universität innsbruck

Do Not Trust Your Eyes

The Semantic Pitfalls of Modern Image Compression

Nora Hofer

Engineering Kiosk Alps Meetup · Innsbruck, Austria · 16 January 2025

Hello!

universität
innsbruck

Research interests:

Security & Privacy Lab Group hike to Viggarspitze, Sept. 2023.
Photo by Benedikt Lorch: Group hike to Viggarspitze, Tyrol, Austria, September 2023.

Digital Image Forensics

Methods for the verification of **image authenticity**, **source attribution**, and the detection of **traces of manipulation**.

Image content

Statistical image properties

2013 Boston Marathon Bombing

One of the suspects, captured by a bystander’s cellphone.

0.2% of all pixels were used to identify the suspect.

Can we rely on digital images if **neural compression** is the default ?

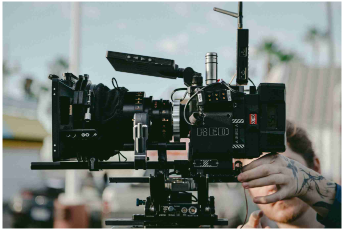
United States Attorney’s Office District of Massachusetts (`https://www.justice.gov/usao-ma/tsarnaev-exhibits-day-2`)

Neural Image Compression

Operators of the lossy compression pipeline are replaced with **learnable elements**.

Neural compression achieves improved **compression rates** at **high quality**.

JPEG



93.6 kB

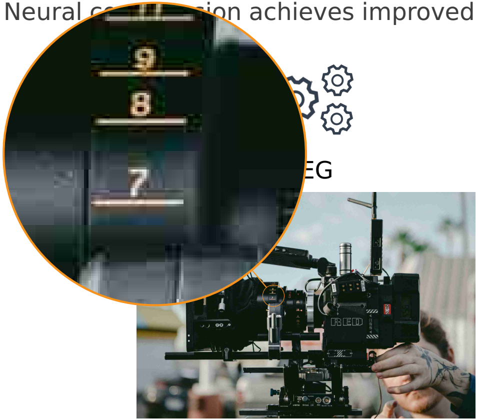
Neural compression

31.2 kB

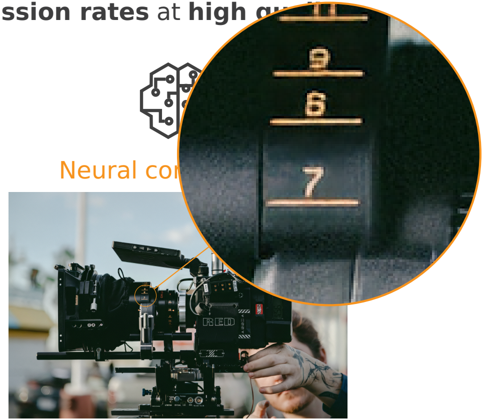
Neural Image Compression

Operators of the lossy compression pipeline are replaced with **learnable elements**.

Neural c… ion achieves improved **compression rates** at **high q**…



93.6 kB



31.2 kB

Miscompressions

Introduced by neural compression

Neural compression jargon for “decompression”

Verbal description of a human observer

Definition

A reconstruction error that results in a difference between the semantic meaning of an original image and its reconstructed version after neural compression

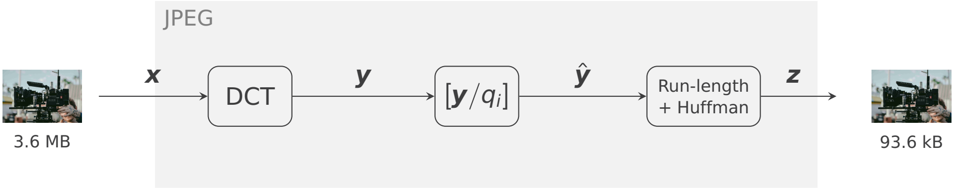
or image detail ($<1\,\%$ of pixels)

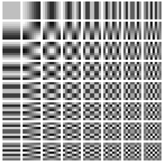
Outline

1. **Primer on neural compression**
2. Our taxonomy of miscompressions
3. Preparing for neural compression

Recall the JPEG Compression Pipeline

Transform | Quantization | Entropy coding





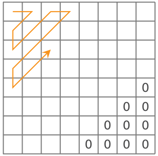
2D base vectors

Horizontal frequencies →

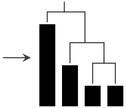
Vertical frequencies ↓

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

quantization table

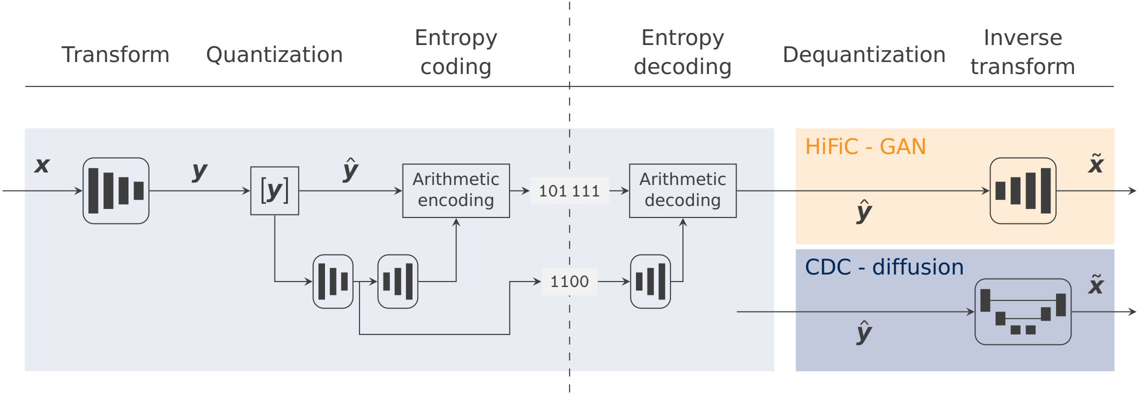


zig–zag arrangement



huffman table

The Neural Compression Pipeline



Ballé, Minnen, Singh, Hwang, and Johnston, “Variational image compression with a scale hyperprior,” in *ICLR*, 2018.
Mentzer, Toderici, Tschannen, and Agustsson, “High-fidelity generative image compression,” *NeurIPS*, 2020.
Yang and Mandt, “Lossy image compression with conditional diffusion models,” *NeurIPS*, 2024.

Outline

1. Primer on neural compression
2. **Our taxonomy of miscompressions**
3. Preparing for neural compression

Method

Manual inspection of the reconstructions of 552 images

Datasets: CLIC2020, DIV2K, Kodak

Neural compression schemes

1. Ballé, Minnen, Singh, Hwang, and Johnston, “Variational image compression with a scale hyperprior,” in *ICLR*, 2018.
2. Minnen and Singh, “Channel-wise autoregressive entropy models for learned image compression,” in *ICIP*. IEEE, 2020.
3. Mentzer, Toderici, Tschannen, and Agustsson, “High-fidelity generative image compression,” *NeurIPS*, 2020.
4. Ballé, Valero, and Eero, "End-to-end optimized image compression." in *ICLR*, 2022.
5. Yang and Mandt, “Lossy image compression with conditional diffusion models,” *NeurIPS*, 2024.

Examples shown in this presentation were produced with

3. HiFiC: Pre-trained GAN; 180 million parameters; intensities: *high, mid, low*

5. CDC: Pre-trained diffusion model; 54 million parameters; optimization ρ: *0, 9*

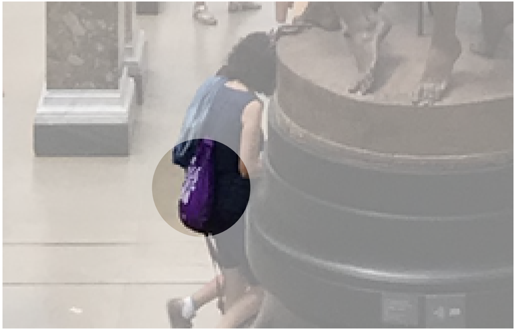
Taxonomy of Miscompressions

Category Amplitude

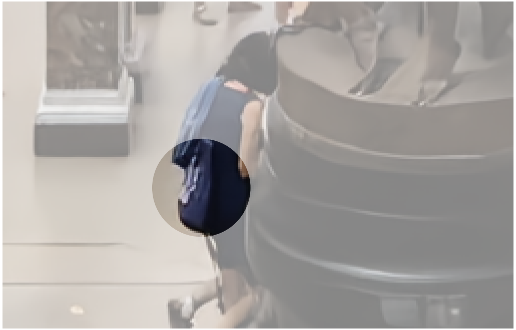
Reconstructions differ in the **amplitude of spatial frequencies** in the signal, affecting attributes such as brightness, color saturation, and the intensity of high frequency components.

Proposal for a Taxonomy

Category Amplitude



Original

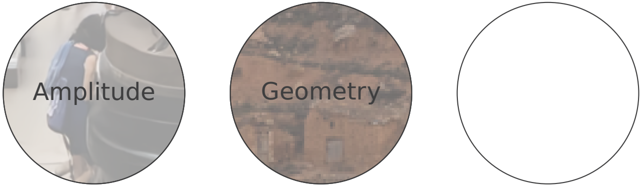


CDC $\rho 0$

Original image 1152×1920. Compressed to 0.17 bpp. Crop: 256×164 (1.89%)

Taxonomy of Miscompressions

Category Geometry



Reconstructions contain **geometric transformations**, such as translation, rotation, scaling, and shearing, including shifted shapes and dissolved contours.

Proposal for a Taxonomy

Category Geometry

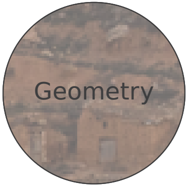
Original

HiFiC lo

Original image 1984×1152. Compressed to 0.18 bpp. Crop: 256×164 (1.84%)

Taxonomy of Miscompressions

Category Shape

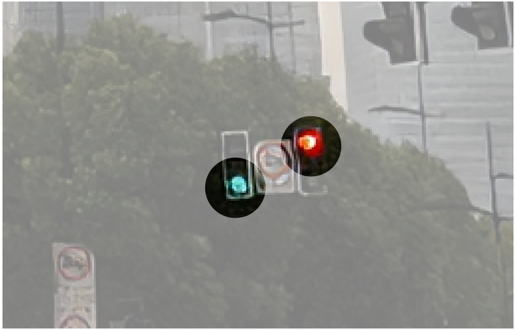


Reconstructions contain changed **contours**.

Proposal for a Taxonomy

Category Shape

Original

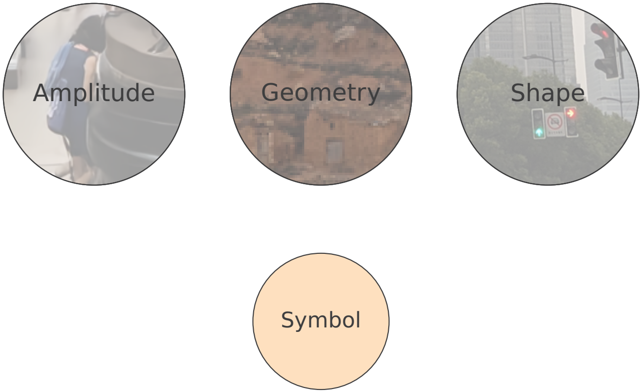


CDC $\rho 0$

Original image 1228×1840. Compressed to 0.15 bpp. Crop: 256×128 (1.86%)

Taxonomy of Miscompressions

Symbol Modifier



Proposal for a Taxonomy

Symbol Modifier

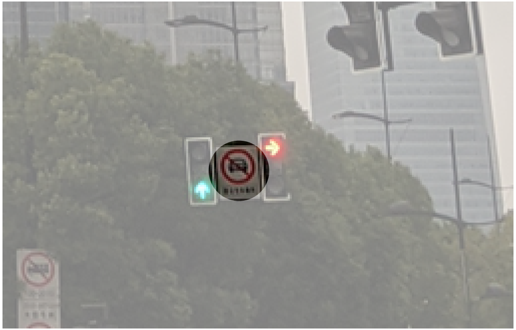
Original

HiFiC lo

Original image 1228×1840. Compressed to 0.15 bpp. Crop: 256×164 (1.85%)

Proposal for a Taxonomy

Symbol Modifier



Original

HiFiC hi

Original image 1228×1840. Compressed to 0.23 bpp. Crop: 256×164 (1.86%)

Outline

1. Primer on neural compression
2. Our taxonomy of miscompressions
3. **Preparing for neural compression**

How to Avoid Miscompressions?

Next steps

1. **Quantify the prevalence** and **identify influencing factors.**

   Needed: Sufficiently large **annotated dataset** of miscompressions.

   Getting the human out of the loop:

   - **OCR models** to detect changes in letters and numbers
   - **Image-to-text models** to compare semantic description of a scene

2. **Tailored detection model** to identify image areas prone to be miscompressed at encoding time

3. Incorporate a **miscompression metric** in the training loss

... in the meantime: We need to deal with the existing risks.

How to Deal with the Risks ?

1. **Document**
   visible watermarks, icons, captions
2. **Annotate** the EXIF data
   JPEG Trust, C2PA
3. **Detect** neural compression

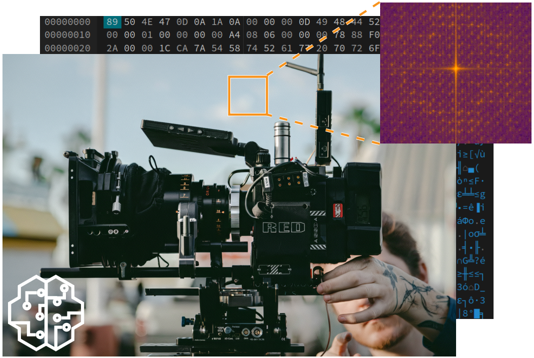


Fig: RED camera. *(Image might contain miscompressions)*

Bergmann et al., “Frequency-domain analysis of traces for the detection of AI-based compression,” in IEEE *IWBF*, 2023.
Bergmann et al., “Forensic analysis of AI-compression traces in spatial and frequency domain,” *Pattern Rec.*, 2024.

Wrap Up

Conclusion

1. Modern image compression algorithms use neural networks.
2. They achieve unprecedented compression rates at very high quality.
3. They can lead to semantic changes in compressed images.

Takeaway

- Consider if the benefit of bandwidth savings is proportionate to potential risks caused by miscompressions.

Publication

Collection of miscompressions

2024 IEEE International Workshop on Information Forensics and Security (WIFS) | 979-8-3503-6442-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/WIFS61860.2024.10810704

A Taxonomy of Miscompressions: Preparing Image Forensics for Neural Compression

Nora Hofer
University of Innsbruck, Austria
nora.hofer@uibk.ac.at

Rainer Böhme
University of Innsbruck, Austria
rainer.boehme@uibk.ac.at

***Abstract*—Neural compression has the potential to revolutionize lossy image compression. Based on generative models, recent schemes achieve unprecedented compression rates at high perceptual quality, but they compromise semantic fidelity. Details of decompressed images may appear optically flawless, but semantically different from the originals, making compression errors difficult or impossible to detect. We explore the problem space and propose a provisional taxonomy of miscompressions. It defines three types of "what happens" and has a binary "high impact" flag indicating miscompressions that alter symbols. We discuss how the taxonomy can facilitate risk communication and research into mitigations.**

***Index Terms*—neural image compression, miscompression, semantic changes, forensics**

(a) Lossless original (4.1 MB, 11.90 bpp) (b) Neural compression (159 kB, 0.46 bpp, [4]) (c) JPEG QF 20 (158 kB, 0.46 bpp)

Fig. 1. State-of-the-art neural compression schemes can alter the semantic in details of the decompressed images. The high fidelity and the lack of visible compression artifacts make false reconstructions look more authentic than JPEG, which introduces visible distortion. (Crop of image 0831 of DIV2K [6], 0.41% of the original.) All figures are best viewed on screen and magnified.

I. Introduction

A turning point in the investigation of the 2013 Boston Marathon bombing was a bystander's cellphone photo that allowed police to identify one of the suspects in a crowd [1] [2]. Remarkably, the relevant part of the image comprised just 0.2 % of all pixels. In this paper, we ask the question whether digital images will continue to serve as reliable sources in a future where neural compression becomes the default.

Neural image compression employs learning-based elements in the image compression pipeline, achieving high perceptual quality at unprecedented compression rates [3]. State of the art schemes use generative networks to synthesize parts of an image [4], [5]. However, a drawback of this approach is that the synthesized details appear plausible and of high perceptual quality, but may be semantically different from the original.

To illustrate this, Fig. 1 compares a small crop (0.4 %) of an uncompressed image (left), to a version from the *HiFiC* neural compression scheme [4] (middle), and the JPEG compressed image at quality factor (QF) 20 (right). This factor was chosen to match the compression rate in bits per pixel between *HiFiC* and JPEG. While the neural compression retains clear readability of the numbers, closer inspection reveals that they differ: the upper row changes from 2264 – 7668 in the original to 2254 – 7664 in the *HiFiC* reconstruction. We confirmed that not only human observers, but also Google's Cloud AI optical character recognition came to this conclusion. Given the high apparent quality of the image, observers unaware of its processing history might be inclined to fully trust the image and its semantic content. By contrast, the visible compression artifacts in the JPEG image not only make the numbers difficult to read, but also signal low reliability and dissuade users from interpreting the numbers with confidence.

In this paper, we propose the term "miscompression" to describe semantic changes resulting from lossy compression. This new[1] phenomenon arose with neural compression and deserves the attention of researchers and forensic practitioners. To facilitate the conversation, we develop a taxonomy of miscompressions based on the explorative visual inspection of three benchmark datasets, examining five different neural compression schemes at different quality settings. The examples used for illustration in this paper are produced using two schemes of the latest generation. We derive implications for forensics and society at large, and outline objectives for research into mitigations.

We emphasize that all authors of recent neural compression schemes generally acknowledge the risk of hallucinations, warn against using their schemes for critical applications [4, p. 10], and occasionally point to examples of text becoming unreadable. However, miscompressions take many forms and pose a risk to society if left unaddressed. We see our work as a first step towards solutions.

The paper is structured as follows. Section II introduces the principles of neural compression by comparing it to the JPEG pipeline. Section III defines miscompressions, describes our empirical approach, and presents the proposed taxonomy using examples. Section IV discusses implications and outlines a research agenda. Section V concludes our paper.

We thank Benedikt Lorch, Martin Benel, Judith Senn, and Verena Lachner for valuable discussions and comments. Computational results were achieved using the LEO HPC infrastructure at the University of Innsbruck.

[1] The closest related work we are aware of is an attempt to introduce copy-evident marks into images which only appear after JPEG compression [9].

Research project SCLIC
Semantic **C**hanges in **L**earning based **I**mage **C**ompression

Funded by: *Tiroler Nachwuchsforscher*innen Förderung*

LAND TIROL

Hofer, N. and Böhme, R., "A Taxonomy of Miscompressions: Preparing Image Forensics for Neural Compression." In IEEE *International Workshop on Information Forensics and Security (WIFS)*. IEEE, Rome, Italy, 2024.

universität innsbruck

Thank You !

Do Not Trust Your Eyes: The Semantic Pitfalls of Modern Image Compression

nora.hofer@uibk.ac.at

Engineering Kiosk Alps Meetup · Innsbruck, Austria · 16 January 2025