Experts-Guided Unbalanced Optimal Transport for ISP Learning from Unpaired and/or Paired Data

Georgy Perevozchikov[1], Nancy Mehta[1]*, Egor Ershov[2], and Radu Timofte[1]

[1] Computer Vision Lab, CAIDAS & IFI, University of Würzburg, John Skilton Str. 4a, 97074 Würzburg, Germany
{georgii.perevozchikov,nancy.mehta,radu.timofte}@uni-wuerzburg.de
[2] Color Reproduction and Synthesis Institute, AXXX, MIRAI
e.i.ershov@yandex.ru

Abstract. Learned Image Signal Processing (ISP) pipelines offer powerful end-to-end performance but are critically dependent on large-scale paired raw-to-sRGB datasets. This reliance on costly-to-acquire paired data remains a significant bottleneck. To address this challenge, we introduce a novel, unsupervised training framework based on Optimal Transport capable of training arbitrary ISP architectures in both unpaired and paired modes. We are the first to successfully apply Unbalanced Optimal Transport (UOT) for this complex, cross-domain translation task. Our UOT-based framework provides robustness to outliers in the target sRGB data, allowing it to discount atypical samples that would be prohibitively costly to map. A key component of our framework is a novel "committee of expert discriminators," a hybrid adversarial regularizer. This committee guides the optimal transport mapping by providing specialized, targeted gradients to correct specific ISP failure modes, including color fidelity, structural artifacts, and frequency-domain realism. To demonstrate the effectiveness of our approach, we retrained existing state-of-the-art ISP architectures using our paired and unpaired setups. Our experiments show that while our framework, when trained in paired mode, improves upon the original paired methods across all metrics, our unpaired mode concurrently achieves quantitative and qualitative performance that provides a competitive alternative to the original paired-trained counterparts. The code and pre-trained models are available at: `https://github.com/gosha20777/EGUOT-ISP.git`.

1 Introduction

Modern smartphone photography is a triumph of computational imaging, relying on a complex Image Signal Processor (ISP) [24, 34, 47] to overcome the physical limitations of small sensors and compact optics [39, 80]. The ISP converts noisy, mosaicked, linear, camera-specific raw data into the perceptually-pleasing sRGB images users expect [24, 32]. While traditional hardware ISPs use

* Corresponding Author

Fig. 1: A conceptual overview of our framework's robustness to outliers. **a)** Standard GANs (RawFormer [64]), when trained on unpaired datasets, are often sensitive to data outliers (e.g., overexposed photos). This leads to training instability and a collapsed output. **b)** Our proposed EGUOT framework, leveraging Experts Guided Unbalanced Optimal Transport, offers improved robustness against such data outliers and focuses on the high-quality core of the target distribution to produce perceptually better results.

fixed, hand-tuned modules [12], recent work has shifted to learned end-to-end pipelines [42,81,87]. These deep learning models offer superior performance and flexibility, bridging the quality gap between mobile phones and high-end DSLR cameras [40–43,67].

Despite their success, these state-of-the-art learned ISPs [37,40–43,67,79,87] are critically dependent on large-scale, paired training datasets [40,43,71]. This requires a difficult and costly data acquisition process: capturing thousands of scenes simultaneously with a smartphone (raw) and a high-end reference DSLR (sRGB), followed by meticulous alignment and manual photo finishing [28,40,43]. This dependency forms a significant bottleneck, as the entire data collection and alignment effort must be repeated for every new camera sensor [60,64].

The obvious solution is unpaired training to break this reliance on paired data. Recent attempts [6,61,64] have explored this using standard Generative Adversarial Network (GAN) frameworks [62,75,90], but these approaches often struggle with training stability and produce sub-optimal results [61]. We assume that a core reason for their failure in this task is the presence of distributional outliers (e.g., blurry, over/under-exposed images) within the raw-to-sRGB datasets [40,43,71]. Standard distribution-matching losses, including those based on classical Optimal Transport (OT) [31,46,52], are highly sensitive to these outliers [17], as model performance is degraded by attempting to learn from such suboptimal data.

To overcome these challenges, we introduce **EGUOT** (Experts-Guided Unbalanced Optimal Transport), a novel training framework robust to dataset outliers (Fig. 1) that operates in *paired* and *unpaired* modes. Our framework is the first to apply Unbalanced Optimal Transport (UOT) [65] for the raw-to-sRGB translation task. By relaxing classical OT constraints [46,52], our method offers improved robustness against outliers, discounting low-quality data to focus on the high-quality core of the target distribution. While UOT provides robustness [17,65], it is insufficient for ISP high-fidelity demands. We guide the transport mapping with a central component: a Committee of Expert Discriminators. Inspired by the perceptual orthogonality of the Human Visual System [78], this novel hybrid regularizer provides targeted, multi-modal gradients to correct specific ISP failure modes, such as semantic color inaccuracies, lack of structural detail, and unrealistic frequency artifacts.

We demonstrate the power and architecture-agnostic nature of our framework. We use our unsupervised method to retrain existing paired state-of-the-art supervised ISP backbones (e.g., MW-ISPNet [42], Restormer [84]). Experiments confirm our EGUOT framework's flexibility. When trained without paired data, EGUOT achieves a new state-of-the-art for unpaired translation, effectively outperforming prior work [6, 64]. Its performance also proves competitive with, and serves as a strong alternative to, the original fully-paired models. Additionally, in a paired-data setting, our method yields consistent performance enhancements, uniformly improving upon the original models' results across all metrics. Crucially, EGUOT acts purely as a training paradigm, introducing zero additional parameters or latency to the deployed ISP backbone.

Contribution

Our contributions can be summarized as follows:

- We propose a novel framework for raw-to-sRGB translation, trainable in both *unpaired* and *paired* modes. To our knowledge, we are the first to successfully apply Unbalanced Optimal Transport (UOT) [17] to this task, leveraging its robustness to dataset outliers, a common challenge in prior adversarial approaches.
- We introduce a hybrid regularization method, the *Committee of Expert Discriminators*, which guides the UOT mapping. This committee provides targeted, multi-modal perceptual gradients (color, structure, frequency), working in synergy with the UOT loss to ensure a high-fidelity image reconstruction rich in perceptual detail.
- We demonstrate that our *architecture-agnostic framework* can train existing SOTA paired ISP backbones. In *unpaired* mode, our models achieve performance competitive with, and act as an alternative to, their original, fully-paired results. In *paired* mode, our framework improves upon these original results on standard benchmarks.

2 Related Work

Our work is positioned at the intersection of three research domains: (1) learned ISP pipelines, (2) unpaired image-to-image translation, and (3) generative modeling with Optimal Transport.

2.1 Learned Image Signal Processing

The camera Image Signal Processor (ISP) [15, 24, 33, 34, 38, 44, 47, 81, 85, 87] converts noisy raw sensor data into sRGB images [24, 32]. Traditional hand-tuned hardware modules [1, 9, 12, 21, 34, 35, 45, 47, 72] are increasingly being replaced by end-to-end deep learning approaches [40–43, 59, 67, 81, 87, 89]. Foundational U-Net-based methods [43, 68, 69] were subsequently enhanced by Discrete Wavelet Transforms for detail preservation [23, 42], attention mechanisms [40, 41, 79, 87], and Transformers [10,26,33,36,67,84,89]. Concurrently, networks targeted specific ISP subproblems [3,5,13,20,21,70], such as cross-sensor raw-to-raw translation [2,64] and high-fidelity color mapping [4,11,14,22,25,48,54,55,61,82].

Despite their success, these state-of-the-art methods fundamentally rely on large-scale, precisely aligned raw-to-sRGB datasets [8,28,40,43,71]. Acquiring this paired data is notoriously difficult and costly for every new camera sensor, creating a major bottleneck [2,6,61,64]. This limitation explicitly motivates our work in developing a robust unpaired framework.

2.2 Unpaired Image-to-Image Translation

The most common approach for unpaired image-to-image translation is the Generative Adversarial Network (GAN) [62]. CycleGAN [90] introduced cycle-consistency, using a pair of generators and discriminators to enforce that a translated image can map back to its original domain. This inspired a wide range of methods based on shared latent spaces (e.g., DualGAN [83], UNIT [56], STARGAN [18], SEAN [91]), adaptive normalization (e.g., U-GAT-IT [49], UVC-GANv2 [75]), contrastive learning (e.g., CUT [63]), and recent [16,19,73,88].

Last years, this unpaired approach was explored for the raw-to-sRGB task [6], proposing the usage of a standard GAN [62] with a multi-term loss, including pre-trained networks (VGG [53], LPIPS [86]) for content preservation. While pioneering, this approach relies on standard adversarial losses [62], which can be sensitive to distributional outliers in real-world datasets [17,61]. We hypothesize that standard GAN losses are negatively impacted by these outliers, sometimes leading to reduced translation quality, thus motivating a more robust adversarial framework.

2.3 Optimal Transport in Generative Modeling

Optimal Transport (OT) [46] offers a powerful theoretically-grounded alternative to traditional GAN losses. While many methods use the OT cost as a loss function (e.g., Wasserstein GANs [7,50]), recent works have shown the OT plan itself can serve as a generative model [29–31,51,52,58,66,76].

However, a well-known limitation of classical OT is its high sensitivity to outliers [17,52,65]. The standard OT problem requires transporting every point in the source distribution to the target distribution. This means a few outlier samples can unduly influence the data [17], drastically altering the transport map and degrading the final result.

To solve this, the theory of *Unbalanced Optimal Transport (UOT)* was proposed [65], relaxing the hard marginal constraints of classical OT [46,52]. This relaxation permits partial transport, which, in practical terms, means the model can learn to *ignore or discard outliers* rather than being forced to match them.

While the UOT theory was powerful, a stable, practical algorithm was needed. The UOTM paper [17] introduced a novel, stable training framework based on the semi-dual formulation of UOT [65]. The authors demonstrated that UOTM is significantly more robust to outliers and converges faster than classical OT models [46,52].

Our work aims to connect these fields. To our knowledge, we are the first to apply an Experts-Guided UOT framework to the paired and unpaired ISP

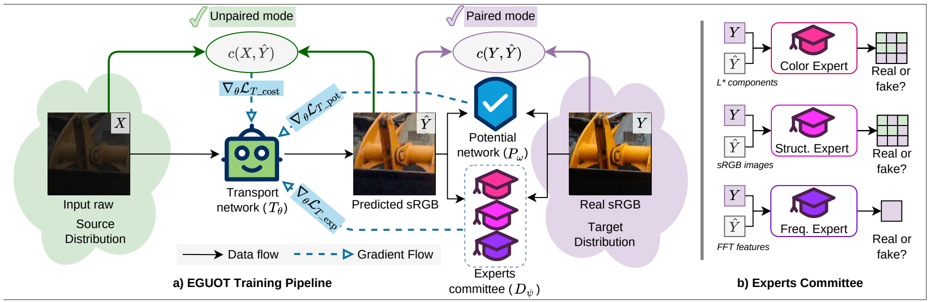

Fig. 2: An overview of our EGUOT training framework. (a) *The main pipeline*, where the generator (T_θ) is trained by three parallel objectives: (1) a content-preserving cost function ($c(\cdot,\cdot)$) for paired and unpaired mode, (2) a robust Potential network (P_ω), and (3) an Experts committee (D_ψ). (b) *The detailed breakdown of the Experts committee*, showing its three specialized discriminators for *Color*, *Structure*, and *Frequency*, each analyzing different features of the real (Y) and predicted ($\hat{Y}$) sRGB images.

problem. Thus, we leverage the robustness of UOT [17] to address a task where prior adversarial methods often experience difficulties due to outlier sensitivity [17].

3 Method

Our goal is to train a raw-to-sRGB mapping network T_θ for *paired* and *unpaired* settings, which we achieve with EGUOT, a flexible training framework (Fig. 2). The framework consists of a generator T_θ trained by a novel hybrid loss system. In unpaired mode, this system uses a cost function $c(x, T_\theta(x))$ for robust content mapping and a Committee of Expert Discriminators D_ψ for perceptual fidelity. In paired mode, the system has a direct pixel loss $c(y, T_\theta(x))$, retaining the expert committee.

3.1 Preliminaries: UOT for Image Translation

Classical OT [46] seeks a transport plan π minimizing the cost $C(\mathbb{X}, \mathbb{Y})$ under strict marginal constraints: all mass from the source $\mathbb{X}$ must be transported exactly to the target $\mathbb{Y}$. This strictness makes standard OT highly sensitive to data outliers.

Unbalanced Optimal Transport (UOT) To overcome this, UOT [65] relaxes these strict marginals using soft penalties based on Csiszár-divergences D_Φ:

$$C(\mathbb{X}, \mathbb{Y}) = \inf_{\pi \in \mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) + D_{\Phi_1}(\pi_0|\mathbb{X}) + D_{\Phi_2}(\pi_1|\mathbb{Y}) \right],$$

where $\mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ denotes positive joint measures with marginals π_0, π_1, and $c(x, y)$ is the transport cost.

For deep learning, UOTM [17] utilizes the semi-dual formulation:

$$C(\mathbb{X}, \mathbb{Y}) = \sup_{p \in \mathcal{C}(\mathcal{Y})} \left[\int_{\mathcal{X}} -\Phi_1^*(-p^c(x))d\mathbb{X}(x) + \int_{\mathcal{Y}} -\Phi_2^*(-p(y))d\mathbb{Y}(y) \right],$$

where p is the dual Kantorovich potential, $p^c(x) = \inf_{y\in\mathcal{Y}}[c(x,y) - p(y)]$ is its c-transform, and Φ_i^* are the convex conjugates of Φ_i.

Generative Modeling and Outlier Rejection Following UOTM [17], we parameterize the potential p with a neural network P_ω and approximate the intractable c-transform using a transport network T_θ, such that $p^c(x) \approx c(x, T_\theta(x)) - P_\omega(T_\theta(x))$.

Crucially, to achieve outlier robustness, we select the Kullback-Leibler (KL) divergence for D_Φ, meaning its convex conjugate takes an explicit exponential form: $\Phi^*(t) = e^t - 1$. Substituting this into the semi-dual yields our empirical minimax objectives over mini-batches.

The potential network P_ω, augmented with an R_1 gradient penalty γ for stability, minimizes:

$$\begin{aligned}\mathcal{L}_P =&\mathbb{E}_{x\sim\mathbb{X}}[\exp(-c(x,T_\theta(x)) + P_\omega(T_\theta(x)))] \\ &+ \mathbb{E}_{y\sim\mathbb{Y}}[\exp(-P_\omega(y))] + \frac{\gamma}{2}\mathbb{E}_{y\sim\mathbb{Y}}[\|\nabla_y P_\omega(y)\|_2^2].\end{aligned} \tag{1}$$

The transport network T_θ is updated to minimize the cost and maximize the potential:

$$\mathcal{L}_{T\_\text{pot}} = \mathbb{E}_{x\sim\mathbb{X}}[c(x, T_\theta(x)) - P_\omega(T_\theta(x))]. \tag{2}$$

Mechanism of Outlier Rejection: The explicit exponential formulation in $\mathcal{L}_P$ is the core driver of EGUOT's robustness. If a source image x is highly corrupted (an outlier), its mapping cost $c(x, T_\theta(x))$ becomes abnormally large. Consequently, the gradient scaling factor $\exp(\cdots - c(\dots))$ exponentially decays toward zero. This mathematically zeroes out the gradients for outliers, dynamically discarding their mass during training rather than forcing a suboptimal mapping *(see Supp. for more details)*.

UOTM [17] application to the raw-to-sRGB task requires two considerations: (1) a valid cost function $c(\cdot,\cdot)$ must be defined for the paired/unpaired case, as raw and sRGB domains are incompatible; and (2) the basic transport loss (Eq. 2) is insufficient alone for the high-fidelity translation quality. Our EGUOT framework addresses both.

3.2 The EGUOT Framework

We propose the Experts-Guided UOT framework to adapt the continuous UOT optimization [17] to the ISP domain. First, we define domain-specific cost functions to enable Eq. (1) and (2). Second, we introduce a committee of expert discriminators as a hybrid regularizer to guarantee perceptual fidelity.

Cost Functions We define the transport cost $c(\cdot,\cdot)$ for unpaired and paired training.

Unpaired Setting: Raw (4-channel) and sRGB (3-channel) images occupy incompatible domains, precluding direct optimal transport cost computation. To establish a valid metric space, a heuristic *dimension projector* maps the raw image x to a crude 3-channel representation $x_{proj} \in \mathbb{X}_{\text{rgb}}$ by averaging its green channels *(see Supp. for more details)*. The unpaired cost is defined as

$c(x, T_\theta(x)) = \tau \|T_\theta(x) - x_{proj}\|_2^2$, functioning as a cross-domain *identity/content loss* standard in modern unpaired translation [64, 74]. Crucially, this satisfies the fundamental OT requirement to anchor the transport map and prevent geometric degeneracy [46, 52]. Consequently, x_{proj} serves as a coarse spatial/content anchor, leaving the complex, high-fidelity reconstruction of colors and textures entirely to the target distribution gradients guided by P_ω and D_ψ.

Paired Setting: Given strictly aligned ground-truth sRGB targets y, we utilize the L_1 norm as the transport cost: $c(x, y) = \|T_\theta(x) - y\|_1$. Crucially, UOT remains highly beneficial even with explicit pairs: its "mass destruction" dynamically discounts misaligned or corrupted outlier pairs. This implicitly cleans the supervisory signal during training, enabling consistent performance improvements for existing ISP backbones.

Experts-Guided Hybrid Loss While UOTM robustly acts as a *global mass router* to discard outliers, is insufficient to enforce the high-quality raw-to-sRGB mapping. To resolve this, we introduce a committee of expert discriminators $D_\psi = \{D_i\}_{i=1}^M$ acting as *local perceptual regularizers* to enforce color and structural accuracy. Crucially, optimizing these experts via a bounded Hinge loss preserves UOT's theoretical robustness: the exponentially formulated UOT cost easily dominates the bounded adversarial gradients on severe outliers, safely discarding corrupted mass.

The expert discriminators D_ψ are trained to distinguish real sRGB images from fake ones by minimizing $\mathcal{L}_D$:

$$\mathcal{L}_D = \sum_{D_i \in D_\psi} \left(\mathbb{E}_{y \sim \mathbb{Y}}[h(-D_i(y))] + \mathbb{E}_{x \sim \mathbb{X}}[h(D_i(T_\theta(x)))] \right) \tag{3}$$

The transport T_θ is then simultaneously trained to fool this committee by minimizing the expert transport loss $\mathcal{L}_{T\_\mathrm{exp}}$:

$$\mathcal{L}_{T\_\mathrm{exp}} = \sum_{D_i \in D_\psi} \mathbb{E}_{x \sim \mathbb{X}}[-D_i(T_\theta(x))] \tag{4}$$

The final, total loss for our transport is:

$$\mathcal{L}_{T\_\mathrm{total}} = \mathcal{L}_{T\_\mathrm{pot}} + \lambda \cdot \mathcal{L}_{T\_\mathrm{exp}} \tag{5}$$

3.3 Network Architectures

We selected state-of-the-art components for our architecture-agnostic framework. Crucially, both the Potential Network P_ω and the Experts Committee D_ψ are used exclusively during training *(see Supp. for more details)*.

Transport Network (T_θ): A learnable ISP pipeline for the complete raw-to-sRGB mapping. As our framework is architecture-agnostic, we adopt multiple SOTA ISP backbones (e.g., [40, 61]) to serve as T_θ in our experiments.

Potential Network (P_ω): This network serves as the potential function [52]. We use a ConvNeXt [57] backbone, whose features are passed through an MLP head to output a single scalar.

Experts Committee (D_ψ): Functioning as a fixed multi-discriminator pattern [27, 77], the committee is comprised of three specialist networks:

Algorithm 1 ISP Training with EGUOT

Input: Source $\mathbb{X}$ and target $\mathbb{Y}$ distributions; transport T_θ, potential P_ω, experts D_ψ; cost function c; inner iterations N; hyperparameters γ, λ.
Output: Learned transport network T_θ.

repeat
  for $k = 1, \ldots, N$ **do** ▷ Potential and Experts update loop
    Sample mini-batches $X \sim \mathbb{X}$, $Y \sim \mathbb{Y}$;
    Update ω using $\nabla_\omega \mathcal{L}_P$ (Eq. 1);
    Update ψ using $\nabla_\psi \mathcal{L}_D$ (Eq. 3);
  end for
  Sample mini-batch $X \sim \mathbb{X}$; ▷ Transport network update
  Update θ using $\nabla_\theta \mathcal{L}_{T\_\text{total}}$ (Eq. 5);
until *converged*

- **Color Expert:** A frozen ConvNeXt [57] encoder from a pre-trained colorization U-Net [68] that judges semantic color plausibility in the L* channel.
- **Structure Expert:** A multi-scale PatchGAN discriminator [64] operating on the sRGB image to enforce local texture fidelity.
- **Frequency Expert:** A lightweight CNN operating on the 2D log-magnitude FFT spectrum of the grayscale image to penalize high-frequency artifacts.

Design Rationale and Efficiency: Motivated by the perceptual orthogonality of the Human Visual System [78], the decoupled experts evaluate color, structure, and frequency independently. They function as local perceptual regularizers that complement UOT's global mass routing. Training overhead is constrained since the Color Expert is frozen and the Frequency CNN is lightweight. Because the entire committee is discarded post-training, it introduces zero inference latency.

3.4 Training Algorithm

We train our framework with an alternating optimization scheme. We use a discriminator-heavy $N : 1$ training ratio: both the Potential Network P_ω and Experts Committee D_ψ are updated N steps for every single Transport Network T_θ update. The complete EGUOT training procedure is in Algorithm 1.

4 Experiments

To validate our EGUOT framework, we conduct a comprehensive set of experiments. We demonstrate that (1) our framework can train existing supervised ISP backbones in paired and unpaired modes to improve upon (in the paired case) or serve as a strong alternative to (in the unpaired case) their original settings; (2) our framework sets a new state-of-the-art for unpaired raw-to-sRGB translation; (3) the core components of our method—the UOT objective and the expert committee—are both essential for its success *(see Supp. for more details).*

4.1 Datasets

To validate real-world generalization and in-the-wild applicability, we evaluate our framework on three diverse benchmarks: the Zurich raw-to-sRGB (ZRR)

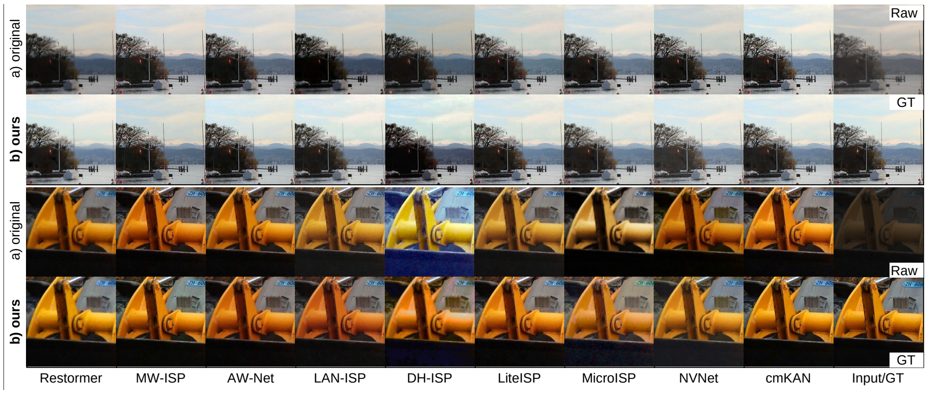

Fig. 3: Our framework's unpaired mode achieves the paired performance across diverse architectures on the Zurich raw-to-sRGB dataset [43]. The top row (a) shows the results of SOTA ISP backbones (e.g., Restormer [64, 84], cmKAN [61]) trained with their *original paired* settings. The bottom row (b) — results of the same backbones trained in *unpaired mode*, highlighting the effectiveness of our method. Best viewed in the electronic version.

dataset [43] with 48K patch pairs (Huawei P20 and Canon 5D Mark IV) which contain distributional outliers such as misalignments and exposure errors [23]; the ISPIW dataset [71] featuring unconstrained conditions with 5.7K patches (Huawei Mate 30 Pro and Canon 5D); and the Mobile AI dataset [42] providing 95K pairs (Sony IMX586 and Fujifilm GFX100 DSLR). For all datasets, we strictly follow their official splits (e.g., 70%/20%/10%) for training, validation, and testing.

4.2 Training Details

All models were trained on a single RTX 4090 for 300 epochs (applying SWA in the last 10) using Adam ($\beta_1 = 0.5, \beta_2 = 0.999$) and a cosine schedule. Initial learning rates were 10^{-4} for T_θ and $2{\cdot}10^{-4}$ for P_ω, D_ψ. Demonstrating EGUOT's robustness, a single set of objective hyperparameters ($\gamma = 1.5, \tau = 10^{-3}, \lambda = 1.0, N = 2$) was fixed across all datasets and architectures (*see Supp. Mat.*). For unpaired tasks, source and target datasets were shuffled independently. Given aligned test pairs, we report full-reference metrics (PSNR, SSIM, ΔE) rather than distribution approximations.

4.3 Results

Our primary results demonstrate EGUOT's performance and flexibility. First, EGUOT can train diverse SOTA ISP architectures in an unpaired manner, achieving competitive performance, while our paired mode achieves superior performance. Finally, our framework sets a new state-of-the-art for unpaired raw-to-sRGB translation. *See supp. materials for additional results.*

ZRR Results We begin our analysis with the ZRR dataset. Table 1 and Figure 3 show the results of applying our framework (paired and unpaired) to nine

Table 1: Results on the Zurich raw-to-sRGB dataset [43], comparing SOTA backbones retrained with our EGUOT framework against their *original paired-data results*. Our *paired* mode consistently outperforms the original, while our *unpaired* mode achieves competitive performance. Better results are highlighted in yellow.

Backbone	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	#Params	FLOPs	Architecture Type
	Paired Mode (original)				Paired Mode (ours)				Unpaired Mode (ours)						
Restormer [84]	21.51	0.83	9.71	0.19	21.73	0.84	9.42	0.18	21.69	0.84	9.93	0.20	5.03M	583G	U-Net-like
MW-ISP [42]	21.88	0.82	10.33	0.21	22.15	0.84	9.61	0.20	21.62	0.83	9.97	0.20	29.22M	3.6T	U-Net-like
AW-Net [23]	21.58	0.75	11.02	0.19	22.02	0.77	9.98	0.18	21.71	0.76	10.17	0.19	54.83M	3.1T	U-Net-like
LAN-ISP [79]	20.76	0.81	10.41	0.22	21.19	0.81	9.36	0.21	20.50	0.80	9.88	0.22	49.2K	56.9G	U-Net-like
LiteISP [87]	22.18	0.83	10.28	0.19	22.29	0.84	9.79	0.19	21.43	0.83	10.04	0.21	9.04M	174G	U-Net-like
DH-ISP [37]	19.68	0.72	11.46	0.26	19.88	0.74	10.99	0.24	19.69	0.74	11.25	0.25	3.1K	31.7G	Forward CNN
MicroISP [41]	20.30	0.78	11.14	0.24	20.61	0.79	11.08	0.24	19.81	0.79	11.65	0.24	105K	37G	Forward CNN
NVNet [40]	20.95	0.80	10.96	0.22	21.13	0.80	10.05	0.21	20.15	0.78	10.31	0.23	3.5K	31.8G	ResNet-like
cmKAN [61]	24.41	0.85	7.27	0.17	24.63	0.86	7.10	0.17	24.46	0.84	8.31	0.18	76.4K	40G	Hyper-network

SOTA ISP backbones. The results are compelling: (1) *our paired* mode consistently surpasses the original paired-trained models across all metrics and architectures, showing the advantage of our expert-guided regularization; (2) *our unpaired* mode consistently rivals its paired counterparts. Notably, our unpaired EGUOT-trained models frequently *match or exceed* the original results in several key metrics, such as SSIM for Restormer [84] and MW-ISP [42], and ΔE for LAN-ISP [79] and AW-Net [23]. This demonstrates our framework is highly architecture-agnostic and serves as an effective alternative to paired data.

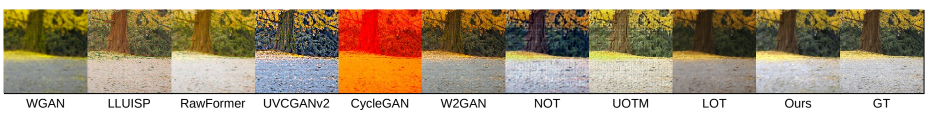

Fig. 4: Qualitative results of *unpaired frameworks* on the Zurich raw-to-sRGB dataset [43]. We compare against modern GAN-based (e.g., RawFormer [64], UVCGANv2 [73, 75]), other OT-based (e.g., NOT [52], UOTM [17]), and recent unpaired ISP methods (LLUISP [6]). Our framework sets a *new state-of-the-art performance across all metrics*. Best viewed in the electronic version.

Next, we establish a new state-of-the-art for unpaired translation on this dataset. Table 2 compares unpaired training frameworks using a fixed Restormer backbone. As shown in Table 2, EGUOT outperforms standard GANs [7, 90], recent specialized methods [6, 64], and pure OT baselines [17, 52], demonstrating our novel experts-guided hybrid loss is critical for achieving high-fidelity results. Crucially, this quality comes without computational bloat: by avoiding dual-generator cycle-consistency, EGUOT's single-pass design maintains a competitive training speed (2.48 it/s for the full optimization cycle, Algorithm 1), with all speeds measured under identical hardware and batch conditions. Figure 4 confirms our superior color and texture fidelity.

ISPIW Results We confirm these findings with the ISPIW dataset. As shown in Table 3 and Figure 5, our unpaired training again achieves performance that is statistically on-par with the original paired training for nearly all architectures.

Table 2: Unpaired translation results on the Zurich raw-to-sRGB dataset [43] using a fixed Restormer backbone. Unlike dual-generator methods (e.g., CycleGAN [90], RawFormer [64]) that inflate computational costs, EGUOT achieves state-of-the-art performance with competitive training speed. Speed and FLOPs are measured on an RTX 4090 (batch 8, $4 \times 224 \times 224$ patches).

Method	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	Train FLOPs / Iter ↓	Train Speed (it/s) ↑	Type
CycleGAN [90]	7.73	0.32	27.8	0.82	2.35T	1.72	GAN-based
UVCGANv2 [75]	16.25	0.61	12.2	0.49	2.38T	1.61	GAN-based
RawFormer [64]	17.09	0.68	11.9	0.45	2.39T	1.58	GAN-based
WGAN [7]	14.13	0.52	14.3	0.67	0.60T	2.71	GAN-based
W2GAN [50]	15.62	0.58	13.9	0.60	0.60T	2.68	GAN-based
LLUISP [6]	19.07	0.72	11.7	0.32	0.72T	1.95	GAN-based
NOT [52]	16.03	0.60	12.6	0.54	0.60T	2.77	OT-based
UOTM [17]	18.96	0.71	11.0	0.31	0.61T	2.62	OT-based
LOT [66]	17.77	0.68	12.2	0.39	0.61T	2.12	OT-based
Ours	21.69	0.84	9.93	0.20	0.64T	2.48	OT-based

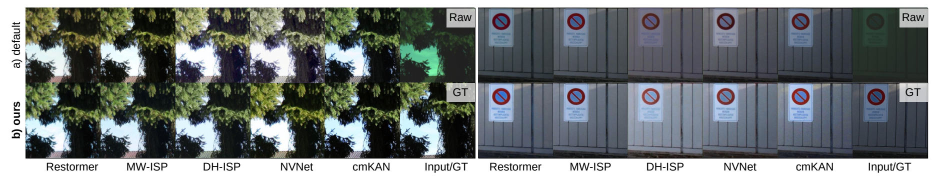

Fig. 5: Qualitative results on the ISPIW raw-to-sRGB dataset [71], demonstrating the architecture-agnostic nature of our EGUOT framework. We apply (a) the original, *paired* training and (b) our proposed *unpaired* training to a wide range of SOTA ISP backbones. Our unpaired mode achieves performance that is highly competitive with the original settings. Best viewed in the electronic version.

For several models (e.g., AW-Net [23], MicroISP [41]), our unpaired method even achieves a higher PSNR or lower ΔE than the paired-data baseline, further highlighting the framework's effectiveness.

Mobile AI Results Finally, we validate our approach on the Mobile AI dataset [40] in Table 4 and Figure 6. The results are consistent with our findings on the other two datasets. Our unpaired mode remains highly competitive with paired settings, and in the case of MW-ISP [42], even outperforms the original model in both PSNR (24.44 vs 24.31) and ΔE (6.09 vs 6.27).

Deployment and Trade-offs While the $N : 1$ update for auxiliary networks (P_ω, D_ψ) adds offline training complexity, lightweight experts keep overhead manageable. At inference, these modules are discarded. Consequently, EGUOT acts as an architecture-agnostic training paradigm: any deployed generator (T_θ) operates with zero added parameters or latency, providing a practical, plug-and-play solution for edge devices.

4.4 Ablation Studies

We conduct a series of ablation studies on the ZRR dataset [43] using the Restormer [84] backbone to validate our core design choices.

Table 3: Results on the ISPIW raw-to-sRGB dataset [71], comparing SOTA backbones retrained with our framework against their *original paired-data results*. Our *unpaired* mode achieves competitive performance. Better results are highlighted in yellow.

Backbone	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓
	Paired Mode (original)				Unpaired Mode (ours)			
Restormer [84]	20.93	0.79	6.85	0.14	20.81	0.79	6.99	0.14
MW-ISP [42]	21.90	0.81	7.03	0.11	21.81	0.81	6.99	0.11
AW-Net [23]	21.75	0.81	6.99	0.12	21.93	0.81	7.30	0.11
LAN-ISP [79]	22.09	0.81	6.97	0.11	21.95	0.81	6.42	0.11
LiteISP [87]	22.14	0.81	6.31	0.11	22.01	0.80	6.88	0.12
DH-ISP [37]	19.92	0.76	7.89	0.17	19.59	0.77	7.40	0.17
MicroISP [41]	20.70	0.77	6.92	0.15	20.79	0.77	6.33	0.15
NVNet [40]	23.80	0.82	5.81	0.10	23.39	0.82	6.08	0.10
cmKAN [61]	24.22	0.83	5.29	0.09	24.21	0.83	5.73	0.09

Table 4: Results on the Mobile AI raw-to-sRGB dataset [40], comparing SOTA backbones retrained with our framework against their *original paired-data results*. Our *unpaired* mode achieves competitive performance. Better results are highlighted in yellow.

Backbone	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	ΔE ↓	LPIPS ↓
	Paired Mode (original)				Unpaired Mode (ours)			
Restormer [84]	23.99	0.88	6.74	0.13	22.51	0.86	7.27	0.14
MW-ISP [42]	24.31	0.88	6.27	0.11	24.44	0.87	6.09	0.11
AW-Net [23]	24.17	0.87	6.35	0.12	24.06	0.87	6.33	0.12
LAN-ISP [79]	23.48	0.87	7.18	0.15	23.15	0.86	7.22	0.14
LiteISP [87]	24.05	0.86	6.43	0.13	24.01	0.86	6.58	0.13
DH-ISP [37]	23.20	0.84	8.52	0.17	22.98	0.84	8.49	0.17
MicroISP [41]	23.87	0.85	7.04	0.16	23.79	0.85	7.19	0.16
NVNet [40]	24.09	0.85	6.19	0.13	23.87	0.84	7.03	0.13
cmKAN [61]	24.51	0.88	5.31	0.10	24.25	0.88	6.07	0.10

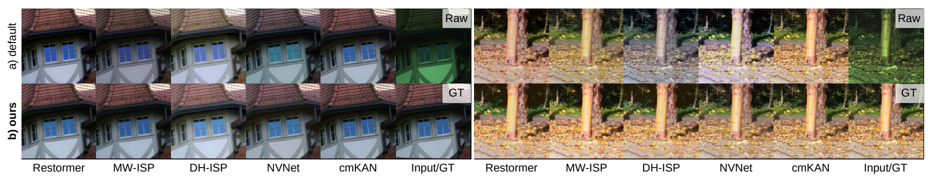

Fig. 6: Qualitative results on the Mobile AI raw-to-sRGB dataset [40], demonstrating the architecture-agnostic nature of our EGUOT framework. We apply (a) the original, *paired* training and (b) our proposed *unpaired* training to a wide range of SOTA ISP backbones. Our unpaired mode achieves performance that is highly competitive with the original settings. Best viewed in the electronic version.

Framework Ablation. We evaluate our hybrid EGUOT framework in Table 5 across four configurations. The standard GAN (C_1) and UOT-Only (C_2) baselines yield sub-optimal fidelity and severe color shifts. Crucially, setting $\tau = 0$ (C_3) leads to mode collapse. Without the cost $c(\cdot, \cdot)$ to penalize mass displacement, the semi-dual objective reduces to maximizing the potential P_ω. This confirms that $c(\cdot, \cdot)$ is essential to anchor the UOT mapping (Sec. 3.2). Finally, our full model (C_4) significantly outperforms all baselines, demonstrating the synergy between the UOT objective and the expert committee.

Robustness to Data Outliers. We corrupt 15% of the ZRR target images [43] with severe color and brightness jitter (Table 6). Standard adversarial approaches (C_1) are forced to fit this entire noisy distribution, leading to a significant 0.58 dB PSNR drop. In contrast, EGUOT (C_4) loses merely 0.28 dB. Its exponential UOT objective dynamically assigns near-zero mass to corrupted samples. By mathematically discounting these outliers, EGUOT guides the expert committee to learn exclusively from the pristine sRGB manifold, confirming its enhanced stability in imperfect real-world environments.

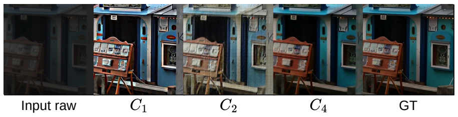

Fig. 7: Visual ablation on the ZRR dataset [43]. Our full method (C_4) produces the best results. The standard GAN (C_1) suffers from artifacts and color issues. The UOT-Only baseline (C_2) fails to reconstruct fine textures. The $\tau = 0$ baseline (C_3) is omitted as the lack of a spatial anchor leads to geometric degeneracy and mode collapse.

Table 5: Framework ablation on the ZRR dataset [43] (Restormer [84] backbone). EGUOT (C_4) significantly outperforms standard GANs (C_1) and pure UOT (C_2). C_3 confirms that a non-zero structural anchor ($\tau > 0$) is strictly required to prevent mode collapse.

Config	Adversarial Obj.	Expert Committee	PSNR ↑	SSIM ↑	ΔE ↓
C_1	Hinge Loss	✓(All)	19.23	0.73	10.8
C_2	UOT-Loss	—	19.15	0.72	12.2
C_3	UOT-Loss (τ=0)	✓(All)	Fails (Mode Collapse)		
C_4 (**Ours**)	UOT-Loss	✓(All)	21.69	0.84	9.93

Table 6: Robustness to 15% data corruption on the ZRR dataset [43]. EGUOT (C_4) remains highly stable, while the standard GAN (C_1) degrades significantly.

Config	Clean ZRR			Dirty ZRR (15% Outliers)		
	PSNR ↑	SSIM ↑	ΔE ↓	PSNR ↑	SSIM ↑	ΔE ↓
C_1 (GAN)	19.23	0.73	10.80	18.65	0.72	11.17
C_4 (**Ours**)	**21.69**	**0.84**	**9.93**	**21.41**	**0.84**	**10.09**

Table 7: Expert committee ablation on ZRR [43] using Restormer [84]. The baseline (C_2) lacks experts, while our full model (C_4) uses all three. Removing any expert—especially Color (E_1)—significantly drops performance.

Config	Color Exp.	Structure Exp.	Frequency Exp.	PSNR ↑	SSIM ↑	ΔE ↓
C_2 (UOT-Only)	—	—	—	19.15	0.72	12.2
E_1 (w/o Color)	—	✓	✓	19.88	0.75	14.8
E_2 (w/o Structure)	✓	—	✓	20.16	0.79	10.4
E_3 (w/o Frequency)	✓	✓	—	20.95	0.82	9.91
C_4 (**Ours**)	✓	✓	✓	21.69	0.84	9.93

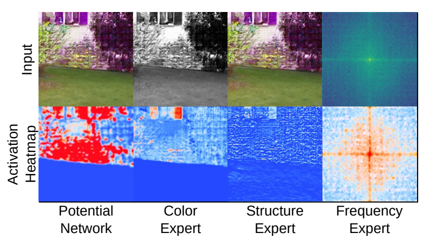

Fig. 8: Activation heatmaps (red=fake, blue=real) showing our potential and the discriminator committee analyzing a fake sRGB image ($\hat{Y}$). Each network learns to spot specific errors: *Potential Network* identifies global implausibility; *Color Expert* highlights semantically incorrect colors; *Structure Expert* detects bad local textures; *Frequency Expert* uses the FFT spectrum to find unnatural high-frequency artifacts (spikes).

Expert Committee Ablation In Table 7, we analyze the contribution of each expert discriminator. We start from the C_2 (UOT-Only) baseline and also show the results of removing one expert at a time from our full model (C_4). The results show that all three experts contribute to the final performance. The UOT-Only model (C_2) is the weakest. Removing any single expert degrades performance, with the Color Expert being the most critical (E_1); its removal causes the ΔE to spike from 9.93 to 14.8, confirming its vital role in maintaining color fidelity. Figure 8 provides a qualitative visualization of this specialization, showing how each component learns to identify distinct error types.

5 Conclusion

In this work, we addressed the critical bottleneck in modern ISP development: the reliance on large-scale, paired raw-sRGB datasets. We identified that prior unpaired methods often experience difficulties due to their sensitivity to the distributional outliers found in real-world target sRGB datasets.

To solve this, we proposed **EGUOT**, a novel and robust framework for raw-to-sRGB translation, designed to operate in both paired and unpaired modes. To our knowledge, our framework is the first to apply Unbalanced Optimal Transport (UOT) for the ISP task, leveraging its enhanced robustness to dataset outliers. To achieve the high fidelity required of a modern ISP, we introduced a Committee of Expert Discriminators. This hybrid regularizer guides the mapping by providing targeted perceptual gradients for color, structure, and frequency, and is a key component in both training modes.

Our comprehensive experiments demonstrated the effectiveness and flexibility of our approach. We showed that EGUOT is architecture-agnostic, capable of training numerous state-of-the-art ISPs in a fully unpaired manner. Our paired training mode consistently improves upon the performance of the original settings across all metrics. Concurrently, our unpaired results are highly competitive with, and serve as a strong alternative to, their original fully-paired counterparts. Furthermore, our unpaired mode achieves state-of-the-art performance for unpaired raw-to-sRGB translation, outperforming prior GAN and OT-based methods. Our ablation studies empirically confirmed our core hypotheses: the UOT objective is essential for outlier robustness, while the expert committee is critical for perceptual fidelity.

We believe this approach will make ISP development more efficient by allowing developers to focus on the desired distribution properties instead of spending extensive time on data collection.

Limitations and Future Work While our expert committee is effective, it adds complexity compared to a single-discriminator framework. Future work could explore methods to prune this committee. Moreover, the demonstrated effectiveness of EGUOT opens promising avenues for applying this framework to other challenging unpaired image restoration tasks, such as low-light enhancement, denoising, and HDR reconstruction, where data outliers are common.

Acknowledgments

This work was partly supported by The Alexander von Humboldt Foundation and the Ministry of Science and Higher Education of the Russian Federation (agreement No. FFNU-2025-0045). The authors also express sincere gratitude to Alexander Korotin for his valuable advice on the theoretical background of Optimal Transport.

References

1. A Sharif, S., Naqvi, R.A., Biswas, M.: Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor. In: CVPR (2021)
2. Afifi, M., Abuolaim, A.: Semi-supervised raw-to-raw mapping. In: BMVC (2021)
3. Afifi, M., Brown, M.S.: Deep white-balance editing. In: CVPR (2020)
4. Afifi, M., Brubaker, M.A., Brown, M.S.: Histogan: Controlling colors of gan-generated and real images via color histograms. In: CVPR (2021)
5. Afifi, M., Derpanis, K.G., Ommer, B., Brown, M.S.: Learning multi-scale photo exposure correction. In: CVPR (2021)
6. Arhire, A., Timofte, R.: Learned lightweight smartphone isp with unpaired data. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 1878–1887 (2025)
7. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International conference on machine learning. pp. 214–223. PMLR (2017)
8. Banić, N., Ershov, E., Panshin, A., Karasev, O., Korchagin, S., Lev, S., Startsev, A., Vladimirov, D., Zaychenkova, E., Iarchuk, D.R., et al.: Ntire 2024 challenge on night photography rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2024)
9. Barron, J.T., Tsai, Y.T.: Fast Fourier color constancy. In: CVPR (2017)
10. Brateanu, A., Balmez, R., Avram, A., Orhei, C., Ancuti, C.: Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement. IEEE Signal Processing Letters (2025)
11. Brateanu, A., Balmez, R., Orhei, C., Ancuti, C., Ancuti, C.: Enhancing low-light images with kolmogorov–arnold networks in transformer attention. Sensors **25**(2), 327 (2025)
12. Brown, M.: Color processing for digital cameras. Fundamentals and Applications of Colour Engineering pp. 81–98 (2023)
13. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage Retinex-based transformer for low-light image enhancement. In: ICCV (2023)
14. Chang, H., Fried, O., Liu, Y., DiVerdi, S., Finkelstein, A.: Palette-based photo recoloring. ACM Transactions on Graphics **34**(4), 139–1 (2015)
15. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018)
16. Cheng, Y., Gan, Z., Li, Y., Liu, J., Gao, J.: Sequential attention gan for interactive image editing. In: Proceedings of the 28th ACM international conference on multimedia. pp. 4383–4391 (2020)
17. Choi, J., Choi, J., Kang, M.: Generative modeling through the semi-dual formulation of unbalanced optimal transport. Advances in Neural Information Processing Systems **36**, 42433–42455 (2023)

18. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: CVPR (2018)
19. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: Stargan v2: Diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8188–8197 (2020)
20. Conde, M.V., Geigle, G., Timofte, R.: Instructir: High-quality image restoration following human instructions. In: European Conference on Computer Vision. pp. 1–21. Springer (2024)
21. Conde, M.V., Vasluianu, F., Vazquez-Corral, J., Timofte, R.: Perceptual image enhancement for smartphone real-time applications. In: CVPR (2023)
22. Conde, M.V., Vazquez-Corral, J., Brown, M.S., Timofte, R.: NILUT: Conditional neural implicit 3D lookup tables for image enhancement. In: AAAI (2024)
23. Dai, L., Liu, X., Li, C., Chen, J.: AWNet: Attentive wavelet network for image ISP. In: ECCV (2020)
24. Delbracio, M., Kelly, D., Brown, M.S., Milanfar, P.: Mobile computational photography: A tour. Annual review of vision science **7**(1), 571–604 (2021)
25. Ding, Z., Li, P., Yang, Q., Li, S., Gong, Q.: Regional style and color transfer. In: CVIDL (2024)
26. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
27. Durugkar, I., Gemp, I., Mahadevan, S.: Generative multi-adversarial networks. arXiv preprint arXiv:1611.01673 (2016)
28. Ershov, E., Savchik, A., Semenkov, I., Banić, N., Belokopytov, A., Senshina, D., Koščević, K., Subašić, M., Lončarić, S.: The cube++ illumination estimation dataset. IEEE access **8**, 227511–227527 (2020)
29. Gazdieva, M., Mokrov, P., Rout, L., Korotin, A., Kravchenko, A., Filippov, A., Burnaev, E.: An optimal transport perspective on unpaired image super-resolution. Journal of Optimization Theory and Applications **207**(2), 40 (2025)
30. Geuter, J., Kornhardt, G., Tomasson, I., Laschos, V.: Universal neural optimal transport. In: Forty-second International Conference on Machine Learning (2025)
31. Gushchin, N., Kolesov, A., Korotin, A., Vetrov, D.P., Burnaev, E.: Entropic neural optimal transport via diffusion processes. Advances in Neural Information Processing Systems **36**, 75517–75544 (2023)
32. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics **35**(6), 1–12 (2016)
33. He, X., Hu, T., Wang, G., Wang, Z., Wang, R., Zhang, Q., Yan, K., Chen, Z., Li, R., Xie, C., et al.: Enhancing RAW-to-sRGB with decoupled style structure in Fourier domain. In: AAAI (2024)
34. Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pająk, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., et al.: Flexisp: A flexible camera image processing framework. ACM Transactions on Graphics **33**(6), 1–13 (2014)
35. Herrmann, C., Bowen, R.S., Wadhwa, N., Garg, R., He, Q., Barron, J.T., Zabih, R.: Learning to autofocus. In: CVPR (2020)
36. Hong, Z., Zhen, D., Xiong, L., Li, X., Lin, Y.: srrsr-net: A new low-light image enhancement network via raw image reconstruction. Applied Sciences **15**(1), 361 (2025)

37. Ignatov, A., Chiang, C.M., Kuo, H.K., Sycheva, A., Timofte, R.: Learned smartphone ISP on mobile NPUs with deep learning, mobile ai 2021 challenge: Report. In: CVPR (2021)
38. Ignatov, A., Chiang, C.M., Kuo, H.K., Sycheva, A., Timofte, R.: Learned smartphone ISP on mobile NPUs with deep learning, mobile AI 2021 challenge: Report. In: CVPRW (2021)
39. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 3277–3285 (2017)
40. Ignatov, A., Perevozchikov, G., Timofte, R., Li, C., Liu, L., Cao, J., Sun, H., Pan, W., Wang, S., Yu, K., et al.: Learned smartphone isp on mobile gpus, mobile ai 2025 challenge: Report. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 1934–1946 (2025)
41. Ignatov, A., Sycheva, A., Timofte, R., Tseng, Y., Xu, Y.S., Yu, P.H., Chiang, C.M., Kuo, H.K., Chen, M.H., Cheng, C.M., et al.: MicroISP: processing 32mp photos on mobile devices with deep learning. In: ECCV (2022)
42. Ignatov, A., Timofte, R., Zhang, Z., Liu, M., Wang, H., Zuo, W., Zhang, J., Zhang, R., Peng, Z., Ren, S., et al.: AIM 2020 challenge on learned image signal processing pipeline. In: ECCVW (2020)
43. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera ISP with a single deep learning model. In: CVPRW (2020)
44. Jeong, W., Jung, S.W.: RAWtoBit: A fully end-to-end camera isp network. In: ECCV (2022)
45. Jiang, Y., Wronski, B., Mildenhall, B., Barron, J.T., Wang, Z., Xue, T.: Fast and high quality image denoising via malleable convolution. In: ECCV (2022)
46. Kantorovitch, L.: On the translocation of masses. Management science **5**(1), 1–4 (1958)
47. Karaimer, H.C., Brown, M.S.: A software platform for manipulating the camera imaging pipeline. In: ECCV (2016)
48. Ke, Z., Liu, Y., Zhu, L., Zhao, N., Lau, R.W.: Neural preset for color style transfer. In: CVPR (2023)
49. Kim, J., Kim, M., Kang, H., Lee, K.: U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. arXiv preprint arXiv:1907.10830 (2019)
50. Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., Burnaev, E.: Wasserstein-2 generative networks. arXiv preprint arXiv:1909.13082 (2019)
51. Korotin, A., Kolesov, A., Burnaev, E.: Kantorovich strikes back! wasserstein gans are not optimal transport? Advances in Neural Information Processing Systems **35**, 13933–13946 (2022)
52. Korotin, A., Selikhanovych, D., Burnaev, E.: Neural optimal transport. arXiv preprint arXiv:2201.12220 (2022)
53. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2017)
54. Li, K., Li, H., Li, H., Wang, P., Guo, C., Jiang, W.: SIRLUT: Simulated infrared fusion guided image-adaptive 3D lookup tables for lightweight image enhancement. In: ACM MM (2024)
55. Liu, C., Yang, H., Fu, J., Qian, X.: 4D LUT: learnable context-aware 4D lookup table for image enhancement. IEEE Transactions on Image Processing **32**, 4742–4756 (2023)

56. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NeurIPS (2017)
57. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)
58. Makkuva, A., Taghvaei, A., Oh, S., Lee, J.: Optimal transport mapping via input convex neural networks. In: International Conference on Machine Learning. pp. 6672–6681. PMLR (2020)
59. Mo, Z., Li, W., Ding, S.: Diffuse and refine latent prior with transformers for neural isp. In: 2025 IEEE International Conference on Image Processing (ICIP). pp. 1456–1461. IEEE (2025)
60. Nguyen, R., Prasad, D.K., Brown, M.S.: Raw-to-raw: Mapping between image sensor color responses. In: CVPR (2014)
61. Nikonorov, A., Perevozchikov, G., Korepanov, A., Mehta, N., Afifi, M., Ershov, E., Timofte, R.: Color matching using hypernetwork-based kolmogorov-arnold networks. arXiv preprint arXiv:2503.11781 (2025)
62. Pang, Y., Lin, J., Qin, T., Chen, Z.: Image-to-image translation: Methods and applications. IEEE Transactions on Multimedia **24**, 3859–3881 (2021)
63. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: ECCV (2020)
64. Perevozchikov, G., Mehta, N., Afifi, M., Timofte, R.: Rawformer: Unpaired raw-to-raw translation for learnable camera ISPs. In: ECCV (2024)
65. Pham, K., Le, K., Ho, N., Pham, T., Bui, H.: On unbalanced optimal transport: An analysis of sinkhorn algorithm. In: International Conference on Machine Learning. pp. 7673–7682. PMLR (2020)
66. Pooladian, A.A., Domingo-Enrich, C., Chen, R.T., Amos, B.: Neural optimal transport with lagrangian costs. arXiv preprint arXiv:2406.00288 (2024)
67. Ren, Y., Niu, X., Jiang, H., Liu, Z., Jiang, T., Liu, G., Liu, S.: Ispformer: Learning raw-to-srgb mappings with wavelet-based self-attention. Neurocomputing p. 131084 (2025)
68. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention (2015)
69. Schwartz, E., Giryes, R., Bronstein, A.M.: DeepISP: Toward learning an end-to-end image processing pipeline. IEEE Transactions on Image Processing **28**(2), 912–923 (2018)
70. Seizinger, T., Conde, M.V., Kolmet, M., Bishop, T.E., Timofte, R.: Efficient multi-lens bokeh effect rendering and transformation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1633–1642 (2023)
71. Shekhar Tripathi, A., Danelljan, M., Shukla, S., Timofte, R., Van Gool, L.: Transform your smartphone into a dslr camera: Learning the isp in the wild. In: European Conference on Computer Vision. pp. 625–641. Springer (2022)
72. Šindelář, O., Šroubek, F.: Image deblurring in smartphone devices using built-in inertial measurement sensors. Journal of Electronic Imaging **22**(1), 011003–011003 (2013)
73. Torbunov, D., Huang, Y., Tseng, H.H., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., Ren, Y.: Rethinking cycleGAN: Improving quality of GANs for unpaired image-to-image translation. arXiv preprint arXiv:2303.16280 (2023)
74. Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., Ren, Y.: UVCGAN: UNet vision transformer cycle-consistent GAN for unpaired image-to-image translation. In: WACV (2023)

75. Torbunov, D., Huang, Y., Yu, H., Huang, J., Yoo, S., Lin, M., Viren, B., Ren, Y.: UVCGAN v2: An improved cycle-consistent GAN for unpaired image-to-image translation. arXiv preprint arXiv:2303.16280 (2023)
76. Vasa, V.K., Qiu, P., Zhu, W., Xiong, Y., Dumitrascu, O., Wang, Y.: Context-aware optimal transport learning for retinal fundus image enhancement. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4016–4025. IEEE (2025)
77. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8798–8807 (2018)
78. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
79. Wirzberger Raimundo, D., Ignatov, A., Timofte, R.: LAN: Lightweight attention-based network for raw-to-rgb smartphone image processing. In: CVPRW (2022)
80. Wronski, B., Garcia-Dorado, I., Ernst, M., Kelly, D., Krainin, M., Liang, C.K., Levoy, M., Milanfar, P.: Handheld multi-frame super-resolution. ACM Transactions on Graphics (ToG) **38**(4), 1–18 (2019)
81. Xing, Y., Qian, Z., Chen, Q.: Invertible image signal processing. In: CVPR (2021)
82. Yang, C., Jin, M., Xu, Y., Zhang, R., Chen, Y., Liu, H.: SepLUT: Separable image-adaptive lookup tables for real-time image enhancement. In: ECCV (2022)
83. Yi, Z., Zhang, H., Tan, P., Gong, M.: DualGAN: Unsupervised dual learning for image-to-image translation. In: ICCV. pp. 2849–2857 (2017)
84. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
85. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: CycleISP: Real image restoration via improved data synthesis. In: CVPR (2020)
86. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: 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 (2018)
87. Zhang, Z., Wang, H., Liu, M., Wang, R., Zhang, J., Zuo, W.: Learning raw-to-sRGB mappings with inaccurately aligned supervision. In: ICCV (2021)
88. Zhao, Y., Wu, R., Dong, H.: Unpaired image-to-image translation using adversarial consistency loss. In: ECCV (2020)
89. Zheng, W., Fu, H., Wang, X., Kang, H., Wang, C., Liu, J., Xu, Z., Zhang, H., Ma, H.: Evraw: Event-guided structural and color modeling for raw-to-srgb image reconstruction. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 209–218 (2025)
90. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
91. Zhu, P., Abdal, R., Qin, Y., Wonka, P.: SEAN: Image synthesis with semantic region-adaptive normalization. In: CVPR (2020)