RISKFUSE: SUPPRESSING UNSUPPORTED DETAIL IN GENERATIVE SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative super-resolution (SR) produces sharp images with rich high-frequency detail, but it can also introduce textures and edges unsupported by the low-resolution (LR) observation. These errors impair visual fidelity and can corrupt the evidence seen by downstream models when SR is used as preprocessing. We introduce RiskFuse, a post-processing method that suppresses unsupported detail while preserving much of the generative output's high-frequency texture. RiskFuse pairs the generative output with a frozen distortion-oriented output. A dense predictor locates regions where the generative result is less faithful; calibration on a disjoint validation split converts its predictions to a stable intervention scale; and LR consistency protects sharp detail that remains compatible with the input. The resulting fusion-weight map drives frequency-aware Laplacian-pyramid fusion without retraining either SR model. Across nine representative SR models spanning generative adversarial network and diffusion formulations, RiskFuse improves average PSNR-Y by 0.96 dB and LR-consistency PSNR by 1.03 dB, while reducing false-edge density by 56.3%. LPIPS increases modestly from 0.1134 to 0.1171. Across downstream case studies in detection and recognition, RiskFuse also improves representative task metrics over both direct LR input and the unmodified generative output. These results show that existing generative SR models can be made markedly more faithful without surrendering the visual detail that motivates their use.

1 INTRODUCTION

Single-image super-resolution (SISR) estimates a high-resolution (HR) image from a low-resolution observation. It supports conventional image-viewing applications such as photo enhancement, digital zoom, and archival enlargement, while also serving purposes beyond human viewing. In aerial imagery, surveillance, traffic scenes, and scene text, objects of interest may occupy only a few pixels. Enlarging such inputs before detection or recognition can expose spatial patterns at a scale better matched to a downstream model. The usefulness of this preprocessing step depends not only on how sharp the SR image appears, but also on whether the added detail is supported by the original observation.

Modern SR methods occupy two broad operating regimes. Distortion-oriented models optimize pixel or feature agreement with the HR reference and achieve strong peak signal-to-noise ratio (PSNR), but their estimates often average multiple plausible HR explanations and therefore appear smooth (Dong et al., 2016a; Lim et al., 2017; Zhang et al., 2018b; Liang et al., 2021). Generative models, first through generative adversarial networks (GANs) and more recently through diffusion priors, move toward natural-image statistics and synthesize sharper textures (Ledig et al., 2017; Wang et al., 2019b; Saharia et al., 2023; Yue et al., 2023; Wang et al., 2024). Yet SR is ill-posed, so visually plausible detail need not be the detail that produced the LR image.

Unsupported texture is consequential in two ways. It may create visible ringing, repeated patterns, or false contours, thereby reducing the credibility of the enlarged image. It may also change the local evidence consumed by a fixed detector or recognizer: an invented edge can split an object, merge adjacent structures, or alter a character stroke. Figure 1 gives examples from object detection on COCO and scene-text spotting on ICDAR2015 (Lin et al., 2014; Karatzas et al., 2015). Although the generative outputs are visually sharp, their altered local evidence leads to missed objects or incorrect

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: Unsupported SR detail can alter downstream predictions. Direct LR is enlarged only for display; each task uses the same frozen downstream model for all inputs. Blue annotations in the HR panels are ground truth. For COCO, green and orange boxes denote matched and unmatched detections; for ICDAR2015, they denote correctly and incorrectly transcribed text regions. The numbers below the panels are per-image AP and end-to-end F1.

text; suppressing unsupported detail recovers useful predictions without returning the entire image to a uniformly smooth solution.

This problem is closely related to the perception–distortion trade-off (Blau & Michaeli, 2018). Favoring perceptual quality generally produces sharper images, whereas favoring lower distortion leads to smoother but more reference-aligned estimates. Distortion metrics do not by themselves prove the absence of hallucination, and perceptual scores do not by themselves certify observation support. Nevertheless, the two regimes provide complementary information: a generative result supplies convincing high-frequency appearance, while a distortion-oriented result supplies a conservative estimate that is less prone to invented texture. We therefore study spatially and frequency-varying fusion of the two outputs.

Existing work addresses the perception–distortion trade-off at several stages of the SR pipeline. Artifact-aware and content-adaptive objectives suppress locally implausible textures during training (Liang et al., 2022; Park et al., 2023a), while structure-guided methods explicitly preserve gradients and boundaries (Ma et al., 2020). Network interpolation and controllable objectives expose global operating points along the perception–distortion continuum (Wang et al., 2019a; Park et al., 2023b; Zhu et al., 2024). Diffusion-based SR methods instead condition synthesis on the LR observation throughout iterative sampling (Saharia et al., 2023; Yue et al., 2023). These approaches improve the training or generation process of a particular SR model, but most control mechanisms remain tied to a specific architecture, objective, or sampling procedure. Global interpolation is inexpensive, yet it applies a single operating point to the entire image even though artifacts are spatially nonuniform. This motivates a post-processing method that estimates local risk, retains generated detail that remains compatible with the LR input, and intervenes selectively across space and frequency.

We propose RiskFuse, a post-processing method. Given an LR image, a generative SR result, and a distortion-oriented SR result, it predicts a dense map of the generative result's excess local error relative to the conservative alternative. Aligned HR references supervise this predictor on one part of the training set; a disjoint validation split then corrects systematic underestimation and determines a common intervention scale. At inference, an LR-consistency term reduces intervention where extra generative detail does not worsen agreement with the observation. A bounded fusion-weight map then blends Laplacian pyramids of the two SR results and assigns less weight to the distortion-oriented result at the coarsest scale. The two SR models remain frozen throughout training, calibration, and inference.

Across nine models spanning GAN- and diffusion-based generative SR and seven standard datasets, RiskFuse raises PSNR-Y and LR consistency while sharply reducing false edges; its LPIPS remains close to that of the original generative output. Local visualizations show that the predicted map

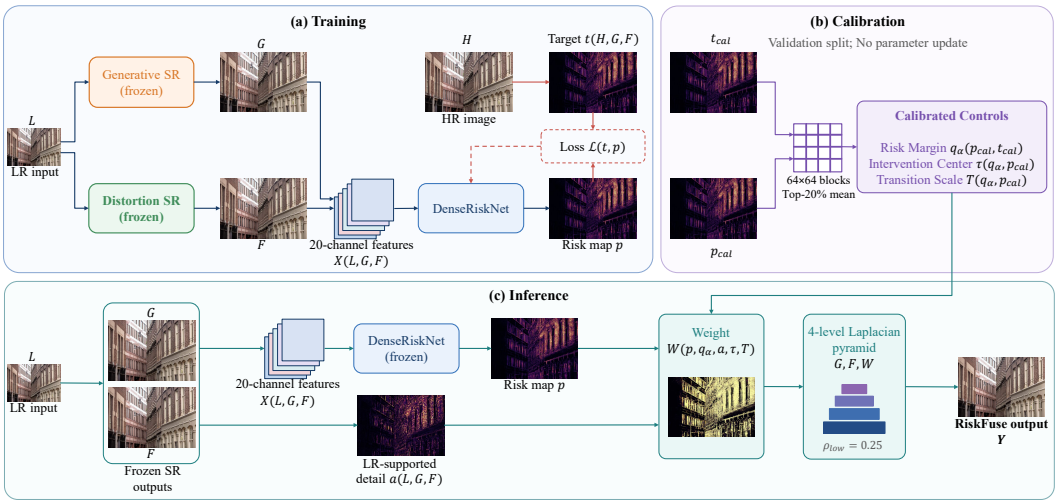

Figure 2: Overview of RiskFuse. Training learns a dense relative-risk predictor from aligned SR/HR crops; calibration estimates the three controls q_α, τ, and T on a disjoint validation split; inference combines calibrated risk with LR-supported detail and performs frequency-aware fusion. The diagram omits implementation constants and budget redistribution for clarity; their complete definitions are given in Eqs. 2–13.

concentrates on unsupported texture instead of smoothing the image uniformly. Compared with fixed-weight full-image output blending and parameter interpolation between the two SR models, spatial and frequency-selective control gives a better fidelity–perception balance at a matched intervention budget. Downstream case studies further show that RiskFuse can provide better inputs than both direct LR and the original generative output in detection and recognition.

Our contributions are threefold:

- A dense predictor of the generative output's excess HR-referenced structural error relative to a distortion-oriented alternative, augmented by local output disagreement.
- A post-hoc fusion procedure combining held-out score adjustment, relative LR-consistency retention, and spatial Laplacian-pyramid blending while keeping both SR models frozen.
- An evaluation across nine generator families and seven synthetic $\times 4$ benchmarks, together with downstream case studies, showing improved reference fidelity at a modest LPIPS cost.

2 Related Work

From Distortion-Oriented to Generative SR. Early learning-based SR established direct mappings from LR to HR images (Dong et al., 2016a). Deeper residual networks and channel or self-attention architectures then improved reference fidelity substantially (Lim et al., 2017; Zhang et al., 2018b; Liang et al., 2021). Because pixel-wise objectives reward the conditional average of plausible HR solutions, these models provide reliable structure but often attenuate fine texture. Perceptual losses and adversarial learning changed the objective from exact pixel recovery toward natural-looking image statistics (Ledig et al., 2017; Wang et al., 2019b). Diffusion models later supplied expressive conditional priors, including iterative, residual-shift, and single-step formulations (Saharia et al., 2023; Li et al., 2022; Yue et al., 2023; Wang et al., 2024). Across these families, improved perceptual realism and improved distortion occupy competing operating points rather than forming a single ordering (Blau & Michaeli, 2018).

The same distinction explains why artifacts are difficult to evaluate. A high-PSNR result may be too smooth, while a perceptually convincing result may contain detail that is plausible in distribution but unsupported for the particular image. Information-theoretic analysis links such hallucination to missing information in ill-posed generative reconstruction (Cohen et al., 2024). Our work uses the

distortion-oriented output not as the desired final image everywhere, but as a conservative source of local evidence when the generative output becomes unreliable.

Artifact Suppression and Perception–Distortion Control. Artifact suppression has been pursued at training, model, and inference levels. Locally discriminative losses distinguish useful high-frequency detail from generated artifacts, while content-aware adversarial objectives vary supervision with image structure (Liang et al., 2022; Park et al., 2023a). Gradient-guided networks explicitly preserve boundaries (Ma et al., 2020). Other methods expose a controllable operating point through parameter interpolation, objective estimation, or multi-objective optimization (Wang et al., 2019a; Park et al., 2023b; Zhu et al., 2024). Diffusion SR can repeatedly couple synthesis to an LR condition during sampling (Saharia et al., 2023; Yue et al., 2023).

These strategies are complementary to RiskFuse: they improve how a particular SR model is trained or sampled, whereas RiskFuse operates after two complete SR results have been produced. This separation makes our method applicable to heterogeneous GAN and diffusion generators. It also enables spatial control. Global interpolation can reduce average error, but it cannot distinguish a well-supported textured region from a nearby unsupported edge. RiskFuse learns that distinction and applies it through a continuous local fusion-weight map.

Risk and Calibration. Selective prediction studies how a model should act when its own score indicates elevated error, while conformal methods use a disjoint validation set to calibrate scores without committing to a parametric error distribution (Geifman & El-Yaniv, 2017; Angelopoulos & Bates, 2023). We adopt the practical principle that a learned risk score should be corrected on unseen data before it controls an intervention. Our calibration is applied to patch summaries because unsupported SR detail is local and concentrated.

3 METHOD

3.1 OVERVIEW

Let $L \in [0,1]^{h\times w\times 3}$ be an LR image. A frozen generative SR model produces $G \in [0,1]^{4h\times 4w\times 3}$, and a frozen distortion-oriented model produces F at the same spatial resolution. We seek an output Y that retains the texture and perceptual character of G wherever possible, while replacing locally unreliable content with evidence from F. During training only, the aligned HR image is denoted by H.

Figure 2 separates RiskFuse into training, calibration, and inference. During training, DenseRiskNet reads L, G, F, and their local differences and predicts a nonnegative relative-risk map p. The HR-supervised target is large where the structural error of G exceeds that of F, so F serves as a conservative local alternative rather than a universally preferred image. On a disjoint validation split, blockwise residuals determine a risk margin q_α, an intervention center τ, and a transition scale T. During inference, LR consistency identifies sharp generative detail that remains compatible with L and lowers the local fusion score. The resulting bounded weight map controls a Laplacian-pyramid blend, with a weaker contribution from F in the low-frequency residual.

3.2 DENSE RELATIVE-RISK ESTIMATION

The predictor needs both scene content and evidence about disagreement between the two SR results. Let $\mathcal{D}_4^{\mathrm{bic}}$ and $\mathcal{U}_4^{\mathrm{bic}}$ denote MATLAB-compatible bicubic downsampling and upsampling by a factor of four, and let $\mathcal{U}_4^{\mathrm{bil}}$ denote bilinear resizing from the LR to the HR grid. We use Δ for the discrete Laplacian, ∇ for gradient magnitude, V_7 for channel-averaged variance in a 7×7 window, and $\langle\cdot\rangle_c$ for averaging over red–green–blue (RGB) channels. The operator cat concatenates feature channels, while angle brackets without a channel subscript denote spatial averaging. The predictor input is the 20-channel tensor

$$X = \mathrm{cat}\big[\mathcal{U}_4^{\mathrm{bic}}(L),\ G,\ F,\ |G-F|,\ \mathcal{U}_4^{\mathrm{bil}}\big(\big\langle|\mathcal{D}_4^{\mathrm{bic}}(G)-L|\big\rangle_c\big),\ \langle|\Delta G|\rangle_c,\ \langle|\Delta F|\rangle_c,\ \langle|\Delta G-\Delta F|\rangle_c,\ \langle|\nabla G|\rangle_c,\ \langle|\nabla F|\rangle_c,\ V_7(G),\ V_7(F)\big]. \tag{1}$$

The first four RGB terms contribute 12 channels; each remaining statistic contributes one. Every channel is standardized using training-set statistics.

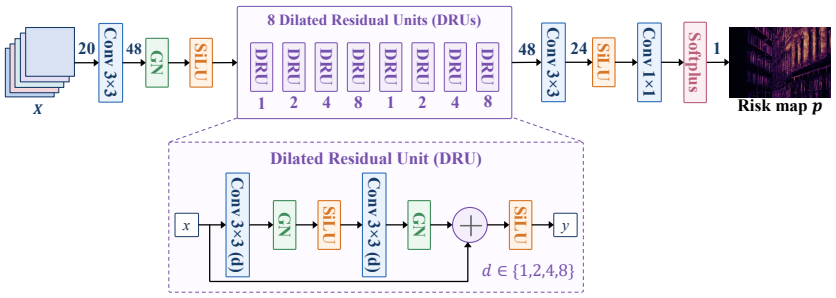

Figure 3: Architecture of DenseRiskNet. The upper row gives the complete network and the lower row expands one dilated residual unit (DRU); d is its dilation rate.

DenseRiskNet begins with a 3×3 convolution from 20 to 48 channels, group normalization (GN) (Wu & He, 2018), and a sigmoid linear unit (SiLU). It then applies eight dilated residual units (DRUs) with dilation rates $1, 2, 4, 8, 1, 2, 4, 8$. Each DRU contains two dilated 3×3 convolutions with GN and SiLU nonlinearities and an identity skip connection. A final 3×3 convolution reduces the width to 24 channels, followed by SiLU and a 1×1 prediction head. If $h(\cdot)$ denotes the head output, the predicted risk is

$$p = 0.1 \ \mathrm{softplus}(h(X)). \tag{2}$$

The network has 353,281 parameters and a 125×125 HR receptive field. The DenseRiskNet architecture is shown in Figure 3.

Training uses H to define a relative target: risk is high where G has more local structural error than F. With $[z]_+ = \max(z, 0)$,

$$\begin{aligned} t = 0.55 \ \big\langle [|\Delta G - \Delta H| - |\Delta F - \Delta H|]_+ \big\rangle_c + \\ 0.30 \ \big\langle [|\nabla G - \nabla H| - |\nabla F - \nabla H|]_+ \big\rangle_c + 0.15 \ \langle |G - F| \rangle_c . \end{aligned} \tag{3}$$

The first two terms activate only when the generative result is less faithful than the distortion-oriented result in Laplacian or gradient structure. The lower-weight third term supplies a disagreement prior in ambiguous regions.

Let $\mathcal{H}_{0.01}$ be the Huber loss with transition point 0.01, and let $\mathcal{T}$ contain the 20% of pixels with the largest target values. We optimize

$$\mathcal{L} = \mathcal{H}_{0.01}(p, t) + \mathcal{H}_{0.01}(p_{\mathcal{T}}, t_{\mathcal{T}}) + 0.02 \, \mathcal{L}_{\mathrm{edgeTV}}. \tag{4}$$

The second data term prevents sparse high-risk regions from being overwhelmed by easy background pixels. For luminance $g = \langle G \rangle_c$, the edge-aware total-variation term is

$$\mathcal{L}_{\mathrm{edgeTV}} = \left\langle e^{-12|\partial_x g|} |\partial_x p| \right\rangle + \left\langle e^{-12|\partial_y g|} |\partial_y p| \right\rangle . \tag{5}$$

It smooths predictions within regions while allowing changes across strong edges in G.

3.3 Calibration on a Disjoint Validation Split

Raw predictions can have different scales across images and generators. For every generator–image pair in the validation split, we sample 32 aligned 192×192 crops and partition each crop into non-overlapping 64×64 blocks. Let p_b and t_b be the predicted and target values restricted to block b, and define

$$P_b = \mathrm{TopMean}_{0.2}(p_b), \qquad T_b = \mathrm{TopMean}_{0.2}(t_b), \tag{6}$$

where $\mathrm{TopMean}_{0.2}$ averages the largest 20% of values. The one-sided block residual is $e_b = [T_b - P_b]_+$. With miscoverage level $\alpha = 0.1$, the finite-sample $(1 - \alpha)$ quantile of $\{e_b\}$ gives the risk margin q_α, and the calibrated risk map is

$$U = p + q_\alpha. \tag{7}$$

Thus U corrects systematic underprediction before the score determines the fusion weights. Each 192×192 crop contributes nine calibration blocks, for 414,720 blocks in total. We sample at most

4,096 predicted pixels per crop to set the remaining controls: the intervention center τ is the 80th percentile of the calibrated samples, and the transition scale T is one quarter of their interquartile range.

3.4 Retaining LR-Compatible Detail

A high learned risk should not automatically remove every sharp component. Some high-frequency detail in G may be both useful and compatible with the measured LR image. We therefore compute a one-sided retention term on luminance. Let L_y, G_y, and F_y denote the luminance channels of L, G, and F, respectively. For a map z, define

$$N_{0.95}(z) = \mathrm{clip}\left(\frac{z}{Q_{0.95}(z)}, 0, 1\right), \tag{8}$$

where $Q_{0.95}$ is the per-image 95th percentile. We define

$$\begin{aligned} d &= N_{0.95}\Big([|\Delta G_y| - |\Delta F_y|]_+\Big), \qquad e_G = \mathcal{U}_4^{\mathrm{bil}}\big(|\mathcal{D}_4^{\mathrm{bic}}(G_y) - L_y|\big), \\ e_F &= \mathcal{U}_4^{\mathrm{bil}}\big(|\mathcal{D}_4^{\mathrm{bic}}(F_y) - L_y|\big), \qquad r = N_{0.95}\big([e_G - e_F]_+\big), \\ a &= d\,\exp(-2r), \qquad s = U - 0.01a. \end{aligned} \tag{9}$$

Here d identifies extra detail supplied by G. The map r becomes large when G is less consistent with the LR observation than F, so a remains large only for sharp detail that does not incur this additional inconsistency. Subtracting $0.01a$ from U protects such detail by lowering the local fusion score.

3.5 Bounded Fusion Weights and Frequency-Aware Fusion

We convert s into an initial weight for the distortion-oriented result:

$$W_{\mathrm{raw}} = 0.4\,\sigma\left(\frac{s-\tau}{T}\right), \tag{10}$$

where σ is the logistic sigmoid and 0.4 caps the local contribution of F. To keep comparisons on a stable intervention budget, we first form a reference map without the retention term,

$$W_0 = 0.4\,\sigma\left(\frac{U-\tau}{T}\right), \qquad W_{\mathrm{base}} = \mathcal{B}\left(\min\left\{1, \frac{0.15}{\langle W_0\rangle}\right\} W_0\right), \tag{11}$$

where $\langle\cdot\rangle$ averages spatial positions and $\mathcal{B}$ is a bilateral filter with $\sigma_{\mathrm{color}} = 0.08$ and $\sigma_{\mathrm{space}} = 8$. We apply the same filter to W_{raw}, then rescale and clip it to $[0, 0.4]$ until its mean matches $\langle W_{\mathrm{base}}\rangle$. The final fusion-weight map W therefore redistributes a fixed image-level intervention budget according to retained detail rather than increasing the total contribution of F.

Direct RGB blending can change color and coarse structure together with texture. Instead, let $\mathcal{P}_\ell(G)$ and $\mathcal{P}_\ell(F)$ be the detail bands of four-level Laplacian pyramids (Burt & Adelson, 1983), and let W_ℓ be the corresponding level of a Gaussian pyramid built from W. The symbol $\odot$ denotes elementwise multiplication. For every detail level ℓ,

$$\mathcal{P}_\ell(Y) = (1 - W_\ell) \odot \mathcal{P}_\ell(G) + W_\ell \odot \mathcal{P}_\ell(F). \tag{12}$$

At the coarsest low-pass level L_p, the contribution of F is attenuated by $\rho_{\mathrm{low}} = 0.25$:

$$\mathcal{P}_{L_p}(Y) = (1 - \rho_{\mathrm{low}} W_{L_p}) \odot \mathcal{P}_{L_p}(G) + \rho_{\mathrm{low}} W_{L_p} \odot \mathcal{P}_{L_p}(F). \tag{13}$$

Reconstructing the pyramid gives the final SR image Y.

4 Experiments

4.1 Experimental Setup

Training protocol. We use the 800-image DIV2K training set (Agustsson & Timofte, 2017) and the corresponding outputs of the nine generative SR models. Of these images, 640 train DenseRiskNet; the remaining 160 form a disjoint validation split for checkpoint selection and calibration. Training uses aligned 192×192 HR crops, a batch size of 16, and AdamW (Loshchilov & Hutter, 2019) with learning rate 2×10^{-4} and weight decay 10^{-4} for 50,000 updates in one run (seed 1234); each fixed configuration is evaluated once. The SR models remain frozen; the code appendix records the exact environment and all final implementation constants.

Table 1: Mean results over seven standard $\times 4$ SR benchmarks. Each entry is the unweighted mean of the seven dataset-level means. Raw denotes the original generative output and +RF denotes RiskFuse. FED is reported in units of 10^{-3}. Bold marks the better result in each Raw/+RF pair.

	PSNR-Y ↑		LR-PSNR ↑		LPIPS ↓		FED ↓	
Generator	Raw	+RF	Raw	+RF	Raw	+RF	Raw	+RF
RRDB-ESRGAN	27.50	**28.73**	44.64	**45.97**	0.109	**0.107**	7.18	**2.96**
CAL-GAN	28.46	**29.24**	42.81	**43.49**	**0.105**	0.110	2.47	**1.10**
LDL	28.33	**29.24**	47.27	**48.69**	**0.101**	0.108	4.03	**1.69**
SPSR	27.66	**28.84**	44.22	**45.64**	**0.106**	0.106	6.41	**2.57**
SROOE	28.57	**29.40**	52.32	**53.34**	**0.094**	0.101	3.99	**1.77**
MOBOSR	29.21	**29.68**	54.06	**54.47**	**0.098**	0.103	1.37	**0.98**
SRDiff	28.17	**29.17**	54.43	**54.62**	**0.130**	0.135	3.87	**1.67**
ResShift	28.58	**29.35**	42.82	**43.91**	**0.116**	0.128	1.49	**0.82**
SinSR	26.52	**28.03**	37.34	**39.03**	0.162	**0.156**	5.37	**2.26**

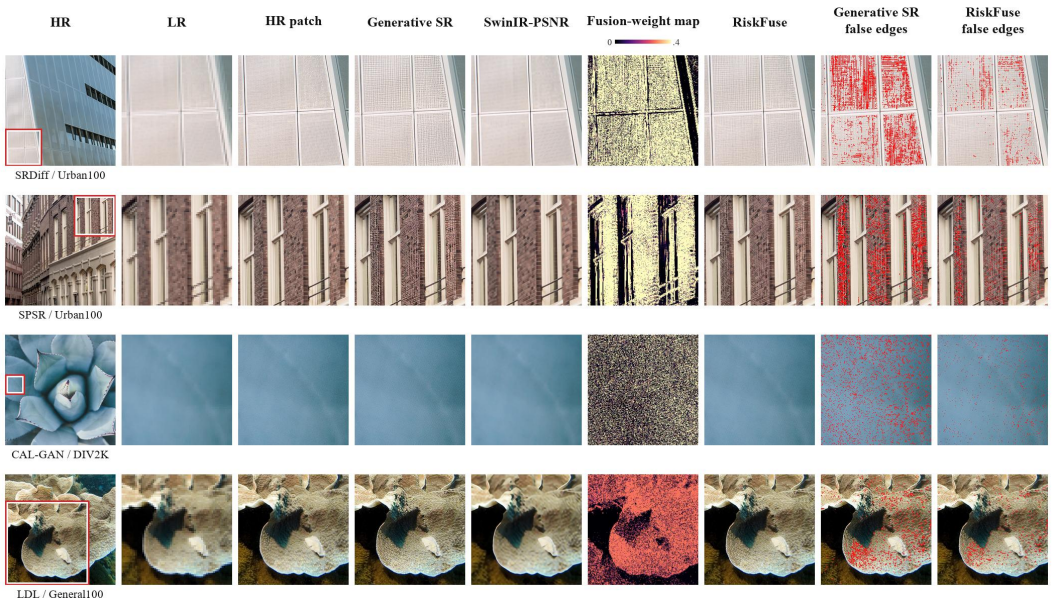

Figure 4: Qualitative evidence of unsupported detail. From left to right: an HR context with the evaluated patch marked in red, the native LR patch enlarged for display, the HR patch, generative SR, SwinIR-PSNR, the RiskFuse fusion-weight map, the RiskFuse output, and false-edge overlays for the generative and RiskFuse outputs. Brighter fusion weights assign more weight to the distortion-oriented result. Red pixels denote output edges without nearby HR-edge support.

SR models and datasets. The generative set comprises six adversarial or perceptual models: RRDB-ESRGAN (Wang et al., 2019b), CAL-GAN (Park et al., 2023a), LDL (Liang et al., 2022), SPSR (Ma et al., 2020), SROOE (Park et al., 2023b), and MOBOSR (Zhu et al., 2024); and three diffusion-based models: SRDiff (Li et al., 2022), ResShift (Yue et al., 2023), and SinSR (Wang et al., 2024). SwinIR-PSNR (Liang et al., 2021) provides the shared distortion-oriented output. We evaluate on the DIV2K validation set (Agustsson & Timofte, 2017), BSD100 (Martin et al., 2001), Urban100 (Huang et al., 2015), Manga109 (Fujimoto et al., 2016), General-100 (Dong et al., 2016b), Set5 (Bevilacqua et al., 2012), and Set14 (Zeyde et al., 2012). Each HR image is mod-cropped and downsampled by a factor of four using MATLAB-compatible bicubic interpolation.

Metrics and aggregation. We report luminance PSNR (PSNR-Y) after cropping four HR-border pixels, and LPIPS on the full RGB image (Zhang et al., 2018a). LR-PSNR compares the input LR image with the SR result downsampled by the same MATLAB-compatible bicubic operator. To measure unsupported contours, let Y_y and H_y be output and HR luminance, and let $|\nabla \cdot|$ be the 3×3

Sobel magnitude. We use the shared image-specific threshold

$$\theta = \max\{Q_{0.90}(|\nabla H_y|), 0.015\}. \tag{14}$$

The output and reference edge masks are $E_Y = \mathbf{1}[|\nabla Y_y| \geq \theta]$ and $E_H = \mathbf{1}[|\nabla H_y| \geq \theta]$. If $\mathcal{D}_2$ denotes radius-two dilation, a false edge is in $E_Y \setminus \mathcal{D}_2(E_H)$. We define false-edge density (FED) as this pixel count divided by image area and report $10^3\times$ FED. For clarity, Table 1 reports the unweighted mean of the seven dataset-level results after first averaging images within each dataset. Complete SR results are provided in the supplementary material.

Table 2: Unweighted mean of the seven dataset-level results on the RRDB pair. Global controls use $w = 0.1$, the nearest tested weight to the average RiskFuse fusion weight. FED is reported in units of 10^{-3}. Bold marks the best adaptive or global method.

Method	PSNR-Y ↑	LR-PSNR ↑	LPIPS ↓	FED ↓
Generative SR	27.50	44.64	0.109	7.18
Global output fusion	27.89	45.42	**0.106**	5.40
Parameter interpolation	28.18	45.64	0.111	4.41
RiskFuse	**28.59**	**45.82**	0.107	**3.06**
Distortion-oriented SR	30.17	55.57	0.216	0.53

4.2 Quantitative Results

Table 1 shows a consistent shift toward evidence-preserving SR. Across the underlying generator–dataset results, RiskFuse improves PSNR-Y and reduces FED in every case, while LR-PSNR improves in nearly all cases. Averaged over the complete grid, PSNR-Y rises from 28.11 to 29.08 dB, LR-PSNR rises from 46.66 to 47.68 dB, and FED falls from 4.02×10^{-3} to 1.76×10^{-3}, a 56.3% reduction. Mean LPIPS changes only from 0.1134 to 0.1171, a difference of 0.0037, with a smaller subset improving LPIPS as well. GAN and diffusion generators exhibit the same overall pattern, showing that RiskFuse is not tied to one synthesis mechanism.

4.3 Qualitative Results

Figure 4 presents representative examples. Generative SR restores sharp texture but may add grids, brick patterns, or surface fluctuations absent from the HR reference. Distortion-oriented SR removes these errors but appears smoother. The learned fusion map assigns it larger weights only on unreliable texture, allowing RiskFuse to retain supported boundaries and generative sharpness. The final columns mark unsupported output edges in red.

4.4 Comparison with Global Controls

Fixed-weight RGB output fusion and deep network interpolation (DNI) (Wang et al., 2019a) can also adjust the perception–distortion preference and suppress some artifacts in a generative output. We compare these global controls with RiskFuse using the same-architecture RRDB-ESRGAN and RRDB-PSNR models as G and F, respectively. The RRDB-pair version of RiskFuse follows the main experiment in its data split, architecture, training target, crop protocol, optimizer, and update schedule; the distortion-oriented branch is changed from SwinIR-PSNR to RRDB-PSNR. Output fusion uses $(1 - w)G + wF$; DNI uses $(1 - w)\theta_G + w\theta_F$, where θ_G and θ_F are the two RRDB parameter sets. We evaluate $w \in \{0.1, 0.2, \ldots, 0.9\}$ for both baselines. Table 2 reports $w = 0.1$, the grid point closest to RiskFuse's image-weighted mean fusion weight of 0.0906, so the comparison holds the overall intervention budget approximately fixed. All methods use the same LR inputs, quantization, and metric implementation.

At this approximately matched global budget, output fusion improves LPIPS slightly but leaves a larger fidelity gap, while parameter interpolation improves PSNR more strongly but changes the whole image. RiskFuse achieves the best PSNR-Y, LR-PSNR, and FED among the adaptive and global variants because it concentrates the distortion-oriented contribution where it is needed.

Table 3: Downstream case studies under the fixed $\times 4$ protocol. One representative metric is shown per task; more tasks and detailed metrics appear in the supplement. FD/img denotes false detections per image.

Method	COCO AP ↑	VisDrone FD/img ↓	DroneVehicle AP-small ↑	ICDAR2015 E2E F1 ↑
Direct LR	0.2296	23.361	0.1061	0.1001
RRDB-PSNR	0.3368	19.900	**0.2074**	**0.2356**
RRDB-ESRGAN	<u>0.3466</u>	<u>17.210</u>	0.1378	0.1944
RiskFuse	**0.3551**	**16.049**	<u>0.1719</u>	<u>0.2074</u>

4.5 Downstream Case Studies

We evaluate RiskFuse as preprocessing for fixed downstream models using the RRDB-pair instance from Table 2. For every task, the HR source is mod-cropped and downsampled by the same MATLAB-compatible $\times 4$ operator used in the SR benchmarks. Direct LR is evaluated at its native resolution, whereas RRDB-PSNR, RRDB-ESRGAN, and RiskFuse are evaluated at the restored $\times 4$ resolution. No downstream model is fine-tuned on any SR output.

The main-paper case studies use COCO val2017 (5,000 images) with YOLO11x (Lin et al., 2014; Jocher & Qiu, 2024), VisDrone validation (548 images) with a fixed YOLOv8s checkpoint (Zhu et al., 2022; Jocher et al., 2023), DroneVehicle validation RGB (1,469 images) with YOLO11x-OBB (Sun et al., 2022; Jocher & Qiu, 2024), and ICDAR2015 test (500 images) with PP-OCRv5 (Karatzas et al., 2015; Cui et al., 2026). Detector input sizes are 1280, 2048, and 1024 pixels for COCO, VisDrone, and DroneVehicle, respectively; input audits confirm that detector preprocessing never reduces the spatial size of an SR input. All detectors use confidence 0.001 and non-maximum-suppression intersection-over-union 0.7. For the VisDrone false-detection audit, predictions with confidence at least 0.25 are matched to ground truth at intersection-over-union 0.5.

Table 3 reports one representative metric per task. Additional downstream tasks and more detailed metrics are reported in the supplementary material. On COCO, RiskFuse raises AP from 0.3466 for RRDB-ESRGAN to 0.3551, compared with 0.2296 for Direct LR. On VisDrone, false detections fall from 17.210 to 16.049 per image, versus 23.361 for Direct LR. For DroneVehicle, AP-small rises from 0.1378 to 0.1719, compared with 0.1061 for Direct LR; on ICDAR2015, end-to-end F1 rises from 0.1944 to 0.2074, compared with 0.1001 for Direct LR. These results show that RiskFuse can improve the evidence presented to fixed downstream models over both direct LR input and unmodified generative SR.

5 Conclusion and Outlook

Generative super-resolution is valuable because it produces clear, detailed images, but some of that detail can be unsupported by the LR observation. RiskFuse addresses this limitation without retraining the generative model. It predicts where the generative output is locally less faithful than a distortion-oriented alternative, calibrates that prediction on a disjoint validation split, preserves LR-compatible sharp detail, and performs bounded frequency-aware fusion. Across a broad set of nine generative SR models spanning GAN and diffusion formulations, this local control improves reference fidelity and LR consistency, sharply reduces false edges, and keeps LPIPS near the original generative output. The visual evidence and downstream case studies further show that RiskFuse can outperform both direct LR input and unmodified generative SR in detection and recognition.

The current formulation assumes a suitable distortion-oriented companion and a fixed scale and degradation protocol. Extending the same principle to blind degradations, larger magnification factors, and adaptive companion selection is a natural next step. More broadly, the results suggest that rich generative detail and observation faithfulness need not be chosen globally: calibrated local control can preserve each where it is most useful.

Reproducibility Statement

Section 3 specifies the model architecture, input features, supervision, calibration, and fusion procedure. Section 4.1 describes the data split, optimization settings, SR models, degradation, metrics, and aggregation; Section 4.5 and Appendix B describe the downstream evaluation settings. Appendices A and C provide the detailed numerical results. The accompanying anonymized code supplement contains the implementation, configurations, an environment lockfile, and scripts for training, calibration, and evaluation. Its `README.md`, `docs/REPRODUCIBILITY.md`, and `docs/REVIEW_ARTIFACT.md` document the execution procedures, code-to-paper mapping, and the scope and limitations of the supplied artifacts. Datasets and model checkpoints are not included in the code archive.

AI Use Statement

We used generative AI tools to assist with implementing the experimental code through an iterative, prompt-based coding workflow. We used these tools to proofread the manuscript and refine its wording for clarity and readability. The authors take full responsibility for the experimental code, the reported results, and the final content of the paper.

References

Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1122–1131, 2017. doi: 10.1109/CVPRW.2017.150.

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5173–5182, 2017.

Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, pp. 135.1–135.10, 2012. doi: 10.5244/C.26.135.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237, 2018. doi: 10.1109/CVPR.2018.00652.

Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. doi: 10.1109/TCOM.1983.1095851.

Regev Cohen, Idan Kligvasser, Ehud Rivlin, and Daniel Freedman. Looks too good to be true: An information-theoretic analysis of hallucinations in generative restoration models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 22596–22623, 2024. doi: 10.52202/079017-0711.

Cheng Cui, Yubo Zhang, Ting Sun, Xueqing Wang, Hongen Liu, Manhui Lin, Yue Zhang, Tingquan Gao, Changda Zhou, Jiaxuan Liu, Zelun Zhang, Jing Zhang, Jun Zhang, and Yi Liu. PP-OCRv5: A specialized 5m-parameter model rivaling billion-parameter vision-language models on ocr tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2467–2476, 2026.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 224–236, 2018. doi: 10.1109/CVPRW.2018.00060.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016a. doi: 10.1109/TPAMI.2015.2439281.

Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision – ECCV 2016*, pp. 391–407. Springer, 2016b. doi: 10.1007/978-3-319-46475-6_25.

Azuma Fujimoto, Toru Ogawa, Kazuyoshi Yamamoto, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. Manga109 dataset and creation of metadata. In *Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding*, pp. 1–5, 2016. doi: 10.1145/3011549.3011551.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4878–4887, 2017.

Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In *Proceedings of the International Joint Conference on Neural Networks*, pp. 715–722, 2013. doi: 10.1109/IJCNN.2013.6706807.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206, 2015. doi: 10.1109/CVPR.2015.7299156.

Glenn Jocher and Jing Qiu. Ultralytics YOLO11. Software, version 11.0.0, 2024. URL https://github.com/ultralytics/ultralytics.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8. Software, version 8.0.0, 2023. URL https://github.com/ultralytics/ultralytics.

Dimosthenis Karatzas, Lluis Gomez-Bigorda, Anguelos Nicolaou, Suman Ghosh, Andrew Bagdanov, Masakazu Iwamura, Jiri Matas, Lukas Neumann, Vijay Ramaseshan Chandrasekhar, Shijian Lu, Faisal Shafait, Seiichi Uchida, and Ernest Valveny. ICDAR 2015 competition on robust reading. In *Proceedings of the 13th International Conference on Document Analysis and Recognition*, pp. 1156–1160, 2015. doi: 10.1109/ICDAR.2015.7333942.

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, 2017.

Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. SRDiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022.

Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020. doi: 10.1016/j.isprsjprs.2019.11.023.

Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5657–5666, 2022.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1132–1140, 2017. doi: 10.1109/CVPRW.2017.151.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755, 2014.

Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17627–17638, 2023.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7769–7778, 2020.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 2, pp. 416–423, 2001. doi: 10.1109/ICCV.2001.937655.

JoonKyu Park, Sanghyun Son, and Kyoung Mu Lee. Content-aware local GAN for photo-realistic super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10585–10594, 2023a.

Seung Ho Park, Young Su Moon, and Nam Ik Cho. Perception-oriented single image super-resolution using optimal objective estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1725–1735, 2023b. doi: 10.1109/CVPR52729.2023.00172.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2023. doi: 10.1109/TPAMI.2022.3204461.

Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11):2298–2304, 2017. doi: 10.1109/TPAMI.2016.2646371.

Yiming Sun, Bing Cao, Pengfei Zhu, and Qinghua Hu. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6700–6713, 2022. doi: 10.1109/TCSVT.2022.3168279.

Wenjia Wang, Enze Xie, Xuebo Liu, Wenhai Wang, Ding Liang, Chunhua Shen, and Xiang Bai. Scene text image super-resolution in the wild. In *Proceedings of the European Conference on Computer Vision*, pp. 650–666, 2020. doi: 10.1007/978-3-030-58607-2_38.

Xintao Wang, Ke Yu, Chao Dong, Xiaoou Tang, and Chen Change Loy. Deep network interpolation for continuous imagery effect transition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1692–1701, 2019a. doi: 10.1109/CVPR.2019.00179.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision Workshops*, pp. 63–79, 2019b. doi: 10.1007/978-3-030-11021-5_5.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. SinSR: Diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25796–25805, 2024.

Yuxin Wu and Kaiming He. Group normalization. In *Computer Vision – ECCV 2018*, pp. 3–19. Springer, 2018. doi: 10.1007/978-3-030-01261-8_1.

Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3974–3983, 2018. doi: 10.1109/CVPR.2018.00418.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. ResShift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, volume 36, pp. 13294–13307, 2023.

Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, volume 6920 of *Lecture Notes in Computer Science*, pp. 711–730. Springer, 2012. doi: 10.1007/978-3-642-27413-8_47.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018a. doi: 10.1109/CVPR.2018.00068.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Computer Vision – ECCV 2018*, volume 11211 of *Lecture Notes in Computer Science*, pp. 294–310. Springer, 2018b. doi: 10.1007/978-3-030-01234-2_18.

Pengfei Zhu, Longyin Wen, Dawei Du, Xiao Bian, Heng Fan, Qinghua Hu, and Haibin Ling. Detection and tracking meet drones challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7380–7399, 2022. doi: 10.1109/TPAMI.2021.3119563.

Qiwen Zhu, Yanjie Wang, Shilv Cai, Liqun Chen, Jiahuan Zhou, Luxin Yan, Sheng Zhong, and Xu Zou. Perceptual-distortion balanced image super-resolution is a multi-objective optimization problem. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 7483–7492, 2024. doi: 10.1145/3664647.3681512.

SUPPLEMENTARY MATERIAL

The supplementary material expands the aggregate results in the main paper. It first reports every generator–dataset result for the four SR metrics and then presents detailed downstream results organized by task family.

A COMPLETE SUPER-RESOLUTION RESULTS

Table S1 expands the aggregate results in the main paper without merging datasets or generator families. Results are grouped by metric; each row is one original generative output or its RiskFuse-controlled counterpart, and Mean is the unweighted mean of the seven dataset-level values. Bold marks the better value within each Raw/RiskFuse pair.

Table S1: Complete per-generator and per-dataset super-resolution results for all four metrics.

Generator	Variant	Set5	Set14	BSD100	DIV2K	General100	Manga109	Urban100	Mean
PSNR-Y (dB) ↑									
RRDB-ESRGAN	Raw	30.460	26.282	25.288	28.170	29.455	28.499	24.361	27.502
	+RiskFuse	**31.397**	**27.463**	**26.453**	**29.415**	**30.505**	**30.134**	**25.760**	**28.732**
CAL-GAN	Raw	31.102	27.293	26.427	29.164	30.317	29.459	25.439	28.457
	+RiskFuse	**31.668**	**28.042**	**27.071**	**29.938**	**30.937**	**30.609**	**26.383**	**29.236**
LDL	Raw	30.998	27.205	26.099	28.950	30.194	29.407	25.477	28.333
	+RiskFuse	**31.680**	**28.053**	**26.946**	**29.883**	**30.956**	**30.659**	**26.473**	**29.236**
SPSR	Raw	30.387	26.650	25.494	28.182	29.479	28.610	24.806	27.658
	+RiskFuse	**31.365**	**27.725**	**26.580**	**29.419**	**30.536**	**30.205**	**26.083**	**28.845**
SROOE	Raw	31.246	27.256	26.172	29.099	30.471	29.901	25.845	28.570
	+RiskFuse	**31.877**	**28.106**	**27.021**	**29.984**	**31.149**	**30.955**	**26.696**	**29.398**
MOBOSR	Raw	31.823	28.174	27.047	29.983	31.113	30.276	26.076	29.213
	+RiskFuse	**32.106**	**28.552**	**27.394**	**30.368**	**31.465**	**31.108**	**26.764**	**29.680**
SRDiff	Raw	30.559	26.974	25.861	28.954	29.986	29.212	25.663	28.173
	+RiskFuse	**31.389**	**27.960**	**26.828**	**29.896**	**30.882**	**30.652**	**26.586**	**29.170**
ResShift	Raw	30.935	27.359	26.469	29.621	30.132	29.875	25.678	28.581
	+RiskFuse	**31.668**	**28.136**	**27.130**	**30.164**	**30.854**	**30.949**	**26.578**	**29.354**
SinSR	Raw	28.153	25.686	25.095	27.741	27.750	27.387	23.795	26.515
	+RiskFuse	**29.874**	**27.042**	**26.280**	**29.030**	**29.226**	**29.419**	**25.334**	**28.029**
LR-PSNR (dB) ↑									
RRDB-ESRGAN	Raw	46.565	43.800	43.878	45.880	45.418	44.002	42.929	44.639
	+RiskFuse	**47.446**	**45.148**	**45.521**	**47.294**	**46.484**	**45.341**	**44.533**	**45.967**
CAL-GAN	Raw	42.696	42.043	41.426	43.099	44.296	44.080	42.029	42.810
	+RiskFuse	**43.186**	**42.716**	**42.223**	**43.836**	**44.840**	**44.839**	**42.811**	**43.493**
LDL	Raw	48.635	46.395	45.296	47.951	48.101	47.902	46.605	47.269
	+RiskFuse	**49.632**	**47.754**	**47.077**	**49.489**	**49.264**	**49.421**	**48.185**	**48.689**
SPSR	Raw	46.382	43.693	42.766	44.861	44.738	44.423	42.697	44.223
	+RiskFuse	**47.548**	**45.065**	**44.354**	**46.333**	**45.915**	**45.940**	**44.314**	**45.638**
SROOE	Raw	53.623	51.557	51.546	53.601	53.232	51.884	50.826	52.324
	+RiskFuse	**54.246**	**52.559**	**52.807**	**54.569**	**53.988**	**53.147**	**52.058**	**53.339**
MOBOSR	Raw	54.731	53.752	53.693	55.144	54.656	53.364	53.098	54.063
	+RiskFuse	**54.820**	**54.072**	**54.260**	**55.591**	**54.885**	**53.979**	**53.702**	**54.473**
SRDiff	Raw	**58.584**	53.619	**55.478**	**56.403**	**56.565**	49.229	51.117	54.428
	+RiskFuse	57.698	**53.887**	54.761	56.283	56.317	**51.188**	**52.213**	**54.621**
ResShift	Raw	43.055	41.731	42.814	44.197	42.727	43.396	41.805	42.818
	+RiskFuse	**43.872**	**42.804**	**44.117**	**45.098**	**43.626**	**44.664**	**43.220**	**43.915**

Continued on next page

Table S1 (continued)

Generator	Variant	Set5	Set14	BSD100	DIV2K	General100	Manga109	Urban100	Mean
SinSR	Raw	36.781	36.596	37.634	39.268	37.118	37.712	36.246	37.337
	+RiskFuse	**38.305**	**38.279**	**39.408**	**40.765**	**38.607**	**39.654**	**38.158**	**39.025**
LPIPS ↓									
RRDB-ESRGAN	Raw	**0.075**	0.134	**0.162**	0.115	0.088	0.065	0.123	0.109
	+RiskFuse	0.080	**0.127**	0.169	**0.115**	**0.085**	**0.055**	**0.117**	**0.107**
CAL-GAN	Raw	**0.067**	**0.139**	**0.167**	**0.103**	**0.085**	0.062	**0.115**	**0.105**
	+RiskFuse	0.073	0.141	0.188	0.108	0.087	**0.055**	0.118	0.110
LDL	Raw	**0.066**	**0.131**	**0.162**	**0.101**	**0.079**	0.055	**0.110**	**0.101**
	+RiskFuse	0.076	0.138	0.185	0.109	0.083	**0.051**	0.114	0.108
SPSR	Raw	**0.065**	**0.132**	**0.161**	0.110	0.086	0.067	0.118	**0.106**
	+RiskFuse	0.072	0.134	0.173	**0.109**	**0.085**	**0.056**	**0.116**	0.106
SROOE	Raw	**0.062**	**0.115**	**0.150**	**0.096**	**0.075**	0.052	**0.107**	**0.094**
	+RiskFuse	0.072	0.121	0.172	0.104	0.080	**0.048**	0.113	0.101
MOBOSR	Raw	**0.061**	**0.124**	**0.151**	**0.099**	**0.078**	0.058	**0.115**	**0.098**
	+RiskFuse	0.064	0.130	0.169	0.106	0.080	**0.054**	0.121	0.103
SRDiff	Raw	**0.099**	**0.150**	**0.201**	**0.138**	**0.104**	0.067	**0.149**	**0.130**
	+RiskFuse	0.105	0.153	0.218	0.145	0.106	**0.064**	0.153	0.135
ResShift	Raw	**0.078**	**0.131**	**0.184**	**0.136**	**0.101**	0.065	**0.121**	**0.116**
	+RiskFuse	0.090	0.144	0.211	0.149	0.107	**0.064**	0.129	0.128
SinSR	Raw	0.123	0.179	**0.216**	0.190	0.144	0.109	0.176	0.162
	+RiskFuse	**0.115**	**0.172**	0.225	**0.185**	**0.133**	**0.094**	**0.167**	**0.156**
FED (10^{-3}) ↓									
RRDB-ESRGAN	Raw	6.214	10.281	9.665	6.471	6.382	3.803	7.445	7.180
	+RiskFuse	**2.394**	**3.931**	**3.425**	**2.634**	**3.026**	**1.829**	**3.474**	**2.959**
CAL-GAN	Raw	1.614	3.807	1.970	2.430	2.268	1.649	3.526	2.466
	+RiskFuse	**0.846**	**1.225**	**0.841**	**0.972**	**1.202**	**0.919**	**1.714**	**1.103**
LDL	Raw	4.047	4.706	4.809	3.870	4.263	2.545	3.943	4.026
	+RiskFuse	**1.741**	**1.638**	**1.671**	**1.629**	**2.125**	**1.223**	**1.835**	**1.695**
SPSR	Raw	5.555	7.077	8.648	6.835	6.697	3.805	6.282	6.414
	+RiskFuse	**2.259**	**2.575**	**2.910**	**2.652**	**3.130**	**1.688**	**2.751**	**2.566**
SROOE	Raw	4.212	5.106	4.993	4.008	3.888	2.154	3.588	3.993
	+RiskFuse	**2.074**	**1.867**	**1.782**	**1.731**	**2.021**	**1.130**	**1.751**	**1.765**
MOBOSR	Raw	1.497	1.261	1.475	1.293	1.466	1.062	1.557	1.373
	+RiskFuse	**1.265**	**0.848**	**0.907**	**0.921**	**1.119**	**0.799**	**1.027**	**0.984**
SRDiff	Raw	3.441	5.650	5.919	3.403	4.382	1.853	2.425	3.868
	+RiskFuse	**2.031**	**2.027**	**1.996**	**1.449**	**2.120**	**0.948**	**1.138**	**1.673**
ResShift	Raw	1.852	2.097	1.838	0.840	1.662	0.685	1.488	1.494
	+RiskFuse	**0.889**	**0.928**	**0.920**	**0.528**	**1.027**	**0.501**	**0.912**	**0.815**
SinSR	Raw	8.385	5.857	5.509	5.102	6.511	1.647	4.563	5.368
	+RiskFuse	**3.161**	**2.215**	**1.931**	**2.329**	**3.201**	**0.852**	**2.132**	**2.260**

B Downstream Evaluation Protocol and Scope

All downstream experiments use the same MATLAB-compatible bicubic $\times 4$ degradation and the same four inputs as the main paper: Direct LR, RRDB-PSNR, RRDB-ESRGAN, and RiskFuse. Detection experiments cover the complete DOTA validation set (2,000 images) (Xia et al., 2018), COCO val2017 (5,000) (Lin et al., 2014), VisDrone validation set (548) (Zhu et al., 2022), GTSDB benchmark (900) (Houben et al., 2013), DIOR validation set (2,000) (Li et al., 2020), and both oriented- and horizontal-box evaluations of the complete DroneVehicle RGB validation set (1,469) (Sun et al., 2022). The remaining tasks are ICDAR2015 scene-text spotting on all 500 test images (Karatzas et al., 2015), CRNN recognition on all 1,343 TextZoom-hard images (Wang et al., 2020; Shi et al., 2017), and geometric matching over all 580 HPatches pairs with SuperPoint and LightGlue

(Balntas et al., 2017; DeTone et al., 2018; Lindenberger et al., 2023). Models, thresholds, image-size handling, and annotations are fixed across the four SR variants within each task.

C Detailed Downstream Results

Results are organized by task family, with one table for each dataset and fixed downstream model. Within every reported task, we retain the complete standard metric suite rather than only the metrics highlighted in the main paper. Bold and underlined values denote the best and second-best results, respectively, using unrounded values. A † marks a metric for which RiskFuse outperforms both Direct LR and RRDB-ESRGAN. To keep the tables compact, RRDB-P and RRDB-G abbreviate RRDB-PSNR and RRDB-ESRGAN.

C.1 Object Detection

For each detector, average-precision and false-positive statistics are paired with the corresponding recall statistics. Subscripts S, M, and L denote small, medium, and large objects.

Table S2: COCO object detection with YOLO11x.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.2296	0.3368	<u>0.3466</u>	**0.3551**†
AP50 ↑	0.3503	0.4875	<u>0.5043</u>	**0.5152**†
AP75 ↑	0.2412	0.3568	<u>0.3666</u>	**0.3770**†
AP_S ↑	0.0784	0.1317	<u>0.1359</u>	**0.1440**†
AP_M ↑	0.2388	0.3686	<u>0.3707</u>	**0.3822**†
AP_L ↑	0.3364	0.4568	<u>0.4786</u>	**0.4853**†
False det./image ↓	**1.586**	2.465	<u>2.076</u>	2.194
False-conf. mass ↓	**0.701**	1.136	<u>0.962</u>	1.020
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.2118	0.2735	<u>0.2783</u>	**0.2826**†
AR@10 ↑	0.3778	0.4776	<u>0.4825</u>	**0.4896**†
AR@100 ↑	0.4232	0.5300	<u>0.5370</u>	**0.5435**†
AR_S ↑	0.2003	0.2941	<u>0.3055</u>	**0.3075**†
AR_M ↑	0.4486	0.5730	<u>0.5803</u>	**0.5884**†
AR_L ↑	0.5355	0.6399	<u>0.6516</u>	**0.6557**†
Small recall@0.50 ↑	0.0439	<u>0.1009</u>	0.0962	**0.1058**†

Table S3: VisDrone object detection with YOLOv8s.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.0676	0.1121	0.1149	**0.1190**†
AP50 ↑	0.1209	0.1912	0.2022	**0.2073**†
AP75 ↑	0.0671	0.1126	0.1133	**0.1178**†
AP_S ↑	0.0362	**0.0619**	0.0585	0.0614†
AP_M ↑	0.1079	0.1705	0.1851	**0.1903**†
AP_L ↑	0.0856	0.1610	0.1695	**0.1746**†
False det./image ↓	23.361	19.900	17.210	**16.049**†
False-conf. mass ↓	9.604	9.147	7.572	**7.204**†
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.0382	0.0564	0.0592	**0.0602**†
AR@10 ↑	0.1105	0.1500	0.1678	**0.1693**†
AR@100 ↑	0.1684	0.2207	0.2429	**0.2463**†
AR_S ↑	0.1221	0.1564	0.1690	**0.1729**†
AR_M ↑	0.2154	0.3040	0.3441	**0.3518**†
AR_L ↑	0.1452	0.2236	0.2392	**0.2424**†
Small recall@0.50 ↑	0.1477	**0.2137**	0.1936	0.1967†

Table S4: DOTA oriented object detection with YOLO11x-OBB.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.3004	**0.3412**	0.3354	0.3399†
AP50 ↑	0.4813	**0.5166**	0.5052	0.5092†
AP75 ↑	0.3131	0.3716	0.3628	**0.3729**†
AP_S ↑	0.1937	0.2723	**0.2751**	0.2750
AP_M ↑	0.3073	0.3515	0.3431	**0.3519**†
AP_L ↑	0.3416	**0.3526**	0.3477	0.3480†
False det./image ↓	**2.894**	4.949	4.296	4.410
False-conf. mass ↓	**1.744**	3.330	2.988	3.083
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.1402	**0.1516**	0.1483	0.1503†
AR@10 ↑	0.3225	**0.3506**	0.3419	0.3466†
AR@100 ↑	0.3693	**0.4044**	0.3949	0.4003†
AR_S ↑	0.2696	0.3448	0.3456	**0.3490**†
AR_M ↑	0.3847	**0.4151**	0.4043	0.4096†
AR_L ↑	**0.4159**	0.4071	0.4055	0.4059
Small recall@0.50 ↑	0.0835	**0.1236**	0.1169	0.1190†

Table S5: DIOR object detection with YOLOv8s.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.2693	0.3041	0.3240	**0.3248**†
AP50 ↑	0.5037	0.5635	0.5997	**0.5997**†
AP75 ↑	0.2554	0.2930	**0.3089**	0.3052
AP_S ↑	0.0647	**0.0952**	0.0940	0.0949†
AP_M ↑	0.2131	0.2459	0.2553	**0.2554**†
AP_L ↑	0.3800	0.4100	0.4379	**0.4404**†
False det./image ↓	**1.038**	1.393	1.262	1.273
False-conf. mass ↓	**0.519**	0.703	0.643	0.645
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.1795	0.1931	**0.2093**	0.2086
AR@10 ↑	0.3368	0.3721	0.3903	**0.3916**†
AR@100 ↑	0.3754	0.4210	0.4383	**0.4397**†
AR_S ↑	0.1083	0.1539	0.1557	**0.1632**†
AR_M ↑	0.3050	0.3439	**0.3593**	0.3579
AR_L ↑	0.5001	0.5308	0.5539	**0.5557**†
Small recall@0.50 ↑	0.1329	0.1729	0.1757	**0.1768**†

Table S6: GTSDB traffic-sign detection with YOLOv8.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.2261	0.2814	0.2888	**0.2894**†
AP50 ↑	0.3475	0.4062	**0.4261**	0.4241
AP75 ↑	0.2710	0.3280	**0.3389**	0.3377
AP_S ↑	0.1715	0.1806	**0.1825**	0.1783
AP_M ↑	0.2801	0.3529	0.3626	**0.3644**†
AP_L ↑	0.3374	**0.3665**	0.3615	0.3642†
False det./image ↓	**0.346**	0.388	0.386	0.386
False-conf. mass ↓	**0.207**	0.254	0.245	0.248
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.2299	**0.2622**	0.2576	0.2604†
AR@10 ↑	0.4132	**0.4554**	0.4318	0.4376†
AR@100 ↑	0.4183	**0.4562**	0.4337	0.4386†
AR_S ↑	0.4216	**0.4247**	0.3882	0.3936
AR_M ↑	0.4214	**0.4792**	0.4673	0.4724†
AR_L ↑	0.4116	**0.4535**	0.4140	0.4140
Small recall@0.50 ↑	0.1002	**0.1099**	0.1062	0.1081†

Table S7: DroneVehicle oriented-box detection with YOLO11x-OBB.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.3426	**0.4116**	0.3343	0.3620†
AP50 ↑	0.6015	**0.6724**	0.5370	0.5879
AP75 ↑	0.3548	**0.4591**	0.3830	0.4092†
AP_S ↑	0.1061	**0.2074**	0.1378	0.1719†
AP_M ↑	0.3863	**0.4500**	0.3722	0.3980†
AP_L ↑	**0.3503**	0.3391	0.2842	0.3048
False det./image ↓	**0.274**	0.452	0.275	0.363
False-conf. mass ↓	0.126	0.222	**0.126**	0.170
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.0378	**0.0425**	0.0374	0.0390†
AR@10 ↑	0.2567	**0.2880**	0.2440	0.2600†
AR@100 ↑	0.4106	**0.4713**	0.3877	0.4174†
AR_S ↑	0.2147	**0.3206**	0.2374	0.2753†
AR_M ↑	0.4492	**0.5034**	0.4187	0.4472
AR_L ↑	**0.4211**	0.4087	0.3544	0.3730
Small recall@0.50 ↑	0.0951	**0.1407**	0.0955	0.1137†

Table S8: DroneVehicle horizontal-box detection with YOLOv8s.

Precision and false positives				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AP ↑	0.3163	**0.3585**	0.2995	0.3067
AP50 ↑	0.6019	**0.6885**	0.5543	0.5738
AP75 ↑	0.3038	**0.3263**	0.2868	0.2891
AP_S ↑	0.2037	**0.2210**	0.1610	0.1784
AP_M ↑	0.3520	**0.3889**	0.3307	0.3363
AP_L ↑	0.1726	**0.2577**	0.2339	0.2410†
False det./image ↓	2.323	1.843	**0.648**	0.722
False-conf. mass ↓	1.034	0.921	**0.320**	0.360
Recall				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
AR@1 ↑	0.0381	**0.0382**	0.0346	0.0350
AR@10 ↑	0.2454	**0.2634**	0.2290	0.2323
AR@100 ↑	0.4139	**0.4487**	0.3809	0.3886
AR_S ↑	0.3548	**0.3747**	0.3004	0.3206
AR_M ↑	0.4315	**0.4664**	0.3987	0.4039
AR_L ↑	0.2439	**0.3636**	0.3305	0.3347†
Small recall@0.50 ↑	0.2155	**0.2274**	0.1634	0.1769

C.2 Scene Text Detection and Recognition

ICDAR2015 evaluates joint scene-text detection and recognition, whereas TextZoom-hard evaluates cropped-word recognition. Corpus-level metrics are paired with image-level or normalized counterparts to make the relation between the two views explicit.

Table S9: ICDAR2015 scene-text detection and recognition with PP-OCRv5.

Corpus-level metrics				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
Detection precision ↑	0.5452	**0.7088**	0.6775	0.6909†
Detection recall ↑	0.1103	**0.4617**	0.4299	0.4348†
Detection F1 ↑	0.1834	**0.5592**	0.5261	0.5337†
E2E precision ↑	0.2976	**0.2986**	0.2504	0.2686
E2E recall ↑	0.0602	**0.1945**	0.1589	0.1690†
E2E F1 ↑	0.1001	**0.2356**	0.1944	0.2074†
False boxes/image ↓	**0.3820**	0.7880	0.8500	0.8080
Matched NED ↑	**0.8411**	0.7347	0.6786	0.6967
Image-level and matched-text metrics				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
Mean image det. precision ↑	0.2942	**0.6699**	0.6523	0.6552†
Mean image det. recall ↑	0.1627	**0.5304**	0.5124	0.5170†
Mean image det. F1 ↑	0.1933	**0.5648**	0.5505	0.5541†
Mean image E2E precision ↑	0.1525	**0.2958**	0.2485	0.2577†
Mean image E2E recall ↑	0.0886	**0.2486**	0.2024	0.2109†
Mean image E2E F1 ↑	0.1033	**0.2593**	0.2156	0.2240†
Matched CER ↓	**0.1881**	0.2897	0.3499	0.3291
Recognition confidence ↑	**0.8260**	0.7890	0.7456	0.7618

Table S10: TextZoom-hard word recognition with CRNN.

Raw transcription				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
Exact accuracy ↑	0.7640	**0.7811**	0.7714	0.7736†
CER ↓	0.1089	**0.1059**	0.1097	0.1080†
NED ↑	0.8925	**0.8960**	0.8923	0.8941†
Recognition confidence ↑	0.9797	**0.9803**	0.9782	0.9788
Normalized transcription				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
Exact accuracy ↑	0.7848	**0.7990**	0.7915	0.7937†
CER ↓	0.1005	**0.0982**	0.1014	0.0995†
NED ↑	0.9011	**0.9036**	0.9008	0.9027†

C.3 GEOMETRIC FEATURE MATCHING

HPatches measures local correspondence quality and the geometric consistency of the resulting matches. The two panels separate point-level matching statistics from robust geometric verification.

Table S11: HPatches geometric matching with SuperPoint and LightGlue.

Local matching				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
Matches ↑	871.5	**1057.1**	929.6	943.8†
Mean reproj. error ↓	8.4495	7.0635	**7.0196**	7.1668
Median reproj. error ↓	7.7772	6.2542	**6.2035**	6.5211
MMA@1 ↑	**0.2624**	0.2519	0.2348	0.2387
MMA@3 ↑	0.6865	**0.7429**	0.7340	0.7375†
MMA@5 ↑	0.8779	0.9005	0.9009	**0.9017**†
Geometric verification				
Metric	Direct LR	RRDB-P	RRDB-G	RiskFuse
MMA@10 ↑	0.9653	0.9697	**0.9716**	0.9710
False-match rate@5 ↓	0.1221	0.0995	0.0991	**0.0983**†
RANSAC success ↑	**1.0000**	**1.0000**	**1.0000**	**1.0000**
RANSAC inliers ↑	635.9	**814.5**	691.9	710.8†
RANSAC inlier ratio ↑	0.6878	**0.7526**	0.7395	0.7481†
Homography corner error ↓	44.5033	28.0539	**14.8508**	18.5308