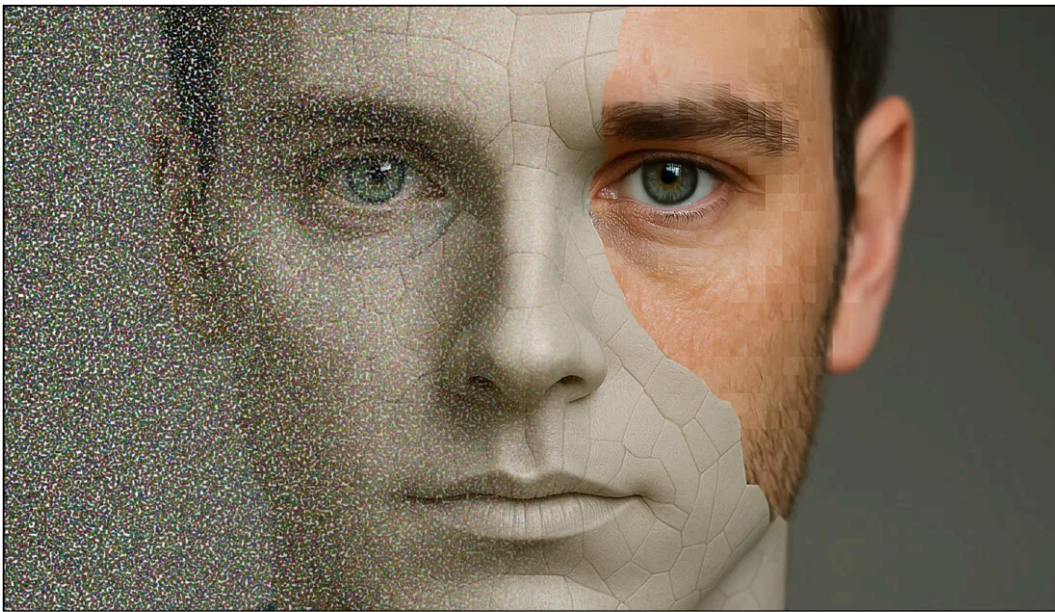


JPEG AI Blurs the Line Between Real and Synthetic

Originally published at <https://www.unite.ai/jpeg-ai-blurs-the-line-between-real-and-synthetic/>



In February of this year, the [JPEG AI international standard](#) was published, after several years of research aimed at using machine learning techniques to produce a smaller and more easily transmissible and storable image codec, without a loss in perceptual quality.



From the official publication stream for JPEG AI, a comparison between Peak Signal-to-Noise Ratio (PSNR) and JPEG AI's ML-augmented approach.

Source: <https://jpeg.org/jpegai/documentation.html>

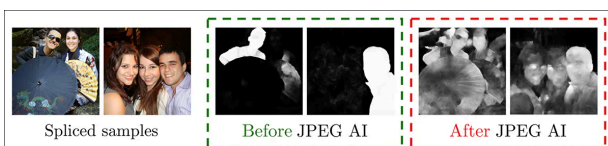
One possible reason why this advent made few headlines is that the core PDFs for this announcement were (ironically) not available through free-access portals such as Arxiv. Nonetheless, Arxiv had already put forward a number of studies examining the significance of JPEG AI across several aspects, including the method's uncommon [compression artifacts](#) and its [significance for forensics](#).



One study compared compression artefacts, including those of an earlier draft of JPEG AI, finding that the new method had a tendency to blur text – not a minor matter in cases where the codec might contribute to an evidence chain. Source:

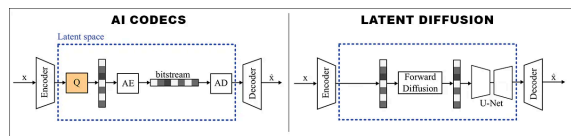
<https://arxiv.org/pdf/2411.06810>

Because JPEG AI alters images in ways that mimic the artifacts of synthetic image generators, existing forensic tools [have difficulty differentiating](#) real from fake imagery:



After JPEG AI compression, state-of-the-art algorithms can no longer reliably separate authentic content from manipulated regions in localization maps, according to a recent paper (March 2025). The source examples seen on the left are manipulated/fake images, wherein the tampered regions are clearly delineated under standard forensic techniques (center image). However, JPEG AI compression lends the fake images a layer of credibility (image on far right). Source: <https://arxiv.org/pdf/2412.03261>

One reason is that JPEG AI is trained using a model architecture similar to those used by generative systems that forensic tools aim to detect:



The new paper illustrates the similarity between the methodologies of AI-driven image compression and actual AI-generated images.

Source: <https://arxiv.org/pdf/2504.03191>

Therefore both models may produce some similar underlying visual characteristics, from a forensic standpoint.

Quantization

This cross-over occurs because of *quantization*, common to both architectures, and which is used in machine learning both as a [method of converting](#) continuous data into discrete data points, and as an [optimization technique](#) that can significantly slim down the file-size of a trained model (casual image synthesis enthusiasts will be familiar with the wait between an unwieldy official model release, and a community-led quantized version that can run on local hardware).

In this context, quantization refers to the process of converting the continuous values in the image's [latent representation](#) into fixed, discrete steps. JPEG AI uses this process to *reduce the amount of data needed* to store or transmit an image by simplifying the internal numerical representation.

Though quantization makes encoding more efficient, it also imposes structural regularities that can resemble the artifacts left by generative models – subtle enough to evade perception, but disruptive to forensic tools.

In response, the authors of a [new work](#) titled *Three Forensic Cues for JPEG AI Images* propose interpretable, non-neural techniques that detect JPEG AI compression; determine if an image has been recompressed; and distinguish compressed real images from those generated entirely by AI.

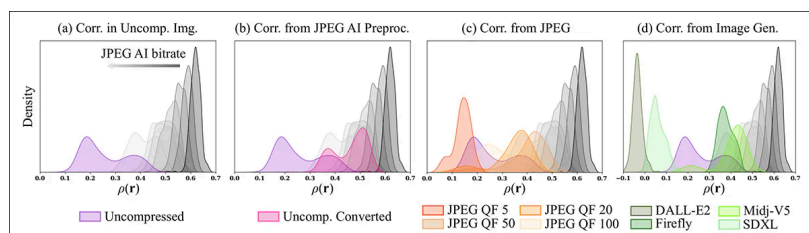
Method

Color Correlations

The paper proposes three 'forensic cues' tailored to JPEG AI images: *color channel correlations*, introduced during JPEG AI's preprocessing steps; *measurable distortions in image quality* across repeated compressions that reveal recompression events; and *latent-space quantization patterns* that help distinguish between images compressed by JPEG AI and those generated by AI models.

Regarding the color correlation-based approach, JPEG AI's preprocessing pipeline introduces statistical dependencies between the image's color channels, creating a signature that can serve as a forensic cue.

JPEG AI converts RGB images to the [YUV color space](#) and performs 4:2:0 chroma subsampling, which involves downsampling the [chrominance channels](#) before compression. This process leads to subtle correlations between the high-frequency residuals of the red, green, and blue channels – correlations that are not present in uncompressed images, and which differ in strength from those produced by traditional JPEG compression or synthetic image generators.



A comparison of how JPEG AI compression alters color correlations in images..

Above we can see a comparison from the paper illustrating how JPEG AI compression alters color correlations in images, using the red channel as an example.

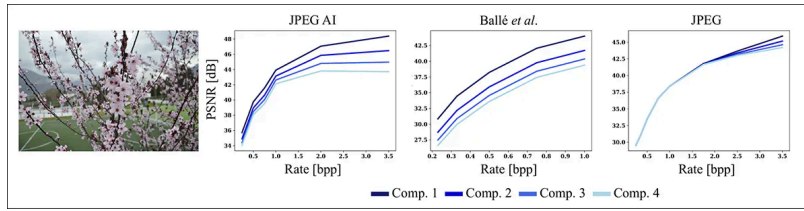
Panel A compares uncompressed images to JPEG AI-compressed ones, showing that compression significantly increases inter-channel correlation; panel B isolates the effect of JPEG AI's preprocessing – just the color conversion and subsampling – demonstrating that even this step alone raises correlations noticeably; panel C shows that traditional JPEG compression also increases correlations slightly, but not to the same degree; and Panel D examines synthetic images, with [Midjourney-V5](#) and [Adobe Firefly](#) displaying moderate correlation increases, while others remain closer to uncompressed levels.

Rate-Distortion

The rate-distortion cue identifies JPEG AI recompression by tracking how image quality, measured by [Peak Signal-to-Noise Ratio](#) (PSNR), declines in a predictable pattern across multiple compression passes.

The research contends that repeatedly compressing an image with JPEG AI leads to progressively smaller, but still measurable, losses in image quality, as quantified by PSNR, and that this gradual degradation forms the basis of a forensic cue for detecting whether an image has been recompressed.

Unlike traditional JPEG, where earlier methods tracked changes in specific image blocks, JPEG AI requires a different approach, due to its neural compression architecture; therefore the authors propose monitoring how both bitrate and PSNR evolve over successive compressions. Each round of compression alters the image less than the one prior, and this diminishing change (when plotted against bitrate) can reveal whether an image has gone through multiple compression stages:



An illustration of how repeated compression affects image quality across different codecs, featuring results from JPEG AI and a neural codec developed at <https://github.com/valardev/ai-jpeg>. Both exhibit a steady decline in PSNR with each additional compression, even at lower bitrates. By contrast, traditional JPEG compression maintains relatively stable quality at higher bitrates.

In the image above, we see charted rate-distortion curves for JPEG AI; a second AI-based codec; and traditional JPEG, finding that JPEG AI and the neural codec show a consistent PSNR decline across all bitrates, while traditional JPEG only shows noticeable degradation at much higher bitrates. This behavior provides a quantifiable signal that can be used to flag recompressed JPEG AI images.

By extracting how bitrate and image quality evolve over multiple compression rounds, the authors similarly constructed a signature that helps flag whether an image has been recompressed, affording a potential practical forensic cue in the context of JPEG AI.

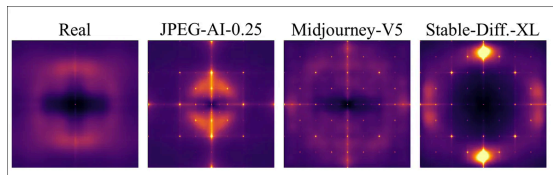
Quantization

As we saw earlier, one of the more challenging forensic problems raised by JPEG AI is its visual similarity to synthetic images generated by diffusion models. Both systems use encoder-decoder architectures that process images in a compressed latent space and often leave behind subtle upsampling artifacts.

These shared traits [can confuse detectors](#) – even those retrained on JPEG AI images. However, a key structural difference remains: JPEG AI applies quantization, a step that rounds latent values to discrete levels for efficient compression, while generative models typically do not.

The new paper uses this distinction to design a forensic cue that indirectly tests for the presence of quantization. The method analyzes how the latent representation of an image responds to rounding, on the assumption that if an image has already been quantized, its latent structure will exhibit a measurable pattern of alignment with rounded values.

These patterns, while invisible to the eye, produce statistical differences that can help separate compressed real images from fully synthetic ones.



An example of average Fourier spectra reveals that both JPEG AI-compressed images and those generated by diffusion models like Midjourney-V5 and Stable Diffusion XL exhibit regular grid-like patterns in the frequency domain – artifacts commonly linked to upsampling. By contrast, real images lack these patterns. This overlap in spectral structure helps explain why forensic tools often confuse compressed real images with synthetic ones.

Importantly, the authors show that this cue works across different generative models and remains effective even when compression is strong enough to zero out entire sections of the latent space. By contrast, synthetic images show much weaker responses to this rounding test, offering a practical way to distinguish between the two.

The result is intended as a lightweight and interpretable tool targeting the core difference between compression and generation, rather than relying on brittle surface artifacts.

Data and Tests

Compression

To evaluate whether their color correlation cue could reliably detect JPEG AI compression (i.e., a first pass from uncompressed source), the authors tested it on high-quality uncompressed images from the [RAISE dataset](#), compressing these at a variety of bitrates, using the JPEG AI reference implementation.

They trained a simple [random forest](#) on the statistical patterns of color channel correlations (particularly how residual noise in each channel aligned with the others) and compared this to a [ResNet50](#) neural network trained directly on the image pixels.

bpp	0.06	0.12	0.25	0.50	0.75	1.0	2.0
ResNet50	0.834	0.859	0.875	0.855	0.801	0.745	0.673
$RF \rho(r)$	0.822	0.797	0.774	0.744	0.738	0.726	0.667
$RF \rho(g)$	0.829	0.794	0.775	0.743	0.716	0.702	0.658
$RF \rho(b)$	0.856	0.829	0.811	0.785	0.761	0.746	0.710

Detection accuracy of JPEG AI compression using color correlation features, compared across multiple bitrates. The method is most effective at lower bitrates, where compression artifacts are stronger, and shows better generalization to unseen compression levels than the baseline ResNet50 model.

While the ResNet50 achieved higher accuracy when the test data closely matched its training conditions, it struggled to [generalize](#) across different compression levels. The correlation-based approach, although far simpler, proved more consistent across bitrates, especially at lower compression rates where JPEG AI's preprocessing has a stronger effect.

These results suggest that even without deep learning, it is possible to detect JPEG AI compression using statistical cues that remain interpretable and resilient.

Recompression

To evaluate whether JPEG AI recompression can be reliably detected, the researchers tested the rate-distortion cue on a set of images compressed at diverse bitrates – some only once and others a second time using JPEG AI.

This method involved extracting a 17-dimensional feature vector to track how the image's bitrate and PSNR evolved across three compression passes. This feature set captured how much quality was lost at each step, and how the latent and [hyperprior](#) rates behave—metrics that traditional pixel-based methods can't easily access.

The researchers trained a random forest on these features and compared its performance to a ResNet50 trained on image patches:

	ResNet50 (Pixel)		RF (Rate-Dist.)	
b_0	Single	Recomp.	Single	Recomp.
2.00	0.797	0.476	0.913	0.511
1.00	0.764	0.475	0.912	0.518
0.75	0.726	0.532	0.879	0.548
0.50	0.644	0.616	0.777	0.725
0.25	0.519	0.726	0.526	0.879
0.12	0.457	0.778	0.465	0.920

Results for the classification accuracy of a random forest trained on rate-distortion features for detecting whether a JPEG AI image has been recompressed. The method performs best when the initial compression is strong (i.e., at lower bitrates), and then consistently outperforms a pixel-based ResNet50 – especially in cases where the second compression is milder than the first.

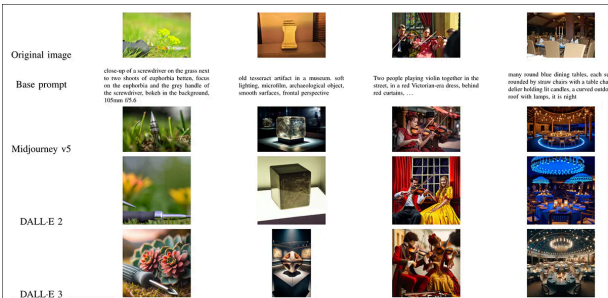
The random forest proved notably effective when the initial compression was strong (i.e., at lower bitrates), revealing clear differences between single and double-compressed images. As with the prior cue, the ResNet50 iteration struggled to generalize, particularly when tested on compression levels it had not seen during training.

The rate-distortion features, by contrast, remained stable across a wide range of scenarios. Notably, the cue worked even when applied to a different AI-based codec, suggesting that the approach generalizes beyond JPEG AI.

JPEG AI and Synthetic Images

For the final testing round, the authors tested whether their quantization-based features can distinguish between JPEG AI-compressed images and fully synthetic images generated by models such as Midjourney, [Stable Diffusion](#), [DALL-E 2](#), [Glide](#), and Adobe Firefly.

For this, the researchers used a subset of the [Synthbuster](#) dataset, mixing real photos from the RAISE database with generated images from a range of diffusion and GAN-based models.



Examples of synthetic images in Synthbuster, generated using text prompts inspired by natural photographs from the RAISE-1k dataset. The images were created with various diffusion models, with prompts designed to produce photorealistic content and textures rather than stylized or artistic renderings. Source: <https://ieeexplore.ieee.org/document/10334046>

The real images were compressed using JPEG AI at several bitrate levels, and classification was posed as a two-way task: either JPEG AI versus a specific generator, or a specific bitrate versus [Stable Diffusion XL](#).

The quantization [features](#) (correlations extracted from latent representations) were calculated from a fixed 256×256 region and fed to a random forest classifier. As a baseline, a ResNet50 was trained on pixel patches from the same data.

Type		ResNet50 (Pixel)	RF (Trunc.)	RF (full)
JPEG AI	bpp 0.06	0.886	0.905	0.985
	bpp 0.12	0.886	0.896	0.982
	bpp 0.25	0.879	0.849	0.977
	bpp 0.50	0.850	0.782	0.959
	bpp 0.75	0.827	0.731	0.917
DM	Midj.-V5	0.850	0.832	0.945
	DALL-E2	0.871	0.789	0.950
	SD.-XL	0.879	0.849	0.977
	SD.-1.4	0.869	0.936	0.980
	Glide	0.661	0.489	0.538
	Firefly	0.708	0.572	0.839
GAN	BigGAN	0.693	0.765	0.767
	StyleGAN2	0.748	0.775	0.853
	StyleGAN3	0.798	0.839	0.859
	ProGAN	0.723	0.854	0.980

Classification accuracy of a random forest using quantization features to separate JPEG AI-compressed images from synthetic images.

Across most conditions, the quantization-based approach outperformed the ResNet50 baseline, particularly at low bitrates where compression artifacts were stronger.

The authors state:

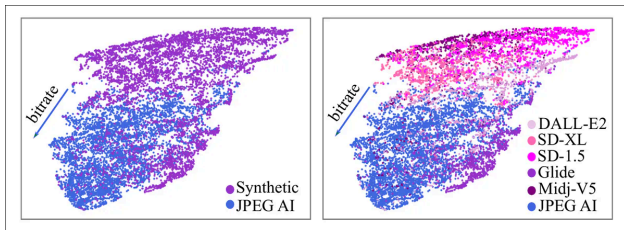
‘The baseline ResNet50 performs best for Glide images with an accuracy of 66.1%, but otherwise it generalizes worse than the quantization features. The quantization features exhibit a good generalization across compression strengths and generator types.

‘The importance of the coefficients that are quantized to zero are shown in the very respectable performance of the truncated [features], which in many cases perform comparable to the ResNet50 classifier.

‘However, quantization features that use the untruncated, full integer [vector] still perform notably better. These results confirm that the amount of zeros after quantization is an important cue for differentiating AI-compressed and AI-generated images.

‘Nevertheless, it also shows that also other factors contribute. The accuracy of the full vector for detecting JPEG AI is for all bitrates over 91.0%, and stronger compression leads to higher accuracies.’

A projection of the feature space using [UMAP](#) showed clear separation between JPEG AI and synthetic images, with lower bitrates increasing the distance between classes. One consistent outlier was Glide, whose images clustered differently and had the lowest detection accuracy of any generator tested.



Two-dimensional UMAP visualization of JPEG AI-compressed and synthetic images, based on quantization features. The left plot shows that lower JPEG AI bitrates create greater separation from synthetic images; the right plot, how images from different generators cluster distinctly within the feature space.

Finally, the authors evaluated how well the features held up under typical post-processing, such as JPEG recompression or downsampling. While performance declined with heavier processing, the drop was gradual, suggesting that the approach retains some robustness even under degraded conditions.

JPEG AI	JPG 70	JPG 80	JPG 90	RS 70	RS 80	RS 90
bpp 0.25	0.931	0.955	0.985	0.908	0.972	0.980
bpp 0.50	0.880	0.885	0.978	0.732	0.908	0.962
bpp 0.75	0.805	0.791	0.840	0.631	0.775	0.928
Midj.-V5	0.817	0.841	0.873	0.690	0.861	0.915
Firefly	0.673	0.738	0.865	0.615	0.648	0.675
DALL-E2	0.720	0.813	0.883	0.715	0.873	0.921

Evaluation of quantization feature robustness under post-processing, including JPEG recompression (JPG) and image resizing (RS).

Conclusion

It's not guaranteed that JPEG AI will enjoy wide adoption. For one thing, there's enough infrastructural debt at hand to impose friction on *any* new codec; and even a 'conventional' codec with a fine pedigree and broad consensus as to its value, such as AV1, [has a hard time](#) dislodging long-established incumbent methods.

In regards to the system's potential clash with AI generators, the characteristic quantization artifacts that help the *current* generation of AI image detectors may be diminished or ultimately replaced by traces of a different kind, in later systems (assuming that AI generators will always leave forensic residue, which is not certain).

This would mean that JPEG AI's own quantization characteristics, perhaps along with other cues identified by the new paper, may not end up colliding with the forensic trail of the most effective new generative AI systems.

If, however, JPEG AI continues to operate as a *de facto* 'AI wash', significantly blurring the distinction between real and generated images, it would be hard to make a convincing case for its uptake.

First published Tuesday, April 8, 2025

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More