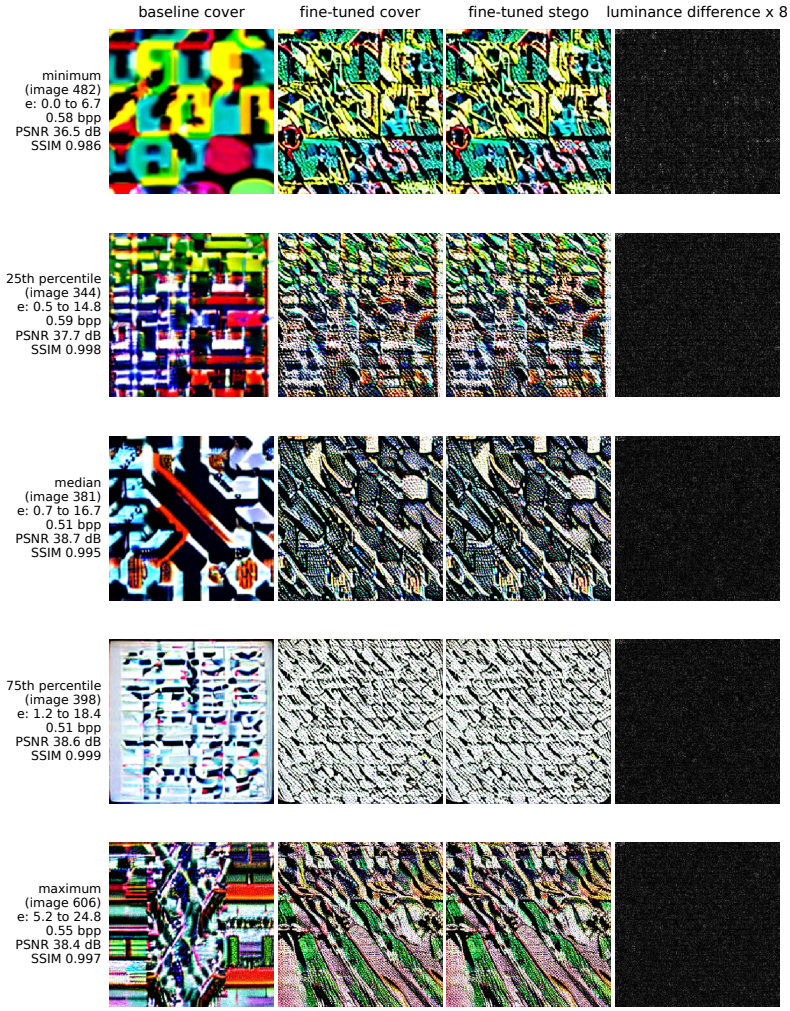
Supplementary Information

Diffusion-enhanced DCT steganography: reward-guided cover generation for increased embedding capacity with exact lossless-channel recovery

MD Ezazul Haque Mahi, Imraj Rabbani, Mustavi Hassan, Samia Tabassum Chowdhury, Md. Fahmid-Ul-Alam Juboraj and Muhammad Iqbal Hossain



Supplementary Figure S1. Rule-selected examples from the 1,000-pair evaluation corpus. Rows are the pairs at the minimum, 25th, 50th and 75th percentiles and maximum of the fine-tuned eligible count $\bar{e}$; no other selection was applied. Columns: baseline cover; fine-tuned cover generated from the same prompt and seed; stego image of the fine-tuned cover; absolute luminance difference between stego and cover, amplified eight times. Row labels give the image index, $\bar{e}$ of the baseline and fine-tuned cover, the embedded payload and the luminance PSNR and SSIM of the fine-tuned stego image (margin-projected embedder, $\mu = 0.15$). The corpus was sampled without classifier-free guidance at 256×256 pixels (Methods).

Supplementary Table S1. Additional processing tests, main mode (margin-projected embedder; the 1,000 fine-tuned and 1,000 baseline stego images of the matched-payload sweep, 0.30 to 0.60 bits per pixel). Payload bits are read at the known slot positions without voting, so a destroyed channel reads about 50%. The non-standard table is an arbitrary 8×8 quantisation table applied to the luminance channel. Shifts and the rotation use border replication; the crop removes a 2-pixel border and resizes bilinearly.

	Fine-tuned covers		Baseline covers
Condition	Bits recovered, % (s.d.)	Exact recovery	Bits recovered, %
Unmodified stego image	100.0 (0.0)	1,000	100.0
JPEG, quality 95	71.5 (2.7)	0	78.9
Gaussian noise, $\sigma = 2$	85.0 (1.5)	0	87.9
Non-standard quantisation table	62.3 (3.2)	0	67.2
Shift by 1 pixel	50.7 (0.7)	0	61.7
Shift by 2 pixels	50.3 (0.5)	0	58.1
Rotation by 1°	50.5 (1.5)	0	54.4
Crop 2 pixels and resize	50.5 (1.8)	0	54.6

Supplementary Table S2. Cover statistics across training checkpoints. Covers were generated from the checkpoint saved at each step with the same prompts and seeds and analysed before embedding, with the same definitions as Table 1 of the main text. $\bar{e}$, mean eligible coefficients per block; K, mean adaptive position parameter, where one of the four baseline covers falls back to the fixed $K = 24$; energy ratio, share of block DCT energy at positions with $3 \leq u + v \leq 10$; capacity, mean raw slot capacity $1022\,(K + 1)$. Values are means over four validation covers per checkpoint, so they vary more than the corpus means of Table 1. Step 0 is the unmodified baseline model. The evaluation corpus in the main text was generated from the final checkpoint.

Step	$\bar{e}$	Mean K	Energy ratio	Capacity (bits)
0	2.67	8.75	0.021	9,964
5,000	14.60	21.25	0.299	22,740
10,000	14.07	19.75	0.259	21,206
15,000	14.26	20.00	0.313	21,462
20,000	14.81	21.25	0.332	22,740
25,000	15.18	21.75	0.329	23,250
30,000	14.98	21.00	0.314	22,484
35,000	15.61	22.50	0.282	24,017
40,000	15.93	22.75	0.278	24,272
45,000	16.73	24.00	0.283	25,550
50,000	15.52	22.25	0.248	23,762
55,000	16.01	23.00	0.282	24,528
60,000	16.42	23.75	0.277	25,294
65,000	15.70	22.50	0.271	24,017
70,000	16.84	24.25	0.281	25,806
75,000	16.08	23.00	0.282	24,528
80,000 (final)	15.70	22.00	0.272	23,506

Supplementary Table S3. Mean adaptive K over a grid of loss weights. Each setting was trained for 5,000 steps, and each value is the mean over ten covers generated from the resulting model (prompts 0 to 9 of the prompt bank, seeds 42 to 51). Every setting was trained on the same filtered dataset, so the row $\lambda_{\text{mid}} = 0$ isolates the effect of the reward on top of the dataset filtering. The mean K includes the fixed fallback $K = 24$ for any cover with fewer than one eligible coefficient per block on average, which inflates settings that produce very few eligible coefficients (notably the row $\lambda_{\text{mid}} = 0$); the table therefore ranks settings but does not measure eligible coefficients directly. The setting used in the main text is $\lambda_{\text{mid}} = 0.005$, $\lambda_{\text{anchor}} = 1$. Values are taken from the authors' ablation record; the generated covers of each setting are archived with it. The record, with links to the image folders, is included in the data repository.

$\lambda_{\text{mid}} \setminus \lambda_{\text{anchor}}$	0	0.1	0.5	1	2
0	12.6	11.8	7.8	6.3	6.0
0.001	24.1	23.9	21.0	20.2	19.9
0.005	30.2	29.6	25.2	22.3	19.0
0.01	34.3	33.6	28.5	24.9	21.9
0.05	35.7	35.7	34.8	33.7	29.7
0.1	36.9	37.1	36.8	36.6	32.9

Supplementary Table S4. Embedding and extraction time per image, margin-projected embedder, all 1,000 pairs (Python with vectorised block transform and colour search; three worker processes on a six-core AMD Ryzen 5 7500F that was shared with other jobs, so single-process times would be shorter). Embedding time includes the verify-fix loop and the colour repair. Median with interquartile range. A version of the colour search that visits one pixel at a time took a median of 15 to 79 s per image at these payloads under the same conditions.

Payload (bpp)	Covers	n	Embedding, median (IQR) (s)	Maximum (s)	Verify-fix passes, median	Extraction, median (s)
0.30 to 0.40	Fine-tuned	330	0.67 (0.44 to 0.90)	22	14	0.019
0.30 to 0.40	Baseline	330	0.69 (0.42 to 1.19)	88	12	0.018
0.40 to 0.50	Fine-tuned	340	1.01 (0.65 to 1.49)	24	15	0.018
0.40 to 0.50	Baseline	340	1.13 (0.64 to 1.97)	44	14	0.018
0.50 to 0.60	Fine-tuned	330	1.46 (0.91 to 2.18)	51	17	0.028
0.50 to 0.60	Baseline	330	1.66 (0.95 to 3.32)	304	15	0.028

Supplementary Table S5. Luminance quantisation table $Q[u, v]$ used for the reward, for eligibility and for embedding. Bold entries ($Q \geq 8$, excluding the DC term) form the set $\mathcal{M}$ of 39 positions. The table is the one defined in the released embedder (`DCT_Adaptive_v5.py`, unchanged from the earlier `DCT_Adaptive.py`), where it is described as a JPEG luminance table of quality approximately 85. It is used as given and does not coincide exactly with the standard libjpeg table at any quality setting.

$u\backslash v$	0	1	2	3	4	5	6	7
0	3	2	2	3	4	6	**8**	**10**
1	2	2	3	4	5	**9**	**10**	**9**
2	3	3	4	5	6	**9**	**11**	**9**
3	3	4	5	6	**8**	**14**	**13**	**10**
4	4	5	7	**9**	**11**	**17**	**16**	**12**
5	5	7	**9**	**10**	**13**	**17**	**18**	**15**
6	**10**	**13**	**12**	**14**	**16**	**19**	**19**	**17**
7	**14**	**17**	**18**	**18**	**19**	**18**	**19**	**17**

Supplementary Table S6. Exploratory robust mode: message bits recovered after processing. Two hundred fine-tuned covers drawn at random (seed 2026), 25 per net payload; the message is framed with a 16 bit byte count and a 16 bit CRC, written over all 39,858 slots in a keyed order and decoded by the reliability-weighted vote (Methods). Entries are the percentage of message bits recovered with the frame length known, and in parentheses the number of messages, out of 25, that the extractor returned exactly with no side information. For the ASCII test messages a decoder that has lost all information still matches about 54% of the bits. This mode does not make the method robust; it shows what redundancy can recover on mild channels at reduced payload.

Net payload (bits per pixel)	0.005	0.02	0.05	0.1	0.2	0.3	0.45	0.6
Mean copies per frame bit	110.7	29.7	12.0	6.1	3.0	2.0	1.4	1.0
Luminance PSNR (dB)	33.1	33.1	33.2	33.2	33.1	33.1	33.1	33.1
None (PNG round trip)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)
Own table, scale 1.0	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (23)	99.9 (0)	99.9 (0)
Own table, scale 0.6	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (20)	99.8 (0)	99.0 (0)	96.1 (0)	94.7 (0)
Own table, scale 0.2	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (19)	99.9 (0)	99.1 (0)	98.7 (0)
JPEG, quality 95	100.0 (25)	100.0 (25)	100.0 (25)	99.9 (0)	98.5 (0)	96.7 (0)	92.4 (0)	90.2 (0)
JPEG, quality 90	100.0 (25)	100.0 (25)	99.9 (0)	98.5 (0)	93.9 (0)	89.3 (0)	85.2 (0)	82.8 (0)
JPEG, quality 70	55.4 (0)	53.9 (0)	55.4 (0)	51.1 (0)	48.5 (0)	45.0 (0)	48.2 (0)	48.2 (0)
JPEG, quality 50	55.5 (0)	55.0 (0)	59.1 (0)	56.2 (0)	54.3 (0)	50.7 (0)	53.1 (0)	53.0 (0)
Gaussian noise, $\sigma = 2$	100.0 (25)	100.0 (25)	100.0 (25)	100.0 (24)	99.9 (0)	99.5 (0)	97.4 (0)	96.3 (0)
Gaussian noise, $\sigma = 5$	100.0 (25)	100.0 (25)	100.0 (18)	99.3 (0)	96.3 (0)	92.8 (0)	88.5 (0)	86.4 (0)
Gaussian noise, $\sigma = 10$	100.0 (25)	98.8 (0)	92.6 (0)	84.9 (0)	77.3 (0)	72.5 (0)	69.6 (0)	68.0 (0)
Gaussian noise, $\sigma = 15$	96.5 (0)	82.2 (0)	72.5 (0)	66.4 (0)	62.1 (0)	59.5 (0)	58.3 (0)	57.7 (0)
Resize to 0.90 and back	59.2 (0)	60.0 (0)	61.1 (0)	58.0 (0)	55.7 (0)	52.6 (0)	53.6 (0)	53.6 (0)
Resize to 0.75 and back	55.5 (0)	54.6 (0)	57.7 (0)	54.7 (0)	53.0 (0)	50.7 (0)	51.9 (0)	51.8 (0)
Resize to 0.50 and back	55.5 (0)	55.0 (0)	58.7 (0)	55.2 (0)	53.4 (0)	50.7 (0)	52.4 (0)	52.2 (0)
Gaussian blur, 3×3	61.6 (0)	61.6 (0)	61.6 (0)	58.9 (0)	55.6 (0)	53.1 (0)	52.7 (0)	51.8 (0)
Crop 5% border and resize back	55.1 (0)	52.8 (0)	54.3 (0)	52.8 (0)	51.4 (0)	50.1 (0)	50.9 (0)	50.9 (0)
Non-standard table	93.3 (0)	86.7 (0)	84.1 (0)	80.8 (0)	76.8 (0)	71.4 (0)	70.7 (0)	69.6 (0)
Shift by 1 pixel	71.5 (0)	61.5 (0)	58.5 (0)	56.6 (0)	54.5 (0)	52.5 (0)	53.0 (0)	52.8 (0)
Shift by 2 pixels	57.4 (0)	53.7 (0)	53.6 (0)	52.6 (0)	51.7 (0)	50.7 (0)	50.9 (0)	51.0 (0)
Rotation by 1°	55.9 (0)	54.5 (0)	56.6 (0)	54.3 (0)	52.5 (0)	50.8 (0)	51.6 (0)	51.6 (0)
Crop 2 pixels and resize	55.7 (0)	54.7 (0)	57.1 (0)	54.8 (0)	52.9 (0)	51.0 (0)	52.0 (0)	51.8 (0)

Supplementary Table S7. Check on prompts outside the evaluation prompt bank. A further 200 baseline and fine-tuned cover pairs were generated from 20 everyday subjects that are not in the prompt bank and are not texture close-ups (for example a city street at night, a dog sitting on grass, a bowl of fruit), each with the ten modifiers of the bank and with seeds used for no other image. The pairs were processed with the protocol of the main text: a target payload drawn uniformly from 0.30 to 0.60 bits per pixel (mean 0.454), K set from the target, the same random ASCII message on both covers, a PNG round trip and a byte for byte comparison. Values are mean ± standard deviation or counts out of 200. The detector rows use the SRNet-KV detector for DEDS of the main text without retraining; it was trained at a margin of 0.25 and, with those training settings, detects DEDS in these 200 fine-tuned covers with AUC 0.993, so the values below measure transfer to the margin of 0.15 used here. Embedding uses the margin-projected embedder ($\mu = 0.15$) with the same targets, K and messages as the earlier run. Luminance PSNR was higher on fine-tuned covers over all pairs ($p = 8.9 \times 10^{-29}$, paired Wilcoxon test) and in every payload bin, by +1.09 to +1.97 dB (all $p < 10^{-4}$); with the bin-centre embedder the two sets had not differed ($p = 0.83$).

	Baseline covers	Fine-tuned covers
Eligible coefficients per block	1.66 ± 1.82	16.89 ± 2.83
Covers on the fallback K	98	0
Cover-derived capacity (bits per pixel)	0.065 (102 covers without fallback)	0.396
Covers with capacity of at least 0.30 bits per pixel	not applicable	181
Exact message recovery	200	200
Verify-fix passes, median (maximum)	14 (21)	15 (22)
Luminance PSNR (dB)	38.33 ± 2.14	39.84 ± 1.46
Luminance SSIM	0.951	0.997
Matched SRNet-KV detector, ROC AUC	0.865 (95% CI 0.815 to 0.911)	0.720 (95% CI 0.695 to 0.751)
Matched SRNet-KV detector, P_E at the validation threshold	0.088	0.455
Matched SRNet-KV detector, P_E at the best threshold	0.070	0.322

Supplementary Note 1: procedure, header layout, colour repair and optional compression

Procedure. Embedding: (1) convert the cover to YCrCb and take Y; (2) compute the 8×8 block DCT of $Y - 128$; (3) determine K with the adaptive rule of the main text, or use the value supplied; (4) write the 64 header bits into blocks $(0, 0)$ and $(0, 1)$; (5) write the n message bits into the first n payload slots in the fixed slot order with the margin-projected write rule and margin μ; (6) reconstruct Y by inverse DCT and rounding, re-read all bits and rewrite any that changed with the escalated margin, repeating until none change; (7) convert to BGR and repair the pixels of any block whose bits read wrongly after the colour conversion; (8) save as PNG. Extraction: (1) load the PNG, convert to YCrCb and take Y; (2) compute the block DCT; (3) read and check the header, obtaining n and K; (4) rebuild the slot order from K and the image size; (5) read the parities of the first n slots and pack them into bytes. Embedding and extraction use the same functions for the block transform, the quantised parity and the slot order.

Header. The 64 header bits are written most significant bit first in the order: identifier (16 bits, value 0x4447), payload length in bits (24 bits), K (6 bits), format version (4 bits: 4 for the main mode, 6 for the robust mode; 3 in the bin-centre implementation), integrity check (14 bits). The integrity check is the low 14 bits of the CRC-32 (zlib) of the first 50 bits packed into 7 bytes with 6 trailing zero bits. Bits 0 to 31 are written to block $(0, 0)$ and bits 32 to 63 to block $(0, 1)$, at the 32 positions of $\mathcal{M}$ with the largest quantisation step, ordered by decreasing step and then by row and column index. Header bits are written at bin centres (margin $\frac{1}{2}$), payload bits with the margin μ. The header blocks carry no payload.

Colour repair. After the verify-fix loop has produced a luminance channel whose block DCT carries every bit, the image is converted from YCrCb to BGR with the 8-bit OpenCV transform. This conversion changes the luminance of some pixels by one grey level when the image is read back. The repair pass re-reads the image, finds the 8×8 blocks in which at least one header or payload bit is wrong, and for each pixel in those blocks whose luminance differs from the target searches the colours within a radius of 4 levels in R and B (10, then 40, in later iterations) for the one nearest to the current colour that gives the target luminance under the integer formula of the main text; G is solved for directly. The pass repeats until all bits read correctly, for at most 60 iterations. The search is evaluated for all affected pixels at once and, at each radius, only on the new ring of candidates, because the smaller squares were already found empty; it returns exactly the colour of the pixel-by-pixel search, including ties, which we checked on 12,000 random pixels at radii 1, 4, 10 and 40 and on complete embeddings. Because a grey pixel (v, v, v) has luminance exactly v, every target value can be reached.

Optional compression. The released training pipeline includes a canonical Huffman codec that can be applied to the message before embedding. Its output begins with the original message length (4 bytes), the number of distinct symbols (1 byte) and one (symbol, code length) byte pair per symbol, followed by the packed code words padded to a byte boundary. The codec was not used in any experiment reported in the main text; all payloads there are uncompressed message bits.

Supplementary Note 2: message content

The test messages of the main text are random ASCII characters, so the first bit of every byte is zero and a byte carries about 6.07 bits of information rather than 8. To see whether this matters, every tenth pair of the corpus (100 pairs) was embedded again with uniformly random bits, using the margin-projected embedder with the same K and the same payload length as in the main

run.

Every message was recovered exactly, on all 200 stego images. On fine-tuned covers the message content made almost no difference: luminance PSNR was 40.07 dB with random bits and 40.10 dB with text. On baseline covers it fell from 38.68 to 38.21 dB ($p = 3 \times 10^{-4}$, paired Wilcoxon test), and SSIM from 0.952 to 0.947. The reason is the zero in every eighth bit of text. Most coefficients of a smooth baseline cover quantise to zero, an even value, so with text more slots already held the wanted parity and did not need to move. With random bits that advantage is gone.

The comparison between the two cover types therefore becomes stronger with random bits. The luminance PSNR advantage of fine-tuned over baseline covers rose from 1.42 to 1.87 dB on these pairs and was positive in all 100 of them ($p = 4 \times 10^{-18}$). The results of the main text, which use text messages, slightly understate the benefit of the fine-tuned covers.

Supplementary Note 3: derivations and numerical checks

Notation. A payload coefficient C at a position with quantisation step Q has quantised value $q = \mathrm{round}(C/Q)$ and carries the bit $q \bmod 2$. Its offset in the bin is $x = C - qQ \in [-Q/2, Q/2]$. The admissible part of the bin of an integer q' is $I_\mu(q') = [(q' - \frac{1}{2} + \mu)Q, (q' + \frac{1}{2} - \mu)Q]$, and the write rule (Methods of the main text, parity embedding with a margin) projects C onto $I_\mu(q)$ when the parity matches and onto the nearer of $I_\mu(q \pm 1)$ otherwise.

Result 1 (expected distortion). *If x is uniform on $[-Q/2, Q/2]$ and independent of the message bit, then $E[(C' - C)^2] = (\frac{1}{2} + \mu)^3 Q^2/3$. Derivation.* With probability $\frac{1}{2}$ the parity matches. The coefficient then moves only if $|x| > (\frac{1}{2} - \mu)Q$, an event of probability 2μ, and the move $t = |x| - (\frac{1}{2} - \mu)Q$ is uniform on $[0, \mu Q]$, so $E[t^2] = \mu^2 Q^2/3$ and the contribution is $2\mu \cdot \mu^2 Q^2/3 = \frac{2}{3}\mu^3 Q^2$. With probability $\frac{1}{2}$ the parity must change. Let $s = Q/2 - |x|$, uniform on $[0, Q/2]$, be the distance to the nearer edge; the nearer neighbouring bin is entered across that edge, and projection onto its admissible part moves the coefficient by $s + \mu Q$. The neighbour on the far side is never closer, so it is never chosen. Hence the contribution is $\frac{2}{Q}\int_0^{Q/2}(s+\mu Q)^2 ds = \frac{2}{3}[(\frac{1}{2}+\mu)^3 - \mu^3]Q^2$. Averaging, $\frac{1}{2}\cdot\frac{2}{3}\mu^3 Q^2 + \frac{1}{2}\cdot\frac{2}{3}[(\frac{1}{2}+\mu)^3 - \mu^3]Q^2 = (\frac{1}{2}+\mu)^3 Q^2/3$. At $\mu = \frac{1}{2}$ the admissible intervals are the bin centres and the value is $Q^2/3$; the gain of the margin rule over bin-centre writing is $G(\mu) = 10\log_{10}[(1/3)/((\frac{1}{2} + \mu)^3/3)] = -30\log_{10}(\frac{1}{2} + \mu)$ dB.

Result 2 (pixel domain). The 8×8 DCT used here is orthonormal, so for any change ΔC of a block, $\sum_{u,v} \Delta C_{uv}^2 = \sum_{i,j} \Delta y_{ij}^2$ for the unrounded pixel change Δy (Parseval). Result 1 therefore also gives the squared pixel error before rounding.

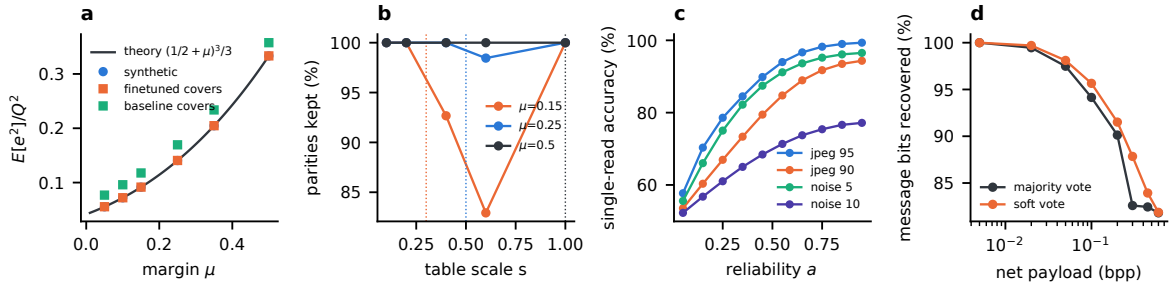
Result 3 (tolerance). *A written coefficient keeps its parity under any later change e with $|e| < \mu Q$.* Every written coefficient lies in some $I_\mu(q')$, whose points are at least μQ from both edges of the bin of q', so $\mathrm{round}((C' + e)/Q) = q'$. Two applications: (i) rounding the inverse transform adds to each pixel an error uniform on $[-\frac{1}{2}, \frac{1}{2}]$, variance $1/12$; with errors close to independent, each coefficient of an orthonormal transform receives variance $\sum_{ij} b_{ij}^2/12 = 1/12$, standard deviation 0.289. (ii) Requantising with a table of steps Q_s moves a coefficient by at most $Q_s/2$, so all parities survive if $Q_s/2 < \mu Q$; with the embedder's table scaled by s, $Q_s \approx sQ$ and the condition is $\mu > s/2$. This is a statement about the coefficients before the requantised image is rounded back to pixels. That rounding adds about 0.29 per coefficient again, so when the margin left over, $(\mu - s/2)Q$, is small, some parities flip in the stored image even though the condition holds. With the unscaled table every coefficient goes to the centre $q'Q$ of its own bin, which has the same parity, for any μ.

Result 4 (margin escalation). A slot that fails for the f-th time is rewritten with $\mu_f = \min(\frac{1}{2}, \mu 2^{f-1})$, so $\mu_f = \frac{1}{2}$ once $f \geq 1 + \log_2(1/2\mu)$: at most three failures for $\mu = 0.15$ and two for $\mu = 0.25$. From then on the slot is written at the bin centre, the setting with the largest tolerance, which is also the setting of the bin-centre rule.

Result 5 (reliability weight). Let a copy written at a bin centre be perturbed by e with a density p that is symmetric and decreasing in $|e|$, and let the read offset be $f = C/Q - \mathrm{round}(C/Q)$. Considering the true bin and the nearest bin of the other parity, the read is correct if $e = fQ$ and wrong if $|e| = (1 - |f|)Q$, so the posterior odds of a correct read are $p(|f|Q)/p((1 - |f|)Q)$, which decrease as $|f|$ grows and therefore increase with $a = 1 - 2|f|$. For Gaussian p with variance σ^2 the log odds are $[(1 - |f|)^2 - f^2]Q^2/(2\sigma^2) = aQ^2/(2\sigma^2)$, exactly proportional to a. Summing $(2b - 1)a$ over the copies of a frame bit therefore equals a scaled sum of per-copy log-likelihood ratios only when Q and σ are the same for every copy. The copies lie at different positions, so the weighted vote of the main text approximates that sum; weighting each copy by Q^2/σ^2 would make it exact under this model.

Numerical checks. Supplementary Fig. S2 and Supplementary Tables S8 to S11 test Results 1

to 5. Synthetic data: two million coefficients per step $Q \in \{8, 13, 19\}$ with uniform offsets and random bits. Real data: the first write pass on 100 fine-tuned and 100 baseline covers (every tenth pair) at the payloads and K of Table 2, in the coefficient domain before rounding. The vectorised write rule used for these checks returned the same value as the embedder's own function on 2,000 random cases. On the real covers Parseval's identity held to a relative difference of 4×10^{-16}, and the rounding noise of the written images had a standard deviation of 0.288 per coefficient against the predicted 0.289.



Supplementary Figure S2. Numerical checks of the analysis. a, Mean squared coefficient change per slot, in units of Q^2, against the margin μ: theory $(\frac{1}{2}+\mu)^3/3$, synthetic coefficients with uniform offsets, and the first write pass on 100 fine-tuned and 100 baseline covers. **b**, Parities kept after requantisation with the embedder's table scaled by s, coefficient domain; dotted lines mark $s = 2\mu$, below which Result 3 guarantees survival. **c**, Accuracy of single reads in robust mode against the reliability a, for the 200 robust-mode stego images under four operations. **d**, Message bits recovered by the reliability-weighted vote and by a majority vote on the same reads, mean over JPEG 95 and 90 and noise of $\sigma = 5$ and 10.

Supplementary Table S8. Check of Results 1 and 2: distortion of the write rule. $E[e^2]/Q^2$: mean squared coefficient change per slot in units of Q^2. Gain: reduction of the total squared change relative to $\mu = \frac{1}{2}$ (bin-centre writing). Fine-tuned covers match the uniform-offset model closely; 94% of the slots of baseline covers hold $|q| \leq 1$ (48% for fine-tuned covers), where the model's assumption fails.

	$E[e^2]/Q^2$				Gain over $\mu = \frac{1}{2}$ (dB)		
μ	Theory	Synthetic	Fine-tuned	Baseline	Theory	Fine-tuned	Baseline
0.05	0.0555	0.0554	0.0555	0.0769	7.79	7.78	6.46
0.10	0.0720	0.0720	0.0721	0.0960	6.66	6.65	5.52
0.15	0.0915	0.0915	0.0916	0.1177	5.61	5.61	4.67
0.25	0.1406	0.1406	0.1407	0.1695	3.75	3.74	3.13
0.35	0.2047	0.2046	0.2047	0.2337	2.12	2.12	1.79
0.50	0.3333	0.3332	0.3333	0.3576	0.00	0.00	0.00

Supplementary Table S9. Check of Result 3: parities kept (%) after requantisation with the embedder's table scaled by s (coefficient domain, the 200 covers of Table S8). Bold: settings with $\mu > s/2$, where survival is guaranteed; it was 100% in every image. At $s = 1$ every coefficient returns to its own bin centre, so survival is 100% for any μ.

μ	$s = 1.0$	$s = 0.6$	$s = 0.4$	$s = 0.2$	$s = 0.1$
0.15	100.0	82.9	92.7	**100.0**	**100.0**
0.25	100.0	98.5	**100.0**	**100.0**	**100.0**
0.5	100.0	**100.0**	**100.0**	**100.0**	**100.0**

Supplementary Table S10. Check of Result 5: accuracy (%) of single reads in robust mode by reliability a (all 200 robust-mode stego images; bins of a of width 0.1, lower edge given).

Operation	$a \geq 0.0$	$a \geq 0.2$	$a \geq 0.4$	$a \geq 0.6$	$a \geq 0.8$	$a \geq 0.9$
JPEG, quality 95	57.7	78.6	89.9	96.7	99.0	99.3
JPEG, quality 90	53.6	67.0	79.5	88.9	93.5	94.3
Gaussian noise, $\sigma = 5$	55.6	75.0	87.5	93.6	96.1	96.5
Gaussian noise, $\sigma = 10$	52.3	61.0	68.4	73.7	76.6	77.2

Supplementary Table S11. Weighted vote against majority vote on the same reads (robust mode, 25 covers per payload). Entries: message bits recovered (%), with exact recoveries out of 25 in parentheses. Over all 800 decodes (four operations, eight payloads) the weighted vote gave 218 exact recoveries and the majority vote 188.

Operation	Net payload (bpp)	Majority vote	Weighted vote
JPEG, quality 95	0.02	100.0 (25)	100.0 (25)
JPEG, quality 95	0.05	100.0 (13)	100.0 (25)
JPEG, quality 95	0.1	99.2 (0)	99.9 (0)
JPEG, quality 95	0.3	91.3 (0)	96.7 (0)
JPEG, quality 95	0.6	90.2 (0)	90.2 (0)
JPEG, quality 90	0.02	100.0 (24)	100.0 (25)
JPEG, quality 90	0.05	99.6 (0)	99.9 (0)
JPEG, quality 90	0.1	97.2 (0)	98.5 (0)
JPEG, quality 90	0.3	84.1 (0)	89.3 (0)
JPEG, quality 90	0.6	82.8 (0)	82.8 (0)
Gaussian noise, $\sigma = 5$	0.02	100.0 (25)	100.0 (25)
Gaussian noise, $\sigma = 5$	0.05	99.8 (1)	100.0 (18)
Gaussian noise, $\sigma = 5$	0.1	98.1 (0)	99.3 (0)
Gaussian noise, $\sigma = 5$	0.3	86.6 (0)	92.8 (0)
Gaussian noise, $\sigma = 5$	0.6	86.3 (0)	86.4 (0)
Gaussian noise, $\sigma = 10$	0.02	97.8 (0)	98.8 (0)
Gaussian noise, $\sigma = 10$	0.05	90.6 (0)	92.6 (0)
Gaussian noise, $\sigma = 10$	0.1	82.1 (0)	84.9 (0)
Gaussian noise, $\sigma = 10$	0.3	68.4 (0)	72.5 (0)
Gaussian noise, $\sigma = 10$	0.6	68.0 (0)	68.0 (0)

Supplementary Table S12. Real photographs and processed baseline covers, all settings. Top: cover statistics before embedding, judged by the embedder's own rule. $\bar{e}$, mean eligible coefficients per block. Fallback, covers with fewer than one eligible coefficient per block on average, which would get the fixed $K = 24$. Capacity, mean cover-derived capacity over the covers without fallback (their number in brackets). High frequency share, median share of luminance power above 0.25 cycles per pixel. Change to the cover, luminance PSNR against the original baseline cover. Bottom: the same forced K and message on every fifth pair (200 pairs); the photographs use the K and message of pairs 0 to 194 and are compared with the fine-tuned covers of those pairs. Δ, fine-tuned minus the row, with the two sided paired Wilcoxon test. The unsharp mask uses a Gaussian of standard deviation one pixel (Methods).

Covers	$\bar{e}$	Fallback	Capacity (bpp)	At least 0.30 bpp	High frequency share (%)	Change to the cover (dB)
Fine-tuned	16.64	0	0.389 (1,000)	924	16.34	none
Baseline	1.57	434	0.056 (566)	0	0.77	none
Real photographs (195)	1.94	72	0.065 (123)	0	2.92	none
Unsharp mask, amount 0.5	2.59	220	0.074 (780)	3	1.38	34.0
Unsharp mask, amount 1	3.55	123	0.093 (877)	7	2.03	28.8
Unsharp mask, amount 2	5.20	42	0.126 (958)	28	3.26	24.0
Unsharp mask, amount 4	7.69	13	0.182 (987)	116	5.36	20.1
Gaussian noise, $\sigma = 2$	1.59	427	0.056 (573)	0	0.79	45.8
Gaussian noise, $\sigma = 4$	1.64	402	0.056 (598)	0	0.85	40.2
Gaussian noise, $\sigma = 8$	1.92	293	0.057 (707)	0	1.05	34.3

Covers (matched payload)	n	Exact recovery	PSNR_Y (dB)	SSIM_Y	$\Delta\ \text{PSNR}_Y$ (dB), p
Fine-tuned	200	200	39.89	0.996	
Baseline	200	200	38.32	0.946	$+1.58,\ 7 \times 10^{-30}$
Unsharp mask, amount 4	200	200	38.68	0.954	$+1.22,\ 2 \times 10^{-32}$
Gaussian noise, $\sigma = 8$	200	200	40.02	0.980	$-0.13,\ 0.014$
Real photographs	195	195	37.21	0.913	$+2.82,\ 3 \times 10^{-33}$

Supplementary Table S13. Payload bits recovered after processing, main mode ($n = 1{,}000$ stego images of the matched-payload sweep per cover set, 0.30 to 0.60 bpp, margin-projected embedder). Bits are read at the known slot positions without voting, so a destroyed channel reads about 50%. Header intact and exact recovery are numbers of images out of 1,000. Resizing and cropping were followed by bicubic resampling back to 256×256. Further conditions are listed in Supplementary Table S1.

	Fine-tuned covers			Baseline
Condition	Bits recovered (%)	Header intact	Exact	Bits (%)
None (PNG round trip)	100.0	1,000	1,000	100.0
Own table, scale 1.0 / 0.6 / 0.2	99.9 / 78.1 / 93.5	989 / 118 / 513	0	99.9 / 85.9 / 95.1
JPEG, quality 95 / 90	71.5 / 71.3	25 / 0	0	78.9 / 81.5
JPEG, quality 70 / 50	49.1 / 51.2	0	0	52.6 / 53.0
Gaussian noise, $\sigma = 2$ / 5	85.0 / 71.1	176 / 1	0	87.9 / 74.4
Gaussian noise, $\sigma = 10$ / 15	58.4 / 52.9	0	0	60.3 / 53.8
Resize to 0.90 / 0.75 / 0.50 and back	51.4 / 50.6 / 50.7	0	0	61.6 / 56.3 / 52.6
Gaussian blur, 3×3, $\sigma = 0.5$	50.8	0	0	56.4
Crop 5% border and resize back	50.2	0	0	52.5

Supplementary Table S14. How DEDS differs from closely related methods. Reference numbers are those of the main text. Each entry describes the method as its own publication reports it. None of these entries comes from an experiment in this work. n.r., not reported in the publication.

Method	Stego image	Message carrier	Training	Extraction needs	Channel targeted	Code
CRoSS (ref. 13)	generated	a secret image, hidden by diffusion inversion under a private and a public condition	none	the diffusion model and the private condition	lossy (tested on degraded images)	public
GSD (ref. 14)	generated	secret data turned into an image by an invertible diffusion model	its own invertible model	that model	lossless (exact recovery, no lossy tests)	not linked
PSyDUCK (ref. 17)	generated	the latent denoising path (96 to 512 bytes per image)	none	the latent diffusion model and shared keys	lossless (JPEG not tested)	not linked
BAFLE-DCT (ref. 20)	existing image	mid frequency DCT coefficients in low saliency blocks	a small block selection network	n.r.	lossless (tested against YeNet and SRNet)	promised on acceptance
DEDS (this work)	generated cover, then embedded	parities of quantised DCT coefficients	a LoRA adapter, trained once	the stego image only	lossless (PNG)	public

Supplementary Note 4: further cover checks and ablation

For the broader-domain check we generated 300 more pairs with the corpus sampler, 50 prompts in each of six domains outside the prompt bank (landscapes, urban scenes, animals, food and objects, indoor scenes, non-photographic styles), with seeds $300000 + i$, and embedded each at a matched payload drawn uniformly from 0.30 to 0.60 bpp (message seed $320000 + i$). The 5,640 Describable Textures Dataset photographs (Cimpoi et al., 2014) (47 categories) were centre-cropped and resized to 256×256 with area interpolation. A stratified subset of 1,000 was embedded with the

Supplementary Table S15. Cover statistics under other generation settings. The same prompts and seeds sampled by both models with classifier-free guidance (scale 7.5) at 256×256 (prompts 0, 5, ..., 995) and at the native 512×512 (prompts 0, 10, ..., 990), next to the corpus covers of the same prompts. For each model: $\bar{e}$, mean eligible coefficients per block; fallback, covers with fewer than one eligible coefficient per block on average; at least 0.30 bpp, covers whose cover-derived capacity reaches 0.30 bpp; HF, median share of luminance power above 0.25 cycles per pixel (%). Statistics at 512×512 are computed on the 512×512 image with the same block rule.

		Baseline				Fine-tuned			
Sampling	n	$\bar{e}$	Fallback	≥ 0.30	HF	$\bar{e}$	Fallback	≥ 0.30	HF
No guidance, 256×256 (corpus)	200	1.52	94	0	0.78	16.70	0	188	16.66
Guidance 7.5, 256×256	200	2.98	62	2	1.76	16.22	0	164	20.60
No guidance, 256×256 (corpus, same 100 prompts)	100	1.71	43	0	1.05	16.93	0	98	17.20
Guidance 7.5, 512×512	100	4.48	17	3	5.30	9.88	28	40	6.83

K and payload of corpus pair j and compared with fine-tuned cover j carrying the same message (seed $520000 + j$). Eligible-slot capacity is the number of slots at a cover's own K (fallback included) holding an eligible coefficient, divided by $H\,W$.

For the ablation, the released filter accepted 4,246 images from the first 87 training shards (3,317 after the size check); an unfiltered set of every 34th image from the same shards left 3,360. We trained four adapters with the released script for 5,000 steps, by which point the eligible count has settled (Supplementary Table S2), changing the objective and, for the last adapter, the data: the full objective; the reward without the anchor network; the anchor without the reward ($\lambda_{\text{mid}} = 0$); and $\lambda_{\text{mid}} = 0$ on the unfiltered set. The first three used the filtered set of 3,317 images and the fourth the unfiltered set of 3,360. Training resumed from checkpoints after interruptions. Each adapter generated covers for prompts 0, 5, ..., 995 with the corpus seeds and sampler, and we compared them with the corpus baseline and fine-tuned covers by CLIP similarity to the baseline cover and by matched-payload embedding (corpus K, message seed $620000 + i$).

Supplementary Table S16. Cover statistics and matched-payload embedding in six further domains. Three hundred pairs, 50 per domain, from prompts outside the evaluation bank, generated with the corpus sampler (Methods). Base, baseline; FT, fine-tuned; $\bar{e}$, mean eligible coefficients per block; fallback, baseline covers on the fixed K; ≥ 0.30, fine-tuned covers whose cover-derived capacity reaches 0.30 bpp; eligible-slot capacity in bpp; PSNR_Y and SSIM_Y at matched payload (mean 0.454 bpp). All 600 messages were recovered exactly, and $\bar{e}$ was higher on the fine-tuned cover in 300 of 300 pairs.

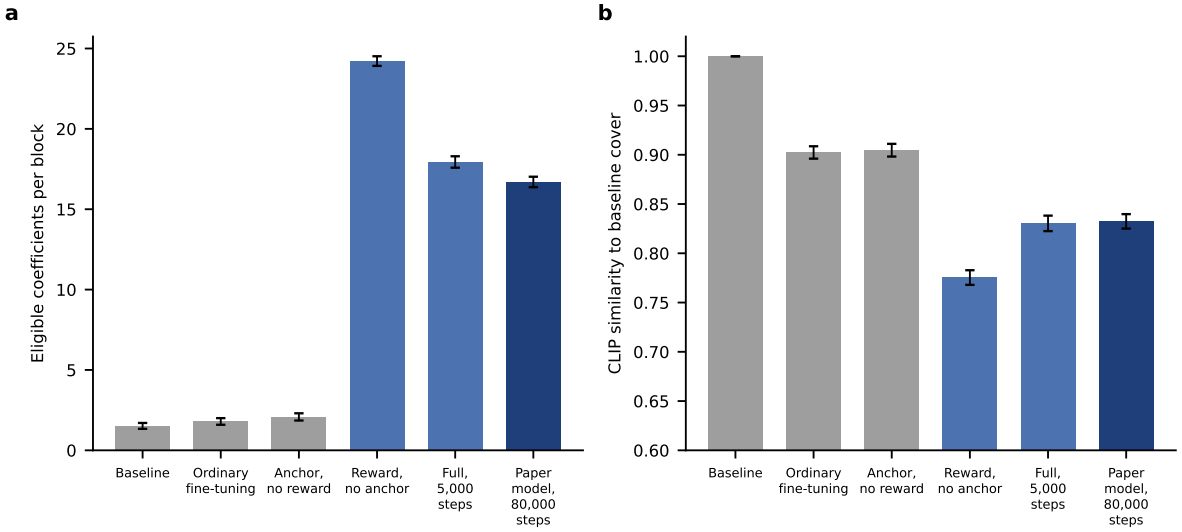
		$\bar{e}$				Eligible-slot (bpp)		PSNR_Y (dB)		SSIM_Y	
Domain	n	Base	FT	Fallback	≥ 0.30	Base	FT	Base	FT	Base	FT
Landscapes	50	1.59	16.46	22	45	0.011	0.213	38.28	39.95	0.954	0.997
Urban scenes, architecture	50	1.22	15.83	30	43	0.010	0.202	38.14	39.83	0.945	0.996
Animals	50	1.65	16.98	21	46	0.011	0.223	38.35	39.86	0.951	0.997
Food and objects	50	1.44	16.39	27	44	0.011	0.214	38.18	39.60	0.951	0.996
Indoor scenes	50	1.67	15.91	22	45	0.012	0.205	38.93	40.41	0.961	0.997
Non-photographic styles	50	1.74	16.45	22	46	0.013	0.213	37.83	39.58	0.948	0.997
All six domains	300	1.55	16.34	144	269	0.011	0.212	38.29	39.87	0.952	0.997

Supplementary Table S17. Texture photographs. The 5,640 Describable Textures Dataset photographs (47 categories, JPEG at source), centre-cropped and resized to 256×256, judged by the embedder's own rule and set beside the corpus covers. Category means of $\bar{e}$ ran from 0.55 to 8.61, and 0.5% of the photographs exceeded the fine-tuned mean of 16.64. At matched payload, each photograph was given the K and message of one corpus pair and compared with that pair's fine-tuned cover (1,000 pairs). Luminance PSNR was 39.17 dB for the photographs and 39.93 dB for the fine-tuned covers (higher on the fine-tuned cover in 711 pairs), and SSIM was 0.962 against 0.996. All 2,000 messages were recovered exactly.

Covers	n	$\bar{e}$	Fallback	Cover-derived ≥ 0.30 bpp	Eligible-slot capacity (bpp)
Texture photographs (DTD)	5,640	2.74	2,279	105	0.022
Fine-tuned corpus covers	1,000	16.64	0	924	0.216
Baseline corpus covers	1,000	1.57	434	0	0.011

Supplementary Table S18. Ablation of the training objective. Four adapters were trained for 5,000 steps with the released script: three on the filtered set of 3,317 images and the last (ordinary fine-tuning) on an unfiltered set of 3,360 images, so they differ in the objective and, for the last, in the data (Methods). They were evaluated on corpus prompts 0, 5, ..., 995 with the corpus seeds. The corpus baseline and the paper's model on the same 200 prompts are shown for reference. $\bar{e}$, mean eligible coefficients per block; fallback, covers on the fixed K; ≥ 0.30, covers whose cover-derived capacity reaches 0.30 bpp; HF, median share (%) of luminance power above 0.25 cycles per pixel; CLIP, cosine similarity of image embeddings to the baseline cover of the same prompt and seed; PSNR_Y and SSIM_Y at matched payload (corpus K, same message for every cover type); exact, messages recovered exactly out of 200. Covers with more eligible coefficients score higher on PSNR_Y and SSIM_Y because the same payload is spread over more slots, not because they are better images.

Covers	$\bar{e}$	Fallback	≥ 0.30	Eligible-slot (bpp)	HF (%)	CLIP	PSNR_Y (dB)	SSIM_Y	Exact
Baseline (no fine-tuning)	1.52	94	0	0.010	0.8	1.000	38.17	0.944	200
Paper model, 80,000 steps	16.70	0	188	0.216	16.7	0.832	39.88	0.996	200
Full objective	17.94	0	195	0.239	18.6	0.830	39.86	0.996	200
Reward without anchor	24.22	0	200	0.364	47.1	0.775	40.17	0.999	200
Anchor without reward	2.08	61	0	0.014	1.4	0.905	38.58	0.955	200
Ordinary fine-tuning (unfiltered)	1.80	80	0	0.012	1.3	0.902	38.54	0.955	200



Supplementary Figure S3. Ablation of the training objective. (a) Mean eligible coefficients per block and (b) CLIP similarity to the baseline cover of the same prompt and seed, over 200 prompts. Grey bars are variants without the reward term (baseline, ordinary fine-tuning, anchor only); blue bars are variants with it (reward only, full objective, paper model). Error bars are 95% confidence intervals of the mean.