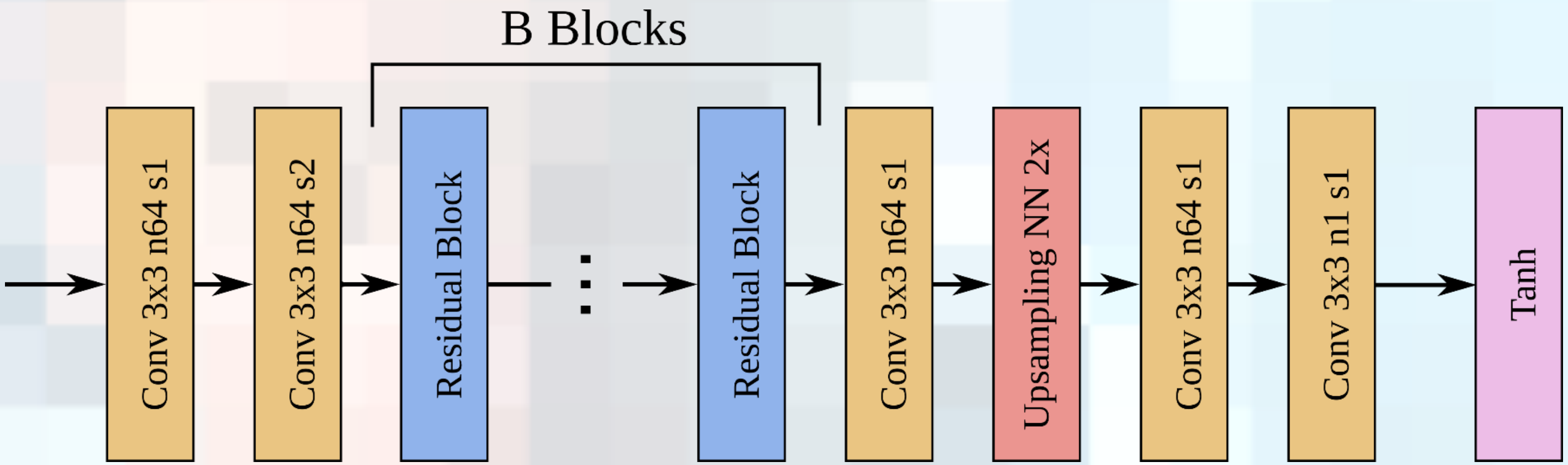
We want to learn a function

$$G(x_{LQ}) \approx \mathcal{C}^{-1}(x_{HQ}; \theta)$$

where θ are codec parameters.

ICCV'17

A Deep Residual Network for Reconstruction



- We use strided convolution to reduce feature map size.
- We avoid checkerboard artifacts with NN upsampling followed by 2 more convolutional layers
- Trained on patches 128x128 pixel extracted from MS-COCO.

Limitations of MSE and SSIM Losses

- SSIM and MSE losses are able to reduce effectively compression artefacts.
- However, reconstructions appear blurry and there are many missing details with respect to the uncompressed version of the image.

JPEG

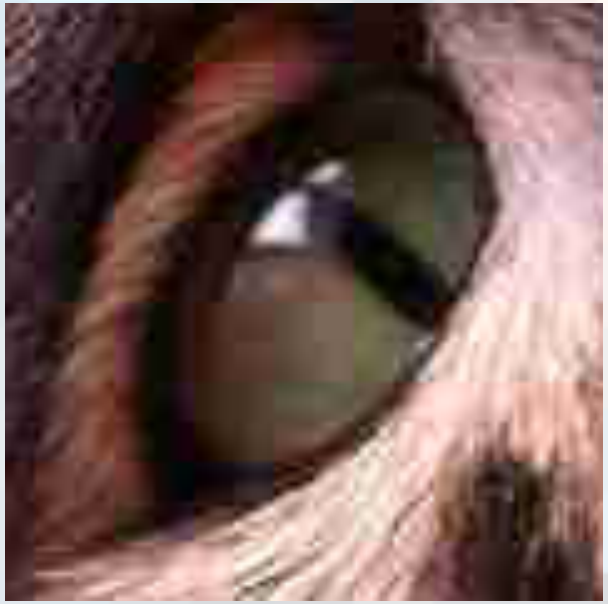
SSIM Loss

GAN

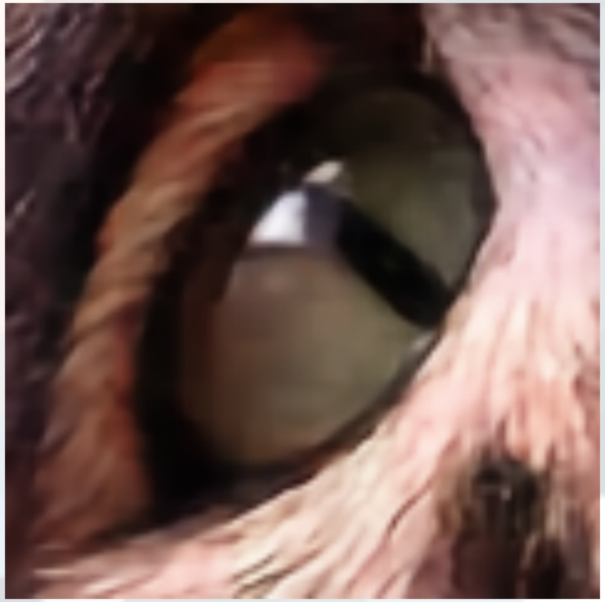
Original

Limitations of MSE and SSIM Losses

- SSIM and MSE losses are able to reduce effectively compression artefacts.
- However, reconstructions appear blurry and there are many missing details with respect to the uncompressed version of the image.

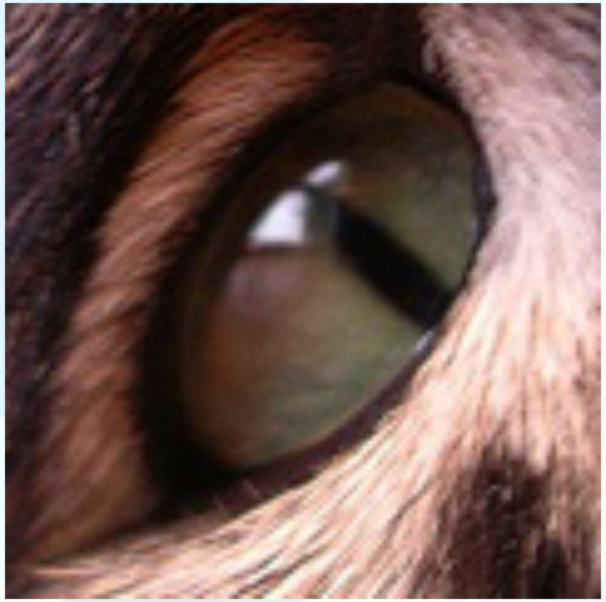


JPEG



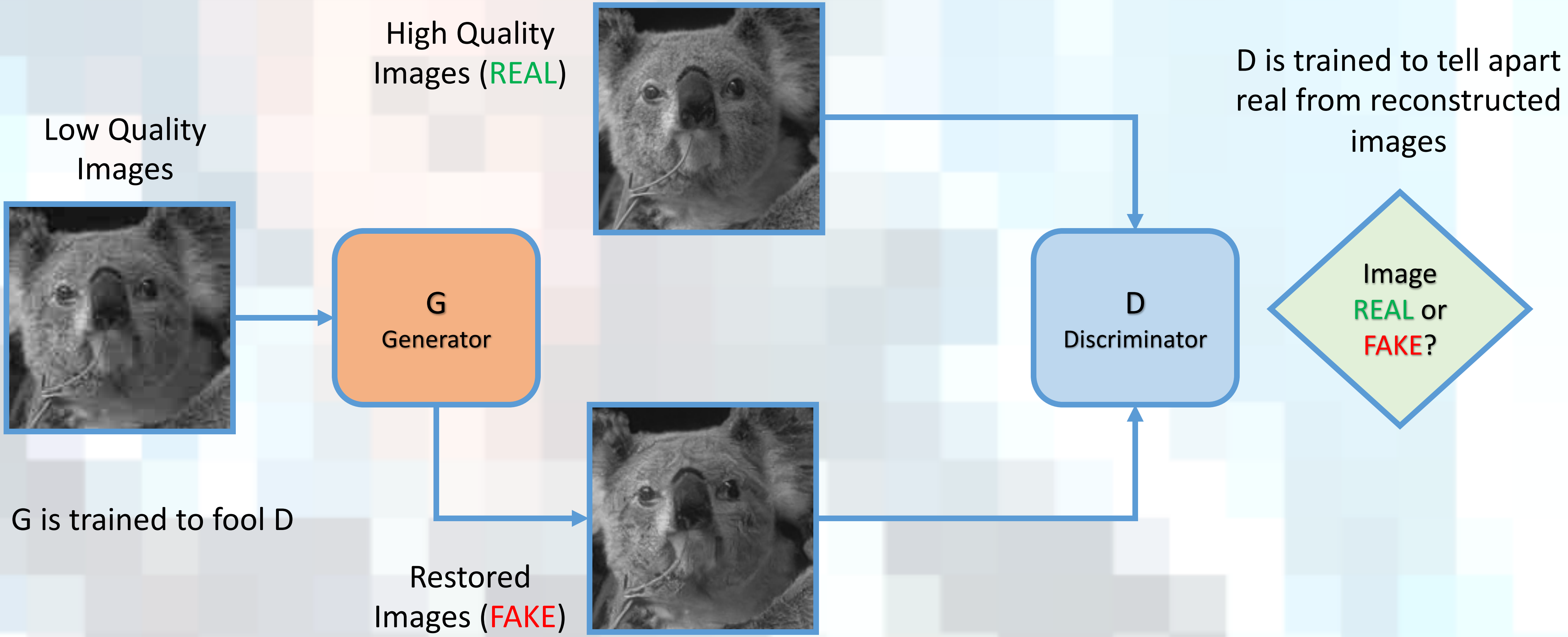
SSIM Loss

GAN



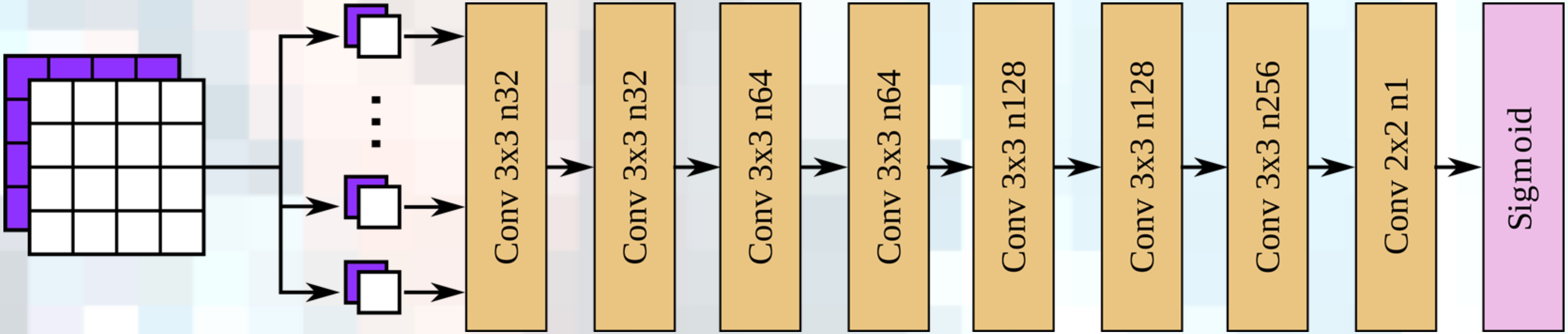
Original

Generative Adversarial Network



ICCV’17

The Sub-Patch Discriminator



- 128 x 128 patches are split into smaller 16x16 sub-patches, concatenated with correspondent input sub-patches and processed by the discriminator.
- The discriminator is trained with a binary cross-entropy loss over all the sub-patches.

ICCV'17

Effect of Sub-Patch Discriminator

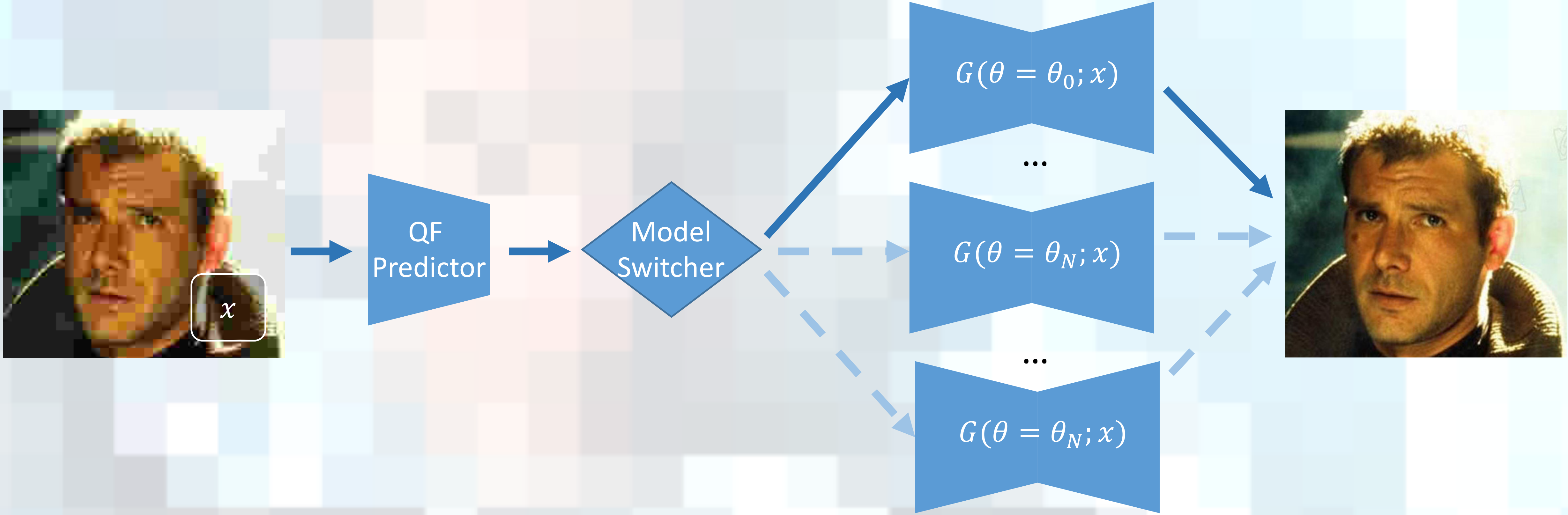
- This technique allows to reduce the mosquito noise present in reconstructions.

W/o Sub-Patch

With Sub-Patch

ICCV’17

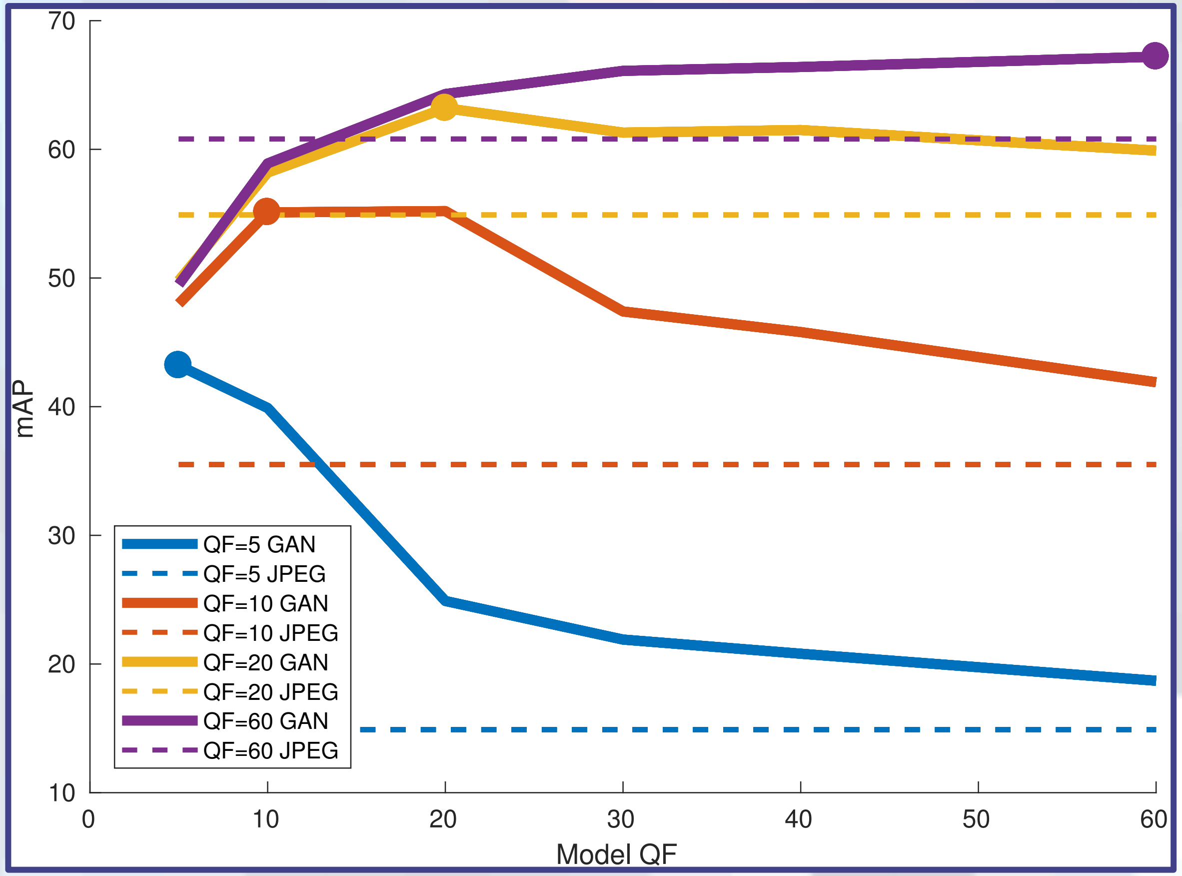
Predicting QF

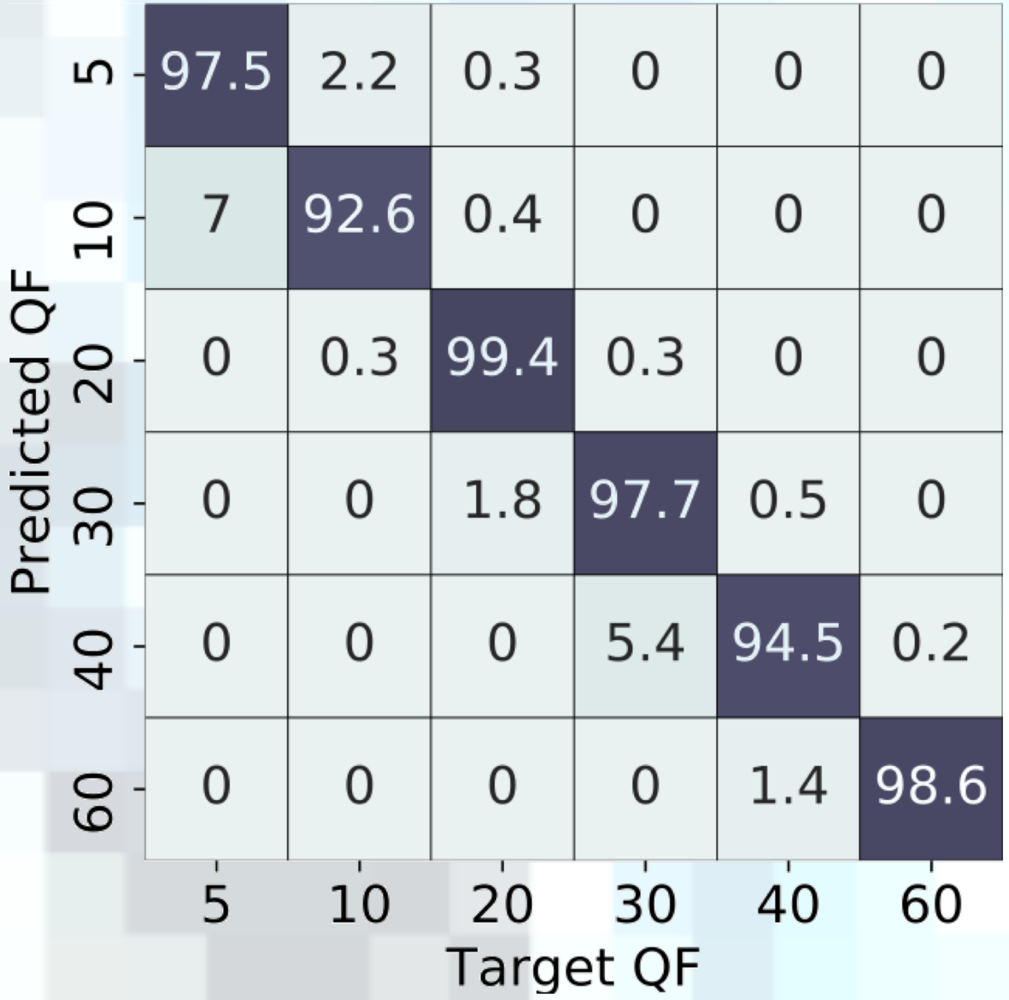


- We train a CNN regressor, named QF predictor, to drive a finite Ensemble of Generators
- We use the most appropriate Generator to restore the image

TMM’19

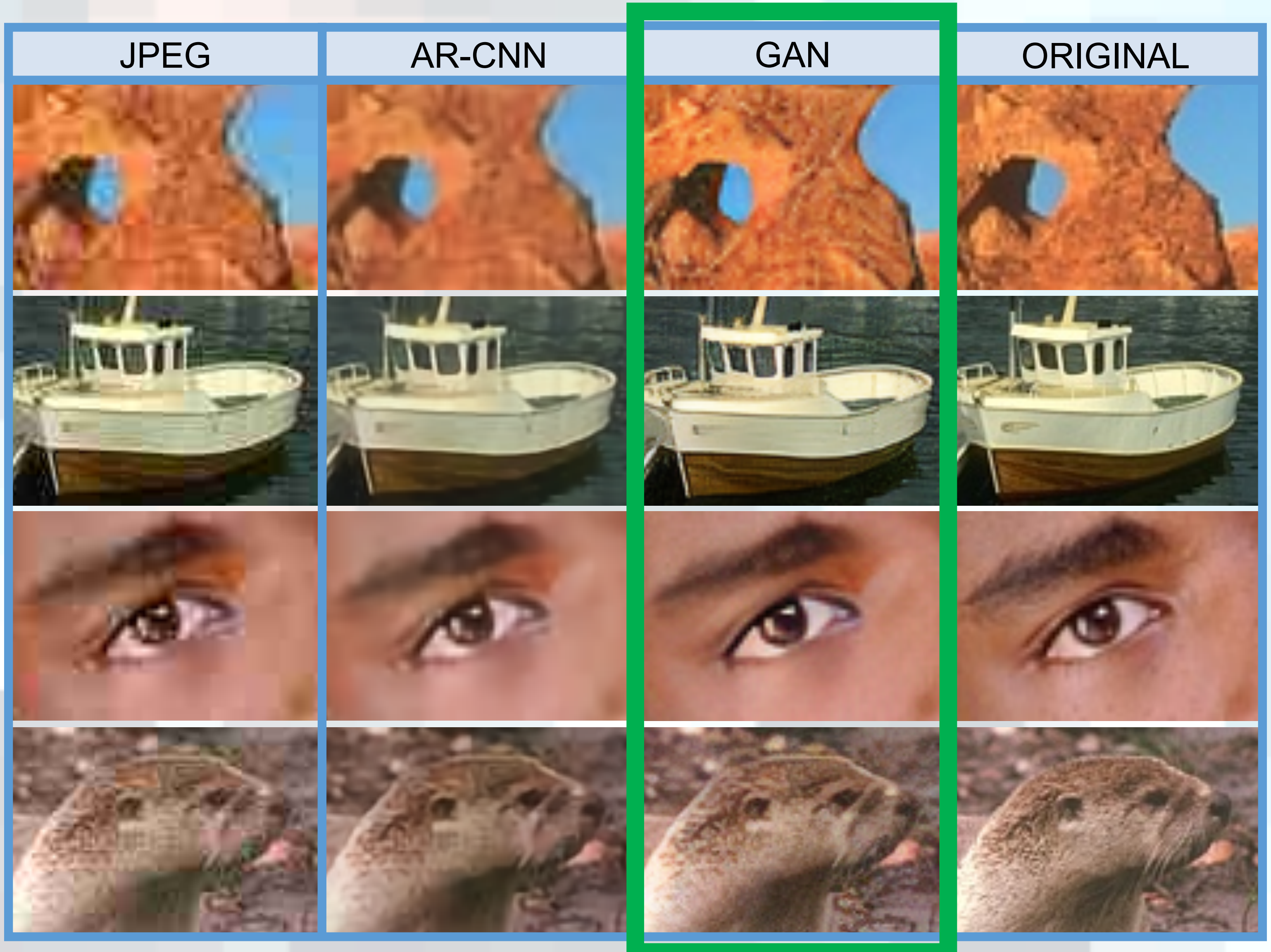
Quality Prediction Results





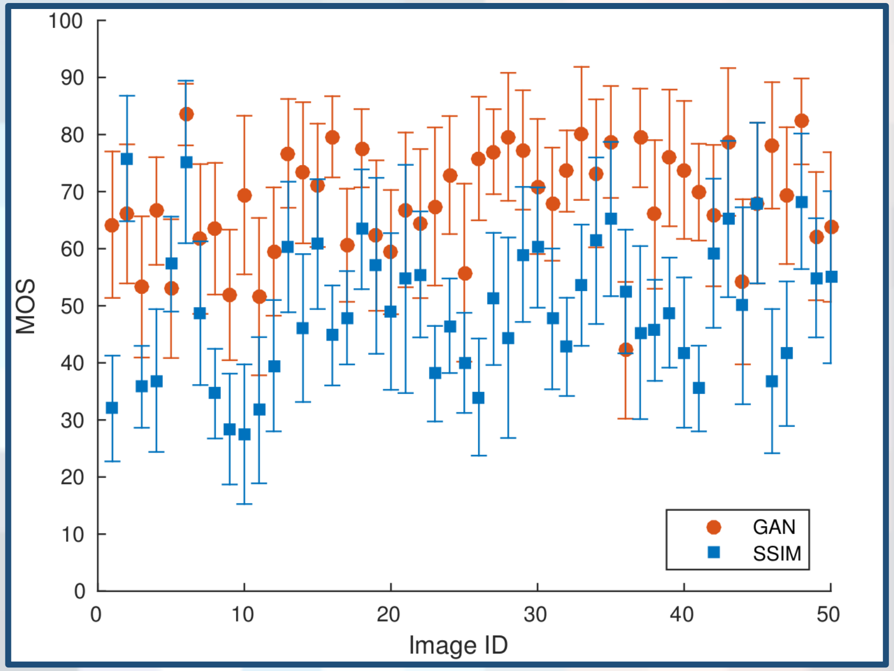
TMM’19

Qualitative Results



TMM'19

Subjective Evaluation



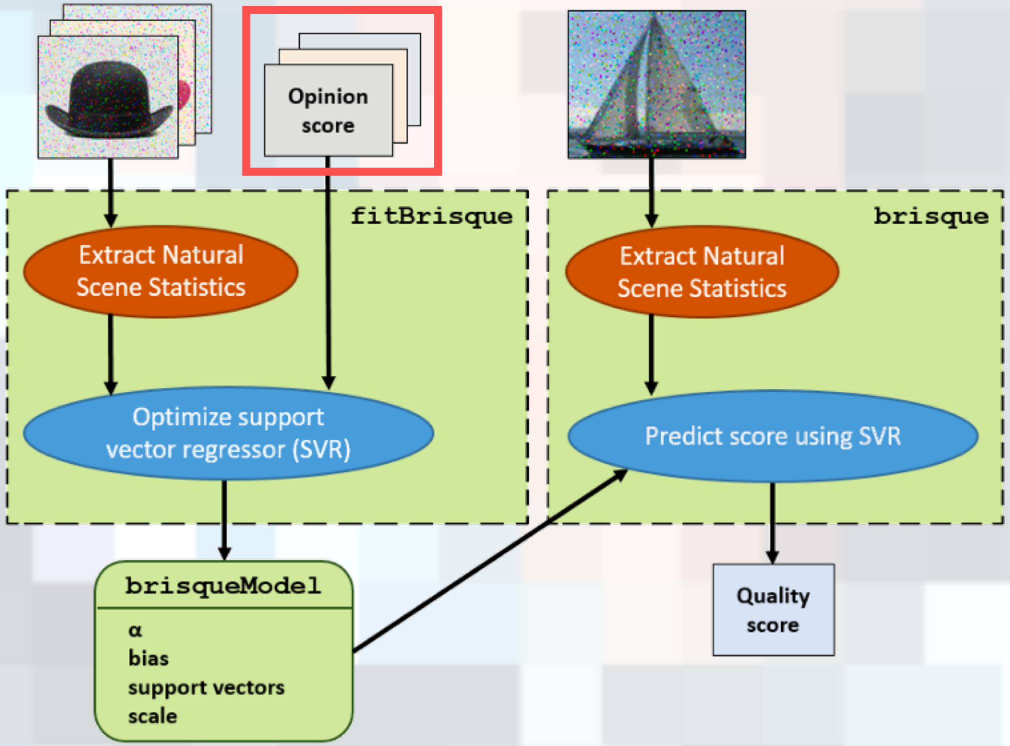
Method	MOS	Std. Dev.
SSIM	49.51	22.72
GAN	**68.32**	20.75

- DSIS setup test image compared to original and similarity scored in 0-100
- We compare SSIM Loss vs Adversarial Training using the same Generator architecture.
- Subjects have a strong preference for GAN restored images over SSIM ones.

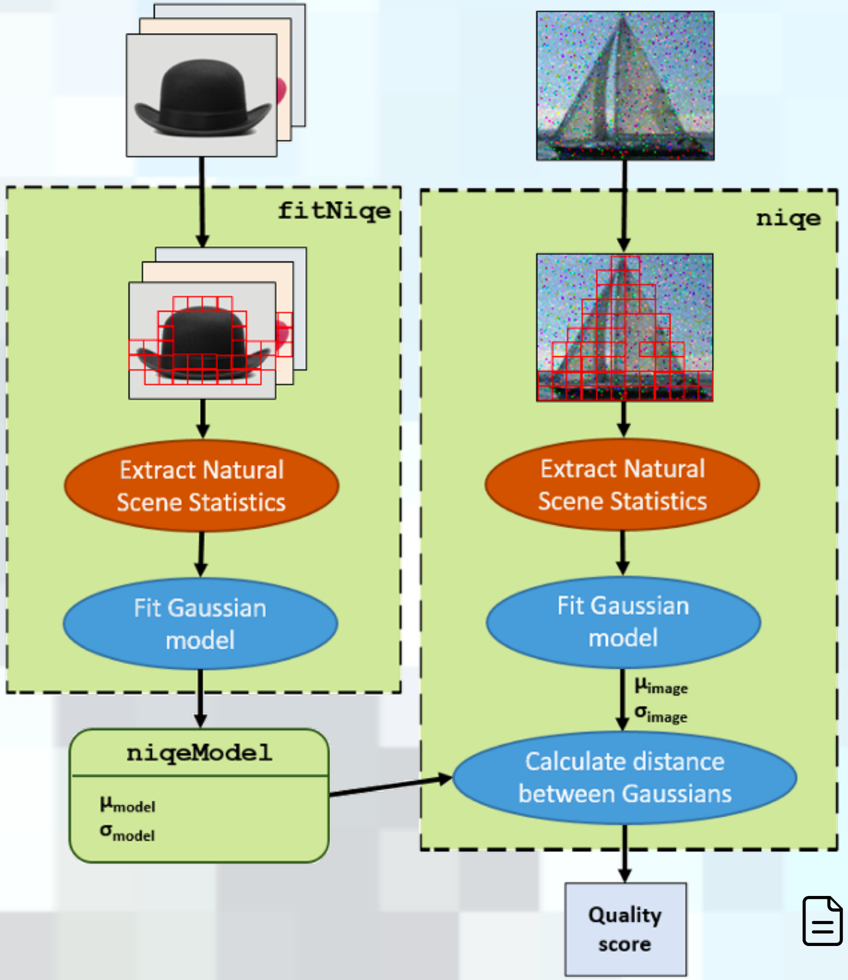
TMM'19

No-Reference Image Assessment

BRISQUE [Mittal'12]



NIQE [Mittal'13]



*workflow diagrams fro www.mathworks.com

No-Reference vs Full Reference

- Our GAN obtains poor scores on Full Reference metrics
- On the other hand NIQE and BRISQUE value GAN images as ’more natural’ the the original ones!

	PSNR	SSIM	NIQE	BRISQUE
JPEG10	**24.8245**	**0.7852**	6.36	53.17
GAN	23.8412	0.7605	**4.27**	**19.65**
ORIG	-	-	4.35	24.32

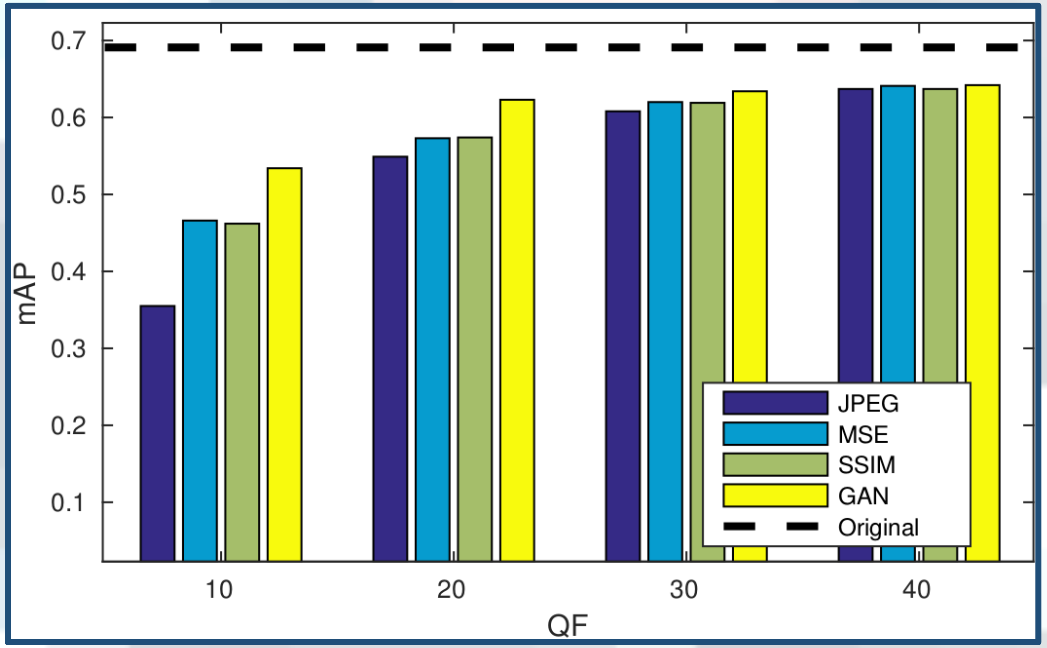
higher is better

lower is better

Under Review

Object Detection Results

- Use an object detector, Faster R-CNN to assess the visual quality of restored images
- Compute mAP on PASCAL VOC using several JPEG quality factors and the correspondent reconstructions.



Class	GAN AP gain @QF 20
Dog	+18.6
Cat	+16.6
Sheep	+14.3
Cow	+12.5

- Large increase in detector performance
- Largest gainers are deformable 'furry' objects such as animals

TMM'19

Evaluation using Language

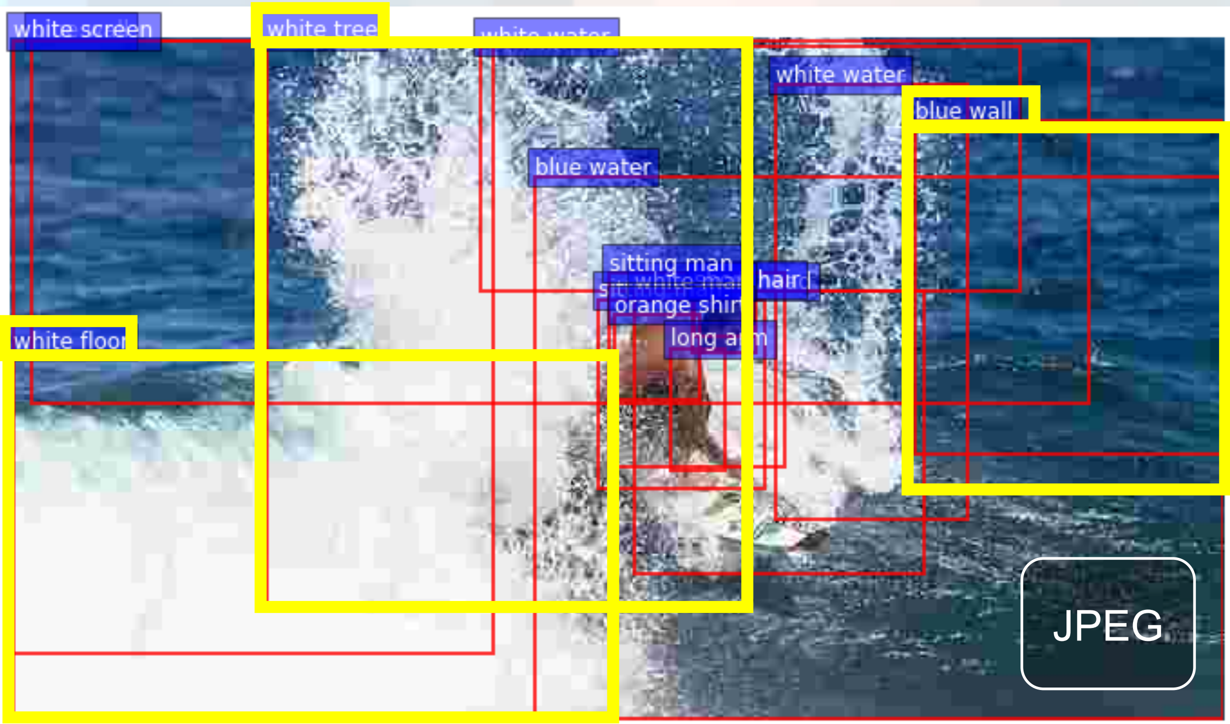
- Use [Anderson et al. CVPR18] image captioning algorithm to evaluate the fine semantics of the image

Quantitative Analysis

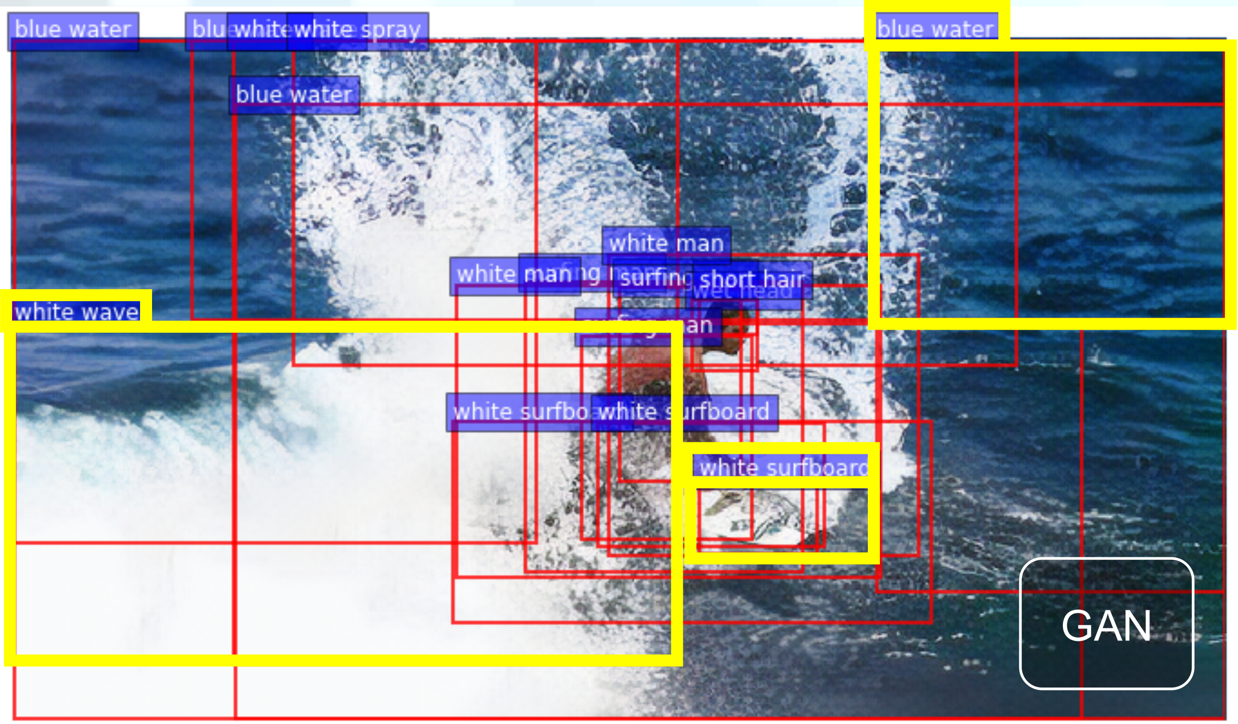
- According all captioning metrics, images enhanced with our GAN are tagged more accurately

	BLEU_1	BLEU_2	BLEU_3	BLEU_4	METEOR	ROUGE	CIDEr	SPICE	VIFIDEL
JPEG	0.685	0.500	0.360	0.250	0.220	0.490	0.810	0.150	0.309
GAN	**0.770**	**0.600**	**0.450**	**0.330**	**0.260**	**0.540**	**1.090**	**0.200**	**0.313**
ORIG	0.800	0.630	0.480	0.360	0.280	0.570	1.200	0.210	0.313

Qualitative Analysis



A couple of people sitting next to a christmas tree.



A man riding a wave on a surfboard in the ocean


Compressed


Restored

Specialized Artifact Removal

- GANs are well known to work well when the distribution is simpler
- Faces are possibly the most interesting object we are willing to transmit
- Here what we can do with a severe degradation and a specialized GAN




Under Review

Specialized Artifact Removal

- GANs are well known to work well when the distribution is simpler
- Faces are possibly the most interesting object we are willing to transmit
- Here what we can do with a severe degradation and a specialized GAN

Under Review

Specialized Artifact Removal

- On a h.264 coded 'talking head' video bandwidth reduced by 150x
- Runs @24 FPS on Iphone X exploiting Neural Engine

Under Review

small pixels

Enable Smart Decompression

Under Review

small
pixels

Enable Smart Decompression

Under Review

Conclusion

- GANs are great for image enhancement. Training allows domain specialization e.g.: faces
- Do not trust standard signal based metrics to evaluate you results
- Humans >> Semantic Tasks > No Reference > Full Reference
- We may in the future see the use of these algorithms to improve user experience

References

- L. Galteri, L. Seidenari, M. Bertini, A. Del Bimbo, 'Deep Generative Adversarial Compression Artifact Removal’, IEEE ICCV 2017
- L. Galteri, L. Seidenari, M. Bertini, A. Del Bimbo, ‘Deep Universal Generative Adversarial Compression Artifact Removal’, IEEE TMM 2019
- L. Galteri, L. Seidenari, M. Bertini, A. Del Bimbo, ‘Towards Real-Time Image Enhancement GANs’, CAIP 2019
- L. Galteri, L. Seidenari, P. Madhyastha, M. Bertini, L. Specia, A. Del Bimbo, ‘Using Language to Evaluate Image Enhancement Tasks’, Language&Vision Workshop@CVPR 2019
- A. Mittal, A. K. Moorthy, and A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. IEEE TIP 2012.
- A. Mittal, R. Soundararajan, and A. C. Bovik. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters 2013
- P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang. Bottom-up and top-down attention for image captioning and visual question answering. CVPR 2018

Thanks

www.small-pixels.com

Thanks

www.small-pixels.com