”ALEXANDRU-IOAN CUZA” UNIVERSITY OF IASI

FACULTY OF COMPUTER SCIENCE IASI

PHD RESEARCH PROPOSAL

Optimizing RAW Image Processing and Understanding Using Deep Learning on Smartphone GPUs and NPUs

proposed by

Andrei Arhire

Contents

Introduction

Today, smartphones represent the dominant photography tool due to their growing imaging capabilities, portability, and affordability. Smartphone camera specifications are key factors that heavily influence the purchase of a new mobile device. Advancements in this field follow two main directions: hardware improvements, including the optical system (lenses) and camera sensors, and progress in computational photography solutions. Notably, the second direction is gaining increasing attention from research groups in both industry and academia, leading to the development of more advanced approaches for various applications. A wide range of solutions for tasks such as super-resolution ([12], [61]), denoising ([51], [32]), color correction ([24], [75]), and more has achieved impressive results, enabled by deep learning models.

The image signal processor (ISP; refer to fig. 1) represents a core component of the smartphone camera. The output of the camera sensor goes through the ISP which refines the data to produce visually pleasing results that are well-suited to human visual system. ISP transformations can be defined as mapping from the RAW image space, characterized by the properties of a specific sensor, to the RGB image space. Generally, a Bayer color filter array (CFA) is placed on top of a silicon photodetector array to capture the scene at different wavelength ranges, enabling the reconstruction of color information. Demosaicing is the process of interpolating the missing red, green, and blue values in the Bayer color filter array to reconstruct a full-color image. Essentially, the functions that make up the ISP pipeline include demosaicing, denoising, white balance, color corection, tone mapping, gamma correction, and compression. Increasingly, traditional, carefully hand-crafted modules of the pipeline are being replaced or augmented by deep learning models. In particular, various neural networks have demonstrated their ability to accurately approximate this complex transformation using real-world RAW-to-RGB datasets. Despite the training paradigm, a learned ISP aims to map a RAW image to a single, high-quality, perceptually pleasing RGB image.

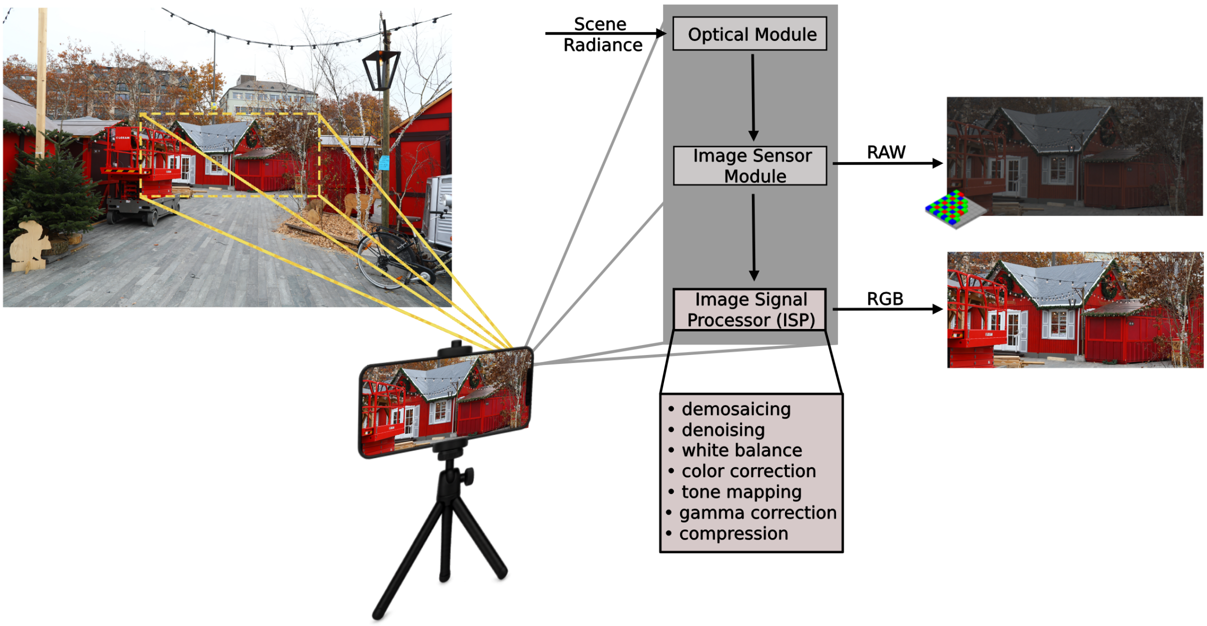


Figure 1: Smartphone imaging pipeline: optical, sensor, and ISP stages.

Developing AI-based solutions such as Learned ISP, Efficient LLMs or Efficient Stable Diffusion, designed to run on edge devices, requires a strong understanding of both the software that can be used and the available hardware with their limitations. Common restrictions include a small amount of RAM and partial support for key components in deep learning frameworks. Such limitations make it impossible to process high-resolution images using standard neural network models, thus demanding resource-efficient redesign strategies for compatibility with mobile target platforms.

Our proposed research topic focuses on 'Optimizing RAW Image Processing and Understanding Using Deep Learning on Smartphone GPUs and NPUs'. Using deep learning technologies, we aim to improve the user experience in smartphone photography through solutions that can be executed in real time using on-device AI accelerators such as Graphics Processing Unit (GPUs) and Neural Processing Units (NPUs). Potential areas of focus include controllability in the image signal processor, low-light object detection, burst RAW restoration, and RAW-to-RAW image translation. Through our approaches, we plan to introduce practical deep learning solutions including architectures and training paradigms, among other strategies, to address real world problems.

Motivation

Smartphones released in the past one to two years, equipped with high-end chipsets, often include powerful AI accelerators. The latest NPUs found in these devices can offer performance comparable to desktop GPUs such as the NVIDIA RTX 5070 or 5080. Although GPUs consume more energy than NPUs and are generally slower in AI-specific tasks, modern GPUs are still faster than NPUs from five years ago, which makes them a valuable resource. This creates a strong motivation to fully leverage the computational capabilities of modern NPUs, or at the very least mobile GPUs, by deploying deep learning models directly on the device.

Since 2020, the idea of developing a learnable ISP capable of reconstructing the RGB output from RAW sensor data has been increasingly studied. In one of the first solutions ([41]), a method was proposed that outperformed the proprietary ISP integrated into the Huawei P20 imaging system. In addition, the Zurich RAW-to-RGB dataset was introduced, consisting of paired RAW and RGB images captured using a smartphone and a DSLR camera. In follow-up research, academic and industry groups have developed additional solutions that process RAW data ([59], [77], [62], [71], [80]), including learnable ISPs ([46], [36], [65], [60], [69]). Some of these efforts that operate on RAW data have also resulted in the release of new datasets ([79], [57], [64], [1], [3]).

Typically, many computer vision tasks are performed by transforming a standard RGB image into an improved RGB version. The original, unaltered RAW Bayer data from the sensor contains richer information, typically stored in 10 to 16 bits per pixel. As the ISP processes RAW data into an RGB version primarily designed to match human perception, it may lose valuable information when performing specific tasks. For example, when trying to detect certain objects in low-light conditions, a stronger solution might be developed relying on the RAW data rather than the processed RGB version. However, compared to RGB datasets, publicly available RAW image datasets are relatively scarce and less diverse.

At the same time, we are witnessing incredible growth in the power of AI accelerators available in our pockets, as well as an increase in the number of training datasets specifically designed for RAW image processing. Since this is a relatively recent development, there are still only a few solutions available for many computer vision applications that work directly with RAW data. Some practical applications have been explored by just one or a handful of research projects, while others are still waiting for someone to break the ice.

The interest in such applications designed for real-time on-device execution attracts significant attention from both industry and university research groups. As I personally experienced through my participation at CVPR as an author, there are many postdocs and PhD candidates interested in specific areas within this broader field. Major tech companies, including Meta, Sony, and others, have shown strong interest in recent research on image signal processing, including our work ([4], [37]) presented at the Mobile AI Workshop in CVPR 2025, and have dedicated research teams focused on this area. As computer vision is one of the largest and most active research fields, numerous top-tier, prestigious, and global conferences are held periodically. Each of these major events accepts many papers on this topic every time. As a result of future research in the coming years, we aim to work on relevant projects and target such conferences for paper submissions.

My deep interest in low-level computer vision supports this research direction. A valuable experience that prepared me for doctoral-level research was the work I did during my master's degree in the Advanced Studies program. During this stage, I had the opportunity to explore computer vision in greater depth and discovered how fascinating and complex the field truly is. I also faced and overcame challenges that I believe are inevitable in any PhD journey, which gave me important insights and a more realistic perspective on research. Because of this, I believe the coming years of doctoral work will be more efficient and focused, as I have already learned valuable lessons from my own experience. The research I conducted under the guidance of my supervisor provided a strong foundation for this. During that time, I contributed to the development of new loss functions and a dynamic loss adaptation strategy. This led to the introduction of the first unsupervised training method applied to an ISP ([4]), along with a supervised training method that achieved better results than those reported by the state of the art last year ([50]), both in terms of inference speed and output quality.

Chapter 1

Overview of the Area

1.1 Workshops and Benchmark Challenges in Mobile AI

Trending practical topics in computer vision are often accompanied by papers written by enthusiastic researchers and published at prestigious international conferences. In particular, these large events frequently host workshops dedicated to specific areas that attract significant interest from industry due to their practical applications. In recent years, major conferences such as the Conference on Computer Vision and Pattern Recognition (CVPR), the European Conference on Computer Vision (ECCV), and the International Conference on Computer Vision (ICCV) have consistently included at least one workshop focused on the latest trends and research in Mobile AI solutions.

It has become a common tradition for such workshops to organize challenges in which participants are invited to develop deep learning approaches aimed at efficiently solving various computer vision tasks on mobile hardware. These activities have facilitated and encouraged the creation of new state-of-the-art solutions, often accompanied by papers ([4], [30], [20], [49], [6], [52], [25]) presented at respective workshops.

In 2025 we can highlight the Mobile AI Workshop at CVPR, New Trends in Image Restoration and Enhacement Workshop (NTIRE) at CVPR and Advancement in Image Manipulation Workshop (AIM) at ICCV. In particular, among the challenges working with RAW data can be mentioned the following:

- Learned Smartphone ISP Challenge ([37])
- RAW Image Reconstruction from sRGB ([13])
- RAW Restoration Challenge (Tracks 1 and 2; [14])

1.2 Development of Mobile AI solutions

If someone decides to design a solution that runs on a mobile device for tasks such as language processing, image classification, image processing, or others, there are generally three main steps that must be followed.

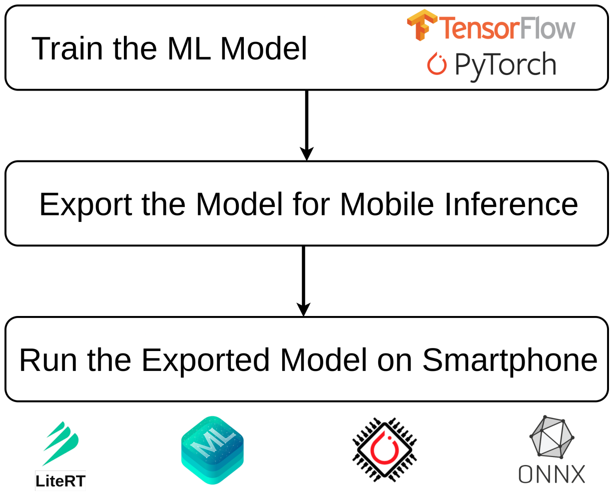


Figure 1.1: Steps for Deploying Machine Learning Models on Smartphones

The first step is to train the original model using standard libraries such as PyTorch or TensorFlow. However, this trained model cannot be executed directly on smartphones because mobile platforms do not support PyTorch or TensorFlow in their native form. As a result, the trained model must be converted into a format that is compatible with a specific mobile machine learning library. Once the conversion is complete, the new model can be executed using the chosen solution. One of the most important aspects of this process is selecting the right mobile machine learning library. The conversion process and supported features are entirely dependent on this choice.

The following introduces the four most popular mobile ML frameworks.

1.2.1 ONNX

One of Open Neural Network Exchange (ONNX) main benefits is that it allows easy conversion of models from PyTorch to ONNX format. In fact, PyTorch provides built-in tools that enable the export of a model with only one line of code. Despite its many advantages, ONNX has a major limitation when it comes to mobile devices: it only supports CPU-based inference. This means that if a model is converted to ONNX and then executed on a mobile device, it will run solely on the device's CPU. As a result, it cannot take advantage of the available GPU or the latest high-performance

NPUs, which are specifically designed to accelerate complex models. Therefore, if inference performance beyond CPU is required, one should consider alternative frameworks better suited for mobile AI acceleration.

1.2.2 ExecuTorch

If a model is developed using PyTorch, ExecuTorch might be the first option to consider, as it is also created by the team behind PyTorch and integrates naturally with its workflow. Unlike ONNX, ExecuTorch provides support for running models on GPUs, NPUs, and DSPs, which is essential to deploy deep learning models efficiently.

Once the PyTorch model is ready, it needs to be exported to a supported format and then compiled using ExecuTorch for the specific target device. This compilation step can apply hardware-specific optimizations that improve runtime performance.

However, compiling for a particular hardware platform is not a simple process. ExecuTorch alone is not enough; it requires the installation of vendor-specific SDKs to generate a working model for each platform. For example, deploying to Qualcomm NPUs involves downloading the Qualcomm SDK, installing the Android NDK, compiling ExecuTorch with Qualcomm support, and building both the model and the runtime library. The resulting model can only be executed on the specific hardware for which it was compiled, using the corresponding version of the compiled library.

This workflow must be repeated for each platform, such as Qualcomm, MediaTek, or Vulkan, making it necessary to prepare separate application versions for different devices. At this stage, developing a universal mobile application is not feasible using ExecuTorch. The only exception is CoreML, which is already supported directly within ExecuTorch and does not require additional vendor SDKs. Due to these limitations and the overall complexity of the process, ExecuTorch is currently not recommended for general mobile AI development.

1.2.3 CoreML (iOS)

On iOS devices, CoreML is the recommended and most practical solution for running machine learning models. CoreML provides a very straightforward model conversion process from PyTorch or TensorFlow, often requiring just two lines of code. Importantly, it does not require any vendor SDKs. It is enough to install the CoreML Tools package using Python via pip, and the conversion can be performed on any platform, including Linux and macOS. Although native support for Windows was discontinued, the process can still be done using the Windows Subsystem for Linux (WSL).

The CoreML converter supports a wide range of models, including large language models and diffusion-based architectures. It allows deployment on Apple's dedicated hardware, such as the Apple Neural Engine (ANE) and Apple GPUs, making it the only way to fully leverage this hardware. Additionally, CoreML includes an efficient CPU backend.

1.2.4 LiteRT (TensorFlowLite)

On Android, the wide variety of hardware vendors creates the need for a unified solution that works on many devices. Fortunately, TensorFlow Lite, which was recently renamed by Google as LiteRT, addresses this challenge. Although the name has changed, the library itself remains identical. The rebranding was mainly a marketing decision, especially as PyTorch has become more popular, while TensorFlow is now mostly used by those already familiar with it.

Previously, only the TensorFlow and Keras models could be converted directly to LiteRT. However, Google recently announced an official PyTorch converter, which allows models to be converted to LiteRT with just one line of code. This makes it possible to build the entire model in PyTorch and easily convert it to Android devices.

A major advantage of LiteRT is that a single converted model can run on many hardware configurations, without requiring separate builds for each device. This is a clear benefit over solutions such as ExecuTorch, which require compiling models specifically for each target platform.

LiteRT supports hardware acceleration on GPUs and NPUs through delegates. The GPU delegate enables inference on most mobile GPUs released in the past ten years, as it requires only OpenCL, which is available on the vast majority of modern Android smartphones. These delegates are plugin modules added to Android applications that enable the model to run directly on specialized hardware, with no additional configuration needed. In addition, LiteRT offers powerful quantization tools, allowing developers to reduce model size and increase performance while preserving good accuracy even with 16-bit or 8-bit precision.

Chapter 2

Existing Work

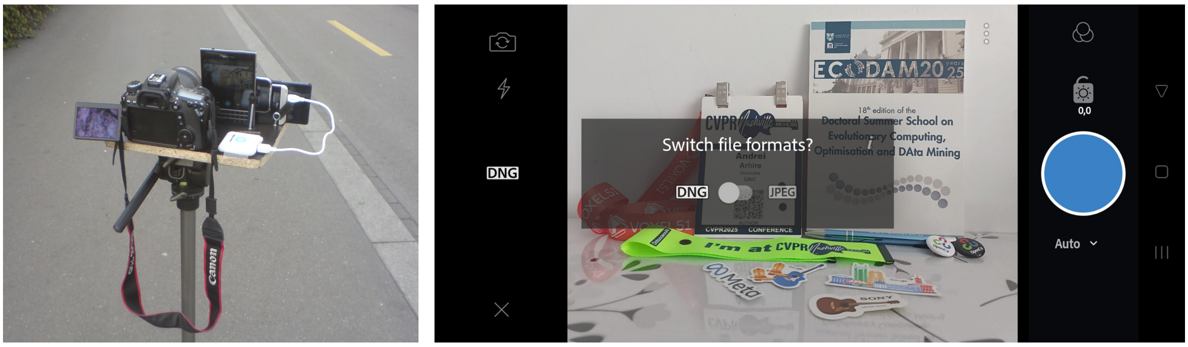
2.1 Available Datasets

There are many public datasets that utilize RAW image data, often captured using various smartphone models. Most modern smartphones support capturing images in RAW format. To verify RAW support, tools like Camera2 API Probe can be used. Once confirmed, applications such as Adobe Lightroom ([2]) can be used to capture RAW images. An example of capturing an image in DNG format using a personal smartphone is shown in fig. 2.1b. For dataset collections that involve a large number of scenes and various camera systems, synchronized imaging setups have been developed, as illustrated in fig. 2.1a.

The following introduces several popular datasets that contain RAW image data.

- Samsung S7 Dataset (DeepISP) [64]. This dataset contains RAW and processed RGB image pairs of 110 different scenes, captured using the Samsung S7 smartphone camera under normal exposure settings. Each image has a resolution of 3024×4032 pixels, taken with the 12MP Sony IMX260 sensor. The image pairs are sufficiently aligned, making them suitable for evaluation with pixel-wise metrics;

- Huawei P20 Pro Dataset (PyNET) [41]. This dataset includes RAW and RGB images captured using the Huawei P20 Pro smartphone with the 12.3 MP Sony Exmor IMX380 sensor. All photos were taken in automatic mode with default settings throughout the entire data collection process. The original resolution of the images is 3840×5120 pixels. However, the RAW and RGB image pairs exhibit a noticeable misalignment, typically exceeding 8 pixels, which may limit the use of strict pixel-wise evaluation;



(a) The 4-camera rig used in [34]
(b) Capturing photos in RAW format using [2]

Figure 2.1: Equipment and interface used for RAW image dataset collection

- SSID Dataset [1]. The dataset provides RAW–RGB image pairs captured using the Samsung Galaxy S6 Edge smartphone with a 16MP Sony IMX240 sensor. All images were taken under normal brightness conditions, and the dataset is well-suited for supervised training and evaluation in RAW image processing tasks;

- RAW-to-RAW Dataset [3]. This dataset includes RAW images of natural scenes along with corresponding RGB outputs rendered by the in-camera ISP of the Samsung Galaxy S9, which uses the Sony IMX345 sensor. It is intended for tasks involving learning the transformations performed by traditional ISPs in a supervised or semi-supervised setting.

- Fujifilm UltraISP Dataset [38]. The dataset contains over 6,000 RAW–RGB image pairs captured with a 12MP Sony IMX586 sensor and a professional 102MP Fujifilm GFX100 camera. The images were taken simultaneously using a dual-camera setup under diverse daytime conditions. Pixel-wise alignment was performed using deep learning-based matching, producing 99K image crops (256×256 px), suitable for supervised RAW-to-RGB learning and evaluation.

- GenISP Low-Light Dataset [57]. This dataset contains 7,200 RAW images captured under low-light outdoor conditions using two cameras: Sony RX100 VII (3.2K images) and Nikon D750 (4.0K images). Each image is annotated with bounding boxes for 46,000 instances across three classes: people, bicycles, and cars. The dataset features diverse illumination conditions and is intended for benchmarking low-light RAW image restoration and object detection methods.

2.2 Application Domains

The following sections provide a brief overview of recent work on key tasks aligned with the proposed research direction with a focus on practical applications.

2.2.1 Learned ISP and Reversed ISP

In recent years, deep learning has enabled the development of multiple neural-ISP solutions ([40], [33], [39], [37], [29], [42], [78]), most of which adopt CNN-based U-Net-like architectures ([63]). Early approaches include DeepISP [64] and PyNET [41]. PyNET employs an inverted pyramidal architecture with five branches, each operating at a different image scale and trained sequentially, to extract and fuse global and local features. Subsequent work, such as PyNET-CA [46], incorporated channel attention mechanisms to enhance performance, while lightweight versions like Micro ISP [38] and PyNET-V2 [36] were designed for efficient mobile deployment, achieving a trade-off between latency and output quality on edge devices. More recent methods, including MW-ISPNet [40], AW-Net [20], LAN [60], and RMFA-Net [50], improve image restoration quality by integrating discrete wavelet transform (DWT) and double attention module (DAM) techniques. To overpass the need for collecting paired raw-sRGB datasets for each new camera model, Rawformer [59] introduces a fully unsupervised Transformer-based RAW-to-RAW translation method that enables the reuse of pre-trained neural ISPs across diverse camera domains. To eliminate the need for pixel-aligned supervision, in [4] the authors proposes an unpaired training strategy for lightweight ISPs, using perceptual and adversarial losses.

Existing RGB-to-RAW methods typically fall into two categories: metadata-based and learning-based. Metadata-based methods rely on camera-specific ISP parameters to reverse the processing pipeline. In contrast, learning-based methods use deep neural networks to directly learn the RGB-to-RAW mapping without metadata, often using encoder-decoder architectures, multi-branch fusion, attention mechanisms, and customized losses to handle channel differences and generalize across devices. These approaches have achieved state-of-the-art performance in recent benchmarks, including the AIM 2022 [15] and NTIRE 2025 RGB-to-RAW reconstruction challenges [13].

2.2.2 RAW Image Restoration and Super-Resolution

RAW image restoration is an important topic in computational photography, addressing key image quality challenges such as low resolution, noise, and motion blur in portable devices. Although modern sensors offer higher bit depths, limitations in size, power, and optics still impact image quality. Smaller sensors reduce light capture and signal-to-noise ratio, making it harder to achieve sharp, high-resolution results.

Unlike RGB images processed by the ISP, RAW data preserves a linear response to scene radiance and retains a wider bit range, typically 12–14 bits. This makes it a better input for tasks like denoising, deblurring, and super-resolution.

Recent research has advanced RAW image super-resolution by directly addressing the limitations of RGB-based methods, which suffer from information loss due to ISP processing. Xu et al. ([73], [74]) were among the first to highlight the benefits of using linear RAW data for real-scene SISR. They addressed the limitations by proposing a data generation pipeline that simulates realistic RAW space degradation, such as variable blur, heteroscedastic noise, and downsampling, and a dual-branch CNN that combines RAW input for structural detail recovery with RGB guidance for color correction, enabling more accurate restoration under real-world conditions.

Building on this, the BSRAW method [16] advances blind super-resolution in the RAW image domain by introducing a realistic and controllable degradation pipeline that simulates complex real-world factors such as noise, blur, exposure inconsistencies, and downsampling, all applied directly to linear RAW sensor data. Unlike conventional SISR methods that operate in the sRGB domain and are hindered by the non-linearities introduced by the ISP, BSRAW enhances images prior to ISP processing, thereby preserving the original scene radiance and structural details more effectively. To facilitate robust training and evaluation, the authors introduce a comprehensive dataset combining DSLR images from MIT-Adobe FiveK [8] with a new DSLM dataset captured using modern mirrorless cameras.

The NTIRE 2024 and 2025 challenges ([17], [14]) established comprehensive benchmarks for this task, further pushing the state-of-the-art by highlighting top-performing solutions trained on RAW patches with synthetic noise and blur modeled after BSRAW. The best solution, including RawRTSR and RBSFormer [14], often employed Transformer-based or dual-stage architectures with dedicated loss functions and training strategies, balancing high fidelity and computational efficiency.

2.2.3 RAW Object Detection

RAW images offer significant advantages over standard RGB images thanks to their higher dynamic range and linear noise profile, making them especially valuable in low-light and different weather conditions. They are usually five to ten times larger than compressed RGB images, and large-scale RAW datasets containing more than 100,000 images are currently unavailable [7]. However, recent studies have shown that even with limited RAW data, models specifically designed for the RAW domain can achieve improved results in tasks such as image classification ([54], [55]), object detection ([19], [22], [53], [70], [72], [76]), semantic segmentation [19], and instance segmentation [10].

Several recent works have tackled the challenge of object detection on RAW images by proposing novel adaptations to either the detection pipeline or the ISP. Ljungbergh et al. [53] investigated the limitations of conventional ISP pipelines for object detection and introduced learnable pre-processing operations, such as a differentiable Yeo-Johnson transformation, which improved detection performance in poor lighting conditions. Berdan et al. [7] presented ReRAW, a reverse ISP that reconstructs sensor-specific RAW images from labeled RGB datasets using a multi-head gamma-based architecture and a stratified sampling strategy, allowing effective training of lightweight object detectors directly on synthetic and real RAW data. Wang et al. [70] proposed AdaptiveISP, a reinforcement learning-based framework that dynamically configures ISP pipelines for each image to optimize object detection accuracy and efficiency in diverse lighting scenarios. Xu et al. [72] developed a synthetic RAW degradation model and a dual-branch convolutional network that combines RAW input with RGB guidance for improved structural detail recovery under realistic distortions. Yoshimura et al. [76] introduced DynamicISP, which adjusts ISP parameters in real time based on feedback from previous recognition outputs, leading to improved performance in low-light video sequences. Dutta et al. [22] demonstrated that continual learning with RAW images allows models to adapt better to challenging conditions such as darkness and weather variability, outperforming conventional RGB-based detectors. Morawski et al. present GenISP [57], a minimal neural ISP optimized for low-light cognition by incorporating color space transformations and expert-inspired color correction modules. GenISP is trained under the guidance of a fixed detector, avoids assumptions about human perception, and generalizes well across sensors and detectors.

Chapter 3

Relevant experience

I have experience in developing deep learning applications, particularly in the field of computer vision, as a researcher. At the same time, I have been an active member of the national competitive programming community. My involvement in teaching and developing efficient solutions has equipped me with the skills and knowledge necessary to drive innovation in this domain and address real-world problems.

3.1 Teaching and Mentoring Experience

Experience in competitive programming demonstrates the ability to work in a team, solve complex problems, perform under pressure, manage time and deadlines, and minimize errors. It also reflects discipline, focus, and speed, which are indispensable skills. As a participant in various international competitions, both individually and as part of a team, I found this experience extremely valuable to develop critical thinking and optimize my workflow in many areas of my future work. This is why, from the beginning of my bachelor's program and as I advanced in this field, I was consistently involved in organizing competitions and activities aimed at developing the local and national community, including university and high school students. In particular, I was a problem-setter in various competitions, including the high school section of the National Olympiad in Informatics, INFOPRO, and FIICode. Among my results as a participant, one notable achievement is a Special Mention (17th place) in SEERC, the Southeastern European Regional Contest, which is a regional phase of the International Collegiate Programming Contest (ICPC). In addition, since 2021, I have been teaching at the Iași County Center of Excellence, while also mentoring students who have earned medals in the National Olympiad in Informatics.

I had the opportunity to teach several classes for one year at the High School of the Alexandru Ioan Cuza University of Iasi. It was a highly rewarding experience, where I helped pre-university students develop a passion for computer science through interactive lessons and hands-on activities.

In the future, I would enjoy continuing to contribute to the teaching of university students, especially in courses closely related to my research area, such as neural networks, computer vision, algorithm design, data structures, and others. The only constraint is to ensure that this teaching activity is carefully balanced with my primary research work, so that I have enough resources to perform well in all my activities.

3.2 Research Experience

My research journey began with a strong foundation in algorithms and problem solving, starting with my Bachelor's and Master's theses, which provided a strong foundation for the direction I am now pursuing. In addition, my experience in the Cyber Threat Intelligence Lab at Bitdefender as a Junior Security Researcher was valuable for developing an organized and critical approach to research.

3.2.1 Bachelor Thesis

My bachelor thesis, titled A Heuristic Solver for the Directed Feedback Vertex Set Problem, focused on the NP-complete problem known as the Directed Feedback Vertex Set (DFVS). This problem is one of Richard Karp's famous 21 NP-complete problems ([45]) and asks: "What is the smallest set of nodes (vertices) that can be removed from a directed graph to eliminate all cycles?". The DFVS problem has a wide variety of applications, including deadlock resolution in operating systems, circuit design, and program analysis. My primary objective was to develop an efficient heuristic algorithm capable of producing solutions close to the global optimum in a short time while using minimal computational resources. This problem was the topic of the 2022 edition of the Parameterized Algorithms and Computational Experiments (PACE) Challenge. The organizers introduced a new benchmark based on various patterns inspired by real-world scenarios. I explored exact and heuristic approaches to address the problem and developed an efficient solver with approximation quality close to the top solution.

3.2.2 Master Thesis

During my master's studies, my goal was to find the most suitable area to focus on. I wanted to explore different branches of computer science in depth to understand how each field is evolving. After carefully analyzing the options, I realized that the domain I chose for this research plan offers a great opportunity to apply my existing skills, develop new ones, and create useful solutions with real-world impact.

In my master's thesis, Learned Lightweight Smartphone ISP with Unpaired Data, I explored the feasibility of developing a deep learning model capable of approximating ISP functions without the need for paired data. Previous work ([35], [34]) explored image enhancement using partially unpaired loss functions. However, to our knowledge, no completely unsupervised training method was available for learning ISP.

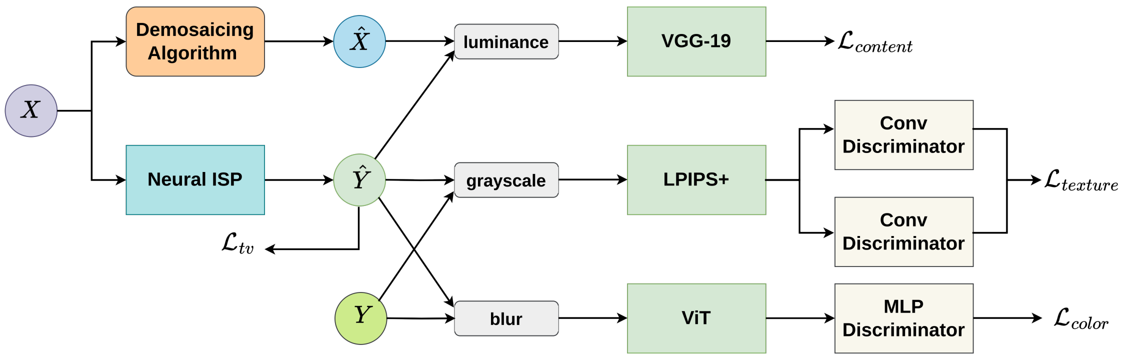


Figure 3.1: Overview of our proposed unpaired training method in [4].

We proposed, to our knowledge, the first unsupervised training method suitable for lightweight deep models to learn ISP transformations (refer to fig. 3.1). We demonstrated the robustness and generalization of our approach through extensive experiments using three different neural network architectures of varying complexity ([39], [50]), evaluated on two real-world RAW-to-RGB datasets ([41], [38]) with multiple hyperparameter configurations. In addition, we ensured that all operations are fully compatible with mobile AI accelerators and provided pre-trained models in TensorFlow Lite format. The results were promising in multiple aspects of image quality, especially in terms of structural metrics, when compared to supervised training. We also trained the same backbones used in the unpaired approach using a custom supervised training method, which achieved results that represent an upper bound for the performance of the unpaired strategy. The supervised-trained models were submitted to different tracks of the Mobile AI Challenge 2025 ([37]), demonstrating competitive performance, while the contributions on the unpaired side were summarized in a paper ([4]) presented at the Mobile AI Workshop, held in conjunction with CVPR 2025.

The novel unpaired training strategy is based on several key contributions that we proposed. In our work, we defined specific loss functions to facilitate the learning of particular attributes in the generated images. The structural consistency was ensured by a loss component that minimized the difference in features maps from the last convolutional layer of pre-trained VGG-19 ([66]). The reference image was obtained by applying a specialized demosaicing algorithm ([56]) to the input. To reduce the influence of misleading colors, both the input and reference images were converted to a different color space, where only the luminance component was retained and replicated before being used as input for the model. The color and texture components were addressed using dedicated loss terms in an adversarial setting. An MLP-based discriminator receives features from a pre-trained Vision Transformer ([21]) applied to slightly blurred images. Additionally, two convolutional discriminators process linear features extracted from different layers of a pre-trained LPIPS+ model ([9]). In the final version, we used a relativistic adversarial loss ([43]) with gradient penalties ([31]) for the discriminators, although we experimented with various alternatives ([27], [5], [28]) during development. To balance training and ensure that each loss contributes with a controllable gradient magnitude for each network update, so that no loss dominates the others, we defined a Dynamic Loss Adaptation strategy [4].

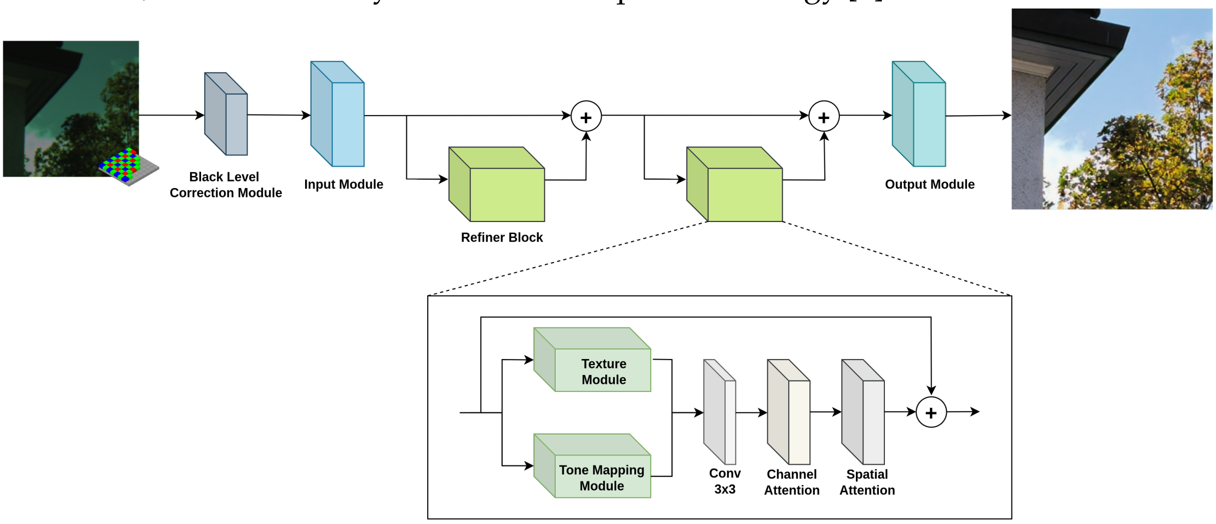


Figure 3.2: Overview of our proposed architecture in supervised setup.

We further adapted the training pipeline for supervised learning and also defined a robust neural architecture (refer to fig. 3.2) that is fully compatible with TensorFlow Lite. The training is divided into four stages, each using different loss functions and hyperparameters. We surpassed the performance reported by last year's state-of-the-art [50] in both output quality and inference speed, and we are currently preparing a research paper to present the results and detail our approach.

To achieve high-quality performance while ensuring fast execution on smartphones, we carefully designed both the architecture and the training process. The architecture is a modified version of RMFA-Net ([50]), in particular the version designed for smartphone inference. The signal passes through a black level correction module, an input module, two refinement blocks, and one output module. In the first module, the input is having its black levels corrected by subtracting learnable per-channel black level offsets, then being scaled by learnable per-channel multipliers. In the input module, the input is processed through two 3x3 convolutional layers to extract features and then passed through a Tanh activation. Each block addresses feature representation by incorporating a dedicated texture module, a tone mapping module, and both channel and spatial attention mechanisms. In the texture module. the input is processed through two parallel branches, a 1x1 convolution for high-frequency details and a 3x3 convolution for low-frequency patterns, then the outputs are concatenated. The tone mapping module processes the input through hierarchical downsampling and upsampling to estimate global illumination features, which are subtracted from the input to produce reflectance according to Retinex theory ([48]). Channel attention enhances important features via global pooling and 1×1 convolutions with sigmoid scaling, while spatial attention refines feature locations using pooled features and a 7×7 convolution. In the last module, the input is upsampled using pixel shuffle, then refined with a final convolution, and mapped to [0,1] range with Sigmoid activation.

3.3 Relevance to PhD Proposal

My work on deep learning solutions applied to the Mobile AI area directly supports the goals of my PhD proposal, which focuses on improving RAW image processing and understanding in real time on edge devices. The activities I was involved in during my Bachelor's studies, along with the research performed for my thesis, helped me develop an efficient and structured approach to perform research. During my Master's program, I discovered my main area of interest and began exploring it in greater depth. This work led to a paper presented at the Mobile AI Workshop at CVPR, where I received valuable feedback from academic researchers and industry professionals. I was also invited by postdoctoral researchers to collaborate on projects aligned with the same direction I am now pursuing. My ongoing research builds on the work of my Master's thesis and provides a strong foundation for my future academic path.

3.4 Dissemination of Research Results

Publications

- **Andrei Arhire** and Radu Timofte. 2025. "Learned lightweight smartphone isp with unpaired data." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, Tennessee, USA (Rank A* in CORE).
- Andrey Ignatov, Georgii Perevozchikov, Radu Timofte, **et al.** 2025. "Learned smartphone isp on mobile gpus, mobile ai 2025 challenge: Report." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, Tennessee, USA (Rank A* in CORE).

Presentations

- **Andrei Arhire** and Radu Timofte. "Learned lightweight smartphone isp with unpaired data." In the Mobile AI Workshop of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, June 2025; and in the 2FII Workshop, Faculty of Computer Science in Iasi, July 2025.
- **Andrei Arhire** and Radu Timofte. "Enhanced Learned Smartphone ISP." In the 2FII Workshop, Faculty of Computer Science in Iasi, February 2025.
- **Andrei Arhire** and Eugen Croitoru. "A Dual Approach to One-Sided Crossing Minimization through Exact and Heuristic Methods" In the 2FII Workshop, Faculty of Computer Science in Iasi, June 2024.
- **Andrei Arhire**, Matei Chiriac, Radu Timofte. "An Exact Twin-Width Algorithm" In the 2FII Workshop, Faculty of Computer Science in Iasi, June 2023.

Chapter 4

Proposed Work

4.1 Requirements

Considering the aspects mentioned below, the scholarship is a highly valuable opportunity that would directly support and enhance my future research activities.

To develop efficient solutions in this area, access to specific hardware is required, particularly a desktop GPU at least. There are a variety of projects in this field, some of which can be trained in a short time on mid-range GPUs ([18]), while others require longer training on more powerful setups ([4]). To address this, I will use my own workstation equipped with a GPU. However, for large-scale or parallel experiments, I will also rely on cloud services. This is especially useful when running multiple experiments simultaneously, such as benchmarking state-of-the-art models on several virtual machines while testing my own methods under different configurations. Nowadays, affordable GPU cloud services are widely available, and in some cases, renting a virtual machine with pre-installed drivers, software, and a specific GPU that becomes accessible within seconds can be much cheaper than only the electricity cost of running the same GPU locally in my home country. For example, constrained by a competition deadline, I completed all the training for the supervised solution in under a week, totaling over 400 hours on GPU, primarily on RTX 3090 and 4090. In this case, cloud services were the only viable option, as I used up to 8 virtual machines running simultaneously on different configurations, each equipped with a dedicated GPU. The unsupervised project was significantly more complex and lasted several months.

4.2 PhD Plan

In my future research, I aim to develop innovative solutions across three key areas within the broader scope of Optimizing RAW Image Processing and Understanding:

1. Learned ISP and Reversed ISP

2. RAW Image Restoration and Super-Resolution

3. RAW Object Detection

The research will progress one step at a time, following a sequence of dedicated six-month projects corresponding to each of the outlined research areas.

The first project is in an advanced stage, featuring a novel architecture optimized for LiteRT with custom loss functions and a four-stage inference pipeline balancing perceptual quality and fidelity. It surpasses last year's state-of-the-art and shows strong preliminary results in this year's ICCV Learned ISP challenge, with final rankings pending. Its scalable design supports both smartphone and GPU deployment. Ongoing work focuses on enhancing tone mapping using NILUT [18], with plans to publish the results.

In the first year, I aim to contribute to ISP-related research by developing a flexible, user-controllable image signal processing framework. Unlike traditional ISPs that generate a fixed RGB output based on a predefined ground truth, my focus is on enabling real-time adjustments to key perceptual attributes such as noise level, sharpness, and tone mapping. This flexibility is crucial, as over-aggressive noise reduction can produce overly smooth results, while insufficient denoising may leave distracting artifacts. To achieve this, I plan to design a compact and expressive latent space for RAW-to-RGB mapping, inspired by models like CRISP [47]. The framework will support intuitive, low-dimensional control over image characteristics such as contrast, detail, and tone mapping. Also, I will explore burst-based RAW restoration methods that leverage multiple aligned frames to enhance quality under challenging conditions, including low light and motion blur. By efficiently fusing temporal information, the pipeline can improve detail preservation and noise reduction without introducing artifacts or relying on computationally intensive architectures.

In the second year, I plan to focus on models based on Implicit Neural Representations ([26], [58], [67]). These models do not store tensors explicitly, but rather learn continuous mappings from coordinates to values, which allows them to represent signals with high fidelity and flexibility. Because of these properties, INRs have been successfully applied to tasks such as super-resolution [11], optical flow [44], and image generation [68]. I am particularly interested in applying INR-based solutions to problems in RAW image enhancement and learned photographic adjustments. Recent methods such as NILUT and INRetouch [23] demonstrate how small INR-based models can be trained to replicate high-quality tone and color edits. I am interested to study how such approaches can be extended to support personalized image enhancement and develop compact representations of complex image operations.

In the second half of the doctoral studies, I aim to develop blind RAW super-resolution solutions that are compatible with ISP tasks and semantic understanding. One of the main objectives is to create a pre-ISP super-resolution module that enhances spatial detail in sensor-domain data without relying on ground-truth high-resolution targets. In addition, I would like to investigate task-specific ISP architectures that replace the standard sequential pipeline with a more flexible design adapted for object detection. The goal is to build modules that can dynamically select or combine ISP functions based on feedback from the detection head, optimizing the image representation specifically for recognition accuracy rather than perceptual quality alone.

In general, a project will start with running baseline pre-trained models and reproducing their performance while conducting a comprehensive literature review to understand the current state of the art. The project will then proceed by analyzing model strengths and weaknesses to identify improvement opportunities and adapt models to new settings. Following this, the research will explore new ideas through prototype implementations and iterate on architectures, training strategies, and loss functions based on preliminary results. The most promising approach will be selected for detailed validation and ablation studies to confirm its effectiveness. Finally, the research will be documented in a scientific paper.

The final phase will involve documenting the research process, experimental results, and conclusions in a well-structured and comprehensive thesis. This will include a detailed analysis of the effectiveness of the proposed methods, challenges encountered, and recommendations for future research directions in the field.

Chapter 5

Conclusion

In the past year, significant progress has been made in making deep learning more accessible for mobile deployment. One of the most impactful developments was the introduction of the first official PyTorch to TensorFlow Lite converter by Google. This tool has major implications, as it enables developers to build powerful models using full capabilities of PyTorch and then easily deploy them on smartphones.

At the same time, the capabilities of AI accelerators integrated into smartphone chipsets have improved substantially, especially over the past two years. The performance of modern NPUs is now comparable to recent desktop GPUs, offering a strong motivation to design solutions that fully leverage this computational power.

In parallel, the research community is increasingly focusing on tasks involving RAW image processing. The number of available datasets for RAW data has increased, along with a growing number of solutions targeting tasks such as enhancement, understanding, and restoration directly on smartphones. Despite these advances, there is still much to explore and improve in this area.

My academic experience during the Advanced Studies Master program helped me develop a structured and efficient approach to research. This foundation has been particularly valuable for my current direction. So far, this activity has led to a paper presented at a highly relevant Mobile AI workshop at CVPR, an experience that enhanced my understanding of the topic and offered valuable interactions with organizers and industry experts, reflecting strong industry interest.

Considering all these recent developments, and based on my current experience and contributions, I believe that continuing along this direction will lead to practical solutions capable of addressing real-world challenges effectively.

Bibliography

[1] A. Abdelhamed, S. Lin, and M. S. Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018.

[2] Adobe Inc. *Adobe Lightroom Mobile App*. Adobe, 2024. Available: `https://www.adobe.com/products/photoshop-lightroom.html`.

[3] M. Afifi and A. Abuolaim. Semi-supervised raw-to-raw mapping. *arXiv preprint arXiv:2106.13883*, 2021.

[4] A. Arhire and R. Timofte. Learned lightweight smartphone isp with unpaired data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025.

[5] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 2017.

[6] M. Ayazoglu and B. B. Bilecen. Xcat-lightweight quantized single image super-resolution using heterogeneous group convolutions and cross concatenation. In *European Conference on Computer Vision*, pages 475–488. Springer, 2022.

[7] R. Berdan, B. Besbinar, C. Reinders, J. Otsuka, and D. Iso. Reraw: Rgb-to-raw image reconstruction via stratified sampling for efficient object detection on the edge. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 11833–11843, 2025.

[8] V. Bychkovsky, S. Paris, E. Chan, and F. Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011.

[9] C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, and W. Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 2024.

[10] L. Chen, Y. Fu, K. Wei, D. Zheng, and F. Heide. Instance segmentation in the dark. *International Journal of Computer Vision*, 131(8):2198–2218, 2023.

[11] Y. Chen, S. Liu, and X. Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021.

[12] Z. Chen, Z. Wu, E. Zamfir, K. Zhang, Y. Zhang, R. Timofte, X. Yang, H. Yu, C. Wan, Y. Hong, et al. Ntire 2024 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6108–6132, 2024.

[13] M. Conde, R. Timofte, R. Berdan, B. Besbinar, and D. Iso. Raw image reconstruction from rgb on smartphones. ntire 2025 challenge report. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1254–1268, 2025.

[14] M. Conde, R. Timofte, Z. Lu, X. Kong, X. Xing, F. Wang, S. Han, M. Park, T. Hao, Y. He, et al. Ntire 2025 challenge on raw image restoration and super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1148–1171, 2025.

[15] M. V. Conde, R. Timofte, Y. Huang, J. Peng, C. Chen, C. Li, E. Pérez-Pellitero, F. Song, F. Bai, S. Liu, et al. Reversed image signal processing and raw reconstruction. aim 2022 challenge report. In *European Conference on Computer Vision*, pages 3–26. Springer, 2022.

[16] M. V. Conde, F. Vasluianu, and R. Timofte. Bsraw: Improving blind raw image super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 8500–8510, 2024.

[17] M. V. Conde, F.-A. Vasluianu, R. Timofte, J. Zhang, J. Li, F. Wang, X. Li, Z. Liu, H. Park, S. Song, et al. Deep raw image super-resolution. a ntire 2024 challenge survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6745–6759, 2024.

[18] M. V. Conde, J. Vazquez-Corral, M. S. Brown, and R. Timofte. NILUT: Conditional neural implicit 3d lookup tables for image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 1371–1379, 2024.

[19] Z. Cui and T. Harada. Raw-adapter: Adapting pre-trained visual model to camera raw images. In *European Conference on Computer Vision*, pages 37–56. Springer, 2024.

[20] L. Dai, X. Liu, C. Li, and J. Chen. Awnet: Attentive wavelet network for image isp. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 185–201. Springer, 2020.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*. OpenReview.net, 2021.

[22] U. K. Dutta. Seeing objects in dark with continual contrastive learning. In *European Conference on Computer Vision*, pages 286–302. Springer, 2022.

[23] O. Elezabi, M. V. Conde, Z. Wu, and R. Timofte. Inretouch: Context aware implicit neural representation for photography retouching. *arXiv preprint arXiv:2412.03848*, 2024.

[24] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal processing*, 129:82–96, 2016.

[25] G. Gendy, N. Sabor, J. Hou, and G. He. Real-time channel mixing net for mobile image super-resolution. In *European conference on computer vision*, pages 573–590. Springer, 2022.

[26] K. Genova, F. Cole, D. Vlasic, A. Sarna, W. T. Freeman, and T. Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7154–7164, 2019.

[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[29] X. He, T. Hu, G. Wang, Z. Wang, R. Wang, Q. Zhang, K. Yan, Z. Chen, R. Li, C. Xie, et al. Enhancing raw-to-srgb with decoupled style structure in fourier domain. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 2130–2138, 2024.

[30] M.-C. Hsyu, C.-W. Liu, C.-H. Chen, C.-W. Chen, and W.-C. Tsai. Csanet: High speed channel spatial attention network for mobile isp. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2486–2493, 2021.

[31] N. Huang, A. Gokaslan, V. Kuleshov, and J. Tompkin. The gan is dead; long live the gan! a modern gan baseline. *Advances in Neural Information Processing Systems*, 37:44177–44215, 2024.

[32] A. Ignatov, K. Byeoung-Su, R. Timofte, and A. Pouget. Fast camera image denoising on mobile gpus with deep learning, mobile ai 2021 challenge: Report. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2515–2524, 2021.

[33] A. Ignatov, C.-M. Chiang, H.-K. Kuo, A. Sycheva, and R. Timofte. Learned smartphone isp on mobile npus with deep learning, mobile ai 2021 challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2503–2514, 2021.

[34] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 3277–3285, 2017.

[35] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool. Wespe: weakly supervised photo enhancer for digital cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 691–700, 2018.

[36] A. Ignatov, G. Malivenko, R. Timofte, Y. Tseng, Y.-S. Xu, P.-H. Yu, C.-M. Chiang, H.-K. Kuo, M.-H. Chen, C.-M. Cheng, et al. Pynet-v2 mobile: Efficient on-device

photo processing with neural networks. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 677–684. IEEE, 2022.

[37] A. Ignatov, G. Perevozchikov, R. Timofte, et al. Learned smartphone isp on mobile gpus, mobile ai 2025 challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025.

[38] A. Ignatov, A. Sycheva, R. Timofte, Y. Tseng, Y.-S. Xu, P.-H. Yu, C.-M. Chiang, H.-K. Kuo, M.-H. Chen, C.-M. Cheng, et al. Microisp: processing 32mp photos on mobile devices with deep learning. In *European Conference on Computer Vision*, pages 729–746. Springer, 2022.

[39] A. Ignatov, R. Timofte, S. Liu, C. Feng, F. Bai, X. Wang, L. Lei, Z. Yi, Y. Xiang, Z. Liu, et al. Learned smartphone isp on mobile gpus with deep learning, mobile ai & aim 2022 challenge: report. In *European Conference on Computer Vision*, pages 44–70. Springer, 2022.

[40] A. Ignatov, R. Timofte, Z. Zhang, M. Liu, H. Wang, W. Zuo, J. Zhang, R. Zhang, Z. Peng, S. Ren, et al. Aim 2020 challenge on learned image signal processing pipeline. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 152–170. Springer, 2020.

[41] A. Ignatov, L. Van Gool, and R. Timofte. Replacing mobile camera isp with a single deep learning model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 536–537, 2020.

[42] W. Jeong and S.-W. Jung. Rawtobit: A fully end-to-end camera isp network. In *European Conference on Computer Vision*, pages 497–513. Springer, 2022.

[43] A. Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018.

[44] H. Jung, Z. Hui, L. Luo, H. Yang, F. Liu, S. Yoo, R. Ranjan, and D. Demandolx. Anyflow: Arbitrary scale optical flow with implicit neural representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5455–5465, 2023.

[45] R. M. Karp. Reducibility among combinatorial problems. In *50 Years of Integer Programming 1958-2008: from the Early Years to the State-of-the-Art*, pages 219–241. Springer, 2009.

[46] B.-H. Kim, J. Song, J. C. Ye, and J. Baek. Pynet-ca: enhanced pynet with channel attention for end-to-end mobile image signal processing. In *European Conference on Computer Vision*, pages 202–212. Springer, 2020.

[47] H. Kim and K. M. Lee. Learning controllable isp for image enhancement. *IEEE Transactions on Image Processing*, 33:867–880, 2023.

[48] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61 1:1–11, 1971.

[49] B. Lee, F. Lei, H. Chen, and A. Baudron. Bokeh-loss gan: multi-stage adversarial training for realistic edge-aware bokeh. In *European Conference on Computer Vision*, pages 619–634. Springer, 2022.

[50] F. Li, W. Hou, and P. Jia. Rmfa-net: A neural isp for real raw to rgb image reconstruction. *arXiv preprint arXiv:2406.11469*, 2024.

[51] Y. Li, Y. Zhang, R. Timofte, L. Van Gool, Z. Tu, K. Du, H. Wang, H. Chen, W. Li, X. Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1905–1921, 2023.

[52] Z. Li, Z. Chen, J. Xu, X. Liu, and J. Jiang. Litedepth: digging into fast and accurate depth estimation on mobile devices. In *European Conference on Computer Vision*, pages 507–523. Springer, 2022.

[53] W. Ljungbergh, J. Johnander, C. Petersson, and M. Felsberg. Raw or cooked? object detection on raw images. In *Scandinavian Conference on Image Analysis*, pages 374–385. Springer, 2023.

[54] B. A. Maxwell, S. Singhania, H. Fryling, and H. Sun. Log rgb images provide invariance to intensity and color balance variation for convolutional networks. In *BMVC*, pages 635–642, 2023.

[55] B. A. Maxwell, S. Singhania, A. Patel, R. Kumar, H. Fryling, S. Li, H. Sun, P. He, and Z. Li. Logarithmic lenses: Exploring log rgb data for image classification. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 17470–17479, 2024.

[56] D. Menon, S. Andriani, and G. Calvagno. Demosaicing with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing*, 16(1):132–141, 2007.

[57] I. Morawski, Y.-A. Chen, Y.-S. Lin, S. Dangi, K. He, and W. H. Hsu. Genisp: Neural isp for low-light machine cognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 630–639, 2022.

[58] T. Müller, A. Evans, C. Schied, and A. Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

[59] G. Perevozchikov, N. Mehta, M. Afifi, and R. Timofte. Rawformer: Unpaired raw-to-raw translation for learnable camera isps. In *European Conference on Computer Vision*, pages 231–248. Springer, 2024.

[60] D. W. Raimundo, A. Ignatov, and R. Timofte. Lan: Lightweight attention-based network for raw-to-rgb smartphone image processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 808–816, 2022.

[61] B. Ren, Y. Li, N. Mehta, R. Timofte, H. Yu, C. Wan, Y. Hong, B. Han, Z. Wu, Y. Zou, et al. The ninth ntire 2024 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6595–6631, 2024.

[62] Y. Ren, H. Jiang, M. Yang, W. Li, and S. Liu. Ispdiffuser: Learning raw-to-srgb mappings with texture-aware diffusion models and histogram-guided color consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 6722–6730, 2025.

[63] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[64] E. Schwartz, R. Giryes, and A. M. Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, 28(2):912–923, 2018.

[65] A. Shekhar Tripathi, M. Danelljan, S. Shukla, R. Timofte, and L. Van Gool. Transform your smartphone into a dslr camera: Learning the isp in the wild. In *European Conference on Computer Vision*, pages 625–641. Springer, 2022.

[66] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[67] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[68] I. Skorokhodov, S. Ignatyev, and M. Elhoseiny. Adversarial generation of continuous images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10753–10764, 2021.

[69] M. Turab. A comprehensive survey on image signal processing approaches for low-illumination image enhancement. *arXiv preprint arXiv:2502.05995*, 2025.

[70] Y. Wang, T. Xu, Z. Fan, T. Xue, and J. Gu. Adaptiveisp: Learning an adaptive image signal processor for object detection. *Advances in Neural Information Processing Systems*, 37:112598–112623, 2024.

[71] C.-T. Wu, L. F. Isikdogan, S. Rao, B. Nayak, T. Gerasimow, A. Sutic, L. Ain-kedem, and G. Michael. Visionisp: Repurposing the image signal processor for computer vision applications. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4624–4628. IEEE, 2019.

[72] R. Xu, C. Chen, J. Peng, C. Li, Y. Huang, F. Song, Y. Yan, and Z. Xiong. Toward raw object detection: A new benchmark and a new model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13384–13393, 2023.

[73] X. Xu, Y. Ma, and W. Sun. Towards real scene super-resolution with raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1723–1731, 2019.

[74] X. Xu, Y. Ma, W. Sun, and M.-H. Yang. Exploiting raw images for real-scene super-resolution. *IEEE transactions on pattern analysis and machine intelligence*, 44(4):1905–1921, 2020.

[75] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)*, 35(2):1–15, 2016.

[76] M. Yoshimura, J. Otsuka, A. Irie, and T. Ohashi. Dynamicisp: dynamically controlled image signal processor for image recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12866–12876, 2023.

[77] M. Yoshimura, J. Otsuka, and T. Ohashi. Pqdynamicisp: Dynamically controlled image signal processor for any image sensors pursuing perceptual quality. *arXiv preprint arXiv:2403.10091*, 2024.

[78] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020.

[79] X. Zhang, Q. Chen, R. Ng, and V. Koltun. Zoom to learn, learn to zoom. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3762–3770, 2019.

[80] Y. Zhang, Y. Zhang, Z. Zhang, M. Zhang, R. Tian, and M. Ding. Isp-teacher: image signal process with disentanglement regularization for unsupervised domain adaptive dark object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7387–7395, 2024.