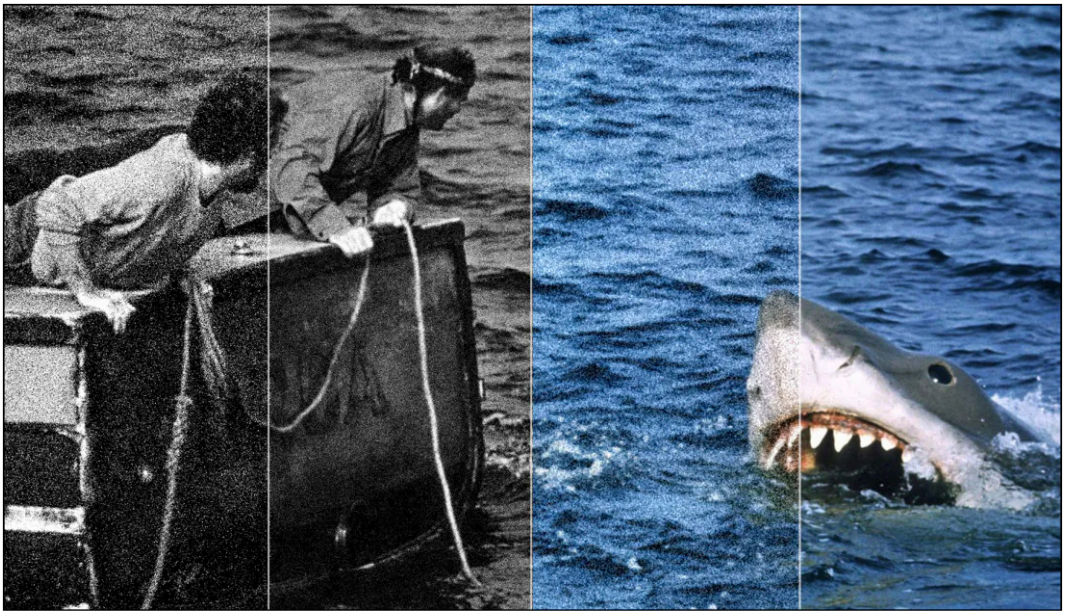
Using AI to Simulate Film Grain

Originally published at https://www.unite.ai/using-ai-to-simulate-film-grain/



Make America Grainy Again: a new AI tool can strip film grain from old footage, compress the video at a fraction of the size, then put the grain back so viewers never notice. It works with existing video standards and cuts bandwidth by up to 90 percent, while keeping the vintage look.

For many of us watching movies or old TV shows, the 'sizzle' of film grain is reassuring; even when we don't consciously register it, grain tells us that what we are watching was made with chemicals, not code, and bonds the experience to the physical world: to stock choice, exposure, lab processes, and bygone eras:

Hollywood's approach to grain has shifted alongside changes in culture and production methods. During the 1960s, evolving camera stocks and photographic practices contributed to the decade's distinct visual identity. Later, directors working in digital began reintroducing grain deliberately. In the mid-1980s director James Cameron selected a particularly coarse Kodak stock for Aliens (1986, lower right in image above), likely to enhance atmosphere while also helping to hide wires from practical VFX miniature work. Source: https://archive.is/3ZSjN (my own most recent article on this topic)

Analog texture comes from a time when producing media cost real money, access was limited, and there was at least a loose sense that only the most capable or determined could get through, acting as a shorthand for realism, and credibility – and, when high resolution capture technologies eliminated it, *nostalgia*.

Christopher Nolan never switched. While most of the industry embraced digital for its speed and flexibility, the acclaimed director dug in, insisting on celluloid as both a discipline and an aesthetic.

Denis Villeneuve, working squarely within digital pipelines, still parses his footage through photochemical processes. For the *Dune* movies, shot digitally, the footage was printed onto film stock and then scanned back to digital, purely for atmosphere and effect.

Fake Grain

Aficionados of film and TV quality associate visible grain with high-resolution, where the bitrate (the amount of data being pushed into each frame) is so high that even the smallest details, such as halide grains, are preserved.

However, if streaming networks really made that kind of bitrate available, it would put severe strain on network capacity, and likely cause buffering and stuttering. Therefore platforms such as Netflix create optimized AV1 versions of their content and use the AV1 codec's capabilities to add grain to the film or episode in an intelligent and apposite way, saving 30% of bandwidth in the process.

AV1 is designed to incorporate artificial film grain, as in these examples. Source: https://waveletbeam.com/index.php/av1-film-grain-synthesis

The 'grain fetish' is a relatively rare digital equivalent to atavistic trends such as the revival of vinyl, and it's hard to say whether it is used by streamers to make highly optimized video look like really expensive 'raw video' (for those viewers who have unconsciously associated those characteristics), making the bitrate seem higher than it is; or to deflect the perceptual quality drop that old 4:3 shows would otherwise take when streaming providers crop them to widescreen aspect ratios; or just to pander to the retro 'Nolan aesthetic' in general.

Grain Siloed

The problem is that grain is also noise. Digital systems hate noise, and streaming codecs such as AV1 scrub it away to save bandwidth, unless grain settings are explicitly configured. Likewise AI upscalers such as the Topaz Gigapixel series treat grain as a flaw to be corrected.

In the field of diffusion-based image synthesis, grain is extremely challenging to generate, since it represents *extreme detail*, and thus would usually only appear in massively overfitted models, because the entire latent diffusion model (LDM) architecture is designed to deconstruct noise (such as grain) into clear imagery, rather than treat grain flecks as implicit properties in media.

Therefore it can be challenging to create convincing grain using machine learning. And even if one could do it, rendering it straight back into an optimized video would just bloat the video's file size right back up.

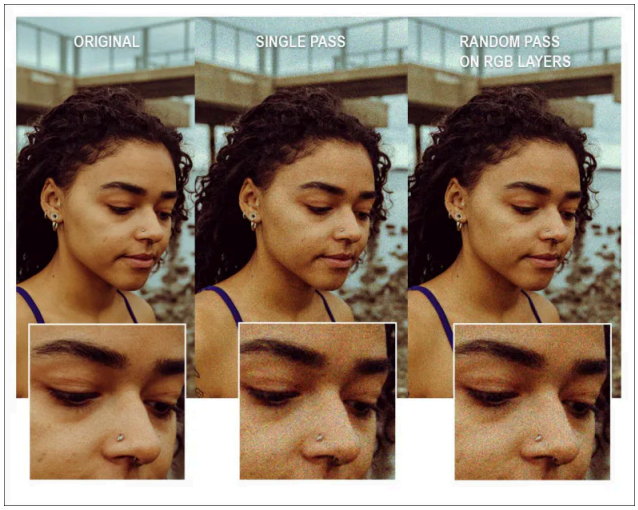
Because of this latter logistical consideration, state-of-the-art video codecs such as Versatile Video Coding (VVC) offer grain as a kind of 'sidecar' service.

VVC compresses the clean, denoised video and discards the grain. Instead of wasting data trying to preserve random high-frequency grain patterns, it analyzes the grain *separately* and encodes a small set of parameters (e.g. amplitude, frequency, and blending mode) that describe how to regenerate similar grain during playback.

These parameters are stored in an *FGC-SEI* (Film Grain Characteristics Supplemental Enhancement Information) stream, which rides alongside the main bit-stream. After decoding, a synthesis module uses these instructions to reapply synthetic grain that mimics the original.

This preserves the 'look' of high-bitrate, grain-rich emulsion, while keeping the actual bitrate low, since the encoder isn't forced to spend resources preserving unpredictable noise.

Additionally, as with discrete subtitles files, this faux 'grain' content is specific to the video in question; haphazardly applying generic grain filters in platforms such as Photoshop or After Effects, or in automated processing pipelines, would not result in 'fitted' grain, but instead an unrelated overlay of noise:

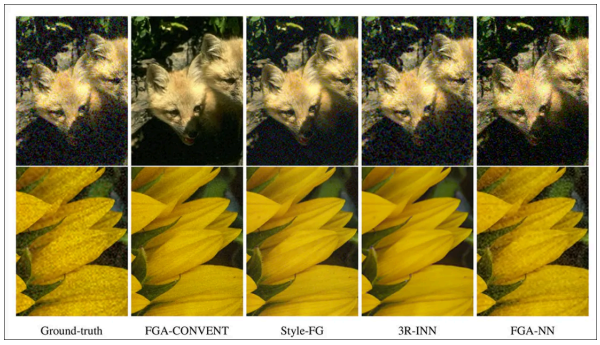


Left: original image. Center: Photoshop Camera Raw Grain applied uniformly across all channels. Right: the same Grain filter applied individually to each channel in sequence. Source image (CC0): https://stocksnap.io/photo/woman-beach-FJCOO6JWDP (via my own previous article)

Photoshop's 'Grain' filter adds uniform random noise; but real film grain comes from halide crystals of varying sizes. Applying the filter to each channel separately (see image above) just creates more chaos, not realism. True film grain reflects how light strikes layered emulsions *at the moment of exposure*. Simulating that would require estimating how different areas of an image would have activated each halide layer, not just splitting the effect across RGB layers.

FGA-NN

Into this specious pursuit comes a new research paper from France – a brief but interesting outing that offers a quantitatively and qualitatively superior method of analyzing and recreating grain:



Comparison between ground-truth grain and results from various analysis and synthesis methods. Source: https://arxiv.org/pdf/2506.14350

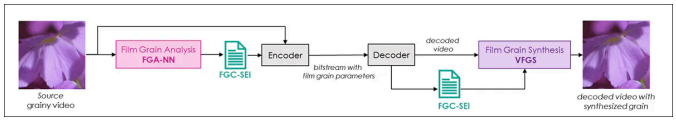
The new system, titled *FGA-NN*, does not depart from the conventional use of conventional Gaussian-based grain synthesis through the standard VVC-compatible method, Versatile Film Grain Synthesis (VFGS). What the system changes is the *analysis*, using a neural network to estimate the synthesis parameters more accurately

Therefore the final grain is still synthesized using the same conventional Gaussian model – but the network feeds better metadata into a standard, rule-based generator, obtaining a state-of-the-art model.

The new paper is titled *FGA-NN: Film Grain Analysis Neural Network*, and comes from three researchers at InterDigital IDCC -2.81% R&D, Cesson-Sévigné. Though the paper is not long, let's take a look at some of the key aspects of the advances that the new method offers.

Method

To recap: the FGA-NN system takes a grainy video as input and extracts a compact description of the grain, outputting parameters in the standardized FGC-SEI format used by diverse modern codecs. These parameters are transmitted alongside the video, allowing the decoder to reconstruct the grain using VFGS, rather than encoding the grain directly.



The schema for analyzing and reapplying film grain in video distribution, using FGA-NN for parameter extraction and VFGS for synthesis.

To train the network, the authors needed pairs of grainy videos and corresponding FGC-SEI metadata. Since most grainy footage lacks this kind of metadata, the researchers created their own dataset by generating FGC-SEI parameters, applying synthetic grain to clean videos, and using these as training examples.

Training data for FGA-NN was created by applying synthetic grain to clean footage from the BVI-DVC and DIV2K datasets. Randomized FGC-SEI parameters were generated and used with the VFGS synthesis tool, allowing each grainy video to be paired with known metadata.

The frequency-based model supported by current video standards was used, with parameter ranges constrained to maintain visual plausibility across luma and chroma channels.

Training data for the new collection was created by applying synthetic grain to clean footage from the BVI-DVC and DIV2K datasets. Randomized FGC-SEI parameters were generated and used with the Versatile Film Grain Synthesis (VFGS) tool, allowing each grainy video to be paired with known metadata.

Parameter	Value
SEIFGCModelID	0
SEIFGCBlendingModeID	0
SEIFGCLog2ScaleFactor	2-7
SEIFGCCompModelPresentComp0,1,2	1
SEIFGCNumIntensityIntervalMinus1Comp0,1,2	0-255
SEIFGCIntensityIntervalLowerBoundComp0,1,2	0-255
SEIFGCIntensityIntervalUpperBoundComp0,1,2	0-255
SEIFGCNumModelValuesMinus1Comp0,1,2	2

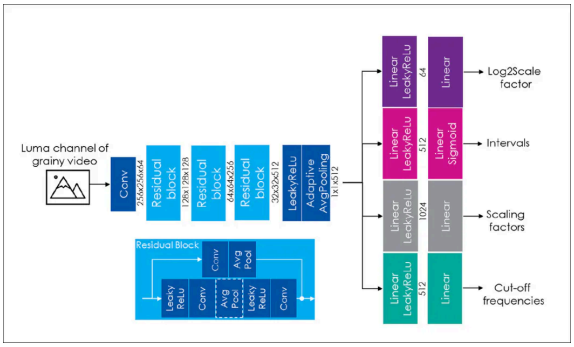
Overview of the randomized FGC-SEI parameter ranges used to generate synthetic grain for training, applied to clean footage from the BVI-DVC and DIV2K datasets. Parameters were constrained to ensure plausible visual results across both luma and chroma channels.

The frequency filtering model, the only synthesis method currently supported in codec implementations such as the VVC Test Model (VTM), was used throughout. Parameter ranges were constrained to preserve visual plausibility across both luma and chroma channels.

Network Effect

FGA-NN features two coordinated models, for luma and chroma, respectively, each designed to predict the specific parameters needed to recreate realistic film grain.

For every input image, the system estimates a set of intensity intervals, the scaling factors linked to each interval, the horizontal and vertical cut-off frequencies, and an overall scale adjustment known as the Log2Scale factor. To handle this, the model uses a shared feature extractor that processes the grainy input and feeds into four separate output branches, each responsible for a different prediction task:



Architecture of the luma version of FGA-NN. A shared backbone extracts features from grainy input frames, followed by four output branches tailored to specific parameter prediction tasks: interval boundaries, scaling factors, cut-off frequencies, and global Log2Scale. The chroma network uses the same structure with adjusted input and output dimensions.

Interval boundaries are predicted using regression, while scaling factors, cut-off frequencies, and the global scale setting are treated as classification problems.

The architecture is adjusted to reflect the complexity of each task, with larger internal layers used for more fine-grained predictions; specifically, the chroma model mirrors the luma structure, but adapts to the different characteristics of color data.

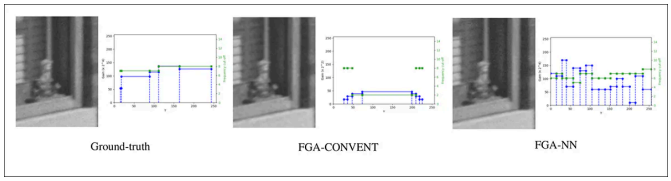
Training and Tests

FGA-NN was trained using four objective functions, each aligned with one of its prediction tasks. For the classification outputs a categorical cross-entropy loss was used to reduce the gap between predicted labels and ground truth.

Interval boundaries were normalized to a 0-to-1 range and optimized using a combined loss: an exponentially scaled L1 loss (expL1) that penalized larger errors more heavily, and a monotonicity penalty that discouraged downward trends. All four losses were combined, with high weights assigned to cut-off and scaling factors, while interval boundaries and Log2Scale were weighted at 1 and 0.1.

Training was undertaken under the Adam optimizer, at a learning rate of 5e-4, across 10,000 iterations, with a batch size of 64.

The only comparable tool suitable for comparative tests was FGA-CONVENT, which also produces values in the FGC-SEI format, and is used for grain-processing. Both systems were tested on UHD sequences from the JVET subjective evaluation set, using footage containing real film grain.



Vertical dashed lines indicate intensity interval boundaries, while the Log2Scale gain is noted in the axis label.

In the image above, we see identical cropped frames generated by VFGS using parameters from each method, compared with the original. Their respective luma estimates are also plotted against ground-truth values set manually using VFGS, which here charts pixel intensity on the X axis (0–255), scaling factors on the blue Y axis (0–255), and cut-off frequencies on the green Y axis (2–14).

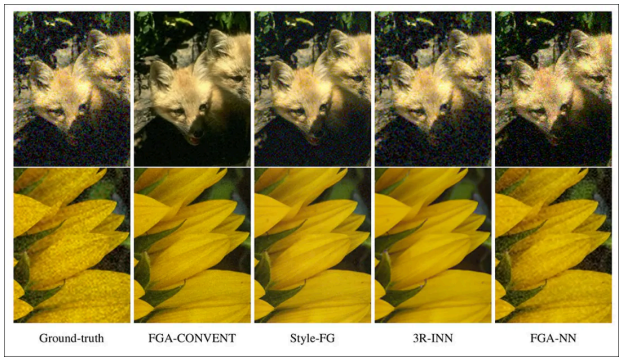
The authors state:

> *'One can observe that FGA-NN accurately captures the overall trend of the ground-truth film grain pattern and amplitude, resulting in synthesized images with perceptually similar film grain to that of the ground-truth ones.*
>
> *'On the other hand, FGA-CONVENT predicts a lower scaling factor, compensated by a correspondingly lower Log2Scale factor as a result of its design, and tends to generate a coarser film grain pattern than the reference, resulting in a distinct yet visually consistent appearance.'*

They note that direct comparison with ground-truth grain parameters is unreliable, since scaling and Log2Scale can compensate for each other, and minor errors often have little visual impact.

Test of Faith

Film grain *fidelity* was benchmarked across four workflows: FGA-NN with VFGS; FGA-CONVENT plus VFGS; Style-FG; and 3R-INN. Tests used both the FGC-SEI and FilmGrainStyle740k datasets, comparing output to ground-truth using Learned Perceptual Similarity Metrics (LPIPS); JSD-NSS; and Kullback–Leibler (KL) divergence.

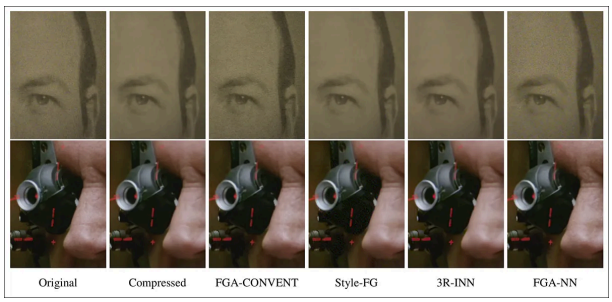


Benchmark results on the FilmGrainStyle740k dataset. Style-FG and 3R-INN outperform others due to being trained on this set, with FGA-NN following closely. FGA-CONVENT underperforms, reflecting its reliance on multi-frame analysis and homogeneous regions – conditions not met by the small, texture-rich inputs used in this case.

Of these results, the authors state:

> *'On the FilmGrainStyle740k test-set, Style-FG and 3R-INN achieve the best results, as these methods were specifically trained on this dataset, with FGA-NN trailing closely behind. The performance of FGA-CONVENT combined with VFGS is suboptimal on both testsets.*
>
> *'This is solely due to the fact that the analysis relies on homogeneous regions and exploits information from multiple frames in a real film grain analysis use-case, whereas in the present evaluation analysis is provided with a single low-resolution image (256×256 to a maximum of 768×512), which often contains significant texture.*
>
> *'This further complicates the challenge for conventional analysis method, making it impossible to apply FGA-CONVENT to such small images.'*

Finally, the authors note that while learning-based methods such as 3R-INN and Style-FG produce strong visual results on curated datasets, their high computational cost renders them unsuitable for deployment on end-user devices.



Comparison of low-bitrate frames enhanced using different analysis and synthesis workflows (third to last columns).

By comparison, the approach proposed in the new paper pairs the lightweight FGA-NN analysis module with the hardware-efficient VFGS synthesis method, which the authors describe as a more viable and deployable solution for reintroducing film grain in compressed video.

They state further that the benefits of FGA-NN are potentially considerable, at scale:

> *'[Encoding] UHD videos with film grain at medium to low bitrates using our film grain analysis and synthesis workflow enables bitrate savings of up to 90% compared to high-bitrate encoding.'*

Conclusion

The obsession with film grain is one of the strangest and most curious conceits of the post-analog age, and it is interesting to note that what was once considered a limitation of the medium has now become a totem of verisimilitude and authenticity in itself, even (perhaps subconsciously) to a new generation of viewers born after the effective decline of emulsion.

It should be noted that none of the state-of-the-art grain-recreation methods, including this latest innovation, can exactly capture the true effect of the way that light effects layers of halides in a true photochemical process, across a range of conditions.

First published Wednesday, June 18, 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