# A Survey to Iamge Signal Processing


Yifan Li
Yuanpei college, Peking University
2100012520@stu.pku.edu.cn


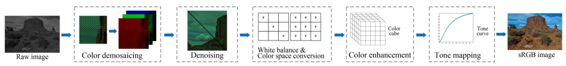


Figure 1. Major components in a traditional camera image signal processing pipeline [6]


## Abstract

*The performance of an ISP plays the key role to improve the quality of sRGB images output from a camera. Traditional image signal processing (ISP) pipeline consists of a set of cascaded hand-crafted image processing modules such as white balance, demosaicing, denoising, sharpening, color space conversion, tone mapping, gamma correction, and others to reconstruct a high-quality sRGB image from the sensor raw data. As the development of deep learnning, many creative methods are being proposed for image signal processing, achieving compelling reconstruction quality. This article summarizes some symbolic classific methods and learning-based approaches proposed for ISP pipeline design.*


## 1. Traditional ISP design

In traditional ISP pipeline design, usually an algorithm is developed for a specific ISP subtask. For example, the demosaicing operation interpolates the raw color filter array (CFA) image with repetitive mosaic pattern into a full color image [1]. As for white balance and color constancy image correction, Cheng *et al*. [2] show that under certain illuminations, a diagonal $3 \times 3$ matrix is capable of correcting not only neutral colors, but all colors in a scene. To improve the resolution, a category of SR approaches learn a mapping between low/high-resolution patches [3, 4, 5].

### 1.1. Tone mapping

Tone mapping transforms the image in the high dynamic range irradiance space to a standard dynamic range image with image structures preserved. In order to deal with tone mapping, algorithms can be categorized into global method and local methods. Global tone mapping methods reproduce an SDR image with a single compressive curve [7, 8]. Some local methods are based on layer decomposition [9], where the base layer is first estimated by an edge-preserving filter and detail layer is the residual between base layer and the original image. Different local tone mapping algorithms mainly differ in the filter design techniques.

Given the fact that a tremendous amount of information is recorded in an HDR radiance map, which part of the information should be assigned a high priority for visual perception is an important question for tone mapping. In psychology, it was found that human vision is more sensitive to edges. Liang *et al*. [10] highlight the importance of the structural information in an image, and impose a structural sparsity prior on the detail layer, that is a hybrid $\ell_1$-$\ell_0$ layer decomposition model. Specifically, the $\ell_0$ term has great piecewise flattening property, better to use for the structural prior. The $\ell_1$ sparsity term has th outlier-rejection nature, the large gradients of the edges in the base layer are preserved, but its piecewice smoothness nature leads to a weak structural prior.

### 1.2. Denoising

Dabov *et al*. [11] propose a sparse 3-D transform-domain collaborative filtering method (BM3D), denominating *grouping* as the concept of collecting similar $d$-dimensional fragments of a given signal into a $d+1$-dimensional data structure named as *group*, to enable the use of a higher dimensional filtering of each group. Filtering between each group exploits the potential similarity, such as correlation and affinity, and estimate the true signal in each of them. Many other approaches have been proposed, such as wavelet-domain processing [12], sparse coding [13, 14], nuclear norm minimization [15].

## 2. Learning-based methods in ISP

Because the traditional ISP is usually designed as a set of hand-crafted modules independently, little attention has been paid to design them as a whole. Such divide-and-conquer strategy is somtime cumbersome and ineffective. Learning-based methods allow to jointly solve multiple tasks, with great potential to alleviate the total computational burden.

### 2.1. Full end-to-end deep neural model for ISP

As pioneering work, Jiang *et al*. impose a local regression framework named $L^3$ [16], where the raw image patches are clustered based on some simple features and then per-class filters are learned to transform the raw patches into the sRGB patches. $L^3$ indicates the key principles of this method: Local, Linear and Learned. This approach however has limited regression performance due to the use of simple parametric models.

Many end-to-end model choose RAW image as their input, for more bits per pixel. Eli Schwartz *et al*. present DeepISP [6], a model learns a mapping from the raw low-light mosaiced image to the final visually compelling image and encompasses low-level tasks, such as demosaicing and denoising, as well as higher-level tasks, such as color correction and image adjustment.

Andrey Ignatov *et al*. demonstrate that even the most sophisticated ISP pipelines can be replaced with a single end-to-end deep learning model trained without any prior knowledge about the sensor and optics used in a particular device. [17]. They present *PyNET*, a pyramidal *CNN* architecture designed for fine-grained image restoration that implicitly learns to perform all ISP steps. Ratnasingam also presents a FCNN to perform the whole ISP pipeline [18].

Imaging in low light is challenging due to low photon count and low SNR. Short-exposure images suffer from noise, while long exposure can induce blur and is often impractical. Chen *et al*. collect a new dataset, *see-in-the-dark* for training and benchmarking, and develop an end-to-end FCN [19].

In many works a single-stage network is straightforwardly trained as an ISP in an end-to-end manner. Liang *et al*. notices that it's hard to optimize the network under this situation, and lead to unsatisfactory learning performance, so they design a two-stage network, called CameraNet [20], to progressively learn the two groups of ISP tasks, restoration and enhancement, giving consideration to both the joint and individual of the whole pipeline.

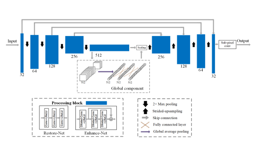

Figure 2. The structure of UNet-like Restore-Net and Enhance-Net modules in the proposed CameraNet system [20]

### 2.2. Deep learning methods for low-level vision

CNN is successful in joint restoration tasks, it is also useful in specific task. Bianco *et al*. use a CNN to predict the scene illumination with color constancy accurately [21]. Zhang *et al*. propose a feed-forward denoising CNN, called DnCNNs [22] with residual learning strategy, which can handle unknown noise level. Gharbi *et al*. also trained a feedforward CNN for joint denoising and demosaicking [23]. Dong *et al*. proposed a deep learning model for single image super-resolution (SR) [24], directly learns an end-to-end mapping between the low/high-resolution images, bring the breakthroughs in accuracy and speed of SR of a single image. Kim *et al*. presented a highly accurate single image super resolution (SR) method, using a very deep convolutional network inspired by VGG-net [25]. They also proposed a simple yet effective training procedure, that is learn residuals only and use extremely high learning rates. Awaring the great performance of deep neural network, some researchers make efforts to modify the architecture of the network. Lim develop an enhanced deep super-resolution network (EDSR) by removing unnecessary modules in conventional residual networks [26]. Tong *et al*. introduces dense skip connections in a very deep network, called SR-DenseNet [27], enable to build short paths from the output to each layer, alleviating the frequent vanishing-gradient problem in deep network. Futhermore, Ledig *et al*. propose SRGAN [28], a generative adversarial network (GAN) with a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes the model's solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. Motivated by perceptual similarity, the content loss replaces the similarity in pixel space.

## 3. Conclusion

Traditional ISP design usually split the ISP pipeline into a set of modules and developing them independently, which requires explicitly heuristics, such as a sparse representation in a redundant dictionary [14], local smoothness [29] and non-local similarity [30]. Those priors may cause error accumulation, and cascaded structure may cause noise amplification. Besides, traditional ISP is difficult to produce high-quality images under challenging scenarios, such as low-light imaging. An advantage of deep learning-based methods is their ability to implicitly learn the statistics of natural images. It needs more investigation on how to train a stable and effective CNN to address the complex mixture of image restoration and enhancement tasks, such as camera ISP learning.

| image enhancement module | traditional methods | deel learning-based methods | | |
|---|---|---|---|---|
| denoising | [11, 12, 13, 14, 15] | [22, 31, 32] | [33] | [16, 6, 17, 18, 19, 20] |
| demosaicing | [1] | [23] | | |
| super resolution | [3, 4, 5] | [24, 25, 26, 27, 28] | | |
| tone mapping | [7, 8, 10, 9] | [34] | | |
| color correction | [2] | [21] | | |

Table 1. Summary of traditional and learning-based methods in different modules.

## References


[1] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016–023016, 2011.

[2] Dongliang Cheng, Brian Price, Scott Cohen, and Michael S. Brown. Beyond white: Ground truth colors for color constancy correction. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[3] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008.

[4] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang. Coupled dictionary training for image super-resolution. *IEEE transactions on image processing*, 21(8):3467–3478, 2012.

[5] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

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

[7] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 13(6):42–48, 1993.

[8] Erik Reinhard and Kate Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE transactions on visualization and computer graphics*, 11(1):13–24, 2005.

[9] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 257–266, 2002.

[10] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[12] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 12(11):1338–1351, 2003.

[13] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *2009 IEEE 12th international conference on computer vision*, pages 2272–2279. IEEE, 2009.

[14] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[15] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014.

[16] Haomiao Jiang, Qiyuan Tian, Joyce Farrell, and Brian A Wandell. Learning the image processing pipeline. *IEEE Transactions on Image Processing*, 26(10):5032–5042, 2017.

[17] Andrey Ignatov, Luc Van Gool, and Radu 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.

[18] Sivalogeswaran Ratnasingam. Deep camera: A fully convolutional neural network for image signal processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[19] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018.

[20] Zhetong Liang, Jianrui Cai, Zisheng Cao, and Lei Zhang. Cameranet: A two-stage framework for effective camera isp learning. *IEEE Transactions on Image Processing*, 30:2248–2262, 2021.

[21] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Color constancy using cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[22] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[23] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016.

[24] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014.

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[26] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[27] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[28] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[29] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[30] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 60–65. Ieee, 2005.

[31] Tal Remez, Or Litany, Raja Giryes, and Alex M Bronstein. Deep class-aware image denoising. In *2017 international conference on sampling theory and applications (SampTA)*, pages 138–142. IEEE, 2017.

[32] Wensen Feng and Yunjin Chen. Fast and accurate poisson denoising with optimized nonlinear diffusion. *arXiv preprint arXiv:1510.02930*, 2015.

[33] Teresa Klatzer, Kerstin Hammernik, Patrick Knobelreiter, and Thomas Pock. Learning joint demosaicing and denoising based on sequential energy minimization. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2016.

[34] Aakanksha Rana, Praveer Singh, Giuseppe Valenzise, Frederic Dufaux, Nikos Komodakis, and Aljosa Smolic. Deep tone mapping operator for high dynamic range images. *IEEE Transactions on Image Processing*, 29:1285–1298, 2019.