Enabling RAW Image Classification Using Existing RGB Classifiers

Rasmus Munksø[1,*], Mathias Viborg Andersen[1,*], Lau Nørgaard[2], Andreas Møgelmose[1] and Thomas B. Moeslund[1]

[1]*Visual Analysis and Perception Lab, Aalborg University, Denmark*

[2]*Phase One A/S, Denmark*

Keywords: RAW, RGB, Transfer Learning, RAW Image Dataset, Classification.

Abstract: Unprocessed RAW data stands out as a highly valuable image format in image editing and computer vision due to it preserving more details, colors, and a wider dynamic range as captured directly from the camera's sensor compared to non-linearly processed RGB images. Despite its advantages, the computer vision community has largely overlooked RAW files, especially in domains where preserving precise details and accurate colors are crucial. This work addresses this oversight by leveraging transfer learning techniques. By exploiting the vast amount of available RGB data, we enhance the usability of a limited RAW image dataset for image classification. Surprisingly, applying transfer learning from an RGB-trained model to a RAW dataset yields impressive performance, reducing the dataset size barrier in RAW research. These results are promising, demonstrating the potential of cross-domain transfer learning between RAW and RGB data and opening doors for further exploration in this area of research.

1 INTRODUCTION

Historically, the use of RAW images has mainly been exploited by photographers wishing to post-process their captures without losing quality before converting to RGB images. Despite the advantages to photographers, RAW files have largely been overlooked in the computer vision community. RAW Bayer sensor images are minimally processed and thus preserve all the details, colors and dynamic range that is captured by the camera's sensor (Yuan and Sun, 2011). The image signal processing (ISP) pipeline responsible for converting the RAW image to RGB involves non-linear operations, such as linearization of the sensor output, white balancing, tone mapping and gamma correction, as displayed in Figure 1, while also typically reducing the bit depth of each channel (Salih et al., 2012; Can and Brown, 2019). This ISP is designed for human consumption, rather than capturing accurate physical descriptions of the scene (Nguyen and Brown, 2017; Wei et al., 2021; Zhang et al., 2021; Nam et al., 2022). Collectively, the ISP results in the RGB image becoming a representation that inevitably loses some of the original capture information and loses a linear relationship to the physical brightness of the scene (Salih et al., 2012; Can and Brown, 2019; Wei et al., 2021).

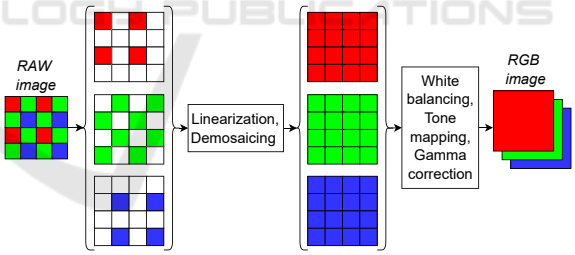

Figure 1: Simplified overview of the processes involved in RAW to RGB image conversion (Kantas et al., 2023).

Given these considerations, it seems glaringly counter-intuitive that information is discarded without careful consideration. Photographers benefit most from RAW image post-processing in challenging scenarios like extreme exposures and low- or high-contrasts, where RAW images retain more details about shadows and highlights, allowing editors to recover otherwise lost information in the RGB images. One might hypothesize that this loss of significant data is not limited to these scenarios, but occurs in every RGB image at different degrees of severity. Consequently, it would seem logical to leverage the additional information in the RAW format in the context

*These authors contributed equally

Munksø, R., Andersen, M., Nørgaard, L., Møgelmose, A. and Moeslund, T.
Enabling RAW Image Classification Using Existing RGB Classifiers.
DOI: 10.5220/0012363200003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 2: VISAPP*, pages 123-129
ISBN: 978-989-758-679-8; ISSN: 2184-4321

of retaining the highest degree of quality in the image data.

RGB is the predominately used image color representation in computer vision. A significant factor is that RGB images have benefited from advancements in image processing techniques applied to RGB images and from wide-spread integration within image processing and deep learning workflows and libraries. Another factor is due to the limited availability of RAW image datasets compared to the presence of large-scale annotated RGB datasets, such as ImageNet (Deng et al., 2009). Furthermore, RAW image data suffers from a lack of standardization, as the RAW formats vary between camera manufacturers and models, whereas the RGB format follows a well-defined standard. Perhaps less important is the fact that RGB images are optimized for human viewing and thus more intuitive to work with than RAW images. However, it is relevant to explore whether RGB images are the optimal image format for all computer vision tasks, especially in domains where preserving fine-grained details and accurate color information is critical. This work aims to address some of these challenges with using RAW images, by employing transfer learning to draw on the abundance of available RGB data, in order to better utilize a small-scale RAW image dataset for the task of image classification. Models trained from scratch are compared to transfer-learned models to verify the change in performance.

Contribution. The main contributions are:

- We introduce an annotated and publicly available dataset consisting of pairwise RAW images and their corresponding RGB counterparts, that can be adapted for a broad range of applications.
- We demonstrate that applying transfer-learning from an RGB-trained model on a small-scale RAW dataset results in surprisingly good performance, while lowering the dataset size barrier for beginning research into using RAW images.

2 RELATED WORKS

In order to work with RAW images effectively, having access to an appropriate RAW image dataset is necessary. A thorough search of the existing publicly available RAW image datasets was conducted and seven datasets were found (Bychkovsky et al., 2011; Omid-Zohoor et al., 2014; Dang-Nguyen et al., 2015; Chen et al., 2018; Zhang et al., 2019; Liang et al., 2020; Morawski et al., 2022; Kantas et al., 2023). However, all of these datasets are limited in size.

Currently, to the best of the authors' knowledge, there is no open-source, large-scale RAW Bayer image dataset, the lack of which severely hinders research into using RAW images. Therefore, it is essential to acquire more data in order to validate findings and expand RAW image usage. Some studies have attempted to address this shortage by so-called RAW image reconstruction, converting existing large-scale RGB datasets into RAW counterparts. (Nguyen and Brown, 2017) propose storing RGB-RAW mapping parameters in JPEG metadata for future RAW reconstruction. (Brooks et al., 2019) reverse the ISP process using camera information like color correction matrices and digital gains. Other lines of research employ learning-based techniques to synthesize RAW data from sRGB images (Liu et al., 2020; Punnappurath and Brown, 2020; Waqas Zamir et al., 2020; Wei et al., 2021). However, these methods rely on the lossy in-camera ISP pipeline, resulting in slight inaccuracies compared to the originals. (Xing et al., 2021) have explored replacing existing ISP pipelines with invertible ISPs, achieving nearly perfect RAW image reconstruction. Although this approach is valid, it does not allow for research into RAW images until more RAW image data has been gathered and processed by their invertible ISP.

Recent research has explored the use of RAW images in specific vision tasks. (Zhang et al., 2019) focuses on improving computational zoom using RAW images. (Chen et al., 2018) aims to train models for low-light image processing from RAW to final output, while (Liang et al., 2020) investigates RAW images for deblurring. These studies collectively demonstrate that RAW images can outperform RGB in tasks relying on linear scene radiance properties. Moreover, object detection with RAW images has been studied and shows promise, either by direct RAW image implementation or by applying learnable non-linear functions as extensions to neural networks (Li et al., 2022; Morawski et al., 2022; Zhang et al., 2022; Ljungbergh et al., 2023). (Kantas et al., 2023) recently explored image classification using RAW images, showing similar performance for RAW and RGB images, while demonstrating a significant speed-up for RAW images by bypassing the ISP pipeline. The tests in (Kantas et al., 2023) were conducted on a custom, minimally processed RAW image dataset. Results were demonstrated by training a ResNet-50 model from scratch on different representations of RAW image data. These results were compared to an identical model trained on a corresponding dataset of 8- and 16-bit RGB images.

To the best of the authors' knowledge, no current research has looked into combining the RAW and

RGB domains by using transfer learning. Using a smaller RAW image dataset together with an RGB to RAW transfer-learning approach to perform a vision task, it can be determined whether the dataset size barrier in RAW image research can be mitigated. And so, this paper will focus on investigating the feasibility of that approach and whether transfer-learning from RGB models can improve performance in RAW image classification.

3 METHODS

In order to investigate the potential of the RAW image format within image classification it would seem logical to classify data only using the RAW image format on models trained from scratch. However, given the scarcity of large-scale RAW image datasets, we hypothesize that the classification performance on a small dataset can be improved by applying RGB-to-RAW transfer learning.

We investigate whether a small-scale RAW dataset can be trained for image classification while utilizing transfer learning from an RGB-trained model. Therefore, some considerations of how to practically use and implement RAW images into a computer vision workflow are in order.

3.1 Using RAW Data

The first challenge associated with using RAW images is the different file formats that are output from the camera. Very few if any of these are directly supported within commonly used image processing libraries, such as OpenCV and Pillow, or deep learning libraries, such as TensorFlow or PyTorch. However, specific image processing libraries exist that do support RAW image formats. In this work, the Python wrapper for LibRaw called RawPy (Riechert, 2014) is used to extract the RAW image data as a simple 2-dimensional data matrix without any processing applied.

The next challenge is storing the RAW image data. While possible, storing a RAW image in a lossless standard RGB format such as PNG or TIFF would be inefficient. Additionally, this work uses a different representation of RAW data obtained by *packing* the original RAW data, where each RAW image is rearranged into the four color channels present in the Bayer pattern. As this results in a 4-channel RAW image representation, standard image formats seem unsuitable. For this reason, both types of RAW images are stored as NumPy array files in this work. The practice of saving RAW image data in NumPy files has the added benefit of making it straightforward to efficiently load data when standardized data loaders in libraries such as TensorFlow or PyTorch do not support the chosen image file format. The approach taken in this work is to load each NumPy file and concatenate the data to one large array, from which a dataset loader can be created in both PyTorch and TensorFlow.

3.2 Dataset

To use the full-resolution RAW images from the datasets (research on datasets can be found in Section 2) as inputs to a classifier, downscaling would be required, which would artificially process the original image, resulting in a lower-quality representation. This would be counter-intuitive for this work since one of the primary reasons for using RAW images is that they contain unaltered capture information. For these reasons, a more relevant dataset for investigating the hypothesis consists of small, unaltered RAW images. Since such a dataset is not publicly available, we create a custom dataset based on classes within the PASCALRAW dataset (Omid-Zohoor et al., 2014). The PASCALRAW dataset has the added benefit of being captured using a Nikon D3200 DSLR camera and therefore also being stored in the same Nikon Raw Image file format (NEF), removing the need for considerations for differences in capture data and image formats.

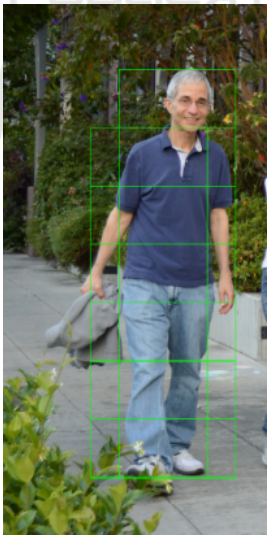

Figure 2: Grid of crops overlapping the semantic mask with at least 50 % of its pixels.

Our dataset is created by extracting 448x448-

https://www.kaggle.com/datasets/mathiasviborg/pascalraw-derived-object-cropped-dataset

sized cropped images of people, bicycles, cars and backpacks that are present in each full-resolution RAW- and corresponding RGB images within the PASCALRAW dataset. The dataset-generating process utilizes YOLOv5 (Jocher et al., 2022) to apply instance segmentation to extract masks of each of the desired four classes that are present in each RGB image. By then placing a grid of cropped images of size 448x448 over each object's entire bounding box, each crop overlapping itself with a stride of 224 pixels along its x- and y-direction, see Figure 2, and only storing those crops that overlap the semantic mask with at least 50 % of its pixels, data samples from both the RGB- and corresponding RAW image are extracted, see Figure 3.

Figure 3: Overview of the first four extracted samples from the object presented in Figure 2. Shown in Original RGB and Original RAW (RAW displayed using Matplotlib).

As one might expect, the majority of these samples do not contain the entire object. However, in this dataset, the aim is to store samples showing approximately a quarter of the object at minimum as a trade-off between maintaining adequate semantic content within each sample and extracting a sufficient number of samples for the dataset. This is achieved by limiting the size of the bounding boxes to be considered and manual inspection of the gathered samples.

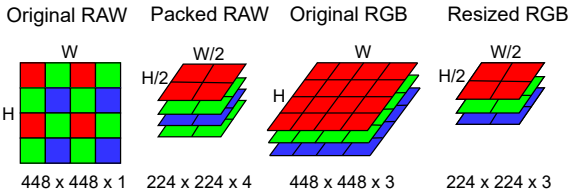

Figure 4: Overview of the different image representations in the dataset as well as their respective dimensions.

After *packing* the RAW image samples the height and width of the samples are halved, which by design corresponds to the input resolution of the chosen classifier, ResNet-50 (He et al., 2016). The RGB samples are simply resized and all the four types of samples are stored in a dataset consisting of data formats as shown in Figure 4.

Figure 5: The 4 classes in the dataset. From left: *Backpack*, *Bicycle*, *Car* and *Person*. Shown in Original RGB and Original RAW (RAW displayed using Matplotlib).

The distribution of the classes within the dataset can be found in Table 1 and examples of samples from each class are shown in Figure 5.

Table 1: Sample distribution of the dataset.

Backpack	Bicycle	Car	Person
3569	5142	4856	4864

3.3 Implementation

ResNet-50 pretrained on ImageNet (Deng et al., 2009) is chosen as the CNN model as it is considered a well-performing general purpose classifier and is trained on RGB data. The transfer-learning fine-tuning is run on an Nvidia A40 GPU. The hyperparameters of the networks, as found in Table 2, are determined by a hyperparameter sweep conducted using the Weights & Biases Bayesian Hyperparameter Optimization tool (Biewald, 2020). The hyperparameters were then tested to ensure performance and stability during training.

Table 2: Overview of the hyperparameters used for transfer learning vs. learning from scratch.

	RAW	PACKED RAW	RGB	RESIZED RGB
Trained via. Transfer Learning				
Optimizer	Adam	Adam	Adam	Adam
Learning Rate	0.0427	0.0489	0.0628	0.0409
Batch Size	8	128	32	32
Trained from Scratch				
Optimizer	Adam	Adam	Adam	Adam
Learning Rate	0.0428	0.0715	0.0629	0.0319
Batch Size	8	8	32	8

The chosen architecture can be seen on Figure 6 and trainable parameters on Table 3.

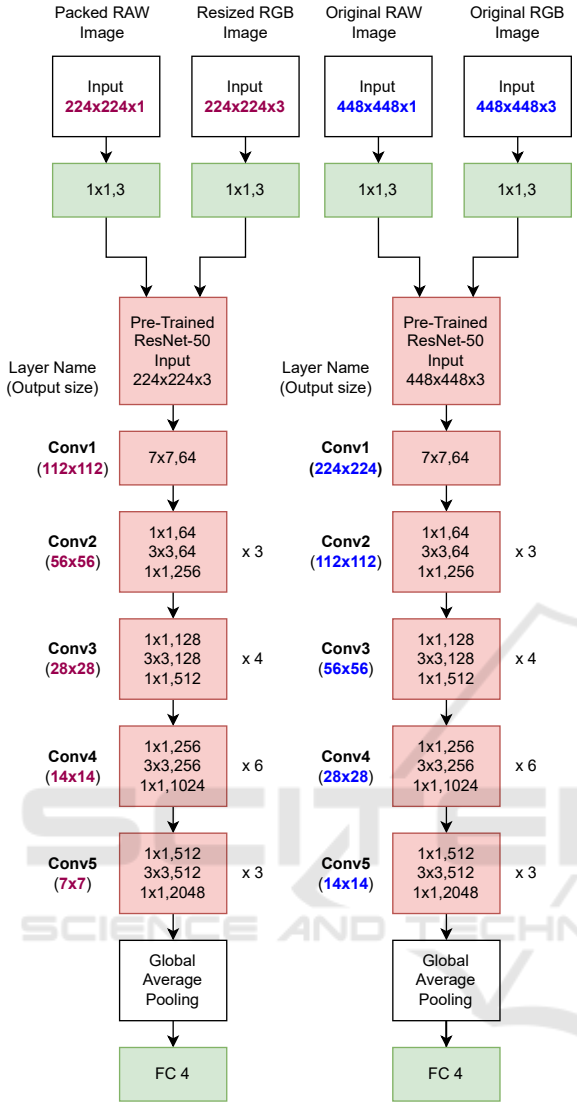

Figure 6: The network architectures for each image representation. Note that models are trained separately for each image representation and that inputs are not fused (simply visualized as such for compactness). Visualized with ResNet-50 architecture (He et al., 2016) shown with the residual units, the size of the filters and the outputs of each convolutional layer. Key: The notation k × k, n in the convolutional layer block denotes a filter of size k and n channels. FC 4 denotes the fully connected layer with 4 neurons representing the 4 classes. The number to the right of the convolutional layer block represents the repetition of each unit. Red colored blocks denote frozen layers that are not updated during training when using transfer learning, while green colored blocks denote layers with learnable parameters.

It can be seen from Figure 6 that the only addition to the architecture apart from the RGB-trained ResNet-50 is a convolutional layer with a 1x1 kernel and 3 filters, resulting in the dimensions (224x224x3) for Packed RAW and Resized RGB, which is the default input dimension for the pre-trained model. Note that the output size from each convolutional layer is different for packed and resized as opposed to original input images, while the learned weights in the pre-trained convolutional layers are identical. It has been experimentally observed that the input dimensions of the Original RAW and Original RGB perform well in this architecture, even though the input dimension to the pre-trained ResNet-50 model is different from what it was trained on. As it performed better than introducing a down-sampling convolutional layer, the architecture was kept as is.

Table 3: Number of trainable parameters for transfer learning vs. learning from scratch.

RAW	PACKED RAW	RGB	RESIZED RGB
Trained via. Transfer Learning			
8.202 K	8.202 K	8.208 K	8.208 K
Trained from Scratch			
23.52 M	23.52 M	23.54 M	23.54 M

4 RESULTS

Experiments are conducted to test whether transfer learning from RGB trained models can be used on RAW images. The impact of transfer learning on improving validation accuracies has been thoroughly investigated and compared to training from scratch. The networks are trained 10 times each on the selected data types (Original-RAW, Packed-RAW, RGB, and Resized-RGB) using the previously stated hyperparameters from Tabel 2. The average of the 10 top-1 accuracies is stored as the final top-1 classification accuracy, see Table 4.

Table 4: Mean top-1 classification validation accuracies as measured from the models with lowest validation loss.

	RAW	PACKED RAW	RGB	RESIZED RGB
Trained via. Transfer Learning				
Mean Top-1 Accuracy	96.04 %	94.27 %	97.25 %	96.79 %
Standard Deviation	0.03	0.12	0.09	0.05
Trained from Scratch				
Mean Top-1 Accuracy	85.64 %	82.92 %	86.54 %	87.49 %
Standard Deviation	0.139	0.013	0.027	0.021

5 DISCUSSION AND FUTURE WORKS

The classification accuracy results found in Table 4 demonstrate that cross-domain transfer learning from RGB-trained models has a significant improvement on RAW image classification, with both RAW image representations gaining more than 10 percentage points. The difference in performance for the different RAW representations is in coherence with the found results in (Kantas et al., 2023) as Original RAW performs better than Packed RAW, although by a higher margin. This raises the question of whether the use of Packed RAW is warranted on its own. However, in previous works, RAW images have been represented by combining Original RAW and Packed RAW by a Bidirectional Cross-Modal (BCA) approach (Liang et al., 2020)RAWInstinct. This representation is not investigated in this work, as the primary research objective is not to outperform RGB in classification tasks. However, it is worth noting that this alternative representation might have improved image classification performance compared to RGB, as demonstrated in prior work (Kantas et al., 2023).

Despite the increased performance from transfer learning, the RAW image data performs similarly to but not surpassing the RGB image data. However, the objective is not for RAW to outperform the RGB image classification results. Instead, these results demonstrate that utilizing RGB-trained models on RAW image data can increase the performance of a small-scale RAW image dataset compared to the performance when training from scratch. This approach effectively lowers the dataset barrier, making RAW image research more accessible and time-efficient. Therefore, these findings are promising, showcasing the effectiveness of cross-domain transfer learning between RAW and RGB data and suggesting potential avenues for further exploration in this area.

In this work, we generate a dataset by extracting samples of backpacks, people, bicycles and cars from an existing RAW image object detection dataset (Omid-Zohoor et al., 2014). Due to the nature of extracting small-sized images from objects within large-scale captures, few of the small-sized crops have full-sized objects within their frame, instead showing a cropped-out portion of it. This adds a built-in way of data augmentation, however, it might not be ideal for the purpose of investigating fine details in RAW images compared to RGB images. Furthermore, only a minority of the images are shot in challenging scenarios in the original PASCALRAW dataset and an argument could therefore be made that the created dataset does not fully explore the advantages of RAW images. This could explain the lack of observable improvement in performance when classifying RAW images compared to RGB in the results presented in this work. As the results found in this paper are only demonstrated on this one dataset, further research on more intricate datasets is necessary to generalize these findings effectively and demonstrate which types of images may benefit the most from RAW image classification.

In future works, researchers may endeavor to construct a dataset specifically designed to highlight the advantages of RAW images. Such a dataset could include images captured in challenging lighting scenarios, resulting in severely under- or overexposed images, images with low contrast, among other scenarios. Guided by the premise that finer details are preserved in the RAW image format compared to the RGB format, leveraging the advantages of RAW images in classifying such challenging scenarios could yield substantial benefits. Therefore it would be interesting to research whether the fine details in the RAW image format could result in better classification accuracies through transfer learning compared to those of corresponding RGB images.

6 CONCLUSION

This work shows that using RGB-trained models for RAW image classification is can be effective. The results indicate that transferring knowledge from RGB models significantly improves accuracy. By tapping into this existing knowledge cross-domain transfer learning is demonstrated to not only enhance accuracy for image classification but also potentially make future research into the use of RAW images more accessible.

ACKNOWLEDGEMENTS

The work was financially supported by the AI:Denmark project funded by the Danish Industry Foundation.

REFERENCES

Biewald, L. (2020). Experiment tracking with weights and biases. Software available from wandb.com.

Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., and T Barron, J. (2019). Unprocessing images for learned raw denoising. In *Conference on Computer Vision and Pattern Recognition*.

Bychkovsky, V., Paris, S., and E. Chan, F. D. (2011). Learning photographic global tonal adjustment with a database of input / output image pairs. In *Conference on Computer Vision and Pattern Recognition*.

Can, H. and Brown, M. (2019). Beyond raw-rgb and srgb: Advocating access to a colorimetric image state. In *Color and Imaging Conference*.

Chen, C., Chen, Q., Xiu, J., and Koltun, V. (2018). Learning to see in the dark. In *Conference on Computer Vision and Pattern Recognition*.

Dang-Nguyen, D.-T., Pasquini, C., Conotter, V., and Boato, G. (2015). Raise - a raw images dataset for digital image forensics. In *ACM Multimedia Systems Conference*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Conference on Computer Vision and Pattern Recognition*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition*.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Yifu, Z., Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvK-itDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., and Jain, M. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation.

Kantas, C., Antoniussen, B., Andersen, M. V., Munksø, R., Kotnala, S., Jensen, S. B., Møgelmose, A., Nørgaard, L., and Moeslund, T. B. (2023). Raw instinct: Trust your classifiers and skip the conversion. In *Conference on Pattern Recognition and Artificial Intelligence*.

Li, Z., Lu, M., Zhang, X., Feng, X., Asif, M. S., and Ma, Z. (2022). Efficient visual computing with camera raw snapshots. In *Conference on Computer Vision and Pattern Recognition*.

Liang, C., Chen, Y., Liu, Y., and Hsu, W. H. (2020). Raw image deblurring. In *IEEE Transactions on Multimedia*, volume 24, pages 61–72.

Liu, Y.-L., Lai, W.-S., Chen, Y.-S., Kao, Y.-L., Yang, M.-H., Chuang, Y.-Y., and Huang, J.-B. (2020). Single-image hdr reconstruction by learning to reverse the camera pipeline. In *Conference on Computer Vision and Pattern Recognition*.

Ljungbergh, W., Johnander, J., Petersson, C., and Felsberg, M. (2023). Raw or cooked? object detection on raw images. In *Scandinavian Conference on Image Analysis*.

Morawski, I., Chen, Y.-A., Lin, Y.-S., Dangi, S., He, K., and Hsu, W. H. (2022). Genisp: Neural isp for low-light machine cognition. In *Conference on Computer Vision and Pattern Recognition*.

Nam, S., Punnappurath, A., Brubaker, M. A., and Brown, M. S. (2022). Learning srgb-to-raw-rgb de-rendering with content-aware metadata. In *Conference on Computer Vision and Pattern Recognition*.

Nguyen, R. and Brown, M. (2017). Raw image reconstruction using a self-contained srgb–jpeg image with small memory overhead. In *International Journal of Computer Vision*, volume 126(6), pages 637–650.

Omid-Zohoor, A., Young, C., and D. Ta, B. M. (2014). Pascalraw: Raw image database for object detection in searchworks catalog. In *Stanford Digital Repository*.

Punnappurath, A. and Brown, M. S. (2020). Learning raw image reconstruction-aware deep image compressors. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 42(4), pages 1013–1019.

Riechert, M. (2014). rawpy: Raw image processing for python, a wrapper for libraw. Software available from https://github.com/letmaik/rawpy.

Salih, Y., bt. Md-Esa, W., Malik, A. S., and Saad, N. (2012). Tone mapping of hdr images: A review. In *International Conference on Intelligent and Advanced Systems*.

Waqas Zamir, S., Arora, A., Khan, S., Hayat, M., Shahbaz Khan, F., Yang, M.-H., and Shao, L. (2020). Cycleisp: Real image restoration via improved data synthesis. In *Conference on Computer Vision and Pattern Recognition*.

Wei, Z., Xiangyu, Z., Hongyu, W. S. G., and Xin, L. (2021). Raw bayer pattern image synthesis for computer vision-oriented image signal processing pipeline design. In *Conference on Computer Vision and Pattern Recognition*.

Xing, Y., Qian, Z., and Chen, Q. (2021). Invertible image signal processing. In *Conference on Computer Vision and Pattern Recognition*.

Yuan, L. and Sun, J. (2011). High quality image reconstruction from raw and jpeg image pair. In *International Conference on Computer Vision*.

Zhang, X., Chen, Q., Ng, R., and Koltun, V. (2019). Zoom to learn, learn to zoom. In *Conference on Computer Vision and Pattern Recognition*.

Zhang, X., Zhang, L., and Lou, X. (2022). A raw image-based end-to-end object detection accelerator using hog features. In *IEEE Transactions on Circuits and Systems I: Regular Papers*, volume 69(1), pages 322–333.

Zhang, Z., Wang, H., Liu, M., Wang, R., Zhang, J., and Zuo, W. (2021). Learning raw-to-srgb mappings with inaccurately aligned supervision. In *International Conference on Computer Vision*.