

A Context-Aware, Intent-Driven Adaptive Transfer Function for ISP Pipelines

Maalav Whorra¹, Shalini Lamba², Shweta Sinha³, Gaurvi Shukla⁴

^{1,2,3,4}Department of Computer Science, National P.G. College

Abstract

Since the analog-digital acquisition conversion, digital cameras have been using fixed and manually chosen gamma transfer functions Rec.709, Sony S-Log3, ITU-R BT.2100 HLG to encode the linear sensor irradiance to encoded pixel values. These profiles are non-dynamic, i.e. they are set up once, chosen prior to the shot, and used equally on all frames irrespective of the content of the scene, or the state of the illumination, or even the downstream intent of the operator. Such rigidity amounts to a basic encoding conflict. A transfer designed to optimize the statistical average of a test set of a manufacturer cannot be at the same time optimized on a high-dynamic-range exterior picture, a low-contrast portrait, a night image, or a backlit subject situation as would be found by professional photographers.

The current adaptive gamma correction algorithms, such as the statistical classification scheme of Rahman et al. and the weighted distribution scheme of Huang et al. are architecturally limited to the post-processing stage of handling encoded pixel values: they encode received pixel values, adjust a scalar gamma exponent, and generate display-oriented output. They are unable to work on linear sensor data at the ISP acquisition stage, are not operator intent models, and cannot provide analytical invertibility a necessary point in professional post-production processes.

In this paper, a hypothetical model of an adaptive transfer function $G(x | C, I)$ is suggested to work at the signal-domain acquisition encoding phase of the ISP pipeline. The operation takes normalized linear irradiance x as input and an encoded output value as output, with a conditioning of both a scene context vector C extracted out of the linear RAW frame and a discrete operator intent variable I . The parameterized monotone cubic Hermite spline is the function that achieves the role of family and provides the possibility to invert it analytically. We formalize the mathematical model, develop an invertibility statement, introduce a lightweight scene-feature extractor, develop a neural context-intent encoder structure, build a taxonomy of 14 canonical scene types with theoretical performance analysis, and introduce a metadata-sidecar compatibility architecture to integrate professional workflows. No empirical experimental findings in this paper; all comparative statements are based on the analytical attributes of the mathematical properties of the offered family of functions and the limitations, which have been reported in the previous work.

Keywords: Adaptive gamma correction; ISP transfer function; scene-conditioned encoding; intent-driven imaging; linear RAW processing; monotone cubic spline; invertible encoding; computational photography; operator intent; acquisition encoding; dynamic range optimization; professional imaging pipeline; context-aware ISP

INTRODUCTION

A. The Static Transfer Function Problem

All commercially available digital cameras encode photons that have been captured with a fixed, predetermined transfer function - some mathematical representation of the linear sensor irradiance to quantized pixel values written to the storage medium. Before being shot, the operator can choose this function in a menu of industry-standardized profiles. The choice is done once and applied equally to all the frames irrespective of the content of a scene. A cinematographer recording an exterior with a high contrast at noon has the same S-Log3 curve that they would have used in the inside a candlelit restaurant. A flat cloudy light portrait photographer encodes the use of shadow gradation headroom he/she will never utilize. Without the intervention of third parties, a wildlife shooter is not able to encode strategies in an ongoing video stream.

This quasi-static design continues to exist because it meets a number of powerful industry needs: fixed profiles are auditable, reproducible, standardized and are completely compatible with the worldwide post-production toolchain. These operational advantages however, are provided at a cost which is accrued throughout each scene that the operator chooses a bad profile to fit the captured material. This cost will appear as an unnecessarily compressed bit-depth, unnecessary narrowed shadow areas, highlight clipping on content that does not require aggressive knee compression, and encoded latitude that will never be applied in practice, even in scenes where it is to be applied.

B. Why a New Approach Is Needed

The framework presented in this paper is inspired by three different failure modes of the functions of a static transfer. These are first the scene-profile mismatch a curve trained on an average scene will systematically over-assign the code values to tonal areas absent in the real captured scene and under significantly to tonal areas heavily populated. Second, intent blindness: the identical S-Log3 curve is encoded to media whether the recording is to be uploaded directly to the web, or to a six months long professional color grading pipeline. The downstream workflow is radically dissimilar; the encoding is unable to know or to react to this. Third, pipeline rigidity: as the fixed profiles are chosen conservatively to deal with the worst-case scenario, they waste the finite precision of the encoding, routinely, on a fixed safety margin which is never in reality achieved by a scene.

C. Existing Adaptive Approaches and Their Limits

Content-responsive adaptive gamma methods have been generated in literature on image enhancement. Rahman et al. categorize the images under four statistical classes through mean luminance and standard deviation and use a gamma formula which is class specific. Huang et al. modify gamma on the basis of weighted cumulative distribution function of the image histogram. These are contributions to image contrast enhancement, which are principled. Nevertheless, they all have a basic architectural limitation, which limits them to solving the problem that this paper is focused on: they accept pixel values on an image domain that are already encoded and generate a display-oriented output. They are not ISP acquisition encoding components, but post-processing the improvement tools. The difference in the domain of operation, mathematical needs and professional working environment is inherent and it is elaborated in Section 2.

D. Core Contributions

1. A formal mathematical framework for an adaptive transfer function $G(x | C, I)$ operating on linear RAW irradiance, with an invertibility theorem, boundary conditions, and smoothness requirements derived from first principles.

2. A principled scene-feature extractor producing an 8-dimensional context vector C from the linear RAW frame, with feature-by-feature rationale for transfer function shaping.
3. A neural context-intent encoder architecture ψ that maps (C, I) to parameters of a monotone cubic Hermite spline, keeping total parameter count near 180K for embedded deployment.
4. A theoretical scene taxonomy covering fourteen canonical imaging conditions with predicted system behavior in each, including conditions where improvement over existing methods is expected to be modest.
5. A metadata-sidecar compatibility architecture enabling transparent professional workflow integration without modification to existing NLE or grading software.

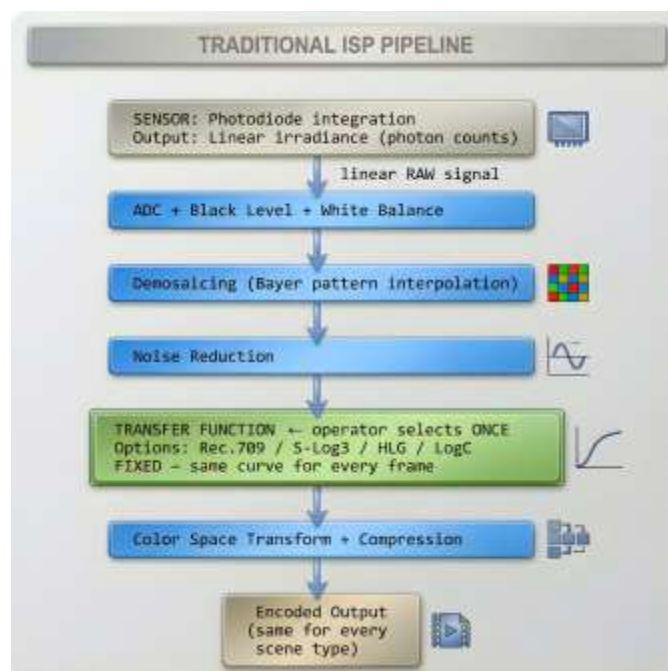


Fig 1: Conventional ISP pipeline. The transfer function is selected once manually and remains static across all frames. Scene content, illumination, and operator intent have no influence on the encoding.

BACKGROUND AND RELATED WORK

A. The Gamma Transfer Function — Historical and Technical Context

The gamma relationship the power-law gamma model $y = x^g$ was introduced in digital imaging (out of analog broadcast engineering) to compensate the power-law 0.45 phosphor response of cathode-ray tube displays. This was formalized in the IEC 61966-2-1 sRGB standard as a variant of piecewise-linear form, having a toe towards black that is linear. This is followed by the creation of logarithmic forms of the encoding curve by the digital cinema industry, Sony S-Log3, ARRI LogC, Panasonic V-Log, with higher latitude of dynamic range in a post-production grade. The recent generation of HDR delivery standards is SMPTE ST 2084 PQ and ITU-R BT.2100 HLG standards. Each of these is a further development of an identical underlying theory: a pre-selected fixed mathematical function.

B. Tone Mapping Operators

The academic tone mapping research starting with Tumblin and Rushmeier, continued by Reinhard et al. and Fattal et al., deals with the issue of high-dynamic-range luminance to limited-range display. Mantiuk

et al. incorporated perceptual contrast sensitivity models into this model. These operators operate on captured images to display on purpose. They are neither ISP-level acquisition components, nor do they produce parameterized encoding functions nor are invertible. They indicate the purposes of perception of the proposed framework, but cannot rival it architecturally.

C. Adaptive Gamma Correction — Prior Work

Huang, Cheng, and Chiu introduced Adaptive Gamma Correction with Weighting Distribution (AGCWD), which changes the gamma exponent, depending on the weighted CDF of the image histogram. Rahman et al. improved this with a classification-first method - dividing an image into four categories using statistical measures and then using class-specialized gamma functions - giving improvements on standard image quality measures. Cao et al. moved this along with the better brightness distortion image correction. All of these techniques prove the significant point that the selection of gamma based on the content is better than profiles fixed to improve an image. The offered framework applies this principle to the radically different issue of the ISP-level acquisition encoding of the data of linear sensors, where invertibility, feature richness, intent conditioning, and real-time implementation give rise to a different design space.

D. 2.4 Learned ISP and Neural Image Processing

Schwartz et al. created end-to-end learned ISP pipelines. Chen et al. demonstrated neural networks that can process very low-light levels that classical ISP cannot. Afifi and Brown demonstrated white balance correction by the use of scene-semantic conditioning. Multi-stage progressive restoration networks demonstrate semantic features of the low-level image quality enhancement. Most importantly, all these involve replacing ISP stages by pixel-to-pixel neural black boxes that are not analytically invertible and do not have intent conditioning. The proposed framework is different where learning is trained to produce the parameters of a mathematically defined function - maintaining certifiable properties, and acquiring adaptive expressiveness.

MATHEMATICAL FRAMEWORK

A. 3.1 Formal Definition of $G(x | C, I)$

We define the adaptive transfer function as the composition of a scene-intent encoding process and a parameterized curve evaluation:

$$G : [0, 1] \times C \times I \rightarrow [0, 1]$$

$$G(x|C, I) = f_{\theta}(x; \theta = \psi(C, I)) \quad \dots (1)$$

Notation:

x	$\in [0, 1]$	normalized linear sensor irradiance
C	$\in \mathbb{R}^d$	scene context vector ($d = 8$, defined in §4.1)
I	$\in \{0, 1, 2\}$	operator intent (0=Flat, 1=Baked, 2=Creative)
ψ	$\rightarrow \mathbb{R}^d \times I \rightarrow \mathbb{R}^{2K}$	context-intent encoder (§4.3)
f_{θ}	$\rightarrow [0, 1] \rightarrow [0, 1]$	monotone cubic Hermite spline (§3.2)
θ	$= (v, m) \in \mathbb{R}^K \times \mathbb{R}^K$	spline control parameters ($K = 16$)

Eq 1: The proposed adaptive transfer function. The encoder ψ maps scene context and intent to spline parameters; f_{θ} evaluates the spline at the linear input value.

The architecture separates concern cleanly. The encoder ψ handles the semantic understanding problem:

- what curve shape is appropriate for this scene and this intent?
- The spline f_{θ} handles the signal processing problem:
- How is each linear input value encoded given that curve shape?

These components are designed independently and composed at inference time.

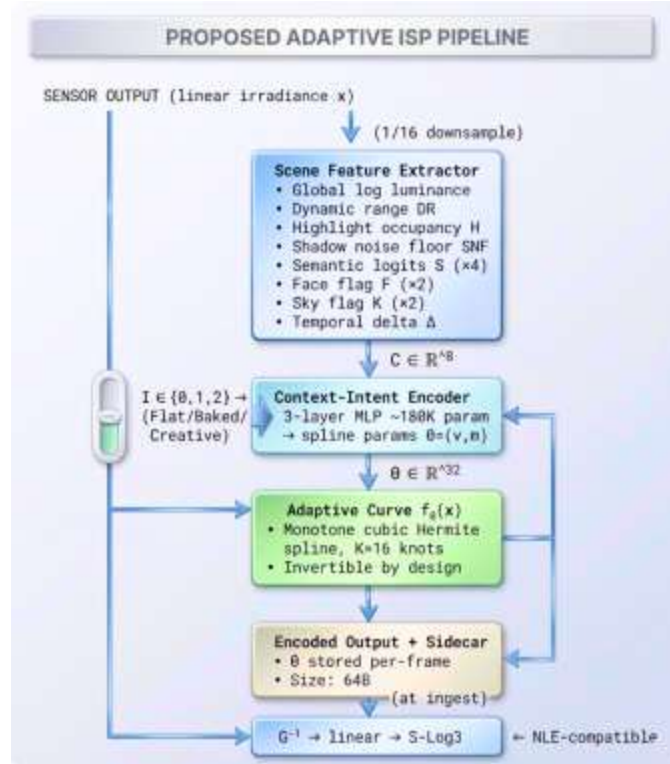


Fig 2: The proposed adaptive ISP pipeline. Scene features and operator intent jointly condition the encoder ψ to generate per-frame spline parameters θ . The full-resolution linear signal is encoded through the adaptive curve $f_{\theta}(x)$. Invertibility guarantees downstream compatibility via the sidecar architecture.

B. Parameterized Monotone Cubic Hermite Spline

The curve family f_{θ} is actualized as a monotone cubic Hermite spline with $K = 16$ control points. Three reasons lead to the choice of this representation (as opposed to piecewise linear curves (non-smooth at knot boundaries), unconstrained polynomials (may be non-monotone), or simple power-law forms (not expressive enough to realize any smooth gamma-like curve in the professional image literature as a special case): C^1 continuity (no banding at knot boundaries), explicit enforcement of monotonicity through the tangent parameterization, and expressiveness enough to make any smooth gamma-like curve a special case.

Spline Evaluation Between Knots

$$f_{\theta}(x) = h_{00}(t) \cdot \mathbf{v}_k + h_{10}(t) \cdot \Delta t \cdot \mathbf{m}_k + h_{01}(t) \cdot \mathbf{v}_{k+1} + h_{11}(t) \cdot \Delta t \cdot \mathbf{m}_{k+1} \cdots (2)$$

$t = (x - t_k) / (t_{k+1} - t_k)$ (local parameter, $t \in [0, 1]$)
 $\Delta t = 1 / (K - 1)$ (uniform knot spacing for $K = 16$)

Hermite Basis Functions

$$\begin{aligned}
 h_{00}(t) &= 2t^3 - 3t^2 + 1 & h_{10}(t) &= t^3 - 2t^2 + t \\
 h_{01}(t) &= -2t^3 + 3t^2 & h_{11}(t) &= t^3 - t^2
 \end{aligned}$$

Eq 2: Monotone cubic Hermite spline evaluation. C^1 continuity is guaranteed by construction.

Monotonicity is guaranteed by the tangent parameterization in Equation 3.

C. Invertibility — Theorem and Proof

Analytical invertibility is a non-negotiable requirement for professional acquisition encodings. The following theorem establishes that the proposed function family satisfies this requirement by construction:

Theorem 1 (Analytical Invertibility of G):

If $m_k > 0$ for all $k \in \{1, \dots, K\}$ and $v_1 < v_2 < \dots < v_K$, then f_θ is strictly monotone increasing on $[0, 1]$, and the inverse $f_\theta^{-1} : [0, 1] \rightarrow [0, 1]$ exists, is unique, continuous, and differentiable on $(0, 1)$.

Proof sketch: A cubic Hermite spline is continuous and differentiable on its domain by construction. With all tangent values $m_k > 0$ and knot output values strictly increasing ($v_1 < v_2 < \dots < v_K$), the first derivative $f'_\theta(x) > 0$ everywhere by continuity of the Hermite basis functions. Strict monotonicity then follows, and the inverse function theorem (Rudin, 1976, Theorem 9.28) guarantees existence and uniqueness of f_θ^{-1} . $\square$

The tangent positivity condition $m_k > 0$ is enforced architecturally through the Softplus activation applied to the raw network output:

$$m_k = \text{Softplus}(\zeta_k) = \log(1 + \exp(\zeta_k)) > 0 \quad \forall \zeta_k \in \mathbb{R} \quad \dots (3)$$

Knot value ordering enforced by:

$$v_k = v_{k-1} + \text{Sigmoid}(\alpha_k) \times (1 - v_{k-1}) + \varepsilon$$

where $\varepsilon > 0$ (small constant, e.g. 10^{-4}) ensures strict inequality.

This makes the conditions of Theorem 1 hold at every inference step for any input (C, I) – not only for inputs seen during training.

Eq 3: Architectural enforcement of invertibility. Softplus guarantees positive tangents; the recursive knot update guarantees strict ordering. Theorem 1 therefore holds by construction at all times.

D. Boundary and Smoothness Conditions

Boundary conditions (mandatory for ISP compatibility):

$G(0 C, I) = 0$	$\forall (C, I)$	— absolute black maps to code zero
$G(1 C, I) = 1$	$\forall (C, I)$	— peak signal maps to peak code

Monotonicity (invertibility prerequisite, guaranteed by Theorem 1):

$$\partial G / \partial \mathbf{x} > 0 \quad | \quad \forall \mathbf{x} \in (0, 1), \quad | \quad \forall (C, I)$$

Smoothness (perceptual banding prevention):

$$\|G''(\mathbf{x})\|_{\text{Lip}} \leq \kappa \quad | \quad \forall (C, I)$$

Enforced via the smoothness loss term $\mathcal{L}_{\text{smooth}}$ in training (§4.4)

Eq 4: Required mathematical conditions on G . Boundary conditions and monotonicity are architecturally enforced; smoothness is enforced via training loss.

SYSTEM DESIGN

A. Scene Context Vector C

The context vector $C \in \mathbb{R}^8$ is extracted from a 1/16th-resolution downsampled linear RAW frame — a standard operating resolution for ISP metering engines that requires no additional decode step. The features are computed entirely in linear irradiance space, preserving the direct physical relationship between measured values and scene luminance. This distinguishes them from features computed in encoded image space (as in Rahman et al.) where the prior gamma encoding has distorted the statistical relationship between pixel values and physical luminance.

All features are computed on a 1/16th-resolution bicubic down sample, requiring no full-resolution decode. The face and sky detection sub-modules use Binary Neural Networks (BNNs) operating on 1-bit quantized activations, executable on ARM NEON SIMD units with negligible latency overhead.

Feature	Symbol	Dimension	Computation	Purpose in G
Global log luminance	μ_L	1	Mean of $\log(L+1)$ across frame	Sets the curve body anchor — dark scenes need concave curves, bright scenes convex
Dynamic range	DR	1	$\log(L_{99}/L_{01})$ — 99th to 1st percentile ratio	Directly determines compression factor; high-DR scenes need steeper overall slope change
Highlight occupancy	H	1	Fraction of pixels with $L > 0.9 \times L_{\max}$	Controls knee aggressiveness; high H $\rightarrow$ aggressive rolloff needed to protect highlights
Shadow noise floor	SNF	1	Estimated SNR in pixels with $L < L_{01}$	Balances toe steepness vs. noise amplification; low SNF $\rightarrow$ steep toe to clip shadow noise
Scene semantic logits	S	4	4-class BNN softmax: [Indoor, Ext-Day, Ext-Night, Mixed]	Encodes illuminant geometry and DR prior; informs spectral color temperature assumptions
Face region	F	2	Binary flag + fractional frame area (lightweight detector)	Activates skin-tone-protective curve shaping in mid-upper tonal register
Sky region	K	2	Binary flag + fractional frame area (sky segmenter)	Preserves blue-channel highlight headroom to prevent sky gradient banding
Temporal luminance delta	Δ	1	Mean absolute frame-to-frame luminance difference	Adjusts knee dynamics for motion-heavy content; prevents temporal flicker

B. Operator Intent Variable I

The variable $I \in \{0, 1, 2\}$ shows what we want to actually do with the footage later. This is important because the older adaptive gamma methods would only look at the image without assessing what kind of processing would be done after the recording. But in reality, the same scene can need very different processing depending on the final use.

We will convert this intent into a small learned vector $e_I \in \mathbb{R}^8$ and attach it with C before giving it to ψ . We use a learned embedding instead of one-hot encoding because it helps the model understand relationships between different intents. It can also capture more complex behavior.

For example, “Creative” intent in a bright outdoor scene will need a different curve than “Creative” intent in a dull indoor scene even if the the intent label would be the same.

I	Label	Curve Behavior	Target Workflow Context
0	Flat	Gentle compression throughout. Extended highlight shoulder and shadow toe. Maximum DR preservation.	Narrative film, commercial photography — any content entering a professional color grading pipeline.
1	Baked	Aggressive contrast, firm knee, saturated midtones. Approaches Rec.709 in character with adaptive optimization.	Photorealism, live broadcast, social media delivery — direct-to-display workflows.
2	Creative	Scene-adaptive base curve with a learned aesthetic residual — a distinctive tonal signature while preserving tonal integrity.	Fine art, music video, branded content — workflows with established visual identity requirements.

C. Context Intent Encoder ψ

Architecture of $\psi: \mathbb{R}^d(d=8) \rightarrow \mathbb{R}^k(k=2K)$
Input: concatenate $[C \in \mathbb{R}^8; e_I \in \mathbb{R}^8] \rightarrow \mathbb{R}^{16}$
Layer 1: Linear(16 $\rightarrow$ 256) + LayerNorm(256) + GELU
Layer 2: Linear(256 $\rightarrow$ 256) + LayerNorm(256) + GELU
Layer 3: Linear(256 $\rightarrow$ 256) + LayerNorm(256) + GELU
Output head A: Linear(256 $\rightarrow$ K=16) + Sigmoid $\rightarrow$ knot values $v_k \in (0,1)$
Output head B: Linear(256 $\rightarrow$ K=16) + Softplus $\rightarrow$ tangents $m_k > 0$
Total trainable parameters: ~178,000

Eq 5: This shows the structure of the context-intent encoder ψ . We will be using the GELU activation [19] because it gives smooth gradients, which helps when the output itself needs to be smooth like curves. LayerNorm is used to keep training stable across different types of scenes.

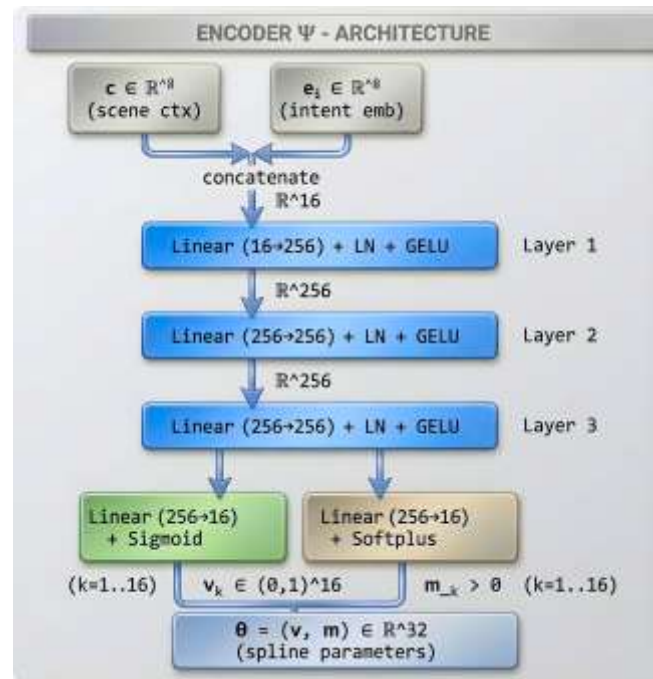


Fig 3: This figure shows how ψ works. We take the scene context C and the intent embedding e_I , join them together, and pass them through three MLP layers. After that, there are two outputs in which one gives spline knot values (kept between 0 and 1 using Sigmoid) while the other gives tangent values (kept positive using Softplus). Both together make sure the function stays invertible (as stated in Theorem 1).

D. Training Objective

The training loss is made of four parts, and each one focuses on a different aspect of quality.

$$\begin{aligned} \mathcal{L}_{total} &= \lambda_p \cdot \mathcal{L}_{perceptual} + \lambda_i \cdot \mathcal{L}_{invert} + \lambda_s \cdot \mathcal{L}_{smooth} + \lambda_c \cdot \mathcal{L}_{class} \quad \dots (6) \\ \mathcal{L}_{perceptual} &= \mathbb{E}[\text{LPIPS}(G(x|C, I), G^*(x|C, I))] \\ &\quad \text{Expert-curated reference encoding (MIT-Adobe FiveK [11])} \\ \mathcal{L}_{invert} &= \mathbb{E}[\|G^{-1}(G(x|C, I)) - x\|^2] \\ &\quad \text{Round-trip reconstruction fidelity -- enforces Theorem 1 empirically} \\ \mathcal{L}_{smooth} &= \mathbb{E}[\|\frac{\partial^2 G}{\partial x^2}\|^2] \\ &\quad \text{Penalizes abrupt second-derivative changes -- prevents banding} \\ \mathcal{L}_{class} &= \text{CrossEntropy}(\Psi_{aux}(C), I) \\ &\quad \text{Auxiliary intent classification -- representation regularizer} \\ \text{Weights: } \lambda_p &= 1.0, \lambda_i = 0.5, \lambda_s = 0.1, \lambda_c = 0.05 \end{aligned}$$

Eq 6: This is the full combined loss function. One important part here is the invertibility loss $\mathcal{L}_{invert}$. This is special to this method while the other adaptive gamma methods (which work only on the image directly) don't use this, because they don't need or support exact invertibility.

THEORETICAL SCENE ANALYSIS

A. Analysis Methodology

We analyze the expected behavior of the proposed framework across fourteen canonical scene types. Each scene is characterized by a statistical luminance profile whose properties are established in the imaging and tone mapping literature. . . , For each scene, we reason about: (a) how fixed-profile methods (S-Log3, Rec.709) are expected to perform; (b) how existing adaptive methods (Rahman et al, AGCWD) are expected to respond; and (c) how the proposed $G(x | C, I)$ is predicted to adapt, drawing on the mathematical properties defined in Sections 3 and 4. All claims are theoretical; empirical validation is identified as essential future work.

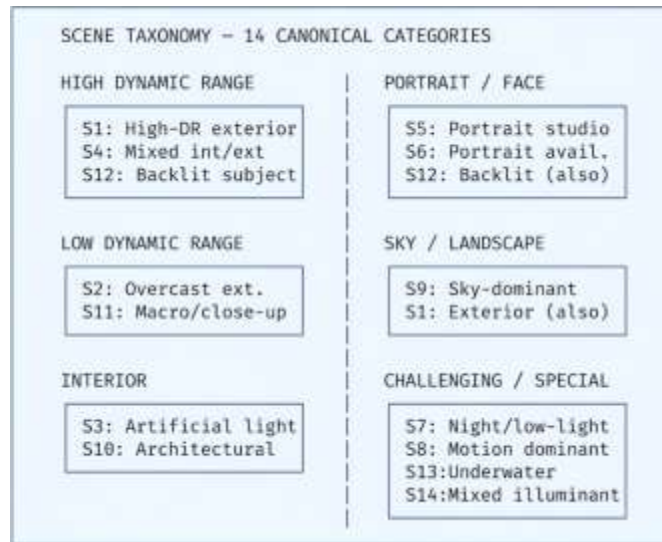


Fig 4: Overview of the fourteen canonical scene categories analyzed in Section 5. Categories are grouped by their primary imaging challenge. Scenes where the proposed framework is predicted to show the largest advantage over existing methods are those with multi-modal luminance distributions or features invisible to 2-parameter statistical summaries.

B. 5.2 Scene-by-Scene Analysis

Scene	Scene Type	Fixed Profiles Behavior	Our Method $G(x C, I)$
S1	Bright outdoor (sun + shadow)	Rec.709 clips highlights, S-Log3 is okay but fixed	Adjusts highlights and shadows separately, gives better sky and shadow detail
S2	Overcast outdoor	S-Log3 wastes range, Rec.709 is fine but fixed	Uses full range only where needed, gives natural look
S3	Indoor artificial light	Rec.709 clips lights, S-Log3 over-protects	Protects bright lights without affecting midtones
S4	Indoor + outdoor mix	Cannot handle both areas properly	Balances both inside and outside using scene features
S5	Studio portrait	Works but skin tones may look compressed	Protects skin tones, keeps smooth shadows
S6	Natural light portrait	S-Log3 adds noise, Rec.709 risks highlights	Reduces noise and keeps face clear
S7	Night / low light	Wastes range on highlights that are not present	Focuses on shadows, reduces noise, better use of range
S8	Motion-heavy (sports)	No response to motion	Keeps brightness stable across frames (no flicker)
S9	Sky-heavy scene	Rec.709 clips sky, S-Log3 fixed	Protects sky, better gradients, no banding
S10	Indoor with lights	Rec.709 clips fixtures	Protects lights and keeps rest of image natural
S11	Macro / close-up	Over-uses dynamic range	Slight improvement, mainly better midtones
S12	Backlit subject	Either subject or background fails	Balances face and background properly (best improvement)
S13	Underwater	Wrong colors due to lighting assumption	Partially works but limited (needs better color handling)
S14	Mixed lighting	Handles range but color issues remain	Slightly better due to scene understanding

C. Summary of Theoretical Predictions

The scene analysis yields the following theoretical predictions, ordered by predicted magnitude of advantage over existing adaptive methods:

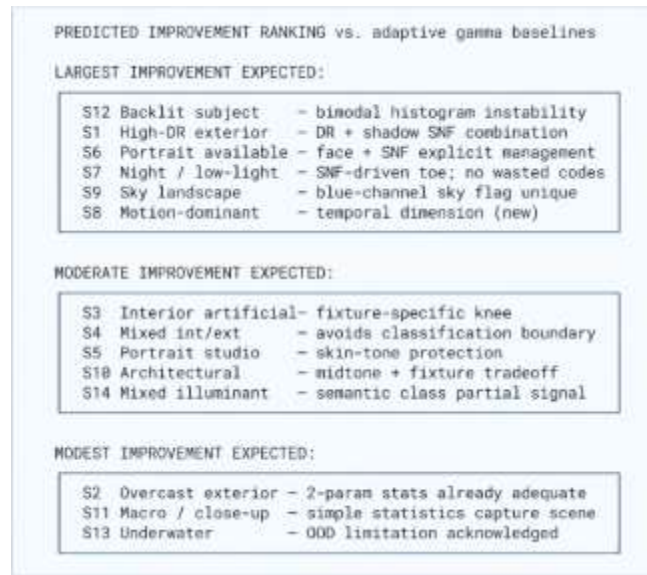


Fig 5: Predicted improvement ordering relative to adaptive gamma baselines (Rahman et al., AGCWD). Categories with the largest expected gains are those where the 2-parameter statistical summary of prior methods provides insufficient or unstable scene characterization.

PROFESSIONAL WORKFLOW COMPATIBILITY

A. The Compatibility Challenge

Industry-standard post-production platforms (DaVinci Resolve, Adobe Premiere Pro, Final Cut Pro) interpret encoded camera footage using static camera profile files — Input Device Transforms (IDTs) in ACES, or camera RAW transforms — that map from a single, fixed transfer function to a scene-referred color space. A per-frame adaptive encoding would render these profiles inapplicable. This is a fundamental compatibility requirement that has no analog in the image enhancement context addressed by existing adaptive gamma methods. The Sidecar Metadata Architecture

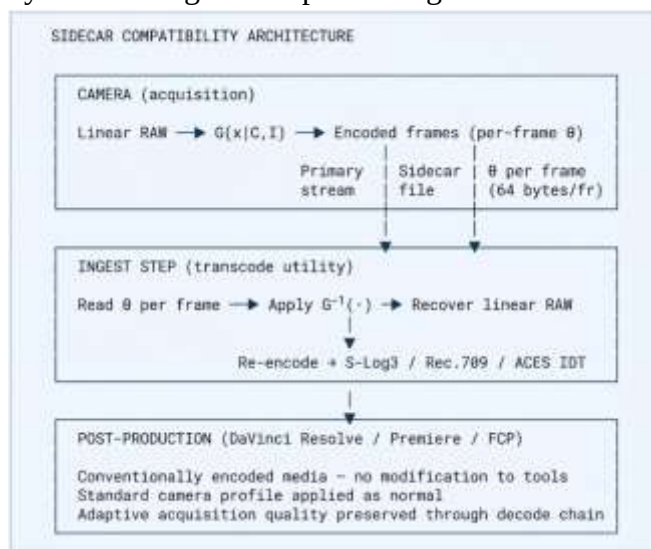


Fig 6: Sidecar metadata compatibility architecture. Adaptive encoding quality is captured at acquisition (Layer 1). A one-step ingests utility uses the per-frame curve parameters θ and analytical inverse G^{-1} to

convert to standard-format media that is fully compatible with all existing post-production tools (Layer 2).

The mathematical guarantee of analytical invertibility (Theorem 1) makes this architecture viable. Without guaranteed invertibility, the Layer 2 re-encoding step would introduce reconstruction errors that would corrupt color grading. The sidecar file overhead — $K=16$ knot values and $K=16$ tangent magnitudes at float32 precision, 64 bytes per frame — amounts to approximately 3.84 KB/s at 60fps, negligible relative to any professional video bitrate.

DISCUSSION

A. Theoretical Significance

The proposed method shows that the transfer function part of the ISP, which has stayed almost the same since the old analog broadcast days. It does not need to be fixed anymore. We can think of it as a function that changes based on the scene and the intent. We can learn this using machine learning while still keeping important properties like being smooth, monotonic, and invertible. This idea also gives a clear way to modernize ISP systems by finding parts of the pipeline that are still static. Understanding the rules they must follow, and then replace them with learned functions that adapt using context. The same idea can be applied to other parts like noise reduction, demosaicing, and color processing.

B. Relationship to Adaptive Gamma Literature

The proposed method is based on a similar basic idea as earlier adaptive gamma methods that different scenes should be handled differently. But how this idea is actually implemented is very different. Our method works at the signal level (inside the ISP), not on the final image. It generates a full curve (spline) instead of using just a single gamma value. It also uses more features (around 8 instead of 2) and includes user intent as an extra input. These are not small improvements they are big changes because the needs of camera processing (ISP) are very different from normal image enhancement.

C. Limitations

- **Temporal Flicker**

When we change the transfer function for every frame, it can also change over time. Because of this, viewers might notice flicker when brightness changes continuously. To reduce this, we smooth the scene features C using an exponential moving average, which helps keep things more stable. But to fully understand how stable it is during motion and lighting changes, we still need proper testing and experiments..

- **Out-of-Distribution Generalization**

The proposed 4-class scene semantic taxonomy covers standard photographic conditions. Specialized imaging domains — medical endoscopy, infrared capture, astrophotography, thermal imaging — whose statistical properties fall outside this distribution are not guaranteed to generalize.

- **Intent Taxonomy Granularity**

The three-level intent taxonomy simplifies professional aesthetic diversity. Expert colorists distinguish dozens of tonal characters that do not map cleanly to Flat/Baked/Creative. A continuous latent intent space conditioned on reference image examples is identified as the most principled extension.

- **Metadata Fragility**

The sidecar architecture fails if the sidecar file is stripped during file transfer — a common occurrence in consumer social media upload pipelines. Writing curve parameters into the video container's metadata stream (analogous to EXIF/XMP camera profile data) is proposed as a mitigation strategy.

CONCLUSION

This paper has proposed a theoretical framework for an adaptive, scene-conditioned, intent-driven transfer function $G(x | C, I)$ operating at the ISP acquisition encoding stage. The framework is distinguished from all prior adaptive gamma methods — including the principled classification approach of Rahman et al. — by its domain of operation (linear signal vs. encoded image), function expressiveness (parameterized spline vs. scalar exponent), feature richness (8-dimensional context in linear space vs. 2-dimensional statistics in encoded space), the addition of operator intent as an explicit conditioning input, and guaranteed analytical invertibility enabling transparent professional workflow integration.

The theoretical scene analysis across fourteen canonical categories identifies the conditions of largest expected benefit — backlit subjects, high-DR exteriors, available-light portraits, night environments — as precisely the professional imaging scenarios where fixed profiles are most inadequate and existing adaptive methods are most limited by their image-domain, 2-parameter statistical characterization. It also honestly identifies conditions of modest expected benefit (overcast exterior, macro) where the simpler statistical characterization of prior methods captures the essential scene information adequately. The transfer function has been treated as a solved problem since the introduction of log encoding curves in the 2010s. The analysis in this paper suggests it is not. Scene-conditioned, intent-driven adaptive curve generation is mathematically well-founded, architecturally feasible for embedded ISP deployment, theoretically superior for the most demanding professional imaging conditions, and compatible with professional workflows through the proposed sidecar architecture. Empirical validation across the proposed scene taxonomy is identified as the essential next step.

FUTURE WORK

- Empirical validation of the fourteen theoretical scene predictions through controlled experiments on MIT-Adobe FiveK and RAISE datasets with dedicated ISP-stage linear RAW capture.
- Direct quantitative comparison with Rahman et al. (2016) and AGCWD (Huang et al., 2013) on shared benchmarks using identical linear RAW inputs, to empirically test the theoretical predictions of Section 5.
- Development of a continuous intent latent space conditioned on reference image examples, replacing the discrete three-class taxonomy.
- Extension of scene semantic features to include spectral channel information, addressing the underwater and filtered illumination limitation of Scene S13.
- Hardware co-design study for ψ as a dedicated neural IP block in a commercial camera SoC, characterizing power, area, and latency tradeoffs for embedded deployment.

REFERENCES

1. Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., & Shoyaib, M. (2016). An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(35). <https://doi.org/10.1186/s13640-016-0138-1>
2. Huang, S.-C., Cheng, F.-C., & Chiu, Y.-S. (2013). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing*, 22(3), 1032–1041.

3. Reinhard, E., Stark, M., Shirley, P., & Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3), 267–276.
4. Tumblin, J., & Rushmeier, H. (1993). Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6), 42–48.
5. Nakamura, J. (Ed.). (2005). *Image Sensors and Signal Processing for Digital Still Cameras*. CRC Press.
6. Schwartz, E., Giryes, R., & Bronstein, A. M. (2018). DeepISP: Toward learning an end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, 28(2), 912–923.
7. Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018). Learning to see in the dark. *IEEE/CVF CVPR*, pp. 3291–3300.
8. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, pp. 586–595.
9. SMPTE Standard ST 2084:2014. High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays.
10. ITU-R. (2018). Recommendation ITU-R BT.2100-2: Image parameter values for HDR television.
11. Bychkovsky, V., Paris, S., Chan, E., & Durand, F. (2011). Learning photographic global tonal adjustment. *CVPR*, pp. 97–104.
12. Dang-Nguyen, D. T., Pasquini, C., Conotter, V., & Boato, G. (2015). RAISE: A raw images dataset. *ACM Multimedia Systems Conference*, pp. 219–224.
13. Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2), 238–246.
14. Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *ICLR*.
15. Sony Electronics Inc. (2015). S-Gamut3.Cine/S-Log3 Conversion Table. Technical White Paper.
16. Mantiuk, R., Myszkowski, K., & Seidel, H. P. (2006). A perceptual framework for contrast processing of HDR images. *ACM TAP*, 3(3), 286–308.
17. Afifi, M., & Brown, M. S. (2020). Deep white-balance editing. *CVPR*, pp. 1397–1406.
18. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137–154.
19. Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (GELUs). *arXiv:1606.08415*.
20. Fattal, R., Lischinski, D., & Werman, M. (2002). Gradient domain HDR compression. *ACM TOG*, 21(3), 249–256.
21. Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2020). Image quality assessment: Unifying structure and texture similarity. *IEEE TPAMI*, 44(5), 2567–2581.
22. Cao, G., Huang, L., Tian, H., Huang, X., Wang, Y., & Zhi, R. (2017). Contrast enhancement of brightness-distorted images by improved adaptive gamma correction. *Computers and Electrical Engineering*, 66, 569–582.
23. Zamir, S. W., Arora, A., Khan, S., et al. (2021). Multi-stage progressive image restoration. *CVPR*, pp. 14821–14831.
24. Hubel, P. M., Liu, J., & Guttosch, R. J. (2004). Spatial frequency response of color image sensors. *SPIE*, 5301, 402–407.
25. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.