COLOUR-INR - IMPLICIT NEURAL REPRESENTATION FOR IMAGE COLORIZATION THROUGH COLOUR GUIDANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Colourization of black and white images can be considered as an inverse problem that searches for pixel-level colour values based on observed intensity mappings. Since the forward problem follows a many-to-one mapping, the inverse admits multiple possible chrominance solutions. However, there is a set of solutions that can be considered realistic and acceptable. When defining this set of solutions, human perception becomes an important factor. Hint-based image colourization provides guidance to constrain this set of solutions. Implicit Neural Representations (INRs) have been applied to multiple inverse tasks such as denoising, super-resolution, and inpainting. However, their use for image colourization remains underexplored. In this research, we show that INRs have the potential to perform image colourization when guided appropriately. Moving further, we introduce Colour-INR, where a coordinate-based INR predicts chrominance from spatial coordinates, sparse colour hints provide chrominance supervision, and grayscale-derived edges guide spatial regularization. On Kodak with 0.5% colour hints, Colour-INR achieves a mean PSNR of 32.64 dB and substantially outperforms the evaluated pretrained learning-based baselines, while remaining competitive with the classical optimization method of Levin et al. Further, we experimentally show that Colour-INR can be optimized directly on high-resolution DIV2K images at their native resolution without changing the network architecture or requiring dataset-level retraining. These results demonstrate the potential of properly guided INRs for image colourization.

1 INTRODUCTION

Image colourization aims to recover plausible colour information from a grayscale image. Unlike conventional image reconstruction problems, however, the missing colour is not uniquely determined by the observed intensity. Different chrominance values can correspond to the same or very similar luminance, and therefore a grayscale image can admit multiple valid colour solutions. Colourization is consequently an under constrained inverse problem in which additional information is required to restrict the solution.

Recent colourization methods largely address this ambiguity through priors learned from large image datasets or pretrained generative models. These approaches can produce visually realistic results, but the selected colours are mainly determined by the learned prior unless additional user guidance is provided. Sparse colour hints provide a direct form of such guidance by specifying the desired chrominance at a small number of image locations. The remaining problem is then to recover a dense colour field from these sparse observations while respecting the spatial structure of the image.

Implicit Neural Representations represent signals as continuous functions of their coordinates and have been successfully applied to several reconstruction and inverse problems. Image colourization presents a different setting: the luminance L is observed, while the chrominance (a, b) is missing and is not uniquely determined by L. The grayscale image nevertheless preserves the spatial structure of the scene, which can be used as a prior on how the unknown chrominance should vary spatially.

In this work, we formulate sparse hint image colourization as an image specific inverse problem. The unknown chrominance field is represented continuously using an INR, while sparse colour

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.

hints provide direct observations of (a, b) at selected coordinates. We treat these observations as *chrominance anchors* that constrain the otherwise ambiguous solution. Grayscale derived structure is then used to regularize the chrominance between the anchor locations.

Based on this formulation, we introduce **Colour-INR**, an image specific coordinate representation that predicts only the missing CIELAB chrominance channels. Sparse chrominance anchors supervise the INR at selected coordinates, while an edge aware smoothness prior encourages locally coherent colour propagation without enforcing smoothness across detected image boundaries. The observed luminance is retained for final reconstruction, and no dataset level training or pretrained colourization model is required.

Our main contributions are:

- We formulate sparse hint image colourization as an image specific inverse problem in which the missing chrominance is not uniquely determined by the observed luminance and is constrained using sparse chrominance anchors.
- We introduce Colour-INR, which represents the unknown chrominance as a continuous coordinate function and combines sparse anchor supervision with edge aware spatial regularization, without requiring dataset level training.
- We evaluate the formulation through smoothness, backbone, and network width ablations, comparisons with representative hint based colourization methods, and native resolution experiments on high resolution DIV2K images.

2 Related Work

Automatic colourization. Early learning based approaches formulated colourization as prediction from large paired grayscale and colour datasets. Deshpande et al. (Deshpande et al., 2015) developed a probabilistic formulation to represent multiple plausible colourizations, while Larsson et al. (Larsson et al., 2016) studied colourization as a representation-learning task. Other approaches combined local image features with global scene-level information (Iizuka et al., 2016), or formulated colour prediction as classification over quantized chrominance values (Zhang et al., 2016). Adversarial approaches were later explored to improve perceptual realism and colour diversity (Vitoria et al., 2020; Wu et al., 2021). More recently, diffusion models have provided powerful generative priors for modelling the multimodality of colourization (Saharia et al., 2022; Li et al., 2024a; Wang et al., 2024; Li et al., 2026). While automatic methods can produce convincing results, they do not directly allow users to specify the desired colour at particular image locations.

Reference guided colorization. User guidance can constrain the set of plausible color solutions without requiring annotations at the pixel level. Approaches based on reference material transfer color information from example images. Deep Example-Based Colorization He et al. (2018), establishes semantic correspondence between the target grayscale image and a reference color image, while later methods improve correspondence through cross-scale and non-local matching Li et al. (2019); Wang et al. (2023). Although such methods provide semantic control, reference images do not directly specify the desired chrominance at individual spatial locations.

Language guided colorization. Language also offers a form of semantic guidance. Previous approaches conditioned colorization networks on textual descriptions Manjunatha et al. (2018); Bahng et al. (2018); Weng et al. (2022), while recent Methods are now making use of pretrained diffusion priors. Diffusing Colors Zabari et al. (2023) fine-tunes a pretrained diffusion model for text-guided colorization, and L-CAD Chang et al. (2023) uses diffusion priors to support descriptions with different levels of detail. COCO-LC Li et al. (2024b) further introduces controllable colorfulness within a framework guided by language. Language guided methods also do not directly specify the desired chrominance at individual spatial locations.

Sparse Hint and Scribble-Guided Colorization. Sparse color hints provide direct control over the otherwise ambiguous colorization problem. Classical optimization based methods propagate user-provided colors according to local image structure Levin et al. (2004). Later learning based approaches replaced handcrafted propagation rules with learned priors and long range feature interactions, including CNN and Transformer based methods Zhang et al. (2017); Huang et al. (2022); Yun et al. (2023). More recent methods have incorporated stronger generative priors and explicit interac-

tive guidance. Diffusion based approaches support color hints, text, strokes, and other conditioning signals Yang et al. (2024); Liang et al. (2025). LGA-Net Lyu et al. (2025) instead formulates sparse scribble colorization as color propagation over learned local and global affinities. These works highlight the importance of effectively propagating sparse user guidance, although most learning-based approaches rely on dataset-level training or pretrained generative priors.

Implicit Neural Representations for Image Colorization. Implicit Neural Representations provide an alternative representation of visual signals in which a neural network parameterizes a continuous function over spatial coordinates. Although INRs have been successfully applied to several inverse imaging problems, their use for image colorization remains limited. Colorization presents a distinct inverse problem, since the observed grayscale image provides luminance information but does not uniquely determine the missing chrominance. We introduce Colour-INR as an INR based inverse problem for image colorization, representing the unknown chrominance as a continuous coordinate function while retaining the observed luminance.Sparse color hints constrain chrominance at selected coordinates, and grayscale derived structural information regularizes its spatial variation. This connects user guided colorization with image specific implicit neural representation without requiring dataset level training or a pretrained colorization prior.

3 METHOD

We first formulate colourization as an under constrained chrominance recovery problem and motivate the role of sparse colour observations. We then introduce Colour-INR, an image specific coordinate based representation that recovers the missing chrominance using sparse anchor supervision and edge aware spatial regularization.

3.1 COLOURIZATION AS AN UNDER CONSTRAINED INVERSE PROBLEM

Consider an image represented in CIELAB space. Given a grayscale observation, the luminance L is available while the chrominance channels (a, b) are unknown. Recovering the original colour therefore requires estimating a dense chrominance field from the observed luminance. However, (a, b) is not uniquely determined by L: different chrominance values can correspond to the same or very similar luminance values. Colourization is therefore an under constrained inverse problem. We further examine this ambiguity in Appendix A.1.

To illustrate how this ambiguity affects an INR, we consider a diagnostic setting in which a coordinate based SIREN predicts all three CIELAB channels $(\hat{L}, \hat{a}, \hat{b})$. We first optimize the network using luminance reconstruction alone. We then repeat the optimization while additionally providing 0.5% sparse chrominance observations. No spatial smoothness regularization is used in either setting, allowing the effect of the colour guidance to be examined separately.

Figure 1 illustrates this behaviour during optimization. With luminance supervision alone, the network progressively fits the observed image structure, but the chrominance remains weakly constrained. When sparse chrominance observations are introduced, the luminance reconstruction remains governed by the same grayscale observation, while the additional colour information constrains the predicted (a, b) field. Quantitative results for this diagnostic experiment are provided in Appendix A.2.

This motivates the formulation used in Colour-INR. We treat each sparse colour observation as a *chrominance anchor*: it directly specifies the missing (a, b) values at a selected spatial coordinate. The remaining task is to recover the dense chrominance field between these anchors. The grayscale image is used as a source of spatial structure rather than direct colour information. We use this structure to regularize local chrominance variation and reduce propagation across detected image boundaries.

Colour-INR therefore separates the two sources of information: sparse chrominance anchors constrain what colours are present, while grayscale derived structure guides how the chrominance field varies spatially. The resulting formulation is described next.



Figure 1: Evolution of the diagnostic INR during optimization. The top row uses luminance reconstruction alone, while the bottom row additionally uses 0.5% sparse chrominance hints. Fitting the observed luminance alone does not constrain the missing chrominance, whereas sparse colour observations guide the INR toward a meaningful colour solution.

3.2 COLOUR-INR

Following the inverse formulation above, Colour-INR represents only the unknown chrominance field while preserving the observed luminance. The INR maps spatial coordinates to predicted CIELAB chrominance values and is optimized independently for each image using sparse chrominance anchors together with edge aware spatial regularization. An overview of the method is shown in Fig. 2.

For a pixel p, let

$$\mathbf{x}_p = (x_p, y_p) \in [-1, 1]^2 \tag{1}$$

denote its normalized spatial coordinate. Colour-INR learns a coordinate based mapping

$$f_\theta(\mathbf{x}_p) = \hat{\mathbf{c}}_p = (\hat{a}_p, \hat{b}_p), \tag{2}$$

where $\hat{a}_p$ and $\hat{b}_p$ denote the predicted CIELAB chrominance values at p. The luminance channel is not predicted by the INR. Instead, the observed luminance L_{obs} is retained from the grayscale image and reused during final reconstruction.

We parameterize f_θ using a sinusoidal representation network (SIREN) (Sitzmann et al., 2020). The network receives only the spatial coordinate (x, y) and directly predicts the two chrominance channels. We use a hidden width of 128 with three hidden sinusoidal layers following the first SIREN layer. The first and hidden layers use frequency parameters $\omega_0 = 60$ and $\omega_0 = 30$, respectively. A linear projection from 128 dimensions to two output channels is followed by a $\tanh$ activation to obtain the normalized chrominance prediction $(\hat{a}, \hat{b})$.

The network input remains purely coordinate based. Neither the grayscale intensity, the hint mask, nor the edge map is provided directly as an input to the INR. Instead, colour information is introduced during image specific optimization through sparse chrominance supervision and spatial regularization.

Let $\mathcal{H}$ denote the set of sparse chrominance anchor locations corresponding to the provided colour hints. Each anchor specifies the target chrominance $\mathbf{c}_p = (a_p, b_p)$ at location $p \in \mathcal{H}$. The INR is supervised at these locations using a mean squared chrominance error,

$$\mathcal{L}_{\text{hint}} = \frac{1}{2|\mathcal{H}|} \sum_{p \in \mathcal{H}} \left[(\hat{a}_p - a_p)^2 + (\hat{b}_p - b_p)^2 \right]. \tag{3}$$

This term anchors the otherwise under-constrained chrominance field at the observed locations. No dense ground-truth chrominance supervision is used during optimization.

Although the sparse hints constrain the colour at selected locations, they do not explicitly determine how chrominance should vary between those locations. We therefore introduce a local ℓ_1 smoothness constraint over the eight connected neighbourhood of each pixel. This encourages nearby

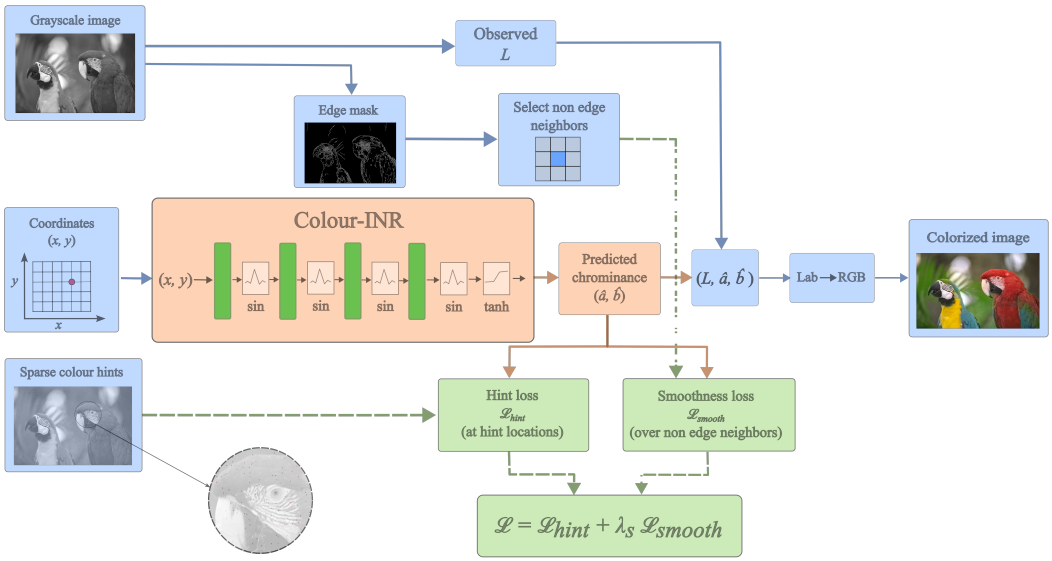

Figure 2: Overview of Colour-INR. A coordinate based SIREN predicts chrominance $(\hat{a}, \hat{b})$ using sparse colour supervision and edge aware smoothness, and the prediction is combined with the observed luminance L_{obs} for RGB reconstruction.

pixels to maintain consistent chrominance where the image structure suggests that such smoothness is appropriate.

Applying the smoothness constraint to every neighbouring pair can, however, propagate colour across object boundaries. To reduce this effect, we compute a Canny edge map E from the observed luminance (Canny, 1986) and use it to identify the neighbour pairs over which the regularization is applied. Let $\mathcal{N}_8(p)$ denote the eight connected neighbourhood of pixel p. The valid neighbour pair set is defined as

$$\mathcal{P} = \{(p, q) \mid q \in \mathcal{N}_8(p),\ E_p = 0,\ E_q = 0\}. \tag{4}$$

A pair therefore contributes to the smoothness term only when both endpoints are non edge pixels. All valid neighbour directions, including the four diagonal neighbours, are treated equally.

The resulting chrominance smoothness loss is

$$\mathcal{L}_{\text{smooth}} = \frac{1}{2|\mathcal{P}|} \sum_{(p,q)\in\mathcal{P}} \left(|\hat{a}_p - \hat{a}_q| + |\hat{b}_p - \hat{b}_q| \right). \tag{5}$$

This term encourages neighbouring non edge pixels to take similar chrominance values while avoiding an explicit smoothness constraint across detected image boundaries.

The final image specific objective combines the sparse hint supervision and the edge aware smoothness regularization,

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{hint}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}}, \tag{6}$$

where the hint coefficient is fixed to one and λ_{smooth} controls the strength of the spatial regularizer. For each image, the INR parameters are optimized independently as

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta). \tag{7}$$

After optimization, the INR is evaluated at every image coordinate to obtain dense predicted chrominance fields $\hat{A}$ and $\hat{B}$. These predictions are combined with the observed luminance L_{obs} and converted back to RGB,

$$\hat{I}_{\text{RGB}} = \text{Lab2RGB}\left(L_{\text{obs}}, \hat{A}, \hat{B}\right). \tag{8}$$

In this way, Colour-INR preserves the luminance structure of the grayscale input and uses the coordinate network only to recover the missing chrominance field from sparse colour guidance.

4 RESULTS

4.1 EXPERIMENTAL SETUP

We evaluate Colour-INR using the Kodak Lossless True Color Image Suite (Kodak24) (Eastman Kodak Company, 1993) as the main dataset for controlled experiments and ablation studies. Kodak contains 24 natural colour images with resolutions of 768×512 or 512×768. For each image, we generate the corresponding grayscale input and retain the original colour image as the ground truth. Sparse colour hints are sampled from the ground truth chrominance channels at a density of 0.5%. The hint locations are generated using a fixed seed and are kept unchanged across compared settings so that the effect of each design choice can be studied under the same colour observations.

Unless otherwise stated, Colour-INR uses a SIREN with a hidden width of 128 and three hidden sinusoidal layers following the first SIREN layer. The first and hidden frequency parameters are set to $\omega_0 = 60$ and $\omega_0 = 30$, respectively. Each image is optimized independently for 2000 epochs using Adam with a learning rate of 10^{-4} and a coordinate batch size of 32,768. The hint loss weight is fixed to $\lambda_{\text{hint}} = 1$. The smoothness coefficient λ_{smooth} is selected through the ablation in Sec. 4.2 and is then kept fixed for the remaining experiments.

We evaluate the colourized images using PSNR (Huynh-Thu & Ghanbari, 2008), SSIM (Wang et al., 2004), MS-SSIM (Wang et al., 2003), LPIPS (Zhang et al., 2018), and CIEDE2000 (Sharma et al., 2005). PSNR, SSIM, and MS-SSIM measure agreement with the reference image, where higher values indicate better reconstruction quality. LPIPS measures perceptual feature distance, where lower values are better, and we compute it using the AlexNet backbone. CIEDE2000 measures perceptual colour difference in CIELAB space, where lower values indicate better colour agreement. Since image colourization is inherently ill-posed and more than one colour solution can be perceptually valid, these reference based metrics are considered together with qualitative comparisons.

4.2 ABLATION STUDIES

We study three design choices of Colour-INR: the strength of the chrominance smoothness regularization, the coordinate network backbone, and the width of the coordinate network. All experiments use the same 0.5% colour hints and the direct ab formulation described in Sec. 3.2.

We first study the effect of the smoothness weight. The hint coefficient is fixed to $\lambda_{\text{hint}} = 1$, while λ_{smooth} is varied over $\{0, 0.05, 0.1, 0.25, 0.5, 1, 1.5, 2, 5\}$. The same network configuration, hint locations, and optimization settings are used for all values. The smoothness loss is calculated over the edge aware eight neighbourhood defined in Sec. 3.2, where all valid neighbour pairs are weighted equally and any pair containing a Canny edge pixel is excluded. Figure 3 shows that a small amount of spatial regularization substantially improves colourization compared with sparse hint supervision alone. Without smoothness, the model achieves a mean PSNR of 29.80 dB and a CIEDE2000 error of 4.90. Setting $\lambda_{\text{smooth}} = 0.05$ improves these to 32.64 dB and 2.85, respectively, while also giving the lowest LPIPS of 0.0667. The highest SSIM is obtained at $\lambda_{\text{smooth}} = 0.25$ (0.9716), although the difference from 0.05 (0.9714) is small. Increasing the smoothness weight further gradually reduces reconstruction quality. At $\lambda_{\text{smooth}} = 5$, the mean PSNR decreases to 30.29 dB and CIEDE2000 increases to 4.53. We therefore use $\lambda_{\text{smooth}} = 0.05$ for the remaining experiments.

We next study the effect of the coordinate-network backbone while keeping the loss formulation, $\lambda_{\text{smooth}} = 0.05$, hint locations, and optimization settings fixed. We compare SIREN (Sitzmann et al., 2020), FINER (Liu et al., 2024), INCODE (Kazerouni et al., 2024), a ReLU MLP, and a ReLU MLP with Gaussian positional encoding (ReLU+PE) (Tancik et al., 2020). All models use a hidden width of 128 where applicable. For a controlled comparison, the same grayscale-derived luminance, sparse colour hints, coordinate batches, optimization schedule, and evaluation procedure are used across the tested backbones. As shown in Table 1, SIREN gives the strongest overall performance, achieving 32.64 dB PSNR, 0.9714 SSIM, 0.0667 LPIPS, and a CIEDE2000 error of 2.849. FINER performs similarly but is consistently slightly lower across the evaluated metrics. INCODE and the ReLU-based alternatives show larger performance reductions under the same sparse-hint optimization protocol. We therefore use SIREN as the coordinate-network backbone for Colour-INR. Figure 4 provides qualitative examples of the chrominance reconstructions obtained with the different coordinate-network backbones. The examples complement the quantitative comparison in

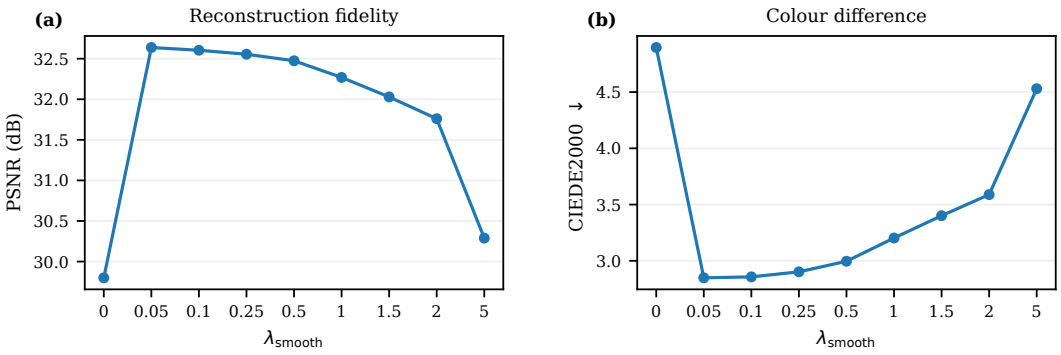

Figure 3: Smoothness weight ablation on the 24 Kodak images with $\lambda_{\text{hint}} = 1$. We select $\lambda_{\text{smooth}} = 0.05$ for subsequent experiments.

Table 1: Backbone ablation on Kodak with 0.5% colour hints.

Backbone	Params	PSNR ↑	SSIM ↑	LPIPS ↓	ΔE_{00} ↓
SIREN	50,178	**32.64**	**0.9714**	**0.0667**	**2.849**
FINER	50,178	32.44	0.9707	0.0673	2.946
INCODE	288,038	31.68	0.9692	0.0725	3.196
ReLU	50,178	31.33	0.9667	0.1070	3.712
ReLU + PE	115,458	27.66	0.9533	0.1200	5.516

Table 1 and illustrate how the choice of coordinate representation affects colour propagation under the same sparse colour guidance.

Finally, we compare SIREN hidden widths of 128 and 256 while keeping the number of layers, frequency parameters, colour hints, loss weights, and optimization settings unchanged. Increasing the hidden width does not improve overall reconstruction quality. As shown in Table 2, the width-128 model achieves a mean PSNR of 32.64 dB, LPIPS of 0.0667, and CIEDE2000 error of 2.848, compared with 32.50 dB, 0.0684, and 2.869 for width 256, respectively. Both configurations give an SSIM of 0.9714 after rounding. Increasing the width also raises the number of trainable parameters from 50,178 to 198,658. We therefore use a hidden width of 128.

4.3 Comparison with Existing Colorization Methods

We compare Colour-INR with representative hint-based colourization methods, including Levin et al. (2004), Zhang et al. (2017), UniColor (Huang et al., 2022), and iColoriT (Yun et al., 2023). All methods are evaluated on the 24 Kodak images using 0.5% sparse colour hints. The same colour observations are used across the methods, with hint locations mapped to the required input resolution when a method operates at a fixed resolution.

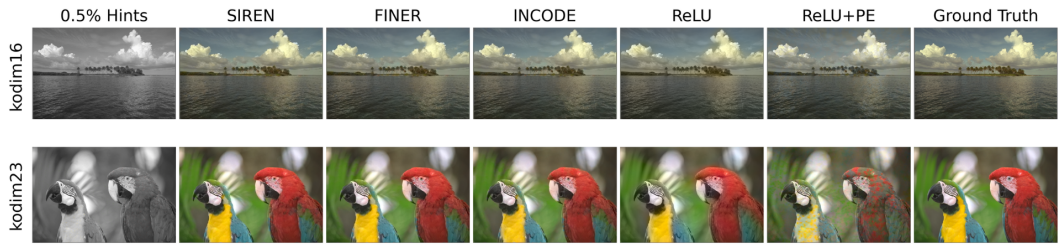

Figure 4: Qualitative comparison of coordinate-network backbones using the same 0.5% colour hints and optimization protocol.

Table 2: SIREN width ablation on Kodak with 0.5% colour hints.

Width	Params	PSNR ↑	SSIM ↑	LPIPS ↓	ΔE_{00} ↓
128	50,178	**32.64**	**0.9714**	**0.0667**	**2.848**
256	198,658	32.50	**0.9714**	0.0684	2.869

Table 3: Comparison with hint-based colourization methods on Kodak at 0.5% hints. Best metric values are shown in bold.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	ΔE_{00} ↓
UniColor	26.40	0.8353	0.9543	0.2658	4.431
Zhang et al.	26.31	0.8327	0.9482	0.2952	4.862
iColoriT	26.42	0.8230	0.9537	0.3175	4.408
Levin et al.	**33.64**	**0.9749**	**0.9777**	**0.0665**	**2.599**
Colour-INR	32.64	0.9714	0.9747	0.0667	2.848

As shown in Table 3, Colour-INR substantially outperforms the evaluated learning-based comparison methods under the sparse-hint setting. Colour-INR achieves a mean PSNR of 32.64 dB, compared with 26.40 dB for UniColor, 26.31 dB for Zhang et al., and 26.42 dB for iColoriT. Similar improvements are observed in SSIM and MS-SSIM.

Figure 5 provides a qualitative comparison on a representative Kodak image under the same 0.5% sparse colour observations.



Figure 5: Qualitative comparison of hint-based colourization methods on a representative Kodak image using the same 0.5% sparse colour observations. Hint locations are geometrically mapped for methods operating at a fixed input resolution.

Levin et al. obtains the strongest reference-based reconstruction scores, with a PSNR of 33.64 dB and a CIEDE2000 error of 2.599, compared with 32.64 dB and 2.848 for Colour-INR. The two methods also give similar LPIPS values (0.0665 for Levin et al. and 0.0667 for Colour-INR). These results show that the proposed coordinate-based formulation remains competitive with the classical optimization-based method while providing substantially stronger reconstruction quality than the evaluated pretrained colourization models under the same sparse colour guidance.

The qualitative comparison complements the dataset-level results in Table 3. In this representative image, the evaluated methods produce visibly different chrominance propagation patterns under the same sparse guidance. Colour-INR produces a spatially coherent colourization through image-specific optimization without dataset-level training.

4.4 High Resolution Colourization

We further evaluate Colour-INR on high-resolution images to examine whether the coordinate-based formulation can be optimized directly at the native image resolution. We use 10 images from the DIV2K dataset (Agustsson & Timofte, 2017) and generate 0.5% sparse colour hints using the same protocol as in the Kodak experiments. Since both Colour-INR and the optimization-based method of Levin et al. (2004) can operate directly on arbitrary image resolutions, we compare them without resizing the input images.

Table 4: High-resolution comparison on 10 DIV2K images at 0.5% hints. Best metric values are shown in bold.

Method	PSNR ↑	SSIM ↑	MS-SSIM ↑	LPIPS ↓	ΔE_{00} ↓
Colour-INR	34.97	0.9623	0.9733	0.0716	0.87
Levin et al.	**37.20**	**0.9759**	**0.9840**	**0.0550**	**0.62**

As shown in Table 4, Colour-INR remains effective when applied directly to high-resolution images, achieving a mean PSNR of 34.97 dB, SSIM of 0.9623, and CIEDE2000 error of 0.87. Levin et al. obtains higher reference-based reconstruction scores, with a PSNR of 37.20 dB and CIEDE2000 error of 0.62. These results show that the proposed coordinate-based formulation can be optimized directly on high-resolution images without changing the network architecture or requiring dataset-level retraining.

5 Conclusion

We introduced Colour-INR, an image specific implicit neural representation that formulates sparse hint colourization as an under constrained inverse problem. Since the missing chrominance is not uniquely determined by the observed luminance, Colour-INR uses sparse colour hints as chrominance anchors to constrain the solution. An edge aware spatial prior regularizes the chrominance between these observed locations, while the original luminance is preserved for final reconstruction.

The experiments show that a small amount of smoothness regularization substantially improves reconstruction over hint supervision alone, and that SIREN provides the strongest coordinate representation among the tested backbones. On Kodak, Colour-INR substantially outperforms the evaluated pretrained learning-based methods and remains competitive with the classical optimization based method of Levin et al. We further demonstrate direct optimization on high resolution DIV2K images without changing the network architecture or requiring dataset level retraining.

These results show that sparse guided colourization can be naturally formulated as image specific recovery of a continuous chrominance field. Future work can investigate richer guidance and adaptive spatial priors while retaining the image-specific nature of the representation.

AI use statement

In this work, we used generative AI tools for grammar checking and correction, LaTeX code generation, and identification of relevant literature. We did not use generative AI tools for research tasks, including the development of theoretical models or conceptual frameworks, formulation of mathematical claims, writing of mathematical proofs, refinement of hypotheses, design of research methodology, implementation of methods. Other required-disclosure tasks, including synthetic dataset generation, translation, dataset cleaning or reformatting, and qualitative or thematic data analysis, were not applicable to this work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

We do not have any concerns.

References

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

Hyojin Bahng, Seungjoo Yoo, Wonwoong Cho, David Keetae Park, Ziming Wu, Xiaojuan Ma, and Jaegul Choo. Coloring with words: Guiding image colorization through text-based palette generation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 431–447, 2018.

John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. doi: 10.1109/TPAMI.1986.4767851.

Zheng Chang, Shuchen Weng, Peixuan Zhang, Yu Li, Si Li, and Boxin Shi. L-cad: Language-based colorization with any-level descriptions using diffusion priors. In *Advances in Neural Information Processing Systems*, volume 36, pp. 77174–77186, 2023.

Aditya Deshpande, Jason Rock, and David Forsyth. Learning large-scale automatic image colorization. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 567–575, 2015. doi: 10.1109/ICCV.2015.72.

Eastman Kodak Company. Kodak lossless true color image suite. `https://r0k.us/graphics/kodak/`, 1993. Accessed: 2026-07-12.

Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan. Deep exemplar-based colorization. *ACM Transactions on Graphics*, 37(4):47:1–47:16, 2018. doi: 10.1145/3197517.3201365.

Zhitong Huang, Nanxuan Zhao, and Jing Liao. Unicolor: A unified framework for multi-modal colorization with transformer. *ACM Trans. Graph.*, 41(6), November 2022. ISSN 0730-0301. doi: 10.1145/3550454.3555471. URL `https://doi.org/10.1145/3550454.3555471`.

Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008.

Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics*, 35(4):110:1–110:11, 2016. doi: 10.1145/2897824.2925974.

Amirhossein Kazerouni, Reza Azad, Alireza Hosseini, Dorit Merhof, and Ulas Bagci. Incode: Implicit neural conditioning with prior knowledge embeddings. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1287–1296, 2024. doi: 10.1109/WACV57701.2024.00133.

Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. *CoRR*, abs/1603.06668, 2016. URL `http://arxiv.org/abs/1603.06668`.

Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23(3):689–694, August 2004. ISSN 0730-0301. doi: 10.1145/1015706.1015780. URL `https://doi.org/10.1145/1015706.1015780`.

Bingchen Li, Zhixin Wang, Fan Li, Jiaqi Xu, Jiaming Guo, Renjing Pei, Xin Li, and Zhibo Chen. Colorflux: A structure-color decoupling framework for old photo colorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15574–15584, 2026.

Bo Li, Yu-Kun Lai, Matthew John, and Paul L. Rosin. Automatic example-based image colorization using location-aware cross-scale matching. *IEEE Transactions on Image Processing*, 28(9):4606–4619, 2019. doi: 10.1109/TIP.2019.2912291.

Gehui Li, Shanshan Zhao, and Tongtong Zhao. Diff-recolor: Rethinking image colorization with a generative diffusion model. *Knowledge-Based Systems*, 300:112133, 2024a. doi: 10.1016/j.knosys.2024.112133.

Yifan Li, Yuhang Bai, Shuai Yang, and Jiaying Liu. Coco-lc: Colorfulness controllable language-based colorization. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 10939–10947. ACM, 2024b. doi: 10.1145/3664647.3680662.

Zhexin Liang, Zhaochen Li, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Control color: Multimodal diffusion-based interactive image colorization. *International Journal of Computer Vision*, 133(11):7897–7923, 2025. doi: 10.1007/s11263-025-02549-6.

Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variableperiodic activation functions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2713–2722, 2024. doi: 10.1109/CVPR52733.2024.00262.

Hongjin Lyu, Bo Li, Paul L. Rosin, and Yu-Kun Lai. Lga-net: Learning local and global affinities for sparse scribble based image colorization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8144–8153, 2025.

Varun Manjunatha, Mohit Iyyer, Jordan Boyd-Graber, and Larry Davis. Learning to color from language. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 2, pp. 764–769. Association for Computational Linguistics, 2018. doi: 10.18653/v1/N18-2120.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–10. ACM, 2022. doi: 10.1145/3528233.3530757.

Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. doi: 10.1002/col.20070.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Patricia Vitoria, Lara Raad, and Coloma Ballester. Chromagan: Adversarial picture colorization with semantic class distribution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2445–2454, 2020.

Han Wang, Xinning Chai, Yiwen Wang, Yuhong Zhang, Rong Xie, and Li Song. Multimodal semantic-aware automatic colorization with diffusion prior. *arXiv preprint arXiv:2404.16678*, 2024.

Hanzhang Wang, Deming Zhai, Xianming Liu, Junjun Jiang, and Wen Gao. Unsupervised deep exemplar colorization via pyramid dual non-local attention. *IEEE Transactions on Image Processing*, 32:4114–4127, 2023. doi: 10.1109/TIP.2023.3293777.

Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pp. 1398–1402 Vol.2, 2003. doi: 10.1109/ACSSC.2003.1292216.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Shuchen Weng, Hao Wu, Zheng Chang, Jiajun Tang, Si Li, and Boxin Shi. L-code: Language-based colorization using color-object decoupled conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 2677–2684, 2022. doi: 10.1609/aaai.v36i3.20170.

Yanze Wu, Xintao Wang, Yu Li, Honglun Zhang, Xun Zhao, and Ying Shan. Towards vivid and diverse image colorization with generative color prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14377–14386, 2021.

Wanyu Yang, Feifan Cai, Yang Shu, Zihao Zhang, Qi Liu, and Youdong Ding. Colorize at will: Harnessing diffusion prior for image colorization. *IEEE Access*, 12:107287–107296, 2024. doi: 10.1109/ACCESS.2024.3435485.

Jooyeol Yun, Sanghyeon Lee, Minho Park, and Jaegul Choo. icolorit: Towards propagating local hints to the right region in interactive colorization by leveraging vision transformer. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1787–1796, 2023. doi: 10.1109/WACV56688.2023.00183.

Nir Zabari, Aharon Azulay, Alexey Gorkor, Tavi Halperin, and Ohad Fried. Diffusing colors: Image colorization with text guided diffusion. In *SIGGRAPH Asia 2023 Conference Papers*. ACM, 2023. doi: 10.1145/3610548.3618180.

Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In *European Conference on Computer Vision (ECCV)*, pp. 649–666. Springer, 2016.

Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. Real-time user-guided image colorization with learned deep priors. *ACM Trans. Graph.*, 36(4), July 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073703. URL https://doi.org/10.1145/3072959.3073703.

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

A Additional Analysis of Colour Ambiguity

A.1 Luminance-Chrominance Distribution Analysis

To examine the ambiguity between luminance and chrominance, we convert each Kodak image to CIELAB space and partition the luminance channel into intervals of width $\Delta L = 5$. For each interval, we analyse the distribution of the corresponding (a^*, b^*) values. If luminance were sufficient to determine chrominance, pixels within a narrow luminance interval would occupy a compact region of the a^*b^* plane.

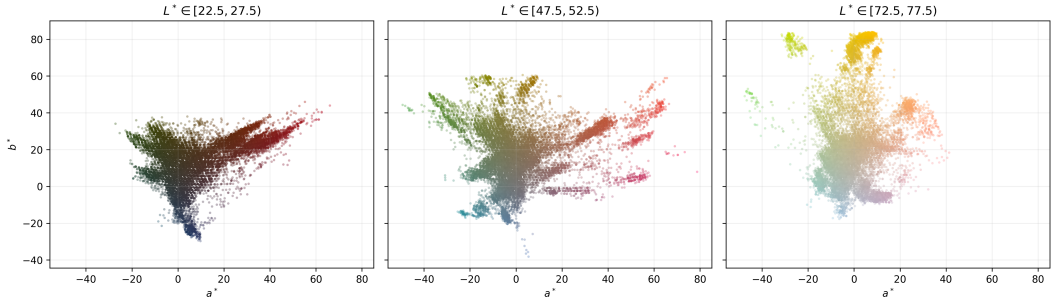

Figure 6: Conditional chrominance distributions for selected luminance intervals on the Kodak dataset. Although each interval covers only a narrow range of luminance values, the corresponding (a^*, b^*) values remain broadly distributed.

As shown in Fig. 6, restricting the luminance to a narrow interval still leaves a wide range of possible chrominance values. Across the 24 Kodak images, $68.7 \pm 14.0\%$ of the total chrominance variance remains within these narrow luminance intervals. This shows that chrominance is not uniquely determined by luminance, making recovery of (a, b) from L alone under constrained.

A.2 Diagnostic INR Experiment

The experiment referenced in Sec. 3.1 examines whether fitting the observed luminance is sufficient for a coordinate network to recover the missing chrominance. In this diagnostic setting, a shared SIREN predicts all three normalized CIELAB channels $(\hat{L}, \hat{a}, \hat{b})$. We compare luminance only optimization with luminance optimization augmented by 0.5% sparse chrominance hints. No spatial smoothness regularization is used in either setting so that the effect of chrominance guidance can be examined separately.

With luminance only optimization, the network reaches a mean luminance PSNR of 28.98 ± 3.24 dB, while the resulting colourization PSNR is only 21.62 ± 3.22 dB. Adding sparse chrominance hints gives a similar luminance PSNR of 28.53 ± 3.11 dB, but increases the colourization PSNR to 26.29 ± 2.39 dB. The chrominance error is also substantially reduced, with the mean ab error decreasing from 19.20 to 9.48 and CIEDE2000 decreasing from 15.18 to 7.50. The paired colourization PSNR improvement is 4.68 ± 2.37 dB, with an improvement observed for all 24 Kodak images.

Table 5: Diagnostic comparison between luminance only optimization and optimization with sparse chrominance hints on Kodak.

Setting	L PSNR ↑	Colour PSNR ↑	ab error ↓	ΔE_{00} ↓
L only	28.98	21.62	19.20	15.18
L + hints	28.53	26.29	9.48	7.50

These results show that accurately fitting the observed luminance does not by itself determine the missing chrominance. The two settings achieve similar luminance reconstruction quality, yet sparse chrominance observations substantially improve the recovered colour. This supports using direct chrominance guidance rather than relying on luminance fitting alone.

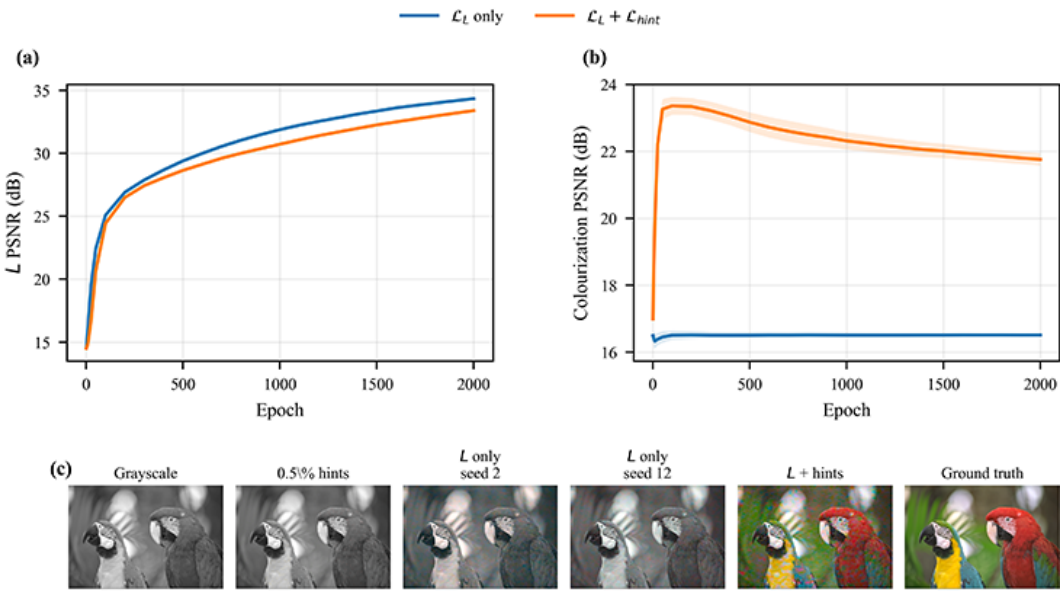

Figure 7: Diagnostic INR experiment comparing luminance only optimization with optimization using sparse chrominance guidance. Curves show mean and standard deviation over five initializations.

B Implementation Details

B.1 Colour Space and Coordinate Normalization

All images are represented in the CIELAB colour space during optimization. The luminance channel is normalized as

$$L = \frac{L^*}{100},$$

while the chrominance channels are normalized as

$$a = \frac{a^*}{128}, \qquad b = \frac{b^*}{128}.$$

Spatial image coordinates are independently normalized to $[-1, 1]$ along both axes. Thus, for an image of arbitrary spatial resolution, the input to the INR remains a two dimensional coordinate $\mathbf{x}_p = (x_p, y_p) \in [-1, 1]^2$.

The INR predicts normalized chrominance values $(\hat{a}, \hat{b})$. After optimization, these are rescaled to native CIELAB units and combined with the observed luminance obtained from the grayscale image before conversion back to RGB.

B.2 Network Architecture

The default Colour-INR backbone is a SIREN with hidden width 128. The network contains one initial sinusoidal layer mapping the two-dimensional coordinate to 128 features, followed by three additional 128-dimensional sinusoidal hidden layers. A final linear layer maps the hidden representation to two chrominance channels, followed by a $\tanh$ activation.

The resulting architecture is therefore

$$2 \to 128 \to 128 \to 128 \to 128 \to 2.$$

The first sinusoidal layer uses $\omega_0 = 60$, while all subsequent sinusoidal layers use $\omega_0 = 30$. The final two-dimensional output is passed through $\tanh$ to produce normalized $(\hat{a}, \hat{b})$ values.

B.3 Sparse Hint Generation

For the controlled Kodak experiments, sparse point hints are generated directly from the ground-truth chrominance channels. A hint density of 0.5% is used unless otherwise stated. Hint locations are generated using a fixed random seed of 123 and are kept identical across all compared configurations and methods whenever possible.

For an image containing N pixels, the target number of hint pixels is

$$N_{\text{hint}} = \text{round}(0.005N).$$

Each selected location provides only its target chrominance (a, b); no unhinted ground-truth chrominance values are used during Colour-INR optimization.

B.4 Edge Aware Smoothness Construction

The edge mask is computed from the observed grayscale-derived luminance using the Canny detector with $\sigma = 1.0$. We use the default Canny low and high thresholds and do not dilate the resulting edge map.

For each sampled coordinate, we consider all eight immediate spatial neighbours: left, right, up, down, and the four diagonal directions. A neighbouring pair contributes to the smoothness loss only when both pixels are classified as non-edge pixels. All retained neighbour pairs are weighted equally; diagonal pairs are not assigned a separate distance-dependent weight.

For a valid pair (p, q), the contribution is

$$\frac{1}{2}\left(|\hat{a}_p - \hat{a}_q| + |\hat{b}_p - \hat{b}_q|\right).$$

The loss is averaged over all valid neighbour pairs in the current optimization step.

B.5 Optimization and Reproducibility

Each image is optimized independently for 2000 epochs using Adam with a learning rate of 10^{-4}. We use a coordinate batch size of 32,768. Up to 2,048 hint coordinates are explicitly included in each optimization step to ensure that sparse colour supervision remains represented in the minibatch.

Unless otherwise stated, we use model initialization seed 2, coordinate-batch seed 2002, and hint-generation seed 123. The hint coefficient is fixed to $\lambda_{\text{hint}} = 1$, and the selected smoothness coefficient is $\lambda_{\text{smooth}} = 0.05$.

The INR receives only normalized spatial coordinates as input. The grayscale luminance, hint mask, hint colours, and Canny edge map are not concatenated to the network input. They affect optimization only through the reconstruction constraints described in Sec. 3.2.

C Additional Ablation Results

C.1 Smoothness Weight

Table 6 reports the complete smoothness-weight sweep corresponding to Fig. 3. All configurations use identical hints, network initialization, coordinate sampling, and optimization settings.

D Additional Comparison Results

D.1 Hint Mapping and Resolution Handling

To keep the sparse colour observations consistent across methods, hint locations are first generated in the coordinate system of the original Kodak image. For methods operating at a different spatial resolution, each hint coordinate is mapped to the corresponding location in the resized image using the same geometric scaling applied to the image. The associated hint colour is preserved from the original ground-truth pixel.

Table 6: Full smoothness-weight ablation on the 24 Kodak images.

λ_{smooth}	PSNR ↑	SSIM ↑	LPIPS ↓	ΔE_{00} ↓
0	29.80	0.9383	0.1632	4.896
0.05	**32.64**	0.9714	**0.0667**	**2.849**
0.10	32.60	0.9715	0.0685	2.857
0.25	32.56	**0.9716**	0.0726	2.902
0.50	32.47	0.9714	0.0779	2.996
1.00	32.27	0.9709	0.0859	3.203
1.50	32.03	0.9704	0.0917	3.401
2.00	31.76	0.9698	0.0968	3.589
5.00	30.29	0.9670	0.1176	4.530

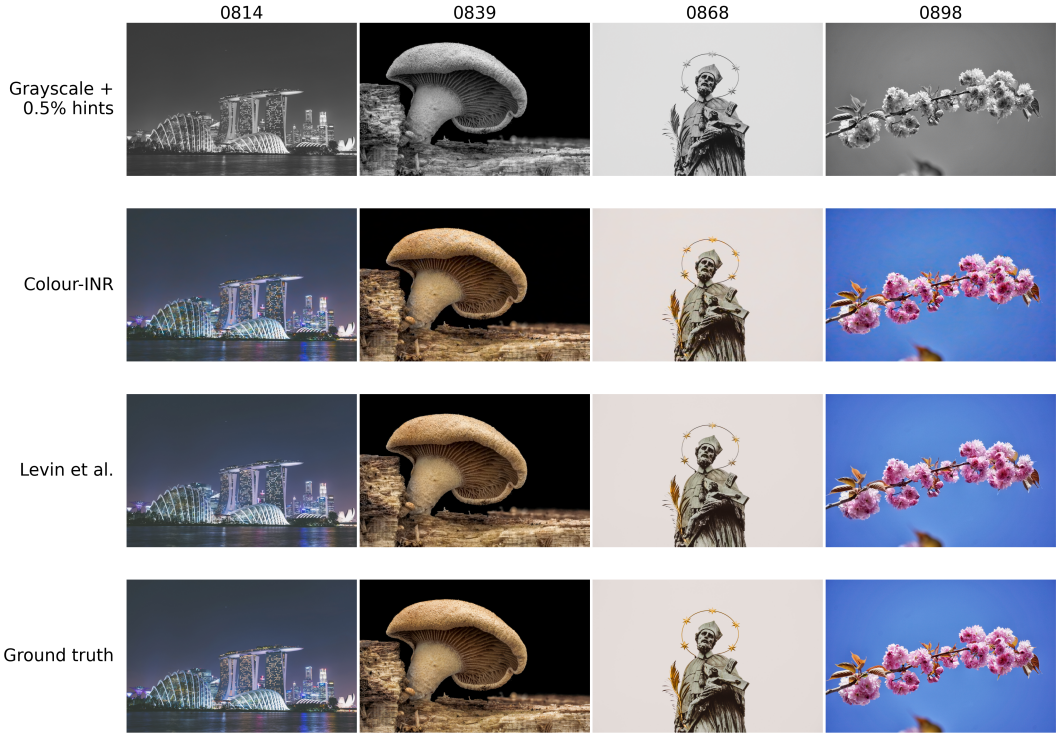

Figure 8: Qualitative high-resolution colourization results on selected DIV2K images using 0.5% colour hints.

Colour-INR and the Levin et al. method are evaluated at the original Kodak resolution. Methods requiring a fixed network input resolution are evaluated using the resolution required by their released implementation. The predicted colour image is subsequently mapped back to the original image resolution for reference-based evaluation where required.

D.2 High Resolution DIV2K Colourization

Figure 8 provides qualitative examples from the high-resolution DIV2K evaluation in Sec. 4.4. Colour-INR and Levin et al. (2004) are applied directly at the native image resolution using the same 0.5% sparse colour observations. The examples include different scene structures and levels of colour ambiguity.