NOKIA

Tampereen yliopisto
Tampere University

BRIDGING THE GAP BETWEEN IMAGE CODING FOR MACHINES AND HUMANS

Paper number: 3129

Nam Le, Honglei Zhang, Francesco Cricri, Ramin G. Youvalari, Hamed R. Tavakoli, Emre Aksu, Miska M. Hannuksela, Esa Rahtu

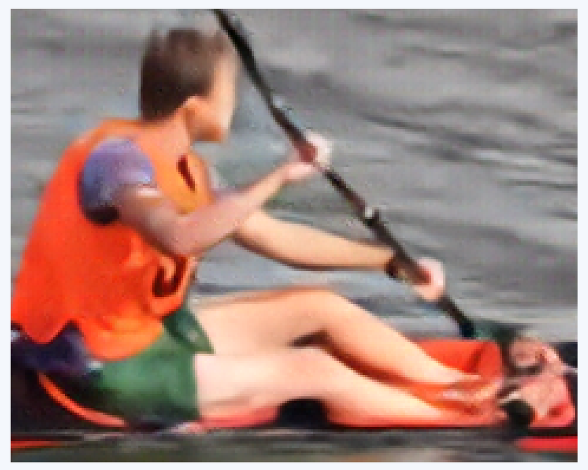
2022 IEEE International Conference in Image processing (ICIP 2022)

OVERVIEW

We propose a simple scheme to finetune CNN-based image codecs to mitigate the "checker-board" artifacts problem.

When finetuned, the codec effectively removes the artifacts **at no extra cost.** At the same time we show that the machine vision performance could be preserved as the visual quality improved.

Output sample with checker-board artifacts



Output sample after codec finetuning

BACKGROUND

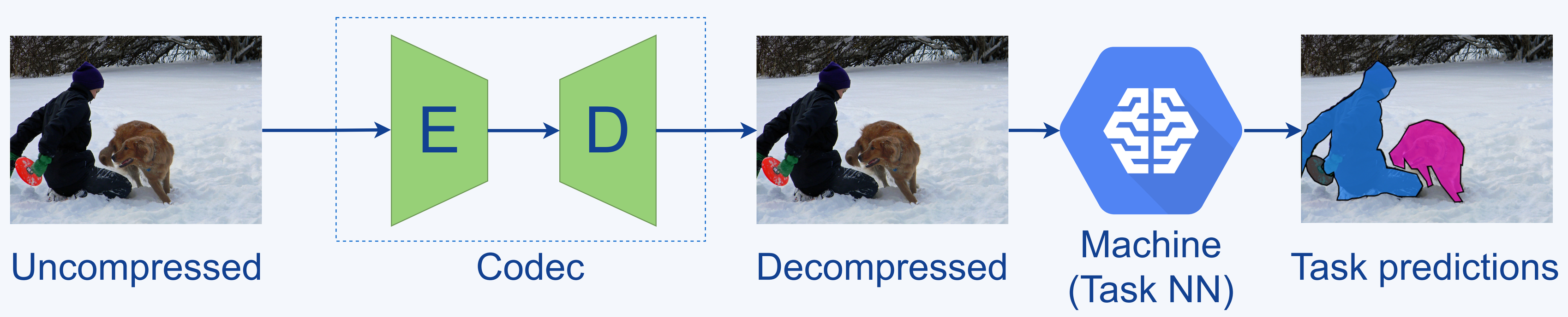


Fig.1: Image coding for machines (ICM)

In Image coding for machines, the primary objective is not pixel fidelity but the machine vision performance. Although these qualities, in many cases, have a strong correlation, in many other they do not.

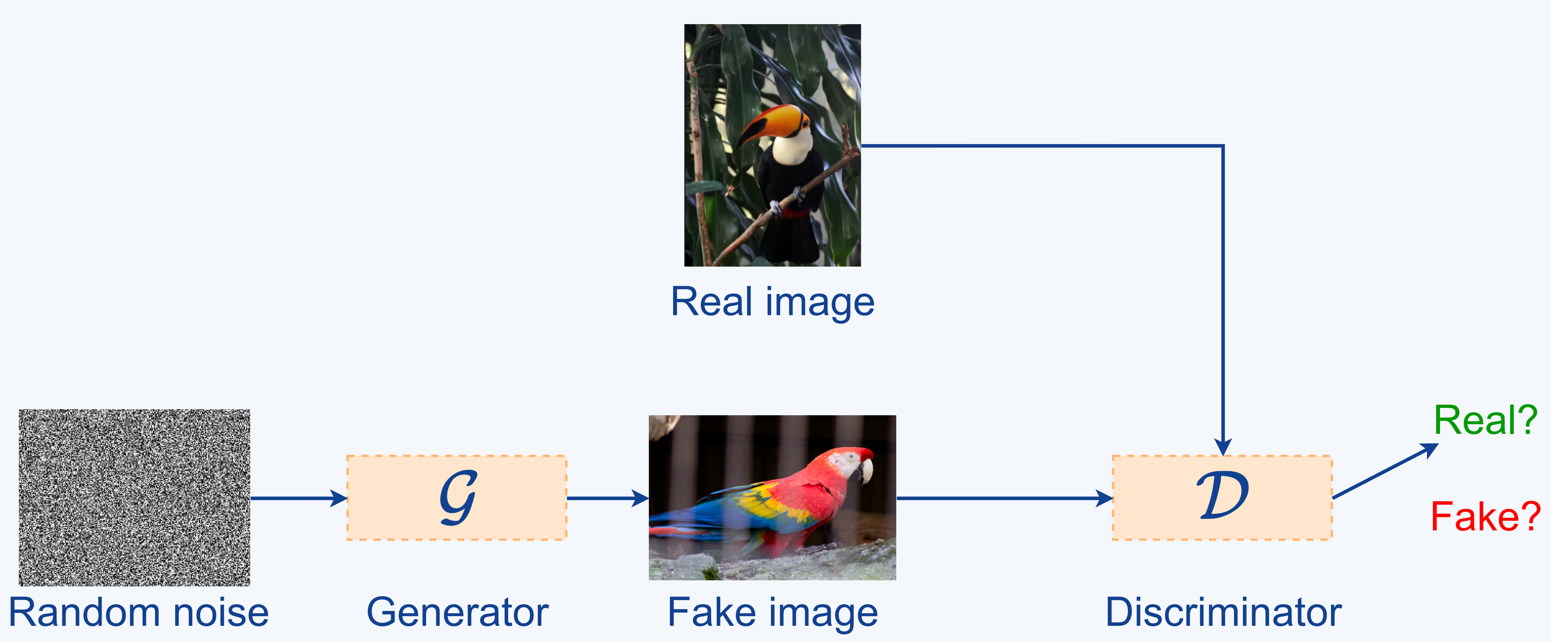


Fig.2: Generative Adversarial Networks (GANs) training

The Generator is trained to fabricate fake images from random noise that resemble samples from a distribution of real images of the training data, while Discriminator's job is to correctly categorize the fake and real images. This adversarial training eventually helps the Generator to learn the distribution of real images.

OUR METHOD

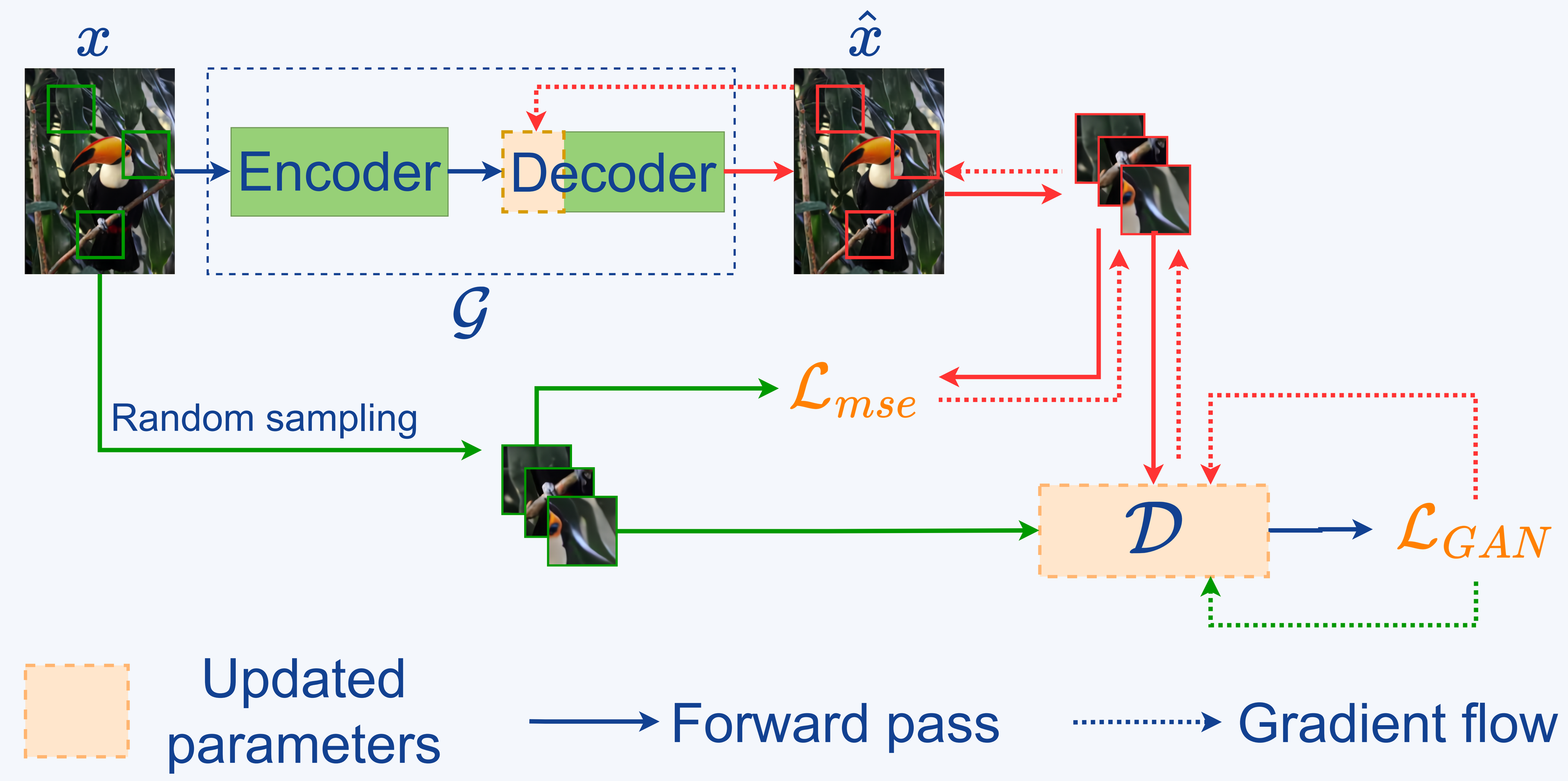


Fig. 3: Proposed finetuning method based on PatchGAN

We finetune the first few layers of the decoder in a PatchGAN-based scheme, where the ICM codec plays the role of the generator, taking uncompressed image as input. Doing this makes our method:

1. a decoder-related technique,
2. finetune only a fraction of parameters of the ICM codec,
3. low computational cost using small image patches,
4. once finetuned, requires no extra computational resources for inference,
5. and free of random hallucinations often found in GANs training.

Achieving higher fidelity or task performance at no extra cost

The proposed method is adoptable for different objectives:

1. For better visual quality: variations of this method are consistently effective against the artifacts, making the outputs more visually appealing to humans.

2. For higher task performance: when a machine is the primary consumer, such as in the case of Image coding for machines (ICM), artifacts are tolerable. We propose a "Low Impact" option where the adversarial dynamics are attenuated, keeping high task performance while still enhancing all of the other metrics.

Table 1. Benchmarks on task instance segmentation (mAP) and other metrics. Configuration with low adversarial impact settings are marked with "*LI*"

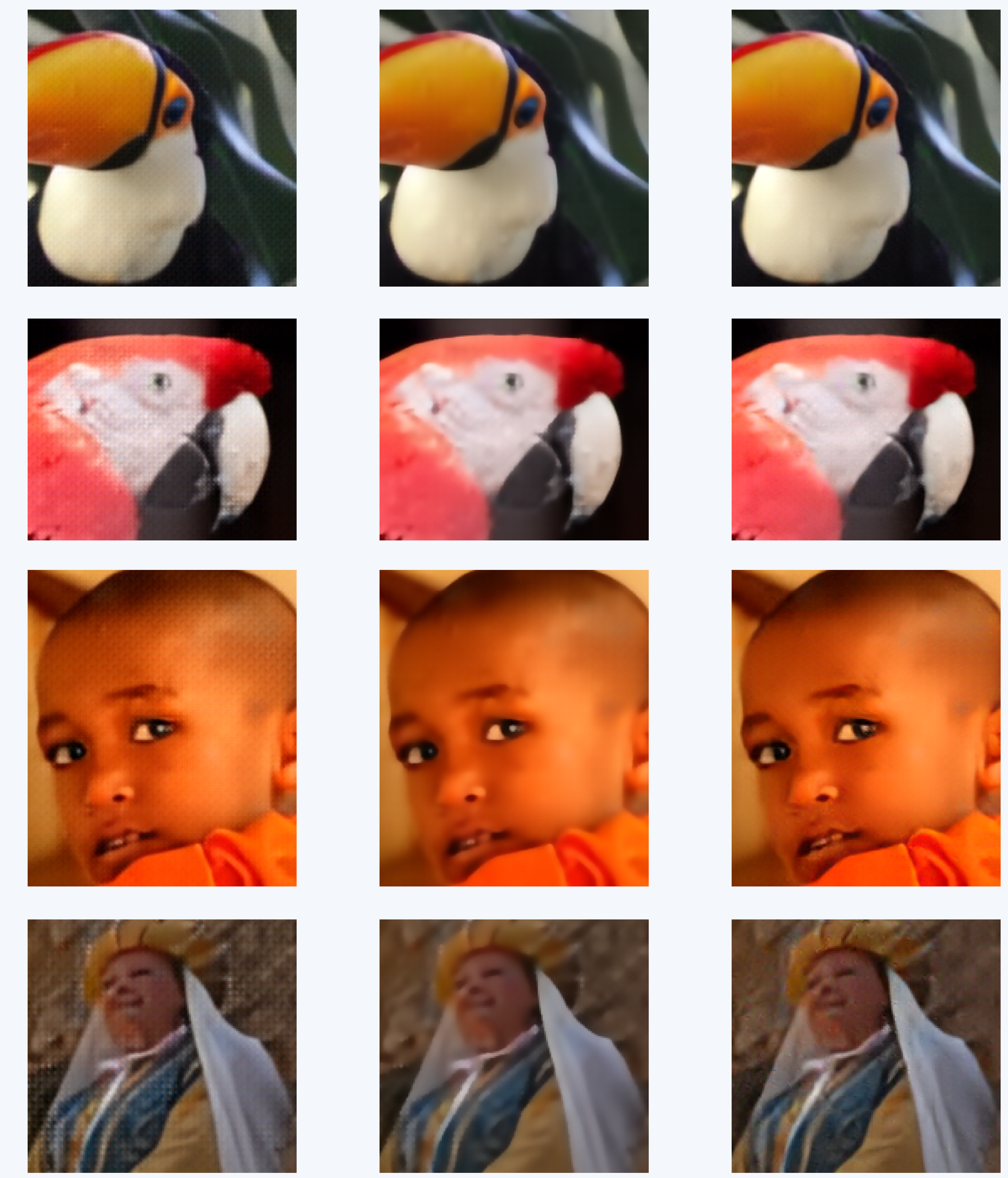
Configuration	mAP↑	PSNR↑	SSIM↑	VGG↓
Base codec	**0.766**	27.695	0.709	0.495
3 patches (64x64)	0.752	28.526	0.764	0.497
3 patches (32x32)	0.750	28.316	0.740	0.537
5 patches (64x64)	0.749	**28.540**	**0.765**	0.499
1 patch (64x64)	0.748	28.518	0.763	0.502
3 patches (128x128)	0.754	28.498	0.763	0.502
3 patches (128x128), *LI*	**0.766**	28.129	0.726	**0.469**

RESULTS

Complete removal of the "checker-board" artifacts

Below are the input and output examples of our method in comparison with the outputs of bilateral filter as the post-processing step.

Base ICM codec | ***Bilateral post-processing*** | ***Finetuned model***



For more output samples please visit our landing page:

https://flysofast.github.io/human-finetuned-icm/