REALSYNZ: NATURALIZING SYNTHETIC IMAGES WITH ENVELOPE VARIATION STATISTICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Generated images still show consistent shortcomings and artifacts compared with natural images (Wang et al., 2025). While they represent scenes well, textures such as unnaturally waxy skin and flat fabric place them in an uncanny valley (Mori et al., 2012). We find that the difference lies in how fine detail is distributed. Three of the five generators we study carry less fine detail than natural images, and the two that carry as much or more distribute it differently. In textured regions, every generator packs fine detail into small areas and leaves the space between them nearly empty. In flat regions, three of the five generators spread fine detail too evenly. We measure this distribution with two statistics of natural images. The moment ratio is the mean absolute value of fine detail in a patch divided by its root mean square, and envelope variation is the coefficient of variation of the local magnitude of fine detail across a patch. Edits that change only the amount of detail leave both statistics far from natural images. Based on this finding, we present REALSYNZ, an independent module that corrects this distribution in the output of any generator without access to the model. REALSYNZ corrects the moment ratio in closed form and trains a small network that decides only where to redistribute fine detail and where to add texture. On 1,000 anachronistic FLUX.2 Klein scenes, REALSYNZ closes 65% of the envelope variation gap, outperforms fifteen baseline methods, and transfers to four other generators without retraining.

1 INTRODUCTION

Text-to-image generators have become increasingly powerful: they now render objects, composition and lighting convincingly (Esser et al., 2024; Labs, 2024; Baldridge et al., 2024; Xie et al., 2025a; Wu et al., 2025). Yet their images remain different from natural images (Wang et al., 2025). Viewers tell them apart above chance (Lu et al., 2023; Kamali et al., 2024), and detectors separate them on low-level cues that transfer across generators (Wang et al., 2020; Corvi et al., 2023b; Ojha et al., 2023; Zhong et al., 2023). The closer a generated picture comes to a natural image, the more its remaining difference stands out, producing an uncanny valley (Mori et al., 2012) of waxy skin, flat fabric and oily water (Kamali et al., 2024; Zhang et al., 2023) (Figure 1).

Historically, researchers have improved the naturalness of generated images with discriminative learning. Energy-based generative-discriminative learning (Tu, 2007) and generative adversarial networks (Goodfellow et al., 2014) train a generator by progressively reducing the largest difference that a discriminator can find between natural and synthetic images. This process has no end point, because a discriminator can still find a decision boundary that separates the images of current generators from natural images (Wang et al., 2025). The reason is that the manifold of natural images is highly complex (Field, 1987; Ruderman & Bialek, 1993; Olshausen & Field, 1996; Zhu & Mumford, 1997; Zhu et al., 1998; Simoncelli & Olshausen, 2001), while each generator explores only a limited family of images. Objects such as cats and dogs now look correct, so generators have captured the statistics that a first glance checks (Kamali et al., 2024; Corvi et al., 2023a; Zhong et al., 2023). These are only a small part of the statistics of natural images. In this work, we therefore take an explicit route: we measure which statistics of natural images a generated image misses, and we correct them in an independent module applied to the finished image.

The most common answer from statistical studies of generated images is that generators produce too little fine detail (Khayatkhoei & Elgammal, 2022; Ricker et al., 2024; Corvi et al., 2023a). Texture-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

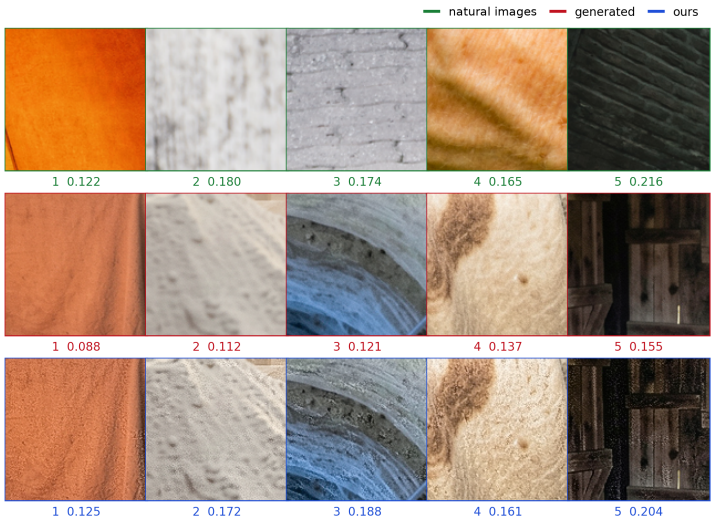
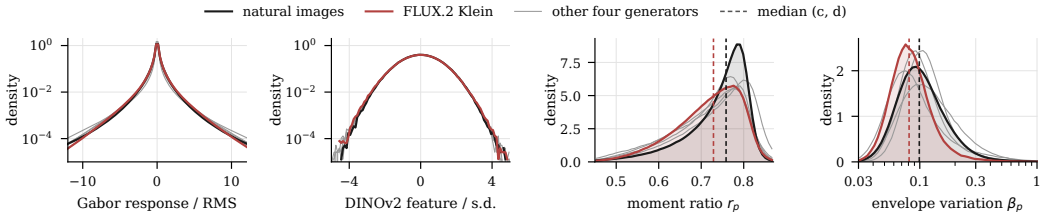


Figure 1: Generated texture that is too even. Each column shows a natural image (top), a generated crop of a similar region (middle) and that crop after REALSYNZ (bottom), labelled with its busyness bin and its envelope variation, the median over its 16×16 patches; low values mean even fine detail. Generated crops are more even; REALSYNZ brings them close.



(a) Gabor filter (b) DINOv2 feature (c) Moment ratio (d) Envelope variation

Figure 2: A classical filter and a deep feature give the same value distribution for natural and generated images, while our two statistics differ; dashed lines mark medians. (a) Gabor responses at the finest wavelength over their root mean square, 1,000 images per pool. (b) DINOv2 ViT-B/14 features (Oquab et al., 2024), each dimension standardised by its natural-image mean and standard deviation. (c) Moment ratio in the busiest bin, over 4,422 natural images and 1,000 per generator: four generators are lower, and Qwen-Image matches. (d) Envelope variation in bin 2: FLUX.2 Klein and SANA-1.5 are lower, and Qwen-Image and FLUX.1-dev higher.

editing tools built on this answer add fine detail back: they sharpen the image (Fattal et al., 2007; Farbman et al., 2008; Paris et al., 2015), add film grain (Newson et al., 2017; Norkin & Birkbeck, 2018; Lesné et al., 2025) or noise (Abdelhamed et al., 2019; Fu et al., 2023), or match the power spectrum of natural images (Durall et al., 2020; Lee et al., 2024). We test whether the amount of fine detail explains the gap. We compare the root-mean-square amplitude of the finest band b, the detail that a one-pixel blur removes, with that of comparable regions in natural images (Section 2.2). Two of the five generators we study carry as much fine detail as natural images or more, and the other three fall short by at most 40% (Appendix D.3). The amount of fine detail therefore does not explain the gap, and adding detail cannot close it.

We find that the difference lies in how fine detail is distributed. In textured regions, every generator packs fine detail into small areas and leaves the space between them nearly empty, whereas a natural image spreads it more evenly. In flatter regions, such as skies and plain walls, the direction depends on the generator: some spread fine detail too unevenly and others too evenly (Figure 1 shows the second kind). Classical models of natural images give two statistics that measure this distribution. The *moment ratio* of a patch is high when its fine detail takes many moderate values and low when it takes a few large ones (Field, 1987; Simoncelli & Olshausen, 2001). *Envelope variation* measures how much the local magnitude of fine detail varies across a patch (Wainwright & Simoncelli, 1999; Portilla et al., 2003). Both statistics separate natural from generated images, while Gabor responses and DINOv2 features do not (Figure 2). Both also stay the same when texture is scaled, so an edit

that only changes the amount of detail cannot correct them. On FLUX.2 Klein, no gain, noise or grain edit lowers the envelope variation gap by more than 1.1 points (Section 3.2).

The difference also varies with how busy a region is, so we compare both statistics within five bins of *busyness*, the image energy at a coarser scale (Bae et al., 2006). In the busiest bin, the envelope variation of every generator exceeds that of natural images by 16.5 to 31.4% (Section 2.4). In bin 2, FLUX.2 Klein, Z-Image and SANA-1.5 fall below natural images and Qwen-Image and FLUX.1-dev above (Figure 2d).

A correction must therefore change how fine detail is distributed, in a direction that depends on the bin, without changing content. Inside a patch, the moment ratio depends on a single exponent, so we solve that exponent from natural images of similar busyness. Across a region, deciding where to add fine detail and where to weaken it has no closed-form solution, so a small network learns it. The resulting method, REALSYNZ, is an independent module that takes only the image and applies to the output of any generator. On 1,000 anachronistic scenes from FLUX.2 Klein (Klein below), it closes 65% of the mean envelope variation gap and brings the mean moment-ratio gap to 0.38%, whereas the best of fifteen baselines closes 46%.

Contributions.

- **A new method.** We introduce REALSYNZ, a post-processing approach that is able to naturalize synthetic images adaptively, and works universally without the need to access the generator model architecture and parameters.
- **A new finding.** Although the gap between natural images and synthetic ones still exists, we found that statistics for envelope variation and the moment ratio serve as helpful metrics to make synthetic images look more natural.
- **A new benchmark.** We build, to our knowledge, the first benchmark for measuring texture naturalism, with 1,000 anachronistic prompts, five generators and fifteen editing methods, scored per busyness bin against the variation among natural images.

2 METHOD

We first describe the data (Section 2.1) and define two statistics of how fine detail is distributed (Section 2.2). Comparing them between natural and generated images within busyness bins (Section 2.3) leads to our finding (Section 2.4), and REALSYNZ corrects both statistics (Section 2.5).

2.1 DATA

We use 22,112 lossless Unsplash images (Unsplash, 2026). A *fit split* of 17,690 images provides every natural-image reference used in training, and a disjoint *score split* of 4,422 images is used for evaluation. We center-crop and resize every natural and generated image to 1024×1024. We generate images with five text-to-image models, FLUX.2 Klein (Labs, 2026; Black Forest Labs, 2025), Qwen-Image (Wu et al., 2025), Z-Image (Z-Image Team et al., 2025), SANA-1.5 (Xie et al., 2025b;a) and FLUX.1-dev (Labs, 2024), each generating one image for each of the same 5,000 training prompts and 1,000 disjoint test prompts. Every prompt describes a physically plausible scene that could not have been captured by a camera, such as people of one era using objects of another, so a generator cannot reproduce memorised images (Appendix A.4).

2.2 THE MOMENT RATIO AND ENVELOPE VARIATION

Sharpening, film grain and spectrum matching control the energy and power spectrum of fine detail, and Gabor filter responses (Daugman, 1985) and DINOv2 features (Oquab et al., 2024) have the same value distribution in natural and generated images (Figures 2a and 2b). The fine-detail values of natural images follow heavy-tailed distributions (Field, 1987; Ruderman & Bialek, 1993; Simoncelli & Olshausen, 2001), so we measure the shape of this distribution, which stays the same when fine detail is scaled (Figures 2c and 2d).

Band b and patches. Given an image $y \in [0,1]^{H \times W \times 3}$, let $G_n y$ denote y smoothed by n passes of the binomial kernel $h = \frac{1}{16}[1,4,6,4,1]$ (Burt & Adelson, 1983); n passes approximate a Gaussian

blur of standard deviation $\sqrt{n}$. The *finest band* $b = y - G_1 y$ is the detail that a single pass removes, averaged over the colour channels. It holds fine texture such as grain, skin pores and fabric fibres. At an object boundary, b records the edge itself, so we exclude *boundary pixels*, where the image gradient is large, and call the remaining pixels *interior* (Appendix A.1). We compute both statistics on a *patch* p, which is a 16×16 block in which at most 2% of the pixels are boundary pixels and whose mean $|b|$ exceeds a tenth of $\mathrm{med}_y\,|b|$, the median of $|b|$ over the image. For any per-pixel quantity x, we write $\mu_p[x]$ and $\mathrm{std}_p[x]$ for its mean and standard deviation over p.

Moment ratio. The moment ratio r_p of patch p is the mean absolute value of band b divided by its root mean square, $r_p = \mu_p[|b|]/\sqrt{\mu_p[b^2]}$. For Gaussian values of b, $r_p = \sqrt{2/\pi} \approx 0.80$, and r_p falls when the same energy is concentrated in a few large values, such as isolated spikes. Figure 4 in Appendix A.1 shows three patches with the same energy and power spectrum whose moment ratios are 0.79, 0.50 and 0.52. Multiplying b by a constant leaves r_p unchanged, so r_p describes the shape of the distribution of b and ignores its amount. The square of r_p is the statistic that Sharifi & Leon-Garcia (1995) use to estimate the shape parameter of a generalised Gaussian.

Envelope variation. The moment ratio pools all values of b in a patch and ignores where they lie. A patch whose texture is strong on one side and faint on the other can therefore have the same moment ratio as a patch with even texture. Envelope variation measures this spatial variation from the *envelope* $e = G_8\,|b\,s|$, the magnitude of b smoothed by eight passes of the binomial kernel h, which gives the local magnitude of fine detail around each pixel. The support mask $s \in [0, 1]$ is close to zero at single-pixel spikes and at boundaries, so the envelope excludes them (Appendix A.2). The envelope variation of patch p is the coefficient of variation of the envelope e over the patch p (Pearson, 1896), $\beta_p = \mathrm{std}_p[e]/\mu_p[e]$. It is low when fine detail is spread evenly over the patch and high when fine detail is concentrated in parts of it. In Figure 4, the patch whose texture is strong in some places and faint in others has $\beta_p = 2.04$, against 0.08 for the Gaussian patch.

2.3 Busyness bins

Both statistics vary with how busy a region is. A flat region such as a plastered wall carries a faint, even grain, and a busy region such as a wool coat carries strong fibres whose magnitude varies from place to place. We therefore compare each patch only with natural-image patches of similar *busyness*. The busyness of patch p is the energy of the image between two and eight blur passes, computed as $\mu_p[(G_2 y - G_8 y)^2]$. Busyness is similar to the textureness measure of Bae et al. (2006); it is low on sky and high on foliage. We divide patches into five *busyness bins* at the 20th, 40th, 60th and 80th percentiles of busyness over the fit-split natural images, so each bin holds a fifth of the natural-image patches. Bin $k(p)$ of patch p runs from 1 (flattest) to 5 (busiest).

2.4 Gap, criterion and natural-image floor

Let t_p denote either statistic of patch p, with $t \in \{r, \beta\}$. The *natural-image level* t_k^{nat} of bin k is the median of t_p over the natural-image patches in that bin. Evaluation uses the levels computed on the score split, and training uses the levels computed on the fit split. For a pool of generated images, t_k^{gen} is the median over its patches in bin k, and the *gap* Δ_k^t is the difference between the two medians in percent of the natural-image level. Our main criterion C_β averages the absolute envelope variation gap over the five bins, and C_r does the same for the moment ratio:

$$\Delta_k^t = 100\,\frac{t_k^{\mathrm{gen}} - t_k^{\mathrm{nat}}}{t_k^{\mathrm{nat}}}, \qquad C_\beta = \frac{1}{5}\sum_{k=1}^{5}\left|\Delta_k^\beta\right|, \qquad C_r = \frac{1}{5}\sum_{k=1}^{5}\left|\Delta_k^r\right|. \tag{1}$$

Most patches of generated images are in the busiest bins, so a mean over patches would mostly measure those bins. We therefore weight the five bins equally (Appendix C). To measure the gap that sampling alone produces, we score the score-split natural images as a generated pool against the fit-split levels. This *natural-image floor* is $C_\beta = 1.07$ and $C_r = 0.14$, and at most 2.1% and 0.4% in any busyness bin (Table 4 in Appendix A.1). In evaluation, we count the gap of a generator in a bin as a difference from natural images only when it exceeds this floor.

Finding. **Generated and natural images differ in how fine detail is distributed, which the moment ratio and envelope variation measure, and energy and power spectrum miss.**

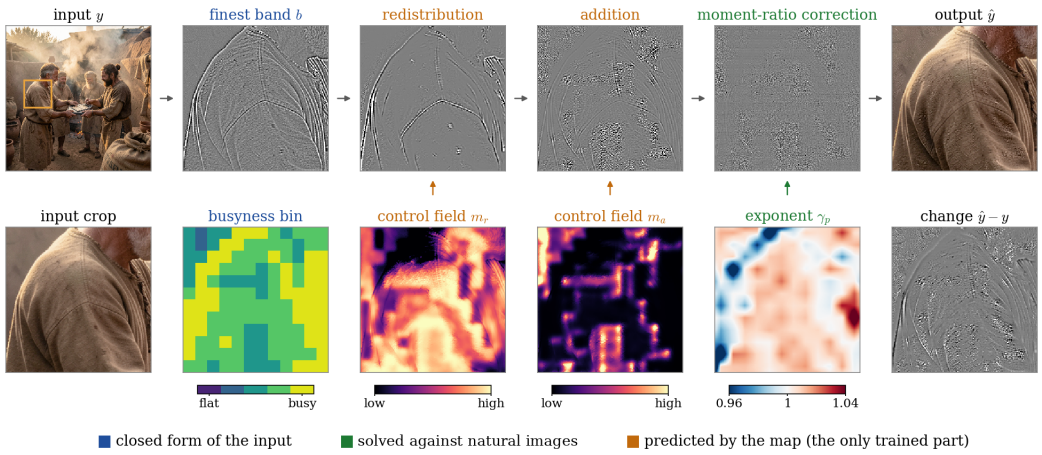


Figure 3: Overview of REALSYNZ on one FLUX.2 Klein test image. The top row follows one crop through the three steps, each shown as the change it makes to band b (mid-grey is zero); the bottom row shows the field that drives each step. The colour of each title shows how the quantity is obtained: blue quantities are computed directly from the input, green ones are solved to match natural images, and orange ones are predicted by the map. The pseudo-code is shown in Alg 1 in Appendix.

Two of the five generators carry as much fine detail in band b as natural images or more, and the other three fall short by at most 40% (Appendix D.3), so the amount does not explain the difference. In the busiest bin, every generator packs fine detail into small areas, with a envelope variation 16.5 to 31.4% above that of natural images, more than fifteen times the mean floor of 1.07 (Table 4 in Appendix A.1). The moment ratio of four generators is also lower there, by 2.6 to 4.0% against a floor of 0.4% (Figure 2c). In each of the two flattest bins, three of the five generators spread fine detail too evenly, with a lower envelope variation than natural images, and the other two too unevenly (Figure 2d). A correction must therefore lower envelope variation in some bins and raise it in others. Both statistics ignore the amount of detail, and no gain, noise or grain edit lowers C_β by more than 1.1 points (Section 3.2).

2.5 REALSYNZ

REALSYNZ corrects the two statistics of our finding with two different mechanisms. The moment ratio of a patch depends only on the shape parameter of a generalised Gaussian (Sharifi & Leon-Garcia, 1995; Mallat, 1989). One number per patch therefore controls it, and REALSYNZ solves one exponent per patch in closed form (Section 2.5.1). Envelope variation depends on how the envelope varies across the pixels of a patch, and no single number controls it. Divisive normalisation (Schwartz & Simoncelli, 2001; Lyu & Simoncelli, 2008) is the closest closed-form operator, and it either leaves every bin below the natural-image level or overshoots every bin (Section 3.2). REALSYNZ therefore corrects envelope variation by redistributing band b and by adding texture (Section 2.5.2). A small convolutional network, which we call the *map*, predicts where and how strongly to apply each of the two operations (Section 2.5.3).

Overview. Given an image y, REALSYNZ computes band b, the interior pixels and the busyness bin of each patch. The map then predicts two per-pixel *control fields* from y: m_r sets how strongly band b is redistributed, and m_a sets how much texture is added. The *operator* applies three steps to band b in order: it redistributes b according to m_r, adds texture according to m_a, and corrects the moment ratio. All three steps change b only at interior pixels, and REALSYNZ adds the change in b to y (Figure 3; Algorithm 1 in Appendix A.2).

2.5.1 CORRECTING THE MOMENT RATIO

We correct the moment ratio with a pointwise power transform of the kind used in texture synthesis (Heeger & Bergen, 1995),

$$b \longmapsto \operatorname{sign}(b)\,|b|^{\gamma}. \tag{2}$$

An exponent γ below one pulls large and small values of b closer and raises the moment ratio, and an exponent above one does the opposite. After the transform, we rescale b so that its local energy is unchanged (Appendix A.2). The correction therefore changes the distribution of b and keeps its amount. REALSYNZ solves one exponent γ_p per patch p to reach a target moment ratio $r_p^\star$.

The targets follow the full distribution of the moment ratio of natural images in the same busyness bin. We rank each patch by r_p among the patches of the same image and bin, and take as its target $r_p^\star$ the natural-image value at the same rank in that bin. This rank matching is one-dimensional optimal transport (Pitié et al., 2007; Peyré & Cuturi, 2019). Starting from $\gamma_p = 1$, we move the exponent in proportion to the relative error $(r_p^\star - r_p)/r_p^\star$, re-measure r_p and repeat for three passes (Heeger & Bergen, 1995) (Equation 8 in Appendix A.2).

2.5.2 CORRECTING ENVELOPE VARIATION

Our finding shows that envelope variation is too high in busy regions and, for some generators, too low in flatter ones. Redistributing band b lowers it, and adding texture raises it.

Redistribution. The redistribution damps band b where fine detail is concentrated and amplifies it in the pixels between. For each pixel, it compares the local envelope with the envelope of the wider surroundings through $\kappa = e_8/e_{64}$, where $e_n = G_n|b|$ is the magnitude of b blurred over n passes. κ is above one where fine detail is concentrated and below one between such places. The redistribution rescales band b as $b \mapsto b\,\kappa^{-1.5\,m_r}$. At $m_r = 0$ the band is unchanged, and a larger m_r makes the envelope more uniform; this rescaling is a form of divisive normalisation (Schwartz & Simoncelli, 2001; Lyu & Simoncelli, 2008).

Addition. Where band b is nearly zero, rescaling it has little effect, so the addition supplies new texture. Its source is a random-phase field d, which is Gaussian noise with the power spectrum of band b in natural images (Galerne et al., 2011). Where the image already carries texture, the source blends in the image's own band b. Dividing the blend by the local envelope gives a pattern u of roughly uniform magnitude (Appendix A.2), and the addition is

$$b \;\longmapsto\; b \;+\; m_a\; c\; u\; \operatorname{med}_y |b|, \tag{3}$$

where $\operatorname{med}_y |b|$ sets the added magnitude in units of the image's own band. The *busyness factor* c is the busyness of the patch around each pixel, normalised to mean one over the image. It makes the addition larger in busy regions such as cloth than in flat regions such as sky, and the control field m_a decides where the addition acts.

2.5.3 LEARNING THE CONTROL FIELDS m_r AND m_a

Within one bin, a patch whose envelope variation is already at the natural-image level needs no redistribution, and a patch with concentrated fine detail needs a strong one. One constant per bin leaves most of the envelope variation gap (Observation 10 in Section 3.4), so REALSYNZ predicts m_r and m_a at every pixel with the map. The map is a small convolutional network. Its input is the image and a one-hot indicator of each pixel's busyness bin, and it outputs $m_r \in [0, 1]$ and $m_a \geq 0$ at every pixel. It ends in one output head per bin, and each pixel takes its control fields from the head of its bin (Jacobs et al., 1991).

We train the map so that every patch reaches the natural-image level of envelope variation in its busyness bin. Generated images have far more patches in the busy bins than natural images (Appendix C). We therefore give each patch a *bin weight* $w_{k(p)}$ inversely proportional to the share π_k of training patches in its bin, $w_k = \pi_5/\pi_k$, so that every bin contributes equally (Cui et al., 2019). The loss is

$$\mathcal{L}_\beta = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} w_{k(p)} \big(\beta_p - \beta^{\text{nat}}_{k(p)}\big)^2, \tag{4}$$

where $\mathcal{P}$ is the set of patches in a training batch, β_p is the envelope variation of patch p after the edit, and $k(p)$ is its bin on the unedited image. Fine detail can be concentrated over areas of different sizes, so we compute this loss on patches of 16, 32 and 64 pixels, written $\mathcal{L}_\beta^{(16)}$, $\mathcal{L}_\beta^{(32)}$ and $\mathcal{L}_\beta^{(64)}$. At the two larger sizes, the term for bin k uses only the 16-pixel sub-patches that fall in bin k, and we call these the *bin-restricted* terms. We train the map through the full operator with the objective $\lambda_\beta \sum_s \mathcal{L}_\beta^{(s)}$ (Appendix A.3).

Table 1: Comparison with existing texture edits on FLUX.2 Klein, 1,000 test images. ↓ means lower is better. SDEdit counts once.

method	C_β, envelope variation ↓	C_r, moment ratio ↓	edit /255
natural images, the other split, the floor	1.07	0.14	–
unedited	12.90	1.07	0
uniform gain (unsharp mask)	13.86	1.22	3.90
AV1 film grain	18.11	1.77	3.87
neural film grain	22.12	1.61	5.50
white noise	22.55	1.75	3.92
pink noise	14.99	1.14	3.91
per-bin level match	11.83	1.91	0.92
textureness transfer	14.72	1.82	3.34
steerable-pyramid matching	27.75	3.47	4.04
Portilla–Simoncelli matching	10.64	1.18	3.91
NeCA-S noise model	30.02	1.27	3.91
STIG	12.43	1.26	3.94
CUT	12.78	1.05	3.97
SDEdit, strength 0.20	14.82	0.86	3.45
SDEdit, strength 0.35	13.85	0.94	3.81
closed-form envelope normalisation	6.96	1.01	3.87
envelope normalisation, per-bin target, best strength	7.32	0.60	0.65
REALSYNZ	**4.49**	**0.38**	3.92

3 EXPERIMENTS

We compare REALSYNZ with fifteen texture edits (Section 3.2), retrain and transfer it across five generators (Section 3.3), compose it with restorers (Section 3.3) and ablate it (Section 3.4).

3.1 SETUP

Implementation. The map of REALSYNZ (Section 2.5.3) has width 64 and 0.28 M parameters. We train it for 700 Adam steps at batch size 16 on 5,000 images of one generator, FLUX.2 Klein by default (Appendix A).

Baselines. We compare with fifteen texture edits: gain, noise and film grain (Norkin & Birkbeck, 2018; Lesné et al., 2025; Bae et al., 2006); classical texture-statistic matching (Portilla & Simoncelli, 2000); learned translation (Fu et al., 2023; Lee et al., 2024; Park et al., 2020); diffusion editing (Meng et al., 2022); and two controls that correct envelope variation with a fixed formula and no training (Appendix B). The *closed-form envelope normalisation* multiplies band b in each patch p by $(\mu_p[e]/e)^\eta$, which damps b where the envelope e is above the patch mean and amplifies it where e is below, with one exponent η per image (Schwartz & Simoncelli, 2001; Lyu & Simoncelli, 2008). The *per-bin-target* version applies the same gain with a separate exponent for each patch, set to move its envelope variation toward the natural-image level of its bin. We also apply REALSYNZ after three diffusion restorers.

Protocol and metrics. We run every method at 1024×1024 on the 1,000 anachronistic test images of each generator (Appendix A.4), and score it by C_β and C_r (Equation 1). Let $\hat{y}$ be the edited image. The *edit size* $\delta = 255\,\overline{|\hat{y} - y|}$ is the mean absolute error between $\hat{y}$ and y (Wang & Bovik, 2009), where the bar averages over pixels and channels; we report its median over test images. Appendix D.1 reports the change in image content.

3.2 COMPARISON WITH EXISTING TEXTURE-EDITING METHODS

***Observation* 1.** Does changing the amount of fine detail bring envelope variation closer to natural images?

No. Of the seven baselines in Table 1 that change the amount of fine detail, six end with a larger C_β than the unedited 12.90. White noise and neural film grain are the worst, at 22.55 and 22.12, because they add the same field to textured and smooth regions; only the per-bin level match improves, to 11.83. This agrees with Portilla & Simoncelli (2000), who reproduced a texture only after matching the correlations of its coefficient magnitudes as well as its spectrum.

Table 2: Ablation on FLUX.2 Klein. No row beats REALSYNZ on both criteria (Appendix D.4).

variant	C_β, envelope variation ↓	C_r, moment ratio ↓	edit /255
REALSYNZ	4.49	0.38	3.92
trained map frozen, one step of the operator switched off			
the moment-ratio correction run alone, redistribution and addition off	11.84	**0.16**	0.56
without the moment-ratio correction	**4.12**	3.91	4.29
without the addition	21.71	0.14	1.31
without the redistribution	5.02	0.53	4.12
map retrained with one design element removed			
without the bin indicator and the per-bin heads	5.59	0.20	4.37
without the bin-restricted patch terms	7.09	0.19	4.09
without the 32- and 64-pixel patch terms	6.23	0.37	3.86
two learned constants per bin in place of the two predicted fields	10.11	0.19	3.63

***Observation* 2.** Is correcting envelope variation with a fixed formula enough?

No. The two envelope-normalisation controls of Section 3.1 are the strongest baselines, and the closed-form version is the best of the fifteen, at $C_\beta = 6.96$. The per-bin-target version reaches its best C_β of 7.32 at an edit of $0.65/255$ and overshoots every bin at an edit of $3.88/255$, where C_β is 55.62 (Appendix B). Both controls lower the envelope variation gap, as expected from divisive normalisation, which removes much of the dependence between neighbouring band magnitudes in natural images (Schwartz & Simoncelli, 2001; Lyu & Simoncelli, 2008). REALSYNZ lowers C_β further, to 4.49.

***Observation* 3.** In which bins does REALSYNZ still leave a envelope variation gap?

In the two flattest bins, where its envelope variation stays 12.8 and 4.4 points above the natural-image level (Figure 8 in Appendix C). REALSYNZ ends within 2.1 points of the level in the three busiest bins, which cover over nine tenths of the patches, and within the natural-image floor on the moment ratio in three of five bins. In the flattest bin, which holds 2.6% of the patches, it overshoots to +12.8, because band b is nearly empty there and the edit raises its energy to six times the natural-image level (Appendix C and Figure 6).

3.3 ACROSS GENERATORS AND RESTORERS

***Observation* 4.** Does REALSYNZ generalize across generators?

Yes. Trained on each generator's own images, REALSYNZ lowers C_β from between 12.41 and 19.85 unedited to between 3.61 and 5.00 on all five generators (Table 7 in Appendix C). It lowers C_r on four of them, and AV1 film grain and pink noise end above it on every generator (Table 8).

***Observation* 5.** Does REALSYNZ trained on one generator transfer to others?

Yes, on three of the four other generators, in line with the finding of Wang et al. (2020) that low-level artefacts learned on one generator recur in others. Applied unchanged, REALSYNZ trained on FLUX.2 Klein lowers C_β by 40%, 64%, 57% and 45% on Qwen-Image, Z-Image, SANA-1.5 and FLUX.1-dev, and on the first three it also beats AV1 film grain and pink noise. On FLUX.1-dev, whose bin composition differs most from Klein's (Appendix C), its C_β of 7.83 is above the 5.69 of film grain and 7.13 of pink noise, at less than half their edit.

***Observation* 6.** Does REALSYNZ still help after a diffusion restorer?

Yes. On the outputs of HYPIR (Lin et al., 2025), InvSR (Yue et al., 2025) and PiSA-SR (Sun et al., 2025), which leave C_β near or above the unedited 12.90, REALSYNZ lowers C_β for all three, to 3.14 after PiSA-SR, while MUSIQ falls by at most 3.0 and NIQE rises by at most 1.9 (Appendix D.2).

3.4 ABLATION

Table 2 switches off one step of the operator with the trained map frozen (first section) or retrains the map without one design element (second section).

***Observation* 7.** Does correcting one statistic correct the other?

No. Run alone, the moment-ratio correction reaches $C_r = 0.16$ but leaves C_β at 11.84 (Table 2), barely below the unedited 12.90, since it keeps the spatial variation of b.

***Observation* 8.** Which step of the operator matters most?

In Table 2, removing the addition costs the most: C_β rises to 21.71, worse than no edit (per-bin gaps in Appendix D.4). Removing the redistribution raises C_β by 0.53, and removing the moment-ratio correction raises C_r from 0.38 to 3.91.

***Observation* 9.** Which parts of the training objective matter?

In the second section of Table 2, C_β rises from 4.49 to 5.59 without the indicator and the per-bin heads, mostly in bins 3 and 4, and to 7.09 without the bin restriction, mostly in the two flattest bins. Without the 32- and 64-pixel terms, C_β rises to 6.23 and bins 3 and 4 fall 8 to 9 points below the natural-image level (Table 17), so 16-pixel supervision alone leaves larger patches uncorrected.

***Observation* 10.** Must the edit vary within a busyness bin?

Yes. In the last row of Table 2, we replace the control fields m_r and m_a (Section 2.5.2) with one learned constant per busyness bin for each field. C_β then rises to 10.11, closer to the unedited 12.90 than to 4.49, so correcting envelope variation needs a per-pixel field.

4 Related work

Generation-side methods. Preference, reinforcement-learning and detector-reward finetuning make what a generator draws more realistic (Xu et al., 2023; Kirstain et al., 2023; Wu et al., 2023; Black et al., 2024; Wallace et al., 2024; Ye et al., 2026), and energy-based and adversarial training (Tu, 2007; Jin et al., 2017; Lazarow et al., 2017; Lee et al., 2018; Goodfellow et al., 2014) trains the generator against a discriminator; RealSynz edits the finished image.

Post-hoc editing. Diffusion editing and restoration redraw the image through a generative prior (Meng et al., 2022; Brooks et al., 2023; Yu et al., 2024; Lin et al., 2025; Yue et al., 2025; Sun et al., 2025), and translation networks map generated to natural images (Isola et al., 2017; Zhu et al., 2017; Park et al., 2020; Parmar et al., 2024; Richter et al., 2023; Pasios & Nikolaidis, 2026). Edge-aware detail manipulation (Bae et al., 2006; Fattal et al., 2007; Farbman et al., 2008; Paris et al., 2015; He et al., 2013; Aubry et al., 2014) and film-grain and noise synthesis (Newson et al., 2017; Norkin & Birkbeck, 2018; Lesné et al., 2025; Abdelhamed et al., 2019; Fu et al., 2023) change the amount of fine detail, and STIG (Lee et al., 2024) matches the power spectrum.

Natural-image and texture statistics. Band-pass coefficients of natural images are heavy-tailed (Field, 1987; Mallat, 1989; Ruderman & Bialek, 1993; Olshausen & Field, 1996) and follow Gaussian scale mixtures (Wainwright & Simoncelli, 1999; Simoncelli & Olshausen, 2001; Portilla et al., 2003), and FRAME-type models learn such statistics as priors (Zhu & Mumford, 1997; Zhu et al., 1998; Tu, 2007). Our correction follows statistics matching (Heeger & Bergen, 1995; Portilla & Simoncelli, 2000; Gatys et al., 2015) and adds random-phase texture (Galerne et al., 2011).

Metrics and synthetic-image detection. Quality models, full-reference measures and distribution distances score overall quality (Ke et al., 2021; Yang et al., 2022; Chen et al., 2024; Wu et al., 2024; Zhang et al., 2018; Ding et al., 2022; Heusel et al., 2017; Parmar et al., 2022), and band statistics (Lyu & Farid, 2005) and detectors on low-level cues (Wang et al., 2020; Frank et al., 2020; Corvi et al., 2023b; Ojha et al., 2023; Tan et al., 2024; Zhong et al., 2023) separate generated from natural images; RealSynz names two statistics that differ and corrects both.

5 Conclusion and future work

Generated images distribute fine texture differently from natural images, and the moment ratio and envelope variation measure this difference, which energy and spectra miss. On FLUX.2 Klein, RealSynz removes 65% of the envelope variation gap, against at most 46% for fifteen baselines, brings the mean moment-ratio gap to 0.38%, and transfers to four other generators. Future work includes a perceptual study (Appendix B.1), a correction for the flattest regions, where RealSynz overshoots, and leaving natural images unchanged (Appendix D.1).

AI USE STATEMENT

AI tools were used only to assist with visualizing results and to make minor refinements to language and grammar. The authors reviewed the AI-assisted content and take responsibility for the final text, figures, and scientific claims.

REPRODUCIBILITY STATEMENT

Everything needed to reproduce our results is included in the appendix. Appendix A gives the band, the two statistics, the busyness bin edges, the natural-image levels and every step of the operator in closed form, together with the architecture of the map, its loss, and a table of all hyperparameters and constants; every trained map uses seed 101. Appendix A.4 gives the prompt template and fields from which the benchmark is generated, and Appendix B gives the settings of every baseline. The natural images come from the public Unsplash corpus.

REFERENCES

Abdelrahman Abdelhamed, Marcus A. Brubaker, and Michael S. Brown. Noise flow: Noise modeling with conditional normalizing flows. In *ICCV*, pp. 3165–3173, 2019. URL https://doi.org/10.1109/ICCV.2019.00326.

Mathieu Aubry, Sylvain Paris, Samuel W. Hasinoff, Jan Kautz, and Frédo Durand. Fast local laplacian filters: Theory and applications. *ACM Transactions on Graphics*, 33(5):1–14, 2014. ISSN 1557-7368. doi: 10.1145/2629645.

Soonmin Bae, Sylvain Paris, and Frédo Durand. Two-scale tone management for photographic look. *ACM Transactions on Graphics*, 25(3):637–645, 2006. ISSN 1557-7368. doi: 10.1145/1141911.1141935.

B. Balas, L. Nakano, and R. Rosenholtz. A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12):13–13, November 2009. ISSN 1534-7362. doi: 10.1167/9.12.13.

Jason Baldridge, Jakob Bauer, Mukul Bhutani, Nicole Brichtova, Andrew Bunner, Kelvin Chan, Yichang Chen, Sander Dieleman, Yuqing Du, Zach Eaton-Rosen, Hongliang Fei, Nando de Freitas, Yilin Gao, Evgeny Gladchenko, Sergio Gómez Colmenarejo, Mandy Guo, Alex Haig, Will Hawkins, Hexiang Hu, Huilian Huang, Tobenna Peter Igwe, Christos Kaplanis, Siavash Khodadadeh, Yelin Kim, Ksenia Konyushkova, Karol Langner, Eric Lau, Shixin Luo, Sona Mokrá, Henna Nandwani, Yasumasa Onoe, Aäron van den Oord, Zarana Parekh, Jordi Pont-Tuset, Hang Qi, Rui Qian, Deepak Ramachandran, Poorva Rane, Abdullah Rashwan, Ali Razavi, Robert Riachi, Hansa Srinivasan, Srivatsan Srinivasan, Robin Strudel, Benigno Uria, Oliver Wang, Su Wang, Austin Waters, Chris Wolff, Auriel Wright, Zhisheng Xiao, Hao Xiong, Keyang Xu, Marc van Zee, Junlin Zhang, Katie Zhang, Wenlei Zhou, Konrad Zolna, Ola Aboubakar, Canfer Akbulut, Oscar Akerlund, Isabela Albuquerque, Nina Anderson, Marco Andreetto, Lora Aroyo, Ben Bariach, David Barker, Sherry Ben, Dana Berman, Courtney Biles, Irina Blok, Pankil Botadra, Jenny Brennan, Karla Brown, John Buckley, Rudy Bunel, Elie Bursztein, Christina Butterfield, Ben Caine, Viral Carpenter, Norman Casagrande, Ming-Wei Chang, Solomon Chang, Shamik Chaudhuri, Tony Chen, John Choi, Dmitry Churbanau, Nathan Clement, Matan Cohen, Forrester Cole, Mikhail Dektiarev, Vincent Du, Praneet Dutta, Tom Eccles, Ndidi Elue, Ashley Feden, Shlomi Fruchter, Frankie Garcia, and Roopal Garg. Imagen 3. *CoRR*, abs/2408.07009, 2024. URL https://doi.org/10.48550/arXiv.2408.07009.

Quentin Bammey. Synthbuster: Towards detection of diffusion model generated images. *IEEE Open Journal of Signal Processing*, 5:1–9, 2024. ISSN 2644-1322. doi: 10.1109/ojsp.2023.3337714.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=YCWjhGrJFD.

Black Forest Labs. FLUX.2-dev, 2025. URL https://huggingface.co/black-forest-labs/FLUX.2-dev.

Tim Brooks, Aleksander Holynski, and Alexei A. Efros. Instructpix2pix: Learning to follow image editing instructions. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18392–18402. IEEE, 2023. doi: 10.1109/cvpr52729.2023.01764.

P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983. ISSN 0096-2244. doi: 10.1109/tcom.1983.1095851.

Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Trans. Image Process.*, 33:2404–2418, 2024. URL https://doi.org/10.1109/TIP.2024.3378466.

Riccardo Corvi, Davide Cozzolino, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. Intriguing properties of synthetic images: from generative adversarial networks to diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 973–982. IEEE, 2023a. doi: 10.1109/cvprw59228.2023.00104.

Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. On the detection of synthetic images generated by diffusion models. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023b. doi: 10.1109/icassp49357.2023.10095167.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9260–9269. IEEE, 2019. doi: 10.1109/cvpr.2019.00949.

John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, Jul 1985. doi: 10.1364/JOSAA.2.001160. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-2-7-1160.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(5):2567–2581, 2022. URL https://doi.org/10.1109/TPAMI.2020.3045810.

Ricard Durall, Margret Keuper, and Janis Keuper. Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In *CVPR*, pp. 7887–7896, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Durall_Watch_Your_Up-Convolution_CNN_Based_Generative_Deep_Neural_Networks_Are_CVPR_2020_paper.html.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=FPnUhsQJ5B.

Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 27(3):1–10, August 2008. ISSN 1557-7368. doi: 10.1145/1360612.1360666.

Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.*, 26(3):51, 2007. URL https://doi.org/10.1145/1276377.1276441.

David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379, December 1987. ISSN 1520-8532. doi: 10.1364/josaa.4.002379.

Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz. Leveraging frequency analysis for deep fake image recognition. In *ICML*, pp. 3247–3258, 2020. URL http://proceedings.mlr.press/v119/frank20a.html.

Jeremy Freeman and Eero P Simoncelli. Metamers of the ventral stream. *Nature Neuroscience*, 14 (9):1195–1201, August 2011. ISSN 1546-1726. doi: 10.1038/nn.2889.

Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, 16(7):974–981, May 2013. ISSN 1546-1726. doi: 10.1038/nn.3402.

Zixuan Fu, Lanqing Guo, and Bihan Wen. srgb real noise synthesizing with neighboring correlation-aware noise model. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1683–1691. IEEE, 2023. doi: 10.1109/cvpr52729.2023.00168.

Bruno Galerne, Yann Gousseau, and Jean-Michel Morel. Random phase textures: Theory and synthesis. *IEEE Trans. Image Process.*, 20(1):257–267, 2011. URL https://doi.org/10.1109/TIP.2010.2052822.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. In *NIPS*, pp. 262–270, 2015. URL http://papers.nips.cc/paper/5633-texture-synthesis-using-convolutional-neural-networks.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pp. 2672–2680, 2014. URL http://papers.nips.cc/paper/5423-generative-adversarial-nets.

Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013. ISSN 2160-9292. doi: 10.1109/tpami.2012.213.

David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*, SIGGRAPH '95, pp. 229–238. ACM Press, 1995. doi: 10.1145/218380.218446.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NIPS*, pp. 6629–6640, 2017.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976. IEEE, 2017. doi: 10.1109/cvpr.2017.632.

Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, February 1991. ISSN 1530-888X. doi: 10.1162/neco.1991.3.1.79.

Long Jin, Justin Lazarow, and Zhuowen Tu. Introspective classification with convolutional nets. In *NIPS*, pp. 823–833, 2017. URL http://papers.nips.cc/paper/6684-introspective-classification-with-convolutional-nets.

Negar Kamali, Karyn Nakamura, Angelos Chatzimparmpas, Jessica Hullman, and Matthew Groh. How to distinguish ai-generated images from authentic photographs. *CoRR*, abs/2406.08651, 2024. URL https://doi.org/10.48550/arXiv.2406.08651.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *ICCV*, pp. 5128–5137, 2021. URL https://doi.org/10.1109/ICCV48922.2021.00510.

Mahyar Khayatkhoei and Ahmed Elgammal. Spatial frequency bias in convolutional generative adversarial networks. In *AAAI*, pp. 7152–7159, 2022. URL https://doi.org/10.1609/aaai.v36i7.20675.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. In *Advances in Neural Information Processing Systems 36*, NeurIPS 2023, pp. 36652–36663. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023. doi: 10.52202/075280-1594.

Black Forest Labs. Flux. https://github.com/black-forest-labs/flux, 2024.

Black Forest Labs. Flux.2 [klein]: Fast, efficient image generation and editing, 2026. URL https://github.com/black-forest-labs/flux2.

Justin Lazarow, Long Jin, and Zhuowen Tu. Introspective neural networks for generative modeling. In *ICCV*, pp. 2793–2802, 2017. URL https://doi.org/10.1109/ICCV.2017.302.

Kwonjoon Lee, Weijian Xu, Fan Fan, and Zhuowen Tu. Wasserstein introspective neural networks. In *CVPR*, pp. 3702–3711, 2018. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Lee_Wasserstein_Introspective_Neural_CVPR_2018_paper.html.

Seokjun Lee, Seung-Won Jung, and Hyunseok Seo. Spectrum translation for refinement of image generation (stig) based on contrastive learning and spectral filter profile. In *AAAI*, pp. 2929–2937, 2024. URL https://doi.org/10.1609/aaai.v38i4.28074.

G. Lesné, Y. Gousseau, S. Ladjal, and A. Newson. Neural film grain rendering. *Computer Graphics Forum*, 44(2), April 2025. ISSN 1467-8659. doi: 10.1111/cgf.70076.

Xinqi Lin, Fanghua Yu, Jinfan Hu, Zhiyuan You, Wu Shi, Jimmy S. Ren, Jinjin Gu, and Chao Dong. Harnessing diffusion-yielded score priors for image restoration. *ACM Trans. Graph.*, 44 (6):208:1–208:21, December 2025. URL https://doi.org/10.1145/3763346.

Zeyu Lu, Di Huang, Lei Bai, Jingjing Qu, Chengyue Wu, Xihui Liu, and Wanli Ouyang. Seeing is not always believing: Benchmarking human and model perception of ai-generated images. In *Advances in Neural Information Processing Systems 36*, NeurIPS 2023, pp. 25435–25447. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023. doi: 10.52202/075280-1105.

Siwei Lyu and Hany Farid. How realistic is photorealistic? *IEEE Trans. Signal Process.*, 53(2-2): 845–850, 2005. URL https://doi.org/10.1109/TSP.2004.839896(410)53.

Siwei Lyu and Eero P. Simoncelli. Nonlinear image representation using divisive normalization. In *CVPR*, 2008. URL https://doi.org/10.1109/CVPR.2008.4587821.

S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. ISSN 0162-8828. doi: 10.1109/34.192463.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=aBsCjcPu_tE.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2013. URL https://doi.org/10.1109/LSP.2012.2227726.

Masahiro Mori, Karl MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012. ISSN 1070-9932. doi: 10.1109/mra.2012.2192811.

Alasdair Newson, Noura Faraj, Bruno Galerne, and Julie Delon. Realistic film grain rendering. *Image Process. Line*, 7:165–183, 2017. URL https://doi.org/10.5201/ipol.2017.192.

Andrey Norkin and Neil Birkbeck. Film grain synthesis for av1 video codec. In *2018 Data Compression Conference*, pp. 3–12. IEEE, March 2018. doi: 10.1109/dcc.2018.00008.

Utkarsh Ojha, Yuheng Li, and Yong Jae Lee. Towards universal fake image detectors that generalize across generative models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24480–24489. IEEE, 2023. doi: 10.1109/cvpr52729.2023.02345.

Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996. ISSN 1476-4687. doi: 10.1038/381607a0.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt. Featured Certification.

Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. Local laplacian filters: edge-aware image processing with a laplacian pyramid. *Communications of the ACM*, 58(3):81–91, February 2015. ISSN 1557-7317. doi: 10.1145/2723694.

Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *ECCV (9)*, pp. 319–345, 2020. URL https://doi.org/10.1007/978-3-030-58545-7_19.

Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On aliased resizing and surprising subtleties in gan evaluation. In *CVPR*, pp. 11400–11410, 2022. URL https://doi.org/10.1109/CVPR52688.2022.01112.

Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models, 2024.

Stefanos Pasios and Nikos Nikolaidis. Regen: Real-time photorealism enhancement in games via a dual-stage generative network framework. *IEEE Transactions on Games*, 18(2):404–410, 2026. ISSN 2475-1510. doi: 10.1109/tg.2026.3661622.

Karl Pearson. Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, (187):253–318, December 1896. ISSN 2053-9258. doi: 10.1098/rsta.1896.0007.

Gabriel Peyré and Marco Cuturi. Computational optimal transport with applications to data sciences. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, February 2019. ISSN 1935-8245. doi: 10.1561/2200000073.

François Pitié, Anil C. Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2):123–137, 2007. ISSN 1077-3142. doi: 10.1016/j.cviu.2006.11.011.

Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, 2000. URL https://doi.org/10.1023/A:1026553619983.

Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12(11):1338–1351, 2003. URL https://doi.org/10.1109/TIP.2003.818640.

Stephan R. Richter, Hassan Abu Alhaija, and Vladlen Koltun. Enhancing photorealism enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1700–1715, February 2023. ISSN 1939-3539. doi: 10.1109/tpami.2022.3166687.

Jonas Ricker, Simon Damm, Thorsten Holz, and Asja Fischer. Towards the detection of diffusion model deepfakes. In *VISIGRAPP (4): VISAPP*, pp. 446–457, 2024. URL https://doi.org/10.5220/0012422000003660.

Daniel L. Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. In *NIPS*, pp. 551–558, 1993. URL http://papers.nips.cc/paper/835-statistics-of-natural-images-scaling-in-the-woods.

Odelia Schwartz and Eero P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, August 2001. ISSN 1546-1726. doi: 10.1038/90526.

K. Sharifi and A. Leon-Garcia. Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):52–56, 1995. ISSN 1051-8215. doi: 10.1109/76.350779.

Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, March 2001. ISSN 1545-4126. doi: 10.1146/annurev.neuro.24.1.1193.

Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In *CVPR*, pp. 2333–2343, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Sun_Pixel-level_and_Semantic-level_Adjustable_Super-resolution_A_Dual-LoRA_Approach_CVPR_2025_paper.html.

Chuangchuang Tan, Huan Liu, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, and Yunchao Wei. Rethinking the up-sampling operations in cnn-based generative network for generalizable deepfake detection. In *CVPR*, pp. 28130–28139, 2024. URL https://doi.org/10.1109/CVPR52733.2024.02657.

Zhuowen Tu. Learning generative models via discriminative approaches. In *CVPR*, 2007. URL https://doi.org/10.1109/CVPR.2007.383035.

Unsplash. Unsplash: the internet's source of freely usable images. unsplash.com, licence at unsplash.com/license, 2026.

Martin J. Wainwright and Eero P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. In *NIPS*, pp. 855–861, 1999. URL http://papers.nips.cc/paper/1750-scale-mixtures-of-gaussians-and-the-statistics-of-natural-images.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8228–8238. IEEE, 2024. doi: 10.1109/cvpr52733.2024.00786.

Haowen Wang, Guowei Zhang, Xiang Zhang, Zeyuan Chen, Haiyang Xu, Dou Hoon Kwark, and Zhuowen Tu. Exploring the equivalence of closed-set generative and real data augmentation in image classification, 2025.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, pp. 2555–2563, 2023. URL https://doi.org/10.1609/aaai.v37i2.25353.

Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A. Efros. Cnn-generated images are surprisingly easy to spot... for now. In *CVPR*, pp. 8692–8701, 2020. URL https://doi.org/10.1109/CVPR42600.2020.00872.

Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009. doi: 10.1109/MSP.2008.930649.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Shengming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuting Yu, Tingkun Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan Cai, and Zenan Liu. Qwen-image technical report. *CoRR*, abs/2508.02324, August 2025. URL https://doi.org/10.48550/arXiv.2508.02324.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-align: Teaching LMMs for visual scoring via discrete text-defined levels. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=PHjkVjR78A.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *CoRR*, abs/2306.09341, 2023. URL https://doi.org/10.48550/arXiv.2306.09341.

Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, and Song Han. SANA: Efficient high-resolution text-to-image synthesis with linear diffusion transformers. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=N8Oj1XhtYZ.

Enze Xie, Junsong Chen, Yuyang Zhao, Jincheng YU, Ligeng Zhu, Yujun Lin, Zhekai Zhang, Muyang Li, Junyu Chen, Han Cai, Bingchen Liu, Daquan Zhou, and Song Han. SANA 1.5: Efficient scaling of training-time and inference-time compute in linear diffusion transformer. In *Forty-second International Conference on Machine Learning*, 2025b. URL https://openreview.net/forum?id=27hOkXzy9e.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. In *Advances in Neural Information Processing Systems 36*, NeurIPS 2023, pp. 15903–15935. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023. doi: 10.52202/075280-0700.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR Workshops*, pp. 1190–1199, 2022. URL https://doi.org/10.1109/CVPRW56347.2022.00126.

Junyan Ye, Leqi Zhu, Yuncheng Guo, Dongzhi Jiang, Zilong Huang, Yifan Zhang, Zhiyuan Yan, Haohuan Fu, Conghui He, and Weijia Li. *RealGen: Photorealistic Text-to-Image Generation via Detector-Guided Rewards*, pp. 589–607. Springer Nature Switzerland, 2026. ISBN 9783032370921. doi: 10.1007/978-3-032-37092-1_33.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *CVPR*, pp. 25669–25680, 2024. URL https://doi.org/10.1109/CVPR52733.2024.02425.

Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23153–23163. IEEE, 2025. doi: 10.1109/cvpr52734.2025.02156.

Z-Image Team, Huanqia Cai, Sihan Cao, Ruoyi Du, Peng Gao, Steven C. H. Hoi, Zhaohui Hou, Shijie Huang, Dengyang Jiang, Xin Jin, Liangchen Li, Zhen Li, Zhong-Yu Li, David Liu, Dongyang Liu, Junhan Shi, Qilong Wu, Feng Yu, Chi Zhang, Shifeng Zhang, and Shilin Zhou. Z-image: An efficient image generation foundation model with single-stream diffusion transformer. *CoRR*, abs/2511.22699, November 2025. URL https://doi.org/10.48550/arXiv.2511.22699.

Lingzhi Zhang, Zhengjie Xu, Connelly Barnes, Yuqian Zhou, Qing Liu, He Zhang, Sohrab Amirghodsi, Zhe Lin, Eli Shechtman, and Jianbo Shi. Perceptual artifacts localization for image synthesis tasks. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7545–7556. IEEE, October 2023. doi: 10.1109/iccv51070.2023.00697.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pp. 586–595, 2018.

Nan Zhong, Yiran Xu, Sheng Li, Zhenxing Qian, and Xinpeng Zhang. Patchcraft: Exploring texture patch for efficient ai-generated image detection, 2023.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pp. 2242–2251, 2017. URL https://doi.org/10.1109/ICCV.2017.244.

Song Chun Zhu and David Mumford. Prior learning and gibbs reaction-diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(11):1236–1250, 1997. URL https://doi.org/10.1109/34.632983.

Song Chun Zhu, Ying Nian Wu, and David Mumford. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):107–126, 1998. URL https://doi.org/10.1023/A:1007925832420.

Corey M. Ziemba, Jeremy Freeman, J. Anthony Movshon, and Eero P. Simoncelli. Selectivity and tolerance for visual texture in macaque v2. *Proceedings of the National Academy of Sciences*, 113(22), May 2016. ISSN 1091-6490. doi: 10.1073/pnas.1510847113.

A IMPLEMENTATION DETAILS

This appendix gives the details of Section 2 in the order the method uses them. It covers the two statistics and the criterion (Appendix A.1), the operator (Appendix A.2), the map and its training (Appendix A.3), and the benchmark prompts (Appendix A.4). Table 3 collects every symbol with the place in the main text that defines it, and Table 5 lists every constant.

Table 3: Notation. The last column gives the section that defines each symbol.

symbol	meaning	defined in
y, $\hat{y}$	input image and edited image	Sections 2.2, 2.5
G_n	blur by n passes of the binomial kernel	Section 2.2
b	finest band $y - G_1 y$, the detail one blur pass removes	Section 2.2
p, $\mathcal{P}(y)$	a 16×16 patch inside one region, and the set of patches of y	Section 2.2, Appendix A.1
μ_p, std_p	mean and standard deviation over the pixels of p	Section 2.2
$\mathrm{med}_y \lvert b \rvert$	median magnitude of b over the image	Section 2.2
r_p	moment ratio of patch p	Section 2.2
s	support mask, the share of b at a pixel that is texture	Section 2.2, Appendix A.2
e	envelope $G_8 \lvert b\, s \rvert$	Section 2.2
β_p	envelope variation of patch p	Section 2.2
$k(p)$	busyness bin of p, from 1 (flattest) to 5 (busiest)	Section 2.3
t_k^{nat}	natural-image level, the median of statistic t over natural images' patches in bin k	Section 2.4
Δ_k^t	gap of a generated pool in bin k, in percent of the natural-image level	Equation 1
C_β, C_r	mean absolute gap over the five bins	Equation 1
$r_p^\star$, γ_p	ratio target and solved exponent of patch p	Section 2.5.1
κ	local envelope over that of its wider surroundings, e_8/e_{64}	Section 2.5.2
m_r, m_a	redistribution and addition control fields predicted by the map	Sections 2.5.2, 2.5.3
u, c	added pattern and busyness factor	Section 2.5.2
v	interior field, one inside a region and zero at a boundary	Appendix A.2
E_8, ρ, δ	low-pass content change, low-pass correlation and edit size	Equation 12

A.1 THE TWO STATISTICS AND THE CRITERION

Patches. Section 2.2 computes both statistics on patches that lie inside one region. A pixel is a boundary pixel where the luminance gradient, after 16 blur passes, exceeds eight times its median over the image. A patch is a 16×16 square of pixels with at most 2% boundary pixels and a mean $|b|$ above a tenth of $\mathrm{med}_y\, |b|$, so that it holds some texture. Sixteen pixels is small enough that most patches lie inside one material. We compute the boundary test and the busyness bins on the unedited image, so every method is scored on the same patches.

Table 4: The gap between generated images and natural images, per busyness bin from flattest to busiest. For each statistic, the first row is the natural-image level and the second the natural-image floor (Appendix A.1); generator rows give the unedited gap on the test images in percent of that level, and the last column is C_r or C_β (Equation 1).

statistic	row	bin 1	bin 2	bin 3	bin 4	bin 5	mean abs
moment ratio	natural images, level	0.798	0.795	0.780	0.759	0.756	
	natural images, the floor	−0.2	−0.0	−0.1	−0.0	−0.4	0.14
	FLUX.2 Klein	+0.0	+0.4	+0.8	−0.4	−3.7	1.07
	Qwen-Image	−0.9	−0.4	+1.0	+1.5	+0.1	0.78
	Z-Image	−4.0	−0.7	−0.3	−1.4	−4.0	2.08
	SANA-1.5	+0.0	+0.4	−0.3	−1.8	−2.6	1.02
	FLUX.1-dev	−0.2	+0.2	+0.4	−1.0	−3.6	1.08
envelope variation	natural images, level	0.089	0.103	0.154	0.195	0.217	
	natural images, the floor	+2.1	+0.4	+2.1	+0.8	−0.0	1.07
	FLUX.2 Klein	−4.1	−20.2	−21.2	−2.5	+16.5	12.90
	Qwen-Image	+6.3	+6.1	+2.4	+20.0	+27.2	12.41
	Z-Image	+46.4	−2.8	+5.6	+16.6	+27.7	19.85
	SANA-1.5	−14.5	−14.0	+1.8	+14.4	+18.4	12.62
	FLUX.1-dev	−8.9	+6.7	+6.6	+17.4	+31.4	14.22

Natural-image level and floor. The natural-image level of bin k is the median of a statistic over the patches in that bin of the 4,422 score-split natural images. The natural-image floor is the gap of the score-split natural images measured against the fit-split levels. Section 2.4 counts a gap of a generated pool only when it exceeds this floor. The shaded region in Figure 8 measures the same sampling variation in a second way. It is the 90th percentile, over 200 random splits of the score split into halves, of the gap between the two halves. The method uses the fit-split levels, which differ from the score-split levels by at most 1.0% on envelope variation and 0.15% on the moment ratio.

Choice of the two statistics and of per-bin gaps. Both statistics measure how fine detail is distributed and are invariant to its amount, since multiplying band b of a patch by a constant leaves both unchanged. We measure distribution because FLUX.2 Klein already carries at least as much fine detail in band b as natural images in the same busyness bin, at 105 to 157% of their root-mean-square amplitude (Table 15). Energy, the power spectrum and other second-order statistics therefore find no deficit of fine detail in its unedited images (Figure 4). The two statistics follow the Gaussian scale mixture model of natural bands (Wainwright & Simoncelli, 1999), which writes a band as a Gaussian variable times a slowly varying multiplier. The moment ratio measures the distribution of values inside a patch, and envelope variation measures how their multiplier varies across a region.

We compute every gap per busyness bin because generated images and natural images contain different mixes of flat and busy regions. Four of the five generators put 64 to 75% of their patches in the two busiest bins, where natural images put 37%. A statistic pooled over all patches would therefore compare different mixes. Busyness is computed from the image between $n = 2$ and $n = 8$ passes, a range the edit leaves unchanged, so every patch keeps its bin.

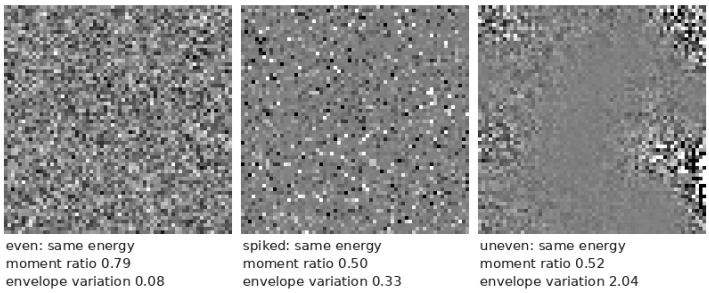


Figure 4: Three synthetic bands of identical energy, shown at 64 pixels and enlarged $4\times$. Left: a Gaussian field. Middle: the same energy concentrated in sparse spikes. Right: the same energy under a slowly varying multiplier. Energy and spectrum are the same for all three. The moment ratio measures 0.79, 0.50 and 0.52, and envelope variation separates the uneven field from the other two at 2.04 against 0.08 and 0.33.

A.2 The operator

This subsection gives the two masks the operator uses and the full form of the redistribution and the addition of Section 2.5.2. It then gives the exponent update of the moment-ratio correction and the inference steps.

Masks. The interior field $v(y) \in [0, 1]$ marks where the edit may act. It is one inside a region and falls to zero at a boundary. We build it from the patch test of Appendix A.1 without the requirement on band b, and shrink it by one patch at every boundary. A guided filter of radius 32 on luminance brings it to pixel resolution, and we clamp it to $[0, 1]$.

The support mask $s(y) \in [0, 1]$ measures how much of band b at a pixel is texture, and Section 2.2 uses it to compute the envelope e. A texture pixel has neighbours of similar magnitude, and a single-pixel impulse or an outline stands out from its neighbours, so

$$s = \min\big(1,\ 2\,\bar{b}_{\mathrm{ring}}/(\bar{b}_{\mathrm{ring}} + |b|)\big)\ w_{\mathrm{int}}, \tag{5}$$

where $\bar{b}_{\mathrm{ring}}$ is the mean $|b|$ over the eight neighbouring pixels. The interior factor $w_{\mathrm{int}} = 1 - B/(B + 1)$ lowers s near outlines. Here B is the smaller of the two gradient magnitudes of the

image blurred at $\sigma = 2$ and at $\sigma = 8$, each divided by its image median, then blurred by eight passes. A single-pixel impulse and an outline both receive $s = 0$.

Redistribution and addition. Both corrections use the envelopes e_n. Each is the magnitude of band b blurred by n passes, computed on the unedited image. Written with the masks, they are

$$\text{redistribute}_{m_r v}(b) = b\,\big[\,\min(\kappa^{-1.5 m_r v}, 1) + s\,\max(\kappa^{-1.5 m_r v} - 1, 0)\big], \tag{6}$$

$$a(y) = \text{cap}\big[\,c(y)\,u(y)\,\text{med}_y\,|b|\,\big]\,v(y), \qquad u = \frac{(1-w)\,b + w\,d}{\max(e_{16},\,0.25\,\text{med}_y\,|b|)}, \tag{7}$$

with $\kappa = \text{clamp}(e_8/e_{64}, \frac{1}{4}, 4)$. Where the factor $\kappa^{-1.5 m_r v}$ is below one, the redistribution damps all of b. Where it is above one, it lifts the texture part of b in proportion to s, so impulses and outlines keep their amplitude.

In the addition, the source u blends the image's own band b with a random-phase field d that has the natural images' spectrum of band b (Galerne et al., 2011). The blend fraction w measures how far the local envelope falls short of the natural-image level, $w = \text{clamp}(1 - \hat{e}_8/\bar{e}_{\text{nat}}, 0, 1)$. Here $\hat{e}_8$ is e_8 divided by its image mean, and $\bar{e}_{\text{nat}}$ is the median of the same quantity over flat interior pixels of natural images, in eight luminance bins. A region that already carries texture therefore keeps its own, and an empty region takes the random field. The busyness factor $c(y)$ is the busyness relative to its image median, normalised to mean one, so busier regions receive more texture. It plays the role of the multiplier field of the Gaussian scale mixture (Wainwright & Simoncelli, 1999).

The cap keeps the added texture within the range natural images show. It multiplies each patch by $\min(1, L_k/r_c)$, smoothed by eight passes. Here r_c is the 99th percentile of the envelope of the addition divided by its image median. L_k is the same ratio measured on natural images in bin k. From flattest to busiest it is 0.63, 0.88, 1.33, 2.47 and 5.51. All three corrections vanish at a boundary. The redistribution control enters as $m_r v$, the addition carries the factor v in Equation 7, and the solved exponent enters as $1 + (\gamma - 1)\,v$.

Exponent update of the moment-ratio correction. We interpolate the target $r_p^\star$ from the nine natural-image deciles of the patch's bin. Starting from $\gamma_p = 1$, each pass sets

$$\gamma_p = \text{clamp}\left(1 - 4\,\text{clamp}\left(\frac{r_p^\star - r_p}{r_p^\star},\,-0.8,\,0.8\right),\,0.2,\,1.6\right), \tag{8}$$

where the factor 4 and the two clamps set the size of one step. We apply Equation 2 to b divided by $G_{16}|b|$. We then rescale the result so that its local energy, averaged by G_{16}, stays unchanged, and multiply it back by $G_{16}|b|$. The correction therefore changes the shape of the values and keeps their amount. Equation 2 divides the shape parameter of a generalised Gaussian by γ. On a Laplacian band at $\gamma = 1$ the derivative of r with respect to γ is exactly $-r/4$, which gives the factor 4 of Equation 8. The inner clamp holds the relative error within 0.8, and the outer one holds γ_p within $[0.2,\,1.6]$. The factor and both clamps are the same for every generator and every experiment. The moment ratio of the transformed band depends on γ_p through the whole distribution of its values. The factor is therefore a linear approximation, and we repeat the correction three times.

Luminance projection. The projection Π_Y applies the edit to luminance only, as one value per pixel. We limit each value so that every pixel stays in $[0, 1]$.

Inference. In this notation, the steps of Section 2.5 are

$$b_1 = \text{redistribute}_{m_r v}(b), \qquad b_2 = b_1 + m_a\,a(y), \tag{9}$$

$$b_3 = \text{even}_{\gamma^*(b_2)}(b_2), \qquad \hat{y} = y + \Pi_Y\big[b_3 - b\big], \tag{10}$$

where even_γ applies Equation 2 with the exponents of Section 2.5.1. We solve the exponent γ^* on the band after redistribution and addition, so the moment-ratio correction acts last. Algorithm 1 lists the same steps. Inference takes the image alone as input and uses no reference natural image, generator label or retrieval.

Algorithm 1 REALSYNZ at inference. Colors follow Figure 3: blue quantities are computed directly from the input, green ones are solved to match natural images, and orange ones are predicted by the map.

Require: generated image y, trained map, natural-image decile curves $\{q_k\}_{k=1}^{5}$
Ensure: edited image $\hat{y}$

$b \leftarrow y - G_1 y$ ▷ finest band b
compute the interior field v, the support mask s, the envelopes e_8, e_{16}, e_{64} and the busyness bin $k(p)$ of every patch p from y
$(m_r, m_a) \leftarrow \mathrm{map}(y, k)$ ▷ control fields m_r, m_a
$\kappa \leftarrow \mathrm{clamp}(e_8/e_{64}, \frac{1}{4}, 4)$
$b_1 \leftarrow \mathrm{redistribute}_{m_r v}(b)$ ▷ redistribution, Equation 6
$b_2 \leftarrow b_1 + m_a\, a(y)$ ▷ addition, Equation 7
$b_3 \leftarrow b_2$
for pass $= 1, 2, 3$ **do** ▷ moment-ratio correction
  for every patch p **do**
    $r_p \leftarrow$ moment ratio of b_3 on p
    $r_p^{\star} \leftarrow q_{k(p)}$ at the rank of r_p among the patches of bin $k(p)$
    $\gamma_p \leftarrow$ Equation 8 ▷ exponent γ_p
  end for
  $b_3 \leftarrow \mathrm{even}_{1+(\gamma-1)v}(b_3)$ ▷ Equation 2 at fixed local energy
end for
$\hat{y} \leftarrow y + \Pi_Y\big[b_3 - b\big]$ ▷ output $\hat{y}$
return $\hat{y}$

A.3 MAP, LOSS AND TRAINING

This subsection gives the architecture and the full loss of Section 2.5.3. The map is a convolutional network of width 64 that works at a quarter of the image resolution. A guided filter of radius 4 brings its output back to full resolution. Its inputs are the image, its low-pass $G_8 y$ and a one-hot indicator of each pixel's busyness bin. Its last convolution and output head have one copy per bin, and the indicator selects the copy for each pixel. It outputs the two control fields m_r and m_a.

The training objective of Section 2.5.3 has two terms. The first is the per-bin envelope variation loss of Equation 4, computed on patches of 16, 32 and 64 pixels. The second matches the median envelope variation of each bin in an image to the natural-image level,

$$\mathcal{L}_{\mathrm{med}} = \frac{1}{|K|} \sum_{k \in K} \big(\mathrm{med}_{p \in \mathcal{P}_k}\, \beta_p - \beta_k^{\mathrm{nat}}\big)^2, \tag{11}$$

where $\mathcal{P}_k$ is the set of 16-pixel patches of one image in bin k and β_k^{nat} is the natural-image level of that bin. The set K holds the bins of that image with at least 64 patches. Table 5 lists every constant, and each takes one value for every image, generator and experiment in the paper.

A.4 BENCHMARK PROMPTS

Section 2.1 describes the two image pools, and this subsection describes how we build the prompts. Each prompt follows a fixed three-sentence template over six closed lists, *A photograph of* {*actors*} {*action with object*} *in* {*setting*}. {*Light*}, {*material*} {*placement*}. *A photorealistic, unretouched photograph.* The actors are 88 groups from one era, and the 36 objects come with actions that place them in another era, so no camera could have captured the scene. Each setting is tagged indoor or outdoor, and the lighting matches that tag. The last list holds the foreground materials. We sample the fields uniformly with the prompt's index as the seed, and every test prompt has an actor-object pair absent from the training prompts. Each generator produces one image for each of the same 5,000 training and 1,000 test prompts. The test-set size matches Synthbuster (Bammey, 2024).

Every table scores all 1,000 test images.

Effect of the anachronistic premise. A FLUX.2 Klein pool generated from ordinary captions at the same resolution measures an unedited C_β of 12.28% and C_r of 1.22%, against 12.90% and

Table 5: Hyperparameters and constants. Every value is shared by all five generators and every experiment.

group	constant	value
statistics	scale of band b	$n = 1$ pass
	busyness scale	$n = 2$ to 8 passes
	busyness bins	5, edges at the fit split's 20th to 80th percentiles
	patch side, boundary share	16 pixels, at most 2%
	boundary scale, factor	16 passes, $8\times$ the median gradient
	envelope variation denominator constant τ	0.1, times the image median of $\lvert b\rvert$
moment-ratio correction	passes	3
	factor and clamps of Equation 8	4, ± 0.8, $[0.2,\ 1.6]$
operator	envelope spreads of κ	8 and 64 passes
	clamp of κ, exponent	$[1/4,\ 4]$, 1.5
	source floor	$0.25\times$ the image median of $\lvert b\rvert$
	blend levels $\bar{e}_{\mathrm{nat}}$	8 luminance bins, from 96 natural images
	cap law L_k	from 200 fit-split natural images
map	stride, width, guided-filter radius	4, 64, 4
	per-bin heads	5
loss	envelope variation term: patch sizes, weight λ_β	16 / 32 / 64 pixels, 29.95
	bin weights	inverse interior-patch shares of the training pool
	median term: weight λ_{med}, minimum patches	60, 64
training	optimiser, learning rate	Adam, 10^{-3}
	batch size, steps, seed	16, 700, 101
	training images, test images	5,000 and 1,000 per generator

1.07% for the anachronistic pool. The premise changes each criterion by less than one point, and it guarantees that no test image can reproduce a natural image.

B BASELINE SETTINGS

Section 3.1 compares REALSYNZ with fifteen texture edits at the same edit size. This appendix gives the setting of each (Table 6), shows what each adds to an image (Figure 5), and describes the baselines whose setting needs more than one line. On FLUX.2 Klein, we solve each strength so that the median edit matches REALSYNZ's $3.92/255$ on the same images. We measure every entry from its 8-bit output. On the other four generators, the controls of Table 8 match the edit size of REALSYNZ trained on that generator.

At the matched edit size of Figure 6, the normalisation leaves all five bins 3 to 13 points below the natural-image level.

The closed-form envelope normalisation. This baseline tests whether envelope variation needs a learned map. In each patch it multiplies band b by $(\mathrm{level}/e)^\eta$, a gain that flattens the envelope e toward the patch's own level. One exponent η holds for the whole image. Here we run it in place of the map with the moment-ratio correction kept on; Table 1 reports the normalisation alone, at $C_\beta = 6.96$. At its own η it under-edits, reaching $C_\beta = 27.8$ and $C_r = 1.43$ at $2.36/255$. At $\eta = 5.3$, which reaches the matched edit, it overshoots the two flattest bins by 53 points. The two corrections undo each other. The normalisation moves the moment ratio that the correction set, and the correction moves the envelope variation that the normalisation set. REALSYNZ applies its moment-ratio correction last, which keeps C_r at 0.38%.

Normalisation with a per-bin target. This variant gives the normalisation the information the map uses, namely each patch's busyness bin and the natural-image level of that bin. It computes one strength η_p per patch. The strength starts at $1 - \mathrm{target}/\mathrm{current}$ clamped to $[-1, 1]$, and three fixed-point iterations refine it. We then average it over the 3×3 neighbouring patches. A global factor λ scales it. At $\lambda = 1$, 2 and 8.84 it reaches C_β of 10.35, 7.32 and 55.6, at edits of 0.30, 0.65 and $3.88/255$, against REALSYNZ's 4.49 at $3.92/255$.

Table 6: Baseline settings. The strength column gives the value solved so that the median edit matches REALSYNZ's, or the reason the entry sits at another size. In the last two columns, "natural" marks a baseline's own unscaled strength or edit.

baseline	operation	strength	edit /255
uniform gain	$b \to (1+a)\,b$, the unsharp mask at $\sigma = 1$	$a = 0.80$	3.90
AV1 film grain	AFGS1 synthesiser fitted on the natural-image corpus	gain 135% (corpus amplitude 177%)	3.87 (5.10)
neural film grain	learned grain synthesiser, released weights	its smallest edit	5.50
white noise	Gaussian field, flat spectrum	amplitude	3.92
pink noise	Gaussian field, $1/f$ spectrum	amplitude	3.91
per-bin level match	scales band b per bin to the natural-image median amplitude	none, an amount match	0.92
textureness transfer	cross-bilateral textureness histogram matched to a natural image of the same busyness	none	3.34
closed-form envelope normalisation	gain $(\mathrm{level}/e)^{\eta}$ on band b per patch	η solved	3.87
normalisation, per-bin target	one η_p per patch from that bin's natural-image level, times λ	$\lambda = 1$ / 2 / 8.84	0.30 / 0.65 / 3.88
steerable-pyramid matching	gradient descent on the sliced-Wasserstein distance of pyramid coefficients, 600 iterations	rescaled from 1.95 at its default setting	4.04
Portilla–Simoncelli matching	gradient descent to the statistic set of one natural image of the same bin, 600 iterations	natural	3.91
NeCA-S noise model	released sRGB noise model, sampling form	level	3.91
STIG	spectrum-matching translator, trained on our pools at 512	blended toward its input	3.94 (unblended 18.28)
CUT	unpaired contrastive translator, trained on our pools at 512	blended toward its input	3.97 (unblended 15.41)
SDEdit	FLUX.2 Klein, 16 steps, guidance 1.0, all-white mask	strength 0.20 / 0.35	3.45 / 3.81
HYPIR, PiSA-SR, InvSR	public release, no prompt; InvSR's $\times 4$ output area-averaged to 1024	none	17.35 / 11.63 / 6.55
REALSYNZ, moment-ratio correction only	REALSYNZ with the map and the addition switched off	none	0.56

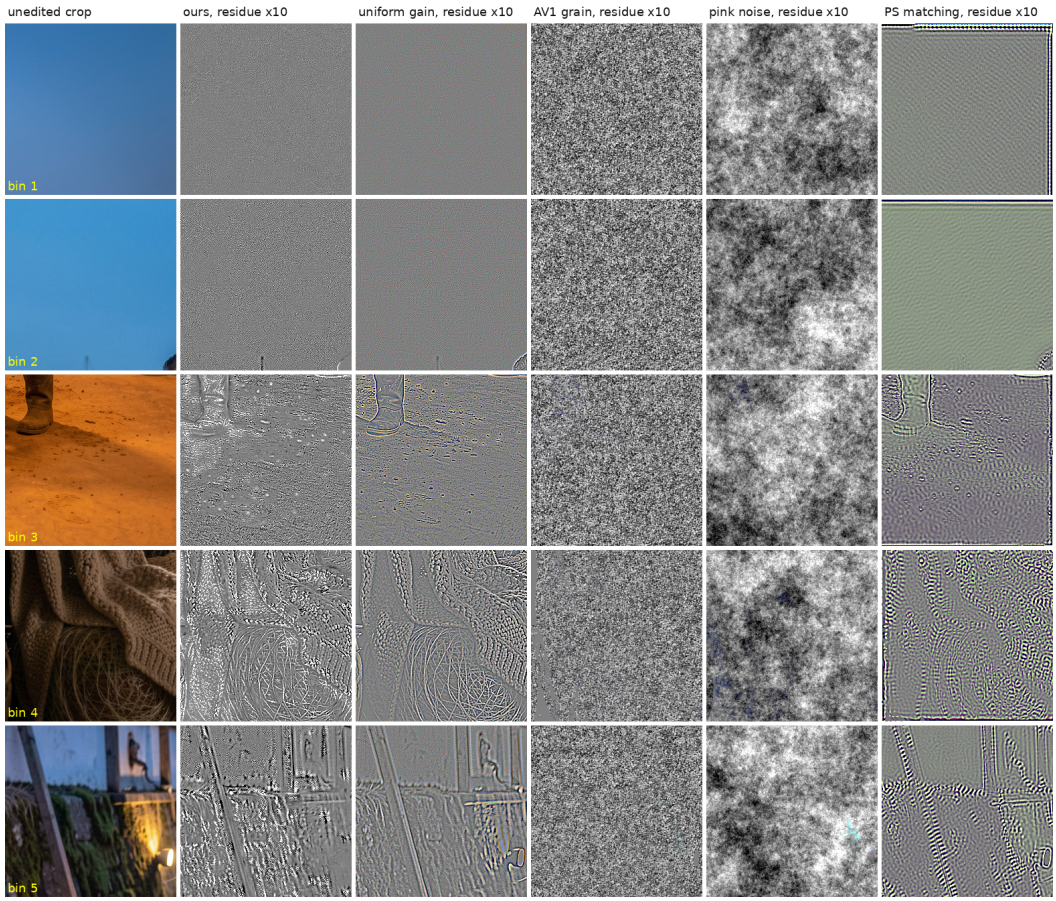


Figure 5: Difference added by each edit. We show one 256-pixel crop per busyness bin, flattest first, and the difference between the edited and the unedited image amplified $10\times$, with mid-grey at zero. The differences of REALSYNZ and the uniform gain follow the image's own texture, the two noises have the same strength everywhere, and Portilla–Simoncelli matching adds the ringing of its pyramid.

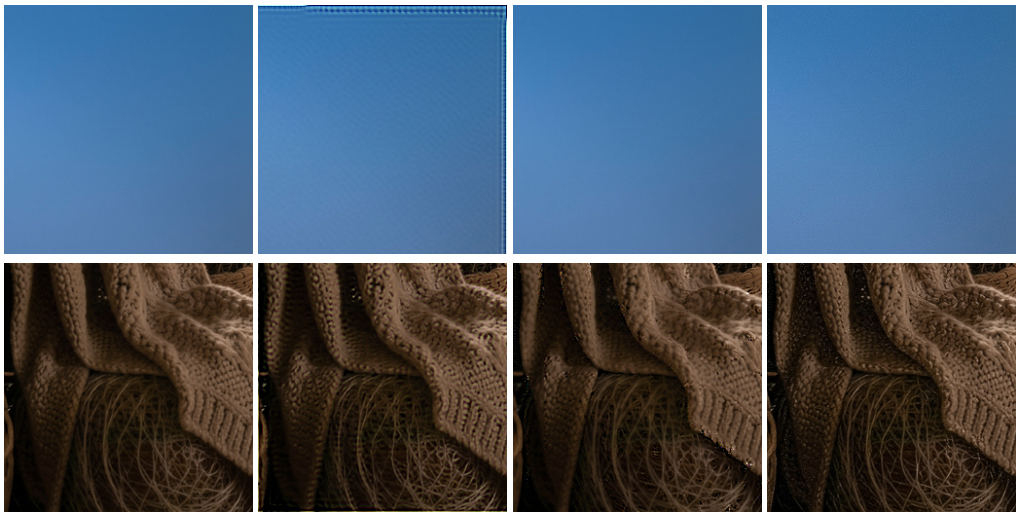


Figure 6: Selected windows at the matched edit size, one flat (top) and one busy (bottom). The middle columns show Portilla–Simoncelli matching, which scores best among the edits without a envelope variation target, and the closed-form envelope normalisation, the strongest analytic control. On the flat window the normalisation changes $0.2/255$ where REALSYNZ changes $1.4/255$, so the closed form keeps the flattest bin close to the generator's output. Amplified differences for every baseline are in Figure 5 and uncurated results in Appendix F.

Per-image optimisation. Steerable-pyramid matching and Portilla–Simoncelli matching optimise each image. They run gradient descent toward the texture statistics of one natural image from the image's own busyness bin. The $\sigma = 8$ low-pass stays fixed, which preserves the low-pass content.

Trained translators and SDEdit. STIG and CUT are image-to-image translators. We train each on 2,000 generated images of our training pool and 2,000 natural images of the fit split, at the authors' 512 resolution, and apply them at 1024. Both edit four to five times more than the matched size. We therefore blend each output toward its input to reach the matched edit in Table 1. Table 11 reports each translator at its own edit size. SDEdit re-noises the image and denoises it again with FLUX.2 Klein in sixteen steps, so strength 0.20 runs three denoising steps and 0.35 runs five. These two strengths give the edits closest to the matched size from below, 3.45 and $3.81/255$. Their low-pass correlations are 0.9994 and 0.9992 (Table 11).

Diffusion restorers. A diffusion restorer re-synthesises a clean, sharp image from its input with a pretrained text-to-image diffusion model. It can therefore change more than band b even when the input is already clean. HYPIR (Lin et al., 2025) and PiSA-SR (Sun et al., 2025) run on Stable Diffusion 2.1 and InvSR (Yue et al., 2025) on SD-Turbo, each at its public release with no prompt. They redraw the whole image, so we leave their edit size δ unmatched. Section 3.3 uses their outputs as inputs to REALSYNZ.

B.1 PERCEPTUAL BASIS OF THE TWO STATISTICS

This subsection gives the perceptual basis for the two statistics. Joint statistics of band-pass coefficients model human texture perception well. Matching a texture on statistics of that family synthesises images that look like the original (Portilla & Simoncelli, 2000; Heeger & Bergen, 1995), and images matched on them within pooling regions are hard to tell apart (Freeman & Simoncelli, 2011; Balas et al., 2009). The second visual area is selective for these dependencies (Freeman et al., 2013; Ziemba et al., 2016). The moment ratio and envelope variation are two summary statistics from that family. The criterion requires them to match in every patch within its busyness bin. The Portilla–Simoncelli baseline matches its statistics over a whole crop and reaches $C_\beta = 10.64$, against REALSYNZ's 4.49. Figure 7 shows windows at several values of envelope variation, so a reader can

check that windows with higher envelope variation look more uneven. We leave a controlled study with human raters to later work.

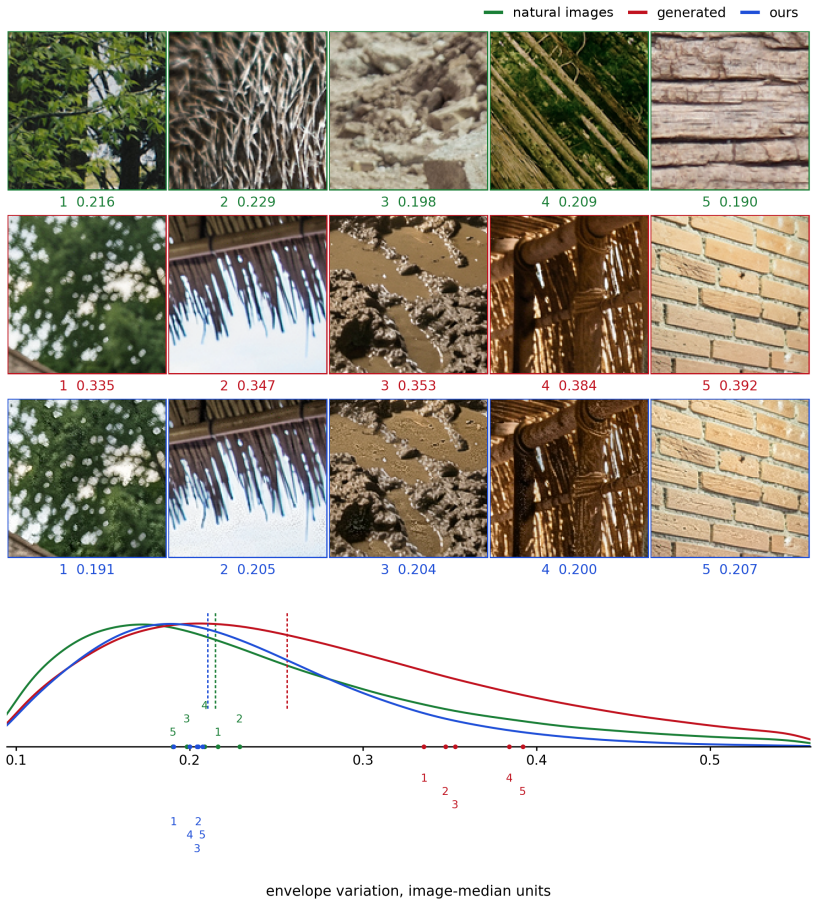


Figure 7: Envelope variation of natural, generated and edited windows in the busiest bin of 1,000 FLUX.2 Klein test images and 1,000 natural images. Each column shows three windows of one material. The top window is a natural image of the same kind of texture at the natural-image level, chosen by CLIP image similarity. The middle window is a generated 128-pixel interior window, and the bottom window is that window after REALSYNZ. Each window is labelled with the median envelope variation of its patches. We chose the generated windows from textured materials in the busier half of the bin, one per material, where REALSYNZ moves envelope variation furthest toward the level. The plot below gives the density of every patch in the bin, with dashed medians at 0.256 for generated images, 0.215 for natural images and 0.210 for REALSYNZ.

C PER-BIN RESULTS

Every entry of Table 1 averages five absolute per-bin gaps (Equation 1). This appendix gives the signed gaps behind those averages, which show the bins where a method overshoots or undershoots.

Table 10 gives the signed gaps on FLUX.2 Klein, flattest bin first, and Figure 8 plots four of the methods. Table 8 gives the two grain and noise controls and the moment-ratio correction on the four other generators, which Table 7 summarises.

Equal-bin mean in the criterion. Generated images and natural images contain different mixes of flat and busy regions (Figure 9). Natural images put about a fifth of their patches in each bin. Four generators put 64 to 75% of theirs in the two busiest bins, and FLUX.1-dev is within 0.08 of the natural images' share in every bin. A mean weighted by patch count would therefore mostly measure

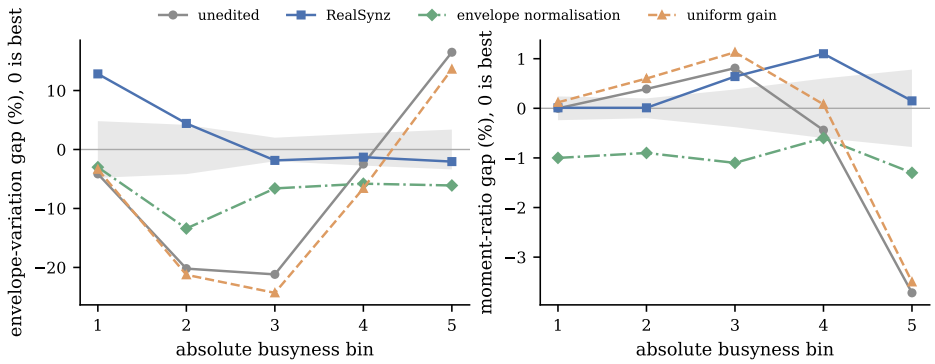


Figure 8: Per-bin gaps on FLUX.2 Klein, in percent of each bin's natural-image level, so zero is best. Left: envelope variation. Right: moment ratio. The shaded band is the 90th percentile of the gap between two random halves of the natural images. Table 10 gives the signed values of every method.

Table 7: Retraining per generator against transferring the Klein map. Cells are C_β / C_r with δ per 255 in parentheses. Grain and noise controls are in Table 8.

generator	unedited	trained on this generator	Klein map, no retraining
FLUX.2 Klein	12.90 / 1.07 (0)	**4.49** / 0.38 (3.92)	same map
Qwen-Image	12.41 / 0.78 (0)	**3.94** / 0.85 (3.98)	7.39 / 0.61 (2.27)
Z-Image	19.85 / 2.08 (0)	**3.74** / 0.59 (3.84)	7.21 / 0.48 (2.21)
SANA-1.5	12.62 / 1.02 (0)	**5.00** / 0.76 (3.92)	5.39 / 0.61 (3.06)
FLUX.1-dev	14.22 / 1.08 (0)	**3.61** / 0.64 (2.47)	7.83 / 1.09 (1.03)

Table 8: Grain and noise controls on the other four generators. Cells are C_β / C_r with δ per 255 in parentheses. AV1 film grain and pink noise are re-solved to the edit size of each generator's own map, and the moment-ratio correction is REALSYNZ with the map and the addition switched off; the unedited and own-map columns repeat Table 7.

generator	unedited	AV1 film grain	pink noise	moment-ratio correction	own map
Qwen-Image	12.41 / 0.78 (0)	11.94 / 2.54 (3.97)	12.44 / 1.49 (3.96)	13.02 / **0.36** (0.53)	**3.94** / 0.85 (3.98)
Z-Image	19.85 / 2.08 (0)	16.16 / 2.46 (3.83)	13.82 / 1.19 (3.84)	17.72 / **0.43** (0.38)	**3.74** / 0.59 (3.84)
SANA-1.5	12.62 / 1.02 (0)	12.92 / 1.44 (3.94)	10.47 / 0.91 (3.92)	10.20 / **0.35** (0.63)	**5.00** / 0.76 (3.92)
FLUX.1-dev	14.22 / 1.08 (0)	5.69 / 1.77 (2.47)	7.13 / 1.05 (2.46)	12.76 / **0.34** (0.27)	**3.61** / 0.64 (2.47)
FLUX.2 Klein	12.90 / 1.07 (0)	18.11 / 1.77 (3.87)	14.99 / 1.14 (3.91)	11.84 / **0.16** (0.56)	**4.49** / 0.38 (3.92)

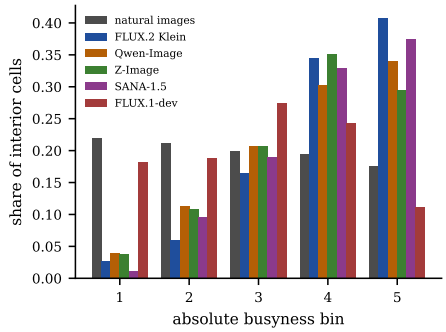


Figure 9: Composition of each pool by busyness bin, as the share of patches in each bin for natural images and for each generator's test pool.

the busiest bins. Table 9 compares the equal-bin mean used throughout with two alternatives. The patch-weighted mean puts three quarters of its weight on the two busiest bins. In the signed mean, an overshoot can cancel an undershoot. REALSYNZ has the smallest envelope variation gap under all three.

Table 9: Three aggregates of the per-bin gaps of Table 10. The equal-bin mean of absolute gaps is used in every table, the patch-weighted mean weights it by each bin's share of patches, and the signed mean keeps the sign of each gap.

	envelope variation gap			moment-ratio gap		
method	equal	patch-weighted	signed	equal	patch-weighted	signed
unedited	12.90	12.35	−6.30	1.07	1.80	−0.58
uniform gain, matched	13.86	12.77	−4.58	1.22	1.68	−0.66
AV1 film grain, matched	18.11	19.05	−18.10	1.77	1.80	+1.78
AV1 grain, corpus amplitude	19.86	23.32	−19.88	2.35	2.69	+2.34
neural film grain	22.12	13.56	−19.12	1.61	0.61	−0.98
white noise	22.55	22.54	−22.54	1.75	1.98	+1.74
pink noise	14.99	14.01	−12.36	1.14	1.70	+0.14
per-bin level match	11.83	11.71	−3.52	1.91	2.08	+0.32
textureness transfer	14.72	18.46	+5.26	1.82	2.81	−1.24
REALSYNZ	**4.49**	**2.21**	**+2.38**	**0.38**	**0.56**	**+0.38**

Overshoot in the flattest bin. The flattest bin holds 2.6% of FLUX.2 Klein's test patches. There the envelope variation gap goes from -4.1% before the edit to $+12.8\%$ after it. In this bin the edit also raises the energy of band b to six times the natural-image level, from 1.4 times before the edit. The criterion is invariant to this amount. A per-bin gain solved to the natural images' energy brings the energy within 23% of the level in every bin, and it raises C_β from 4.5 to 10.2. A gain that differs across patches changes how the fine detail is distributed. We report the overshoot as a limit of the method.

D ADDITIONAL RESULTS

D.1 CONTENT COST AND EDITS OF NATURAL IMAGES

Low-pass content cost. A texture edit should preserve the image content at scales coarser than band b. We measure the change in this content on the blurred image $G_8 y$, which keeps only those coarser scales. The *low-pass error* E_8 computes the mean absolute difference between the blurred edited and unedited images, and the *low-pass correlation* ρ computes their correlation,

$$E_8 = 255\,\overline{|G_8\hat{y} - G_8 y|}, \qquad \rho = \operatorname{corr}\big(G_8 y,\, G_8\hat{y}\big), \tag{12}$$

each reported as the median over the test images (Table 11). The matched baselines share REALSYNZ's edit size δ by construction, so the table shows how much of the same edit reaches the content. REALSYNZ moves the low-pass by $E_8 = 0.90/255$ at $\rho = 0.9998$. At about the same edit size, SDEdit's two strengths move it by 2.59 and $2.74/255$, three times as much. Re-sampling redraws the coarse scales, and a band edit keeps them. REALSYNZ's own $0.90/255$ comes from the range clamp on near-saturated pixels and from the smooth fields that multiply band b.

Edits of natural images. A correction for generated images should leave a natural image nearly unchanged. We run REALSYNZ on 400 random natural images from the score split of Section 2.1 (Table 12), and every FLUX.2 Klein baseline on the same natural images at its generated-image strength (Figure 10). REALSYNZ edits natural images by a median $\delta = 1.36/255$, a third of its edit on generated images. It computes a smaller correction where the texture is already at the natural-image level. It still moves them off the level, raising C_β to 6.44% and C_r to 0.64%, against the natural-image floor of 1.07% and 0.14%. REALSYNZ takes no generator label and corrects every input it receives. We report this shift as a limit of the method.

Every baseline changes natural images as well. The two noises, the two film grains and the learned noise model add a texture field. They reach C_β of 10.9 to 26.8%, all above the 8.12% that RE-

Table 10: Signed per-bin gaps on FLUX.2 Klein, 1,000 test images, in percent of the natural-image level, flattest bin first. Each method is at the edit size given in Table 1.

statistic	method	bin 1	bin 2	bin 3	bin 4	bin 5	mean |gap|
envelope variation	unedited	−4.1	−20.2	−21.2	−2.5	+16.5	12.90
	uniform gain, matched	+9.4	−16.8	−23.1	−6.2	+13.8	13.86
	AV1 film grain, matched	−6.8	−16.1	−32.6	−26.5	−8.5	18.11
	AV1 grain, corpus amplitude	−4.6	−15.2	−33.6	−30.6	−15.4	19.86
	neural film grain	−30.3	−35.4	−29.1	−8.3	+7.5	22.12
	white noise	−15.2	−20.8	−34.9	−28.8	−13.0	22.55
	pink noise	−4.6	−19.0	−29.6	−15.2	+6.6	14.99
	per-bin level match	+2.8	−19.8	−18.1	−0.5	+18.0	11.83
	textureness transfer	+4.3	−22.7	−0.9	+26.9	+18.7	14.72
	closed-form envelope normalisation	−3.0	−13.4	−6.6	−5.8	−6.1	6.96
	normalisation, per-bin target, $\lambda = 1$	−2.0	−19.6	−18.9	−2.4	+8.9	10.35
	normalisation, per-bin target, $\lambda = 2$	+0.9	−14.8	−13.3	−5.0	−2.6	7.32
	normalisation, per-bin target, matched	+63.5	+71.2	+63.6	+47.4	+32.3	55.62
	steerable-pyramid matching	+101.9	+16.8	−2.9	+3.1	+13.9	27.75
	Portilla–Simoncelli matching	−5.8	−17.3	−16.6	+1.4	+12.0	10.64
	NeCA-S noise model	−34.0	−35.8	−41.0	−29.2	−10.2	30.02
	STIG, matched size	−3.2	−19.6	−17.6	+1.6	+20.2	12.43
	CUT, matched size	−7.0	−19.3	−16.2	+2.5	+19.0	12.78
	SDEdit, strength 0.20	−8.0	−21.6	−24.8	−8.7	+11.0	14.82
	SDEdit, strength 0.35	+14.6	−13.8	−21.6	−8.0	+11.2	13.85
	REALSYNZ, moment-ratio correction only	−6.0	−19.0	−20.3	−3.4	+10.5	11.84
	REALSYNZ	**+12.8**	**+4.4**	**−1.9**	**−1.3**	**−2.1**	**4.49**
moment ratio	unedited	+0.0	+0.4	+0.8	−0.4	−3.7	1.07
	uniform gain, matched	−1.2	+0.2	+1.1	+0.1	−3.5	1.22
	AV1 film grain, matched	+1.1	+1.5	+2.7	+3.5	+0.1	1.77
	AV1 grain, corpus amplitude	+1.2	+1.7	+3.1	+4.3	+1.4	2.35
	neural film grain	−6.5	+0.1	+0.6	+0.4	+0.5	1.61
	white noise	+1.2	+0.9	+2.3	+3.6	+0.7	1.75
	pink noise	−0.1	+0.4	+1.5	+1.3	−2.4	1.14
	per-bin level match	+1.8	+2.0	+1.8	−0.2	−3.8	1.91
	textureness transfer	+0.7	+0.7	−0.6	−3.2	−3.8	1.82
	closed-form envelope normalisation	−1.0	−0.9	−1.1	−0.6	−1.3	1.01
	normalisation, per-bin target, $\lambda = 1$	−0.1	+0.3	+0.6	−0.4	−3.1	0.91
	normalisation, per-bin target, $\lambda = 2$	−0.2	+0.1	+0.2	−0.4	−2.1	0.60
	normalisation, per-bin target, matched	−3.5	−3.5	−5.2	−5.9	−6.1	4.84
	steerable-pyramid matching	−6.3	−2.7	−2.1	−2.5	−3.8	3.47
	Portilla–Simoncelli matching	−0.7	−0.1	+0.0	−1.8	−3.4	1.18
	NeCA-S noise model	+0.2	+0.7	+1.9	+2.5	−1.0	1.27
	STIG, matched size	−0.4	+0.2	+0.5	−1.0	−4.2	1.26
	CUT, matched size	−0.5	+0.0	+0.4	−0.9	−3.5	1.05
	SDEdit, strength 0.20	−0.4	−0.4	−0.0	−0.3	−3.2	0.86
	SDEdit, strength 0.35	−1.0	−0.4	+0.2	−0.0	−3.0	0.94
	REALSYNZ, moment-ratio correction only	+0.1	+0.1	+0.1	+0.2	+0.4	0.16
	REALSYNZ	**+0.0**	**+0.0**	**+0.6**	**+1.1**	**+0.2**	**0.38**

ALSYNZ reaches when scored the same way from 8-bit PNG outputs. The two trained translators and Portilla–Simoncelli matching give natural images a smaller C_β than REALSYNZ. They do so at three times the edit size and with a low-pass error of 4.2 to $4.4/255$ against REALSYNZ's 0.27. The moment-ratio correction alone and the per-bin normalisation at $\lambda = 1$ compute their target from the image itself, which a natural image already meets. SDEdit needs a generation prompt, and natural images come without one, so we omit it.

Table 11: Content fidelity on FLUX.2 Klein, 1,000 test images. E_8 is the low-pass error of Equation 12 in units of 255, and ρ is the low-pass correlation. An unchanged image scores zero and one.

method	low-pass error /255	low-pass corr.	edit /255
HYPIR	4.94	0.9926	17.35
STIG, unblended output	16.75	0.9647	18.28
CUT, unblended output	13.08	0.9704	15.41
InvSR ($\times 4$ regime)	3.13	0.9977	6.55
PiSA-SR	3.90	0.9964	11.63
SDEdit, strength 0.20	2.59	0.9994	3.45
SDEdit, strength 0.35	2.74	0.9992	3.81
REALSYNZ	**0.90**	**0.9998**	3.92

Table 12: REALSYNZ on 400 natural images of the score split. We report signed per-bin gaps in percent of the natural-image level, the median edit, and C_β or C_r.

	bin 1	bin 2	bin 3	bin 4	bin 5	edit /255	mean |gap|
envelope variation, edited natural images	+2.3	+11.3	−1.2	−6.3	−11.1	**1.36**	**6.44**
ratio, edited natural images	+0.5	+0.3	+0.6	+0.7	+1.1		**0.64**

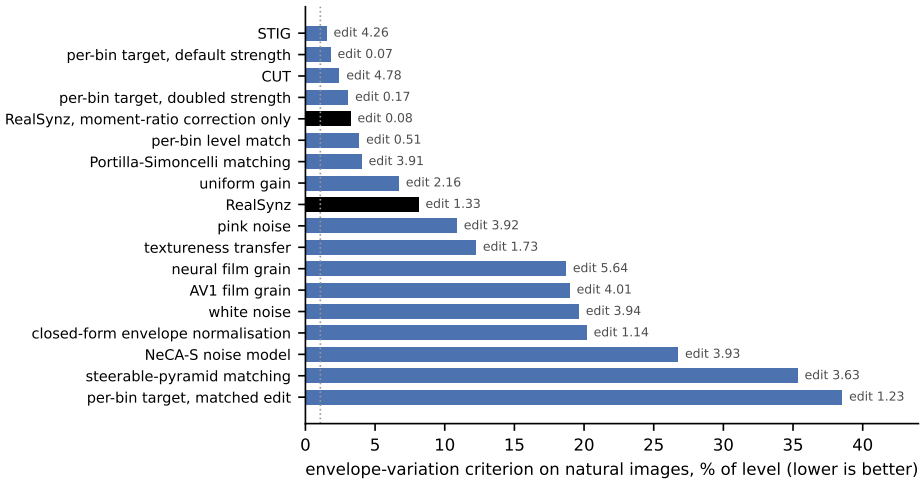


Figure 10: Every FLUX.2 Klein baseline applied to 400 natural images, at the strength of its generated-image entry and scored from 8-bit PNG outputs. Bars are C_β, each labelled with its median edit per 255, and the dotted line marks the natural-image floor, $C_\beta = 1.07$.

D.2 COMPOSITION WITH RESTORERS

No-reference quality models. Section 3.3 reports the two criteria for three diffusion restorers with and without REALSYNZ. This subsection adds four no-reference quality models. MUSIQ (Ke et al., 2021) and MANIQA (Yang et al., 2022) are transformers trained to predict human opinion scores of image quality, and CLIPIQA (Wang et al., 2023) is CLIP's preference for "a good photo" over "a

Table 13: Composition with three diffusion restorers on FLUX.2 Klein, 1,000 test images, each restorer alone and with the frozen REALSYNZ map on top. The edit and E_8 are measured against the unedited generation, so the difference within a pair is not the change REALSYNZ adds, which the text gives; every row scores the patches and bins of the unedited generation. Quality models are in Table 14.

row	C_β, envelope variation ↓	C_r, moment ratio ↓	edit /255	E_8 /255
natural images, the floor	1.07	0.14	–	–
unedited	12.90	1.07	0	0
REALSYNZ	**4.49**	**0.38**	3.92	0.90
HYPIR	12.66	**0.75**	17.35	4.94
HYPIR + REALSYNZ	**5.24**	1.07	19.99	5.19
InvSR	23.27	5.28	6.55	3.13
InvSR + REALSYNZ	**7.98**	**0.87**	9.15	3.36
PiSA-SR	15.62	2.02	11.63	3.90
PiSA-SR + REALSYNZ	**3.14**	**0.58**	14.04	4.16

bad photo". NIQE (Mittal et al., 2013) is the distance between the image's local band statistics and a model fitted on clean natural images, and lower values are closer. Table 13 gives the two criteria and the content quantities of Section 3.3, and Table 14 reports all four quality models beside them. Appendix B gives each restorer's backbone and settings.

The four models disagree with each other and with the criterion, so we report them in the appendix only. They also rate natural images below every generated-image entry, by 8 to 12 points on MUSIQ, so they prefer a smooth generated image to a natural image. Adding REALSYNZ to a restorer lowers MUSIQ by 0.7 to 3.0 points and raises NIQE by 1.6 to 1.9. REALSYNZ produces the same NIQE rise on an unedited image, and the rise carries NIQE past the natural images' own value. CLIPIQA changes by at most 0.012 and MANIQA rises by 0.02 to 0.04.

Table 14: No-reference quality of the restorer compositions on FLUX.2 Klein, 1,000 test images, each restorer alone and with REALSYNZ on top. C_r and C_β are lower-is-better; MUSIQ, CLIPIQA and MANIQA rise with quality and NIQE falls. We measure E_8 and the edit against the unedited image, so the second entry of each pair includes the restorer's own change (Table 13). Bold marks the better value within each pair. Natural images score below every generated-image entry on MUSIQ, CLIPIQA and MANIQA.

row	C_r, moment ratio	C_β, envelope variation	MUSIQ ↑	CLIPIQA ↑	MANIQA ↑	NIQE ↓	E_8 /255	edit /255
natural images	0.14	1.07	63.55	0.618	0.371	3.75	–	–
unedited	1.07	12.90	**73.28**	0.775	0.465	**2.67**	0	0
REALSYNZ	**0.38**	**4.49**	71.40	**0.783**	**0.531**	4.35	0.90	3.92
HYPIR	**0.75**	12.66	**74.20**	**0.764**	0.548	**3.08**	4.94	17.35
HYPIR + REALSYNZ	1.07	**5.24**	71.25	0.752	**0.570**	4.71	5.19	19.99
InvSR	5.28	23.27	**75.84**	0.821	0.645	**3.09**	3.13	6.55
InvSR + REALSYNZ	**0.87**	**7.98**	75.18	**0.823**	**0.685**	4.94	3.36	9.15
PiSA-SR	2.02	15.62	**75.21**	**0.779**	0.546	**2.79**	3.90	11.63
PiSA-SR + REALSYNZ	**0.58**	**3.14**	74.04	0.776	**0.564**	4.57	4.16	14.04

Per-bin gaps of the compositions. Figure 11 shows the per-bin gaps behind Table 14. Each restorer alone pushes the busiest bins further above the natural-image level, to +22 points on HYPIR, +24 to +32 on InvSR and +20 to +22 on PiSA-SR. With REALSYNZ added, the three busiest bins end within 5.5 points of the level in every composition, and the remaining gap sits in the two flattest bins. InvSR's 4× output, area-averaged back to 1024 before scoring, has the largest ratio gap of any entry, 5.28, and REALSYNZ brings it to 0.87. Each composition keeps the content cost of its restorer. Its low-pass error is within $0.26/255$ of the restorer's alone, and its edit is 2.4 to $2.6/255$ larger. HYPIR is the one composition where the moment ratio gets worse, from 0.75 to 1.07. Every bin of this composition stays within 3 points of the natural-image level before and after.

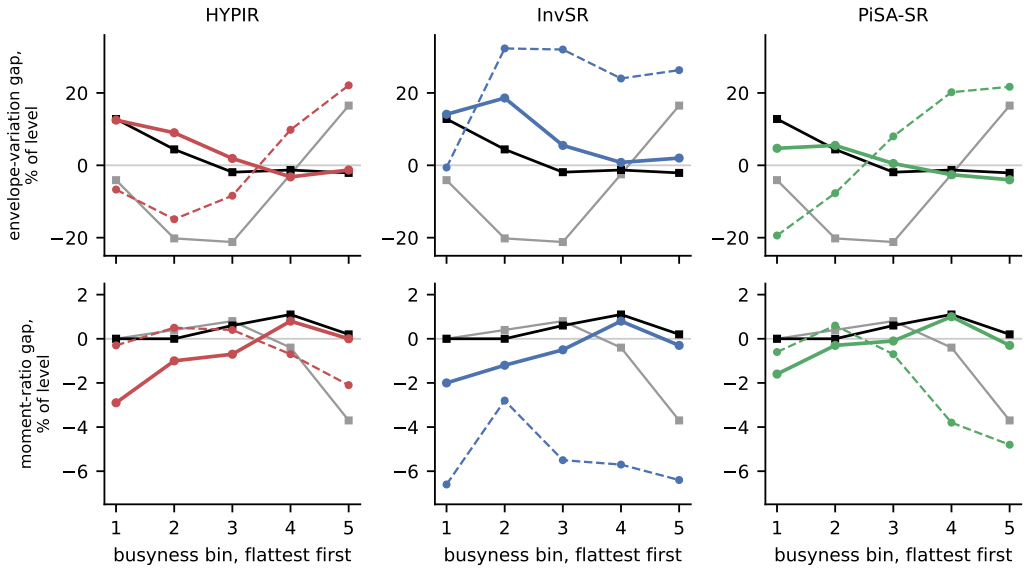


Figure 11: Per-bin gaps of the restorer compositions on FLUX.2 Klein, 1,000 test images, in percent of the natural-image level, one panel per restorer. Each panel shows the restorer alone (dashed) and with REALSYNZ on top (solid), beside the unedited pool (grey) and REALSYNZ alone (black).

D.3 AMPLITUDE AND MOMENT RATIO PER OCTAVE

This subsection compares the amount of fine detail each generator produces with its envelope variation gap. Table 15 gives, per busyness bin, the root-mean-square amplitude of every octave of the FLUX.2 Klein test images in percent of the natural images'. It also gives the moment-ratio gap of each octave. FLUX.2 Klein carries a larger amplitude than natural images at every octave in every bin. On the other four generators band b carries 60 to 75% (Z-Image), 62 to 90% (Qwen-Image) and 72 to 105% (FLUX.1-dev) of the natural images' amplitude. SANA-1.5 carries 100 to 132%. Yet the unedited C_β is 12.4 to 12.9 for FLUX.2 Klein, SANA-1.5 and Qwen-Image across that whole range, and Z-Image, which carries the least, is the worst at 19.85. The amount of fine texture a generator produces is therefore a poor predictor of its envelope variation gap.

Table 15: Amplitude and moment ratio per octave and per busyness bin on FLUX.2 Klein, flattest first, on luminance. Amplitude is the root-mean-square value in percent of the natural images', and the moment-ratio gap is in percent of the natural images' value. Bold marks band b, the only band REALSYNZ edits. In the flattest bin the octaves above $\sigma = 2$ are nearly empty.

	octave	bin 1	bin 2	bin 3	bin 4	bin 5
amplitude, % of natural images	$\sigma = 1$ (band b)	**137**	**157**	**137**	**105**	**105**
	σ 1–2	120	115	118	110	111
	σ 2–4	117	116	115	117	126
	σ 4–8	202	128	121	135	149
	σ 8–16	498	179	144	156	153
moment-ratio gap, %	$\sigma = 1$ (band b)	**+0.0**	**+0.4**	**+0.8**	**−0.5**	**−3.6**
	σ 1–2	+12.5	+1.1	−0.6	−1.3	−1.0
	σ 2–4	+12.4	−0.7	−0.8	+0.0	+1.2
	σ 4–8	+75.0	+0.7	−0.1	+0.1	+1.0
	σ 8–16	+87.2	+0.7	−0.1	−0.0	+0.6

D.4 ABLATION DETAILS

Section 3.4 removes one component at a time. This subsection gives the per-bin gaps behind the inference-side entries of Table 2. It also reports two removals that the main text omits: the blend and the bin definition. Each entry re-runs the trained map with one operation switched off, and the entry with nothing switched off reproduces the REALSYNZ result exactly.

Removal of the blend. At each pixel, the blend mixes the random-phase field with the image's own band b, in proportion to how far the local envelope falls short of the natural-image level (Appendix A.2). Without the blend, the addition uses the raw source, and every envelope variation bin drops by 9 to 12 points. The three busiest bins then end 12 to 14 points below the level.

Fixed busyness bins and per-image quantile bins. The moment-ratio correction looks up its target by the patch's busyness bin (Section 2.5.1). Per-image quantile bins would put a fifth of every image in each bin, however textured the image is. With them in place of the fixed bins, the ratio ends 1.3 to 4.7 points above the level in every bin.

Redistribution and luminance projection. Without the redistribution, envelope variation moves by at most 4.2 points in any bin. We keep it because it rescales band b toward the envelope of its wider surroundings (Equation 6). Without the luminance projection, the criterion falls from 4.49 to 2.65. An edit that carries colour, however, re-tints the near-saturated pixels that the range clamp touches, and leaves coloured specks on 987 of the 1,000 images. We keep both components for effects outside the criterion.

Table 16: Inference-side ablation in full, on FLUX.2 Klein, 1,000 test images. Each entry switches off one operation of the trained map. We report per-bin gaps in percent of the natural-image level, flattest first, and δ per 255. Bold marks the statistic the removed operation controls, and Table 2 aggregates four of the entries.

removed	controls	envelope variation gap per bin	ratio gap per bin	edit
nothing (REALSYNZ)		+12.8 / +4.4 / −1.9 / −1.3 / −2.1	+0.0 / +0.0 / +0.6 / +1.1 / +0.2	3.92
the moment-ratio correction	the ratio	+8.2 / +1.2 / −5.8 / −4.3 / −1.0	**+3.1 / +3.9 / +5.1 / +5.1 / +2.3**	4.29
the addition	envelope variation	**−9.4 / −29.2 / −35.9 / −25.3 / −8.7**	+0.1 / −0.0 / +0.1 / +0.3 / +0.2	1.31
the blend	envelope variation, busy bins	**+4.0 / −6.8 / −14.0 / −12.8 / −12.1**	−0.7 / +0.0 / +0.4 / +0.9 / +0.6	3.64
fixed bins, replaced by per-image quintile bins	the ratio	+10.7 / +2.6 / −5.7 / −4.6 / −1.8	**+1.3 / +1.4 / +3.9 / +4.7 / +1.3**	4.21
the redistribution	envelope	+15.0 / +6.7 / −0.8 / +0.6 / +2.1	+0.1 / +0.1 / +0.9 / +1.3 / −0.3	4.12

Training-side ablation. In Table 17, we retrain the map with one component removed. Without the bin-restricted patch terms, the second bin rises to $+9.1$ points against $+4.4$. Without the 32- and 64-pixel patch terms, the middle three bins fall by 5 to 7 points.

Table 17: Training-side ablation on FLUX.2 Klein. Each entry retrains the map with one component removed. We report per-bin gaps in percent of the natural-image level, flattest first, and the edit per 255. Bold marks the bins the removal moves most.

removed	envelope variation gap per bin	ratio gap per bin	edit
nothing (REALSYNZ)	+12.8 / +4.4 / −1.9 / −1.3 / −2.1	+0.0 / +0.0 / +0.6 / +1.1 / +0.2	3.92
bin indicator and per-bin heads	+12.1 / +2.6 / **−5.9** / **−4.2** / −3.2	+0.1 / −0.2 / −0.2 / +0.4 / +0.1	4.37
bin-restricted patch terms	**+17.1** / **+9.1** / −5.5 / −3.6 / −0.2	−0.1 / −0.1 / +0.2 / +0.3 / +0.2	4.09
the 32- and 64-pixel patch terms	+9.6 / −1.0 / **−8.9** / **−8.0** / −3.6	+0.2 / −0.2 / +0.2 / +1.2 / +0.0	3.86

Table 18: Inference time per 1024×1024 image on one GPU, unoptimised, with every method on the same card in the same session. A baseline's cost is the difference between a 40-image and a 20-image run after a warm-up, ours is the median of 60 calls, and InvSR is timed in its released $\times 4$ regime.

method	ms / image	parameters	training
REALSYNZ	**514** (map 10.8)	**0.28 M**	700 steps, 7 min on 4 GPUs
HYPIR	1,339	1,549 M	diffusion prior, released
InvSR ($\times 4$ regime)	22,894	1,324 M	diffusion prior, released
PiSA-SR	751	1,298 M	diffusion prior, released
uniform gain (unsharp mask)	5.5	0	none
neural film grain	9.1	released	released
white noise	11.9	0	none
pink noise	19.7	0	none
per-bin level match	31.8	0	five constants from the corpus
AV1 film grain	45.5	0	fitted on the corpus
textureness transfer	74.7	0	none
closed-form envelope normalisation	102	0	none
NeCA-S noise model	196	released	released
STIG	338	released	trained on our pools at 512
CUT	1,332	released	trained on our pools at 512
SDEdit on FLUX.2 Klein, strength 0.20 / 0.35	2,000 to 4,000	9,000 M	diffusion prior, released
steerable-pyramid matching (600 iterations)	10,021	0	per-image optimisation
Portilla–Simoncelli matching (600 iterations)	41,333	0	per-image optimisation

E LATENCY

Table 18 reports the inference time per 1024×1024 image of every method, measured on one GPU in the same session. REALSYNZ takes 514 ms, of which the map takes 10.8 ms and the closed-form steps take the rest. The gain, noise and grain edits are seven to ninety-four times faster, and the learned noise model and the two translators are within a factor of three. Per-image optimisation takes 10 to 41 s per image.

F QUALITATIVE RESULTS

This appendix shows six test images per generator, each edited by the map trained on that generator's own images. Each row shows the unedited image, the REALSYNZ output, the edit map in absolute units (full green is $8/255$), and a $4\times$ crop of one region before and after. The images were chosen by the statistic. Each has an interior patch in the busier half of the bins whose unedited envelope variation is at least the natural-image floor away from the level. The edit closes 60 to 100% of that gap. In every region shown the generated fine detail is concentrated in parts of the region with little between them. The edit spreads it the way a natural image of the same kind of material carries it. The natural images of Figure 1 were chosen by eye from the natural-image corpus to match each generated crop in material and busyness.

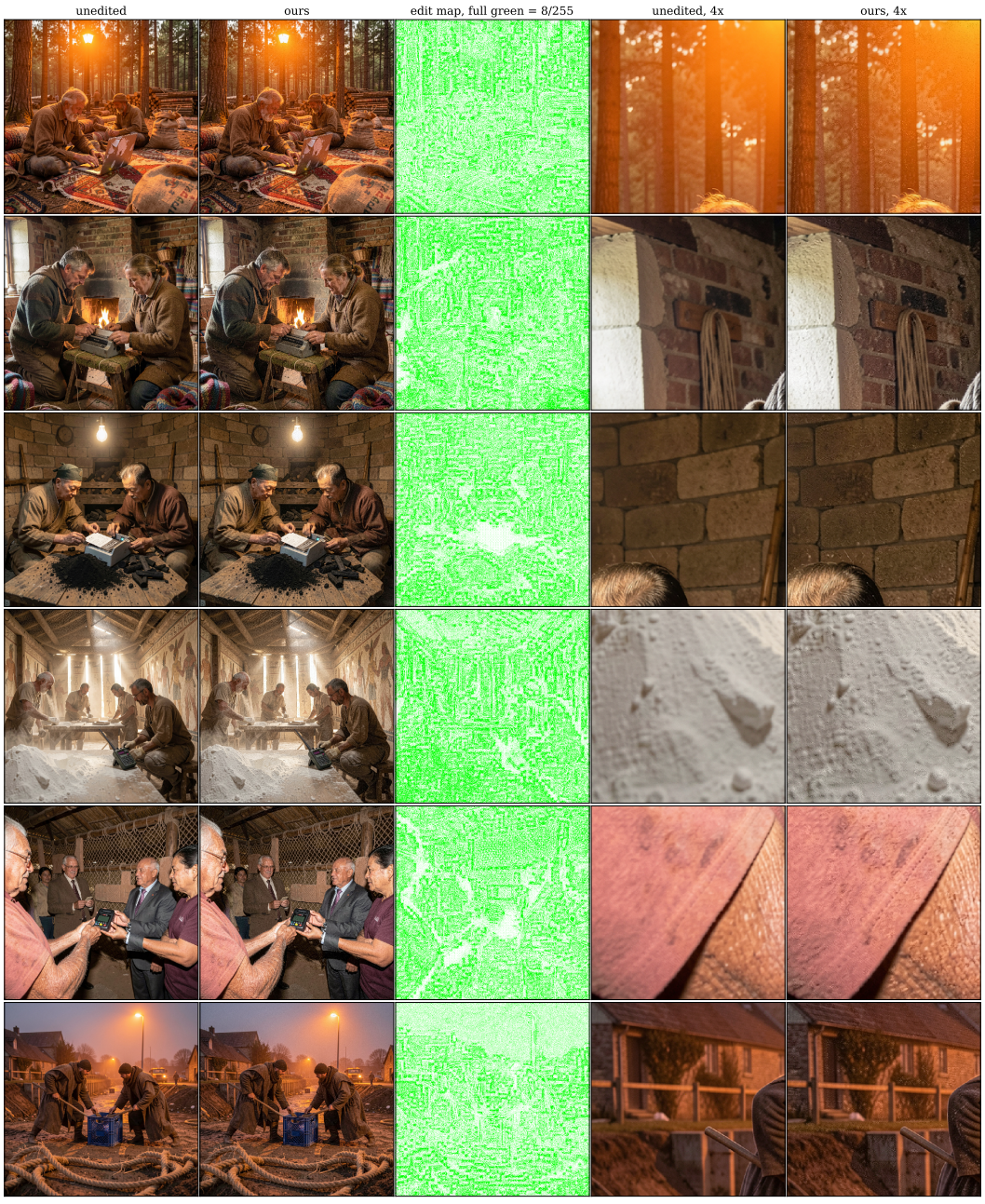


Figure 12: Qualitative results on FLUX.2 Klein. Each row shows the unedited image, the REAL-SYNZ output, the edit map in absolute units (full green is $8/255$), and a $4\times$ crop of one region before and after.

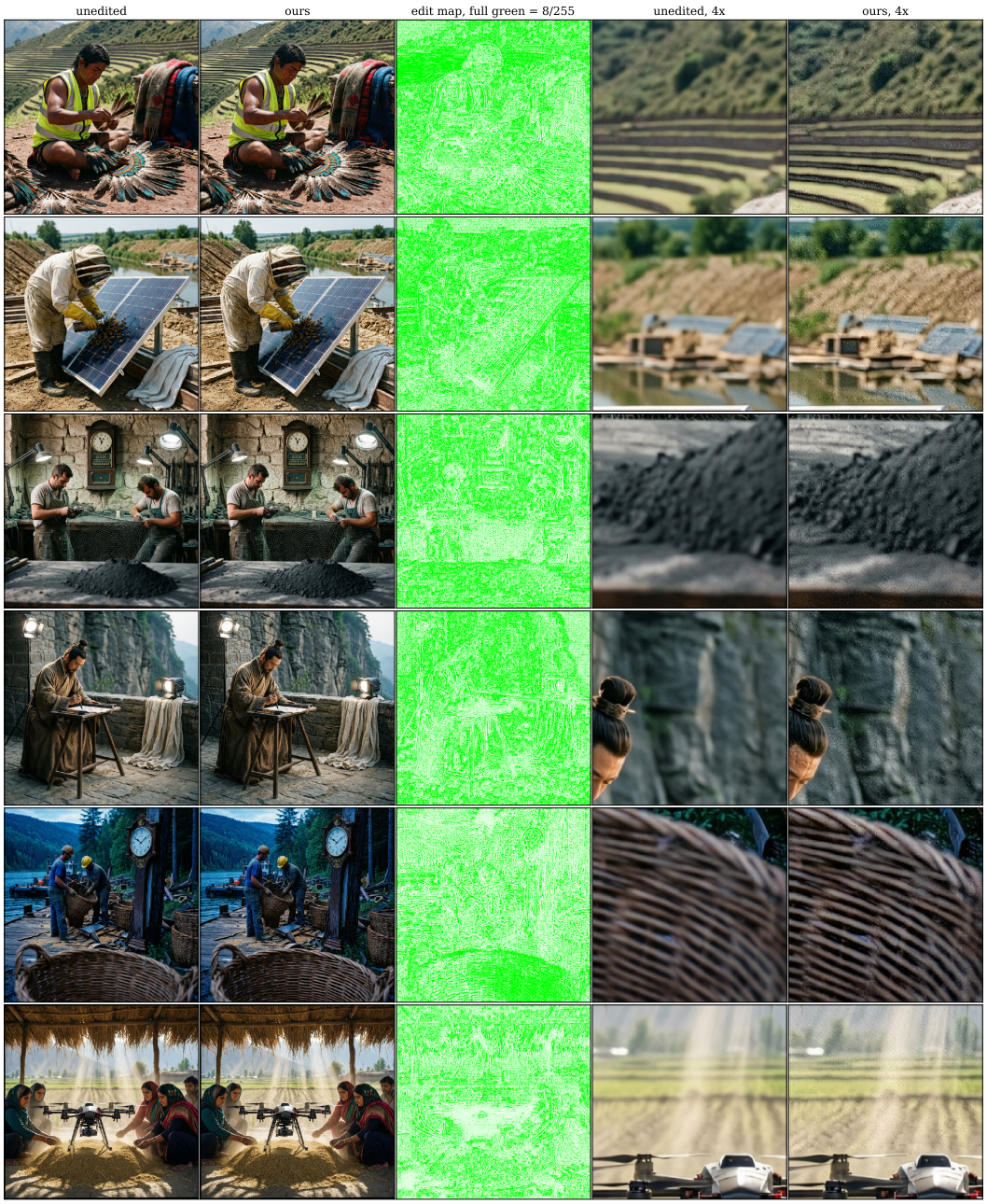


Figure 13: Qualitative results on Qwen-Image, laid out as in Figure 12.



Figure 14: Qualitative results on Z-Image, laid out as in Figure 12.

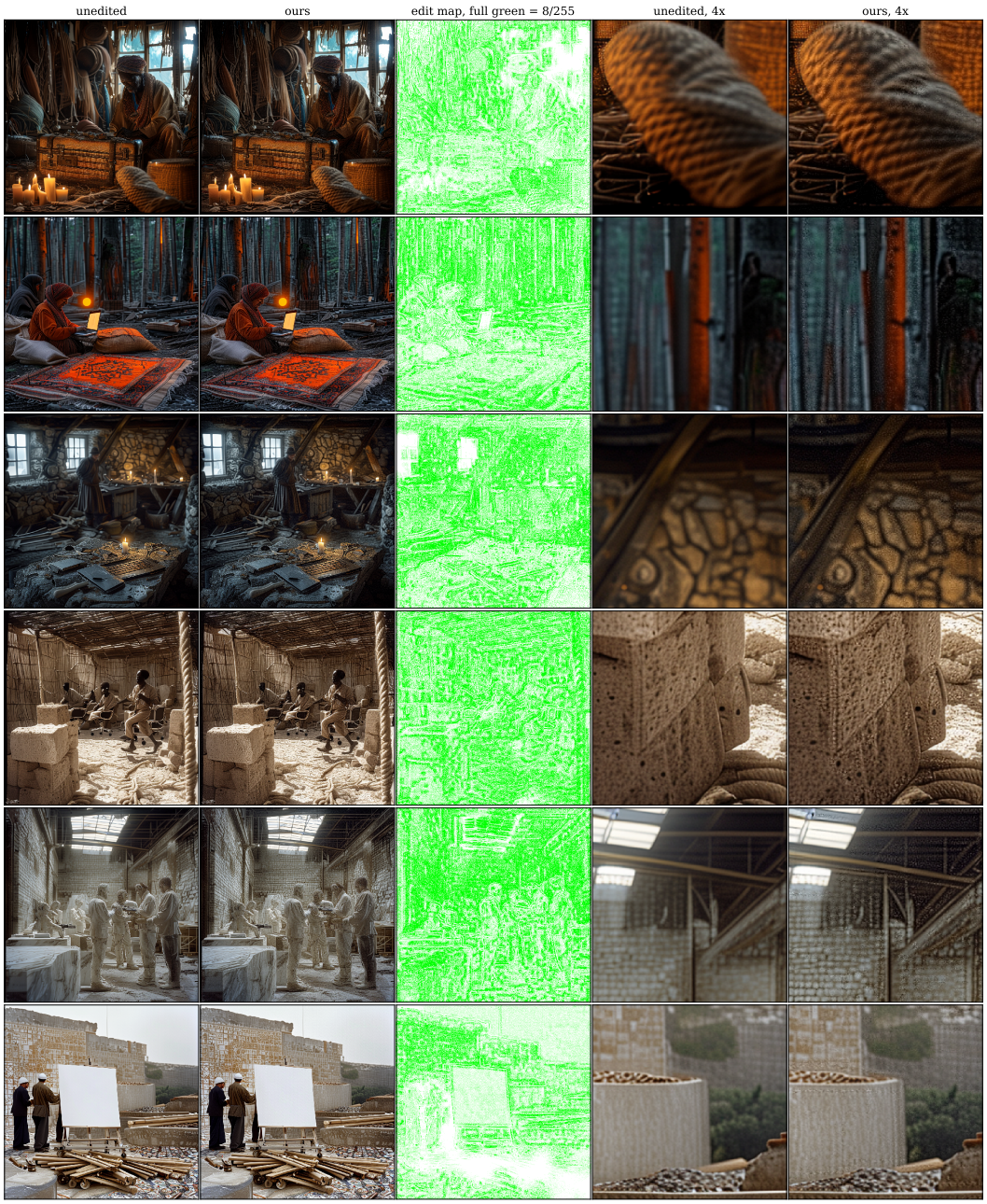


Figure 15: Qualitative results on SANA-1.5, laid out as in Figure 12.

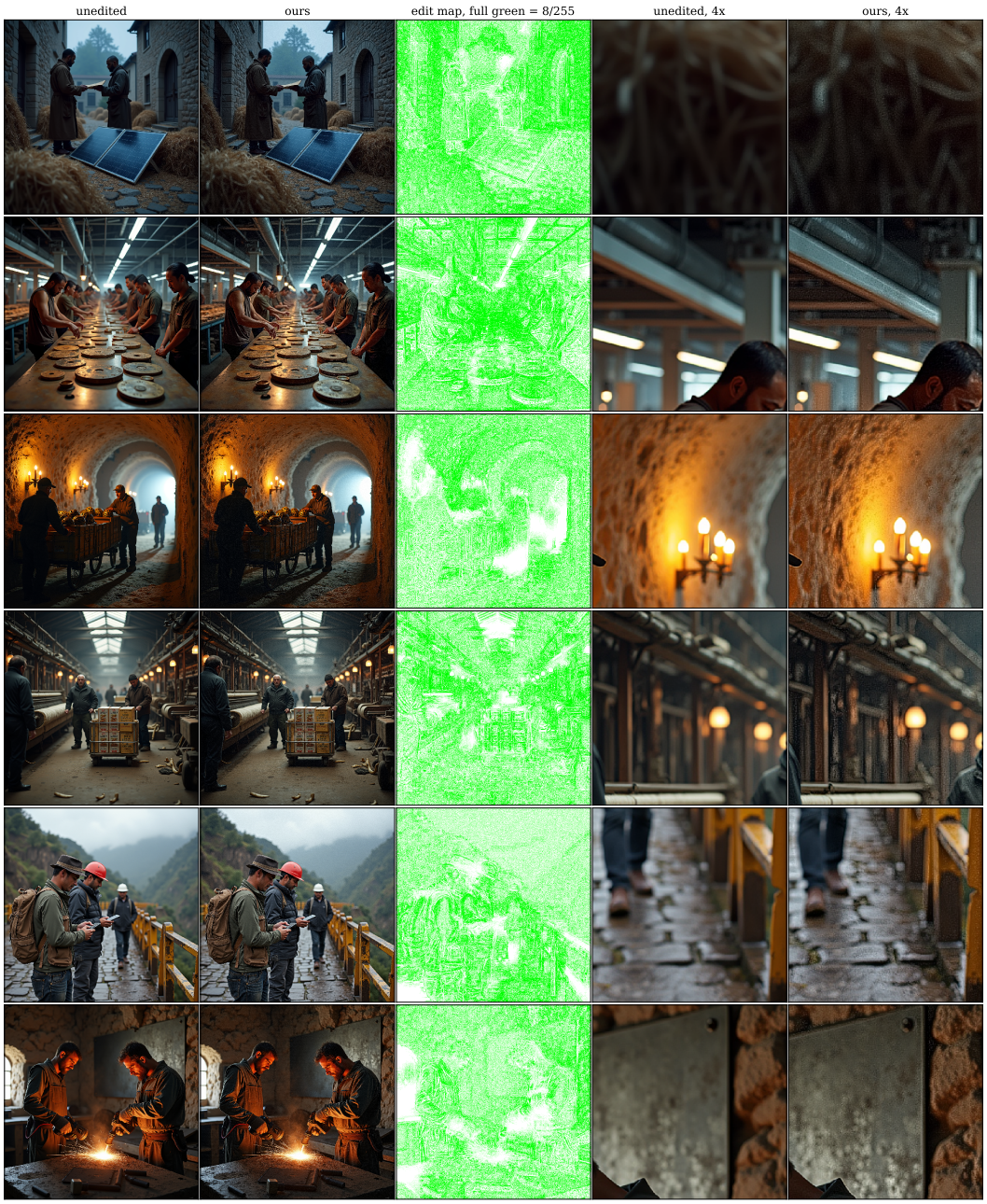


Figure 16: Qualitative results on FLUX.1-dev, laid out as in Figure 12.