

浅析ISP流程：从传统到AI的演进与展望

1. 背景

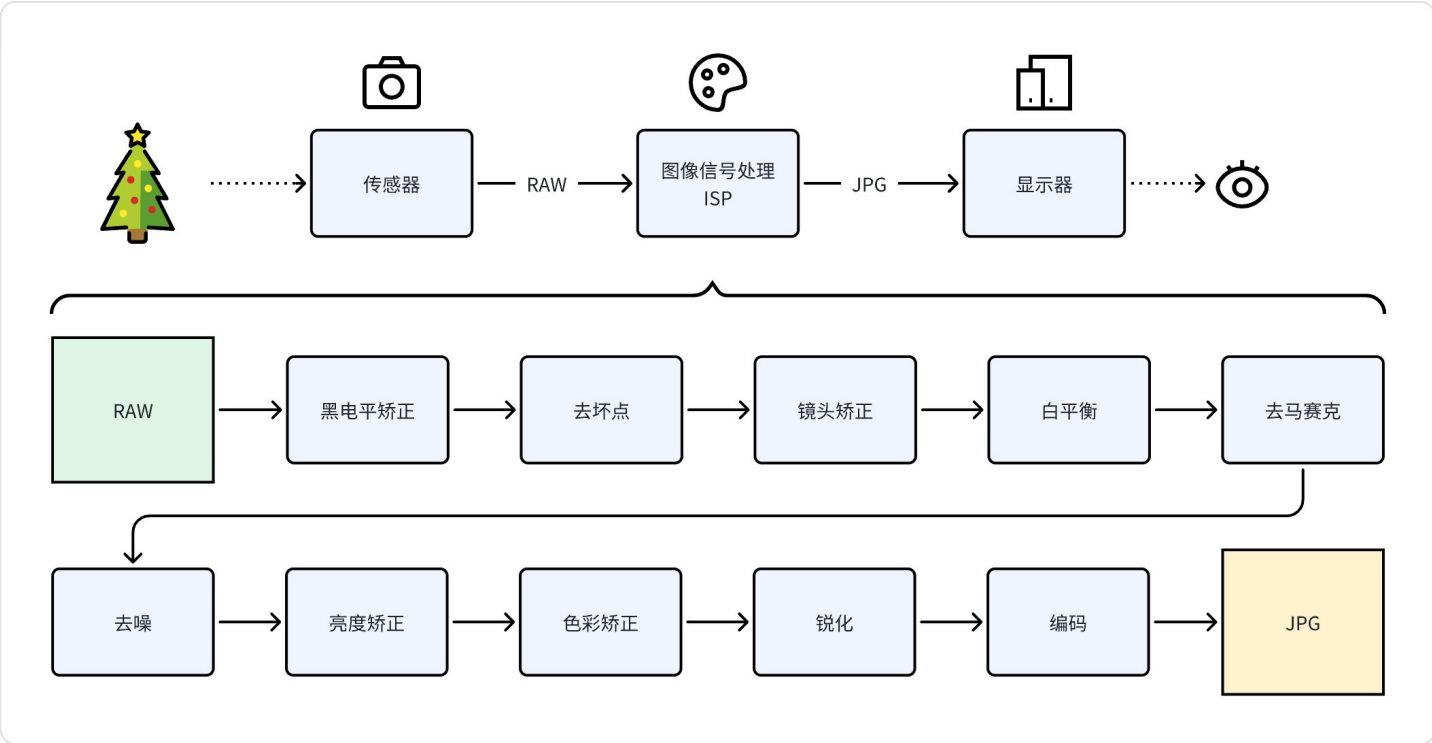
ISP（Image Signal Processor）指的是对图像或视频信号进行实时处理和优化的专用处理器或专用硬件模块。它通常集成在手机、摄像机、无人机、监控设备、车载系统等影像采集系统中，为图像/视频信号提供高质量、低延迟的处理能力。核心目的：对传感器采集的原始图像（RAW）数据进行处理，输出质量更高、细节更丰富的可用图像或视频流（如YUV、RGB等格式）。

传统ISP的发展历程

在20世纪末至21世纪初，ISP主要依赖于固定算法和硬件流水线处理。这些算法包括白平衡校正、色彩插值、降噪、锐化等，广泛应用于数码相机和早期手机中。传统ISP的优势在于其稳定性和高效性，但其固定的处理流程难以适应复杂场景和多样化需求。

AI ISP的崛起

近年来，随着AI的爆发和新兴智能终端的崛起和需求，AI ISP逐渐成为行业热点。通过将神经网络与传统ISP相结合，AI ISP不仅继承了传统ISP的稳定性，还具备了神经网络的自适应性和智能化能力。AI ISP的出现标志着ISP技术进入了一个全新的阶段。



ISP流程

2. ISP的基础介绍

2.1 传统ISP

传统ISP的软件部分依赖于经典算法，其处理流程通常分为以下几个步骤：

1. 黑电平校正（Black Level Correction）

黑电平形成的原因有多种，主要的形成原因如下面两点：

(1) CMOS传感器采集的信息经过一系列转换生成原始RAW格式数据。以8bit数据为例，单个pixel的有效值是0~255，但是实际AD芯片（模数转换芯片）的精度可能无法将电压值很小的一部分转换出来，因此，sensor厂家一般会在AD的输入之前加上一个固定的偏移量，使输出的pixel value在5（非固定）~255之间，目的是为了让暗部的细节完全保留，当然同时也会损失一些亮部细节，由于对于图像来说，我们的关注度更倾向于暗部区域，ISP后面会有很多增益模块（LSC、AWB、Gamma等），因此亮区的一点点损失是可以接受的。

(2) sensor的电路本身会存在暗电流，导致在没有光线照射的时候，像素单位也有一定的输出电压，暗电流这个东西跟曝光时间和gain都有关系，不同的位置也是不一样的。因此在gain增大的时候，电路的增益增大，暗电流也会增强，因此很多ISP会选择在不同gain下减去不同的bl的值。

如多sensor输出raw数据中附加的黑电平值，需要在ISP最前端去干净。如果不去干净，干扰信息会影响后端ISP各模块的处理，尤其会导致AWB容易不准，出现画面整体偏绿或者整体偏红现象。



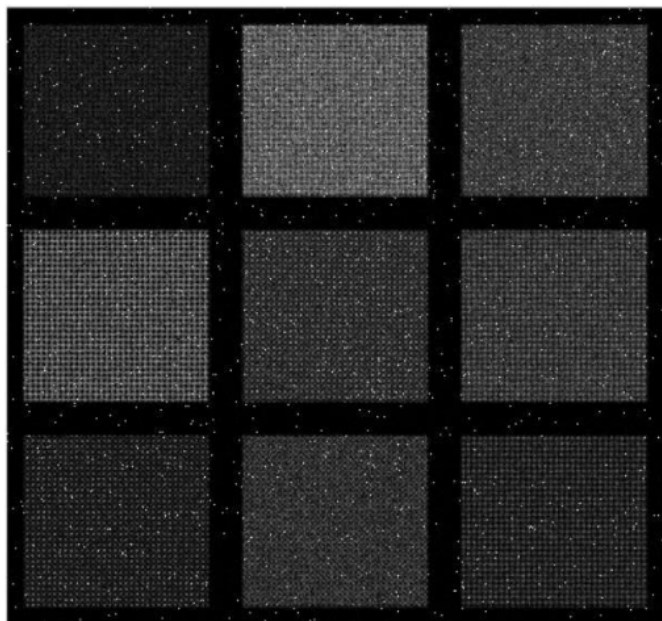
矫正前



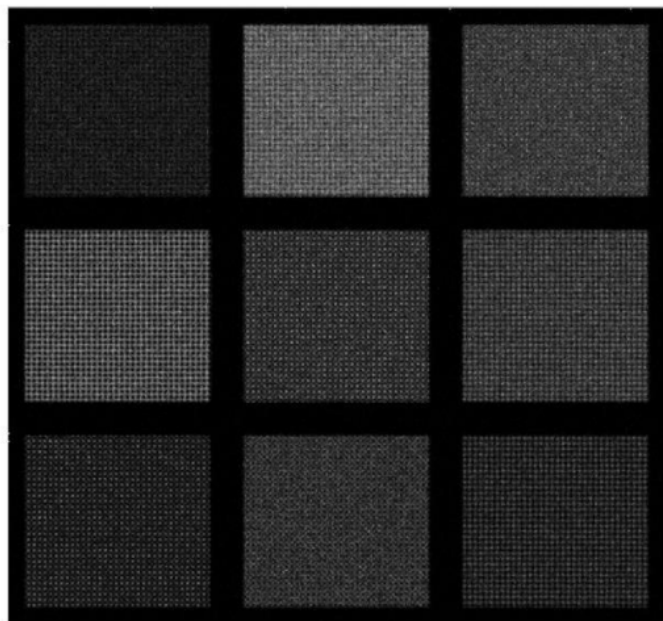
矫正后

2. 去坏点（Defect Pixel Correction）

由于传感器是高精密一起，里面包含的像素点非常多，难免会出现做工问题导致局部像素产生坏点。



去坏点前



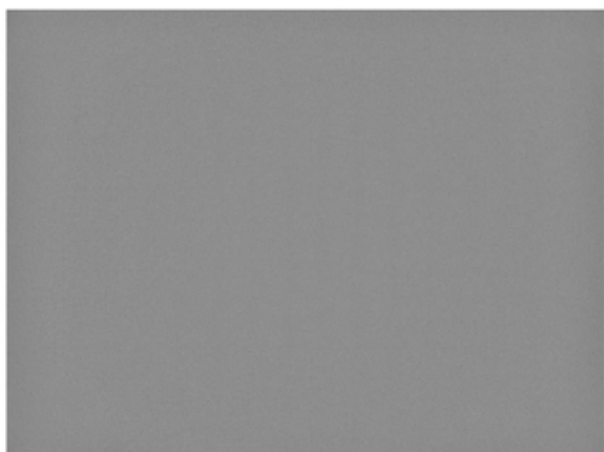
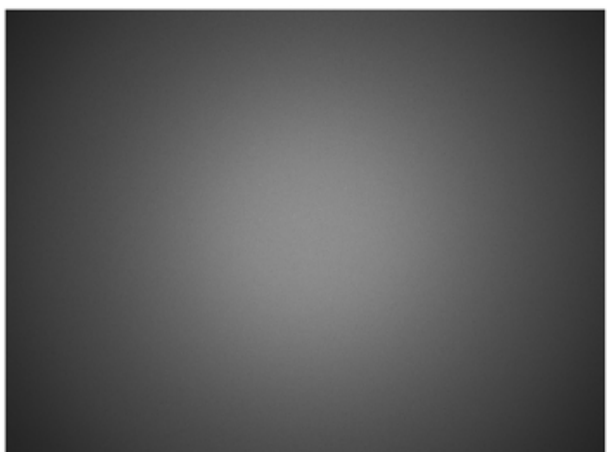
去坏点后。

3. 镜头阴影校正 (Lens Shading Correction)

由于镜头天然的缺陷导致成像会出现暗角，一般分为luma shading和color shading。

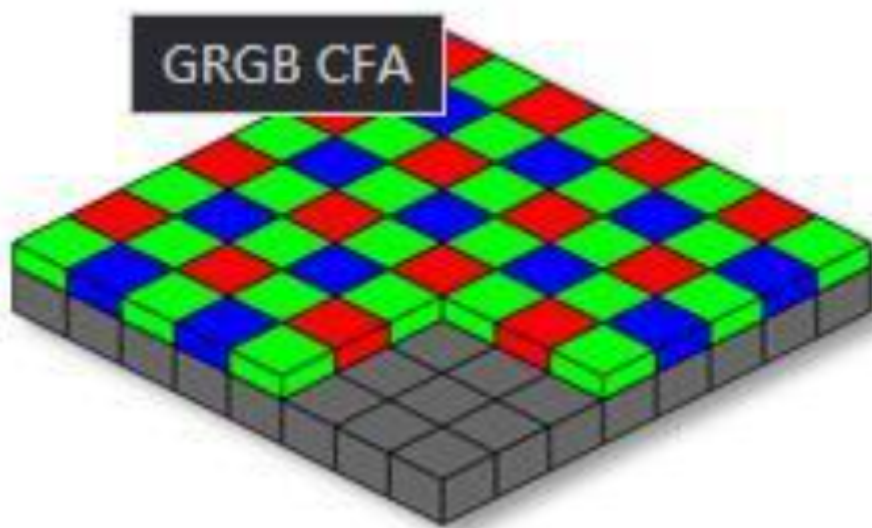
Luma shading的主要原因是镜头中心到边缘的能量衰减导致的，如图所示，蓝色和绿色用相同的数量线条表示能量，中心位置的蓝色几乎所有能量都能达到最右侧的成像单元，但是边缘的绿色由于有一定角度射入，经过镜头的折射，有一部分光（最上方的几条绿色线条）就没法达到成像单元，因此成像单元中心的能量就会比边缘的大，表现在亮度上就是亮度向边缘衰减变暗。

color shading主要是因为不同颜色的光的折射率不同，导致白光经过镜头后达到成像面时不同颜色的光的位置不同导致偏色。当然偏色还会和CRA有关，但是一般镜头选型的时候都会注意和sensor的CRA进行匹配，一般两者不会相差太大，所以CRA导致的偏色不做重点讨论。



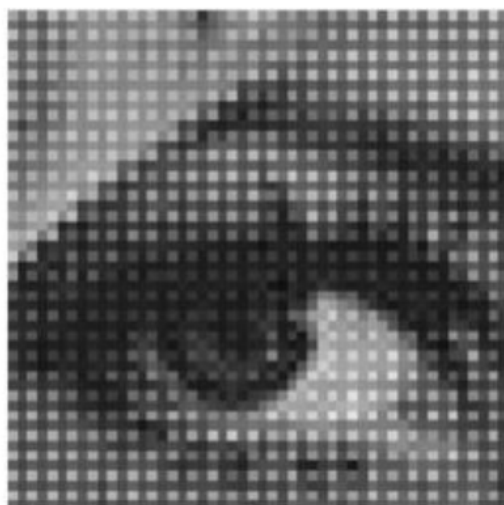
4. 解马赛克 (Demosaicing)

如今几乎所有彩色sensor输出的都是**bayer pattern**的数据，即每个像素点仅有一种颜色分量数据。demosaic(解马赛克)就是为每个像素点找回另外2个缺失的颜色分量，使每个像素点都具备r,g,b三个通道的数据，最终将拜耳阵列中恢复全彩图像。

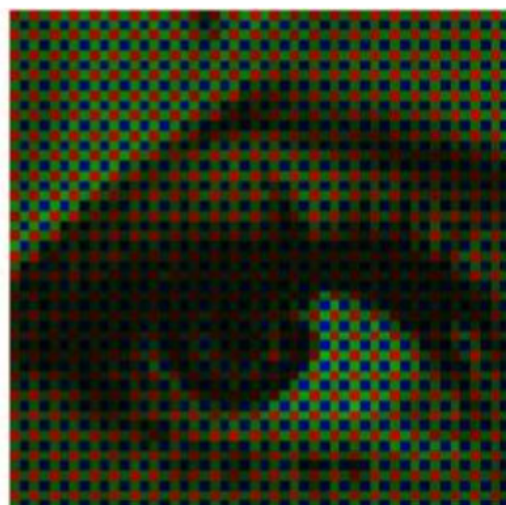


R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B

(a) RGGB模式的CFA



(b) raw图的灰度显示



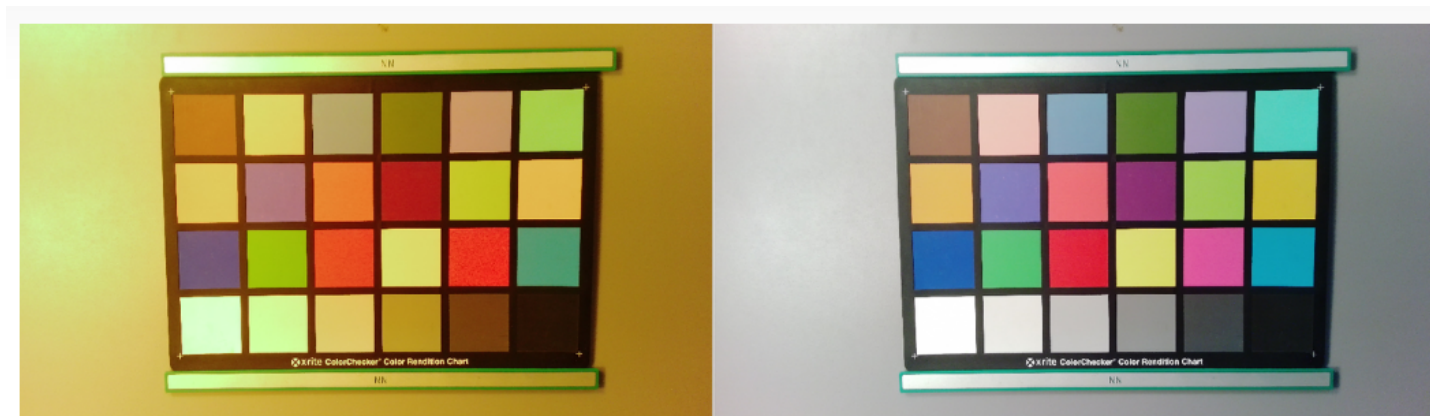
(c) raw的颜色通道



(a) demosaic后的结果

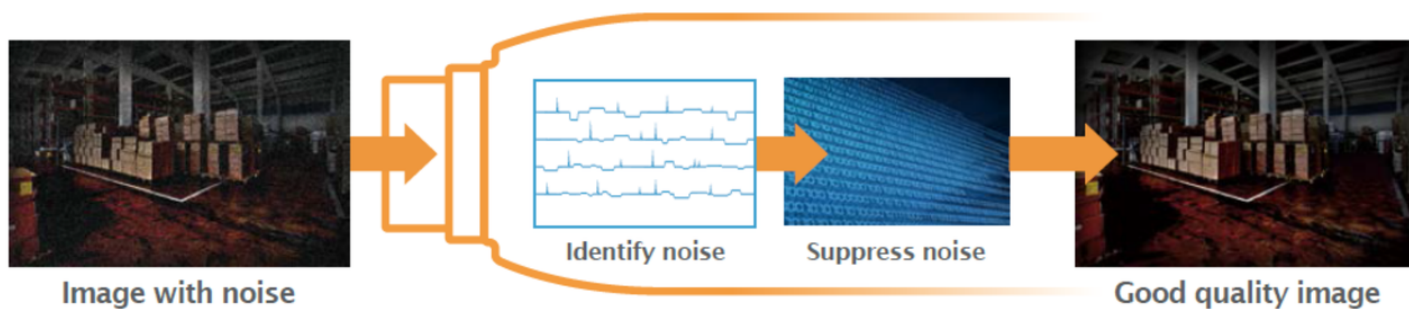
5. 白平衡校正 (White Balance)

人眼具有颜色恒常性，可以避免光源变化带来的颜色变化，但是图像传感器不具备这种特性，从而造成色偏，白平衡就是需要校正这个颜色的偏差

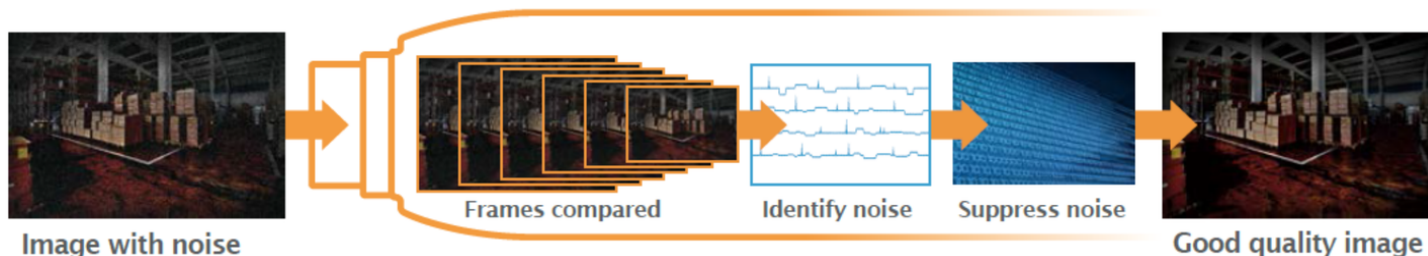


6. 降噪 (Noise Reduction)

空域降噪：空域降噪是一种2D降噪方法，它只处理一帧图像内部的噪声。降噪算法根据实现原理不同可以分成很多种类型，比如线性/非线性、空域/频域，频域又包括小波变换、傅里叶变换或其他变换。在camera ISP 中使用的降噪算法需要足够简单、快速、节省内存、适合硬件实现。经典的低通滤波器如中值滤波或高斯滤波虽然足够简单但是容易破坏图像中的边缘，所以主流ISP产品一般会使用某个加强了边缘保持特性的改进版本，如引导滤波 (guided filter)，双边滤波 (bilateral filter) 等。



时域降噪：时域降噪是一种3D降噪方法，它的主要思想是利用多帧图像在时间上的相关性实现降噪。一种最简单的实现方法是时域均值滤波，即将相邻几帧图像做加权平均。由于累加后噪声的增长速度（根号关系）小于信号的增长速度（线性关系），所以图像的信噪比会提高。这种方法的主要问题在于只适合处理静态图像，如果画面中存在运动的物体则会出现伪影 (ghost effect)。

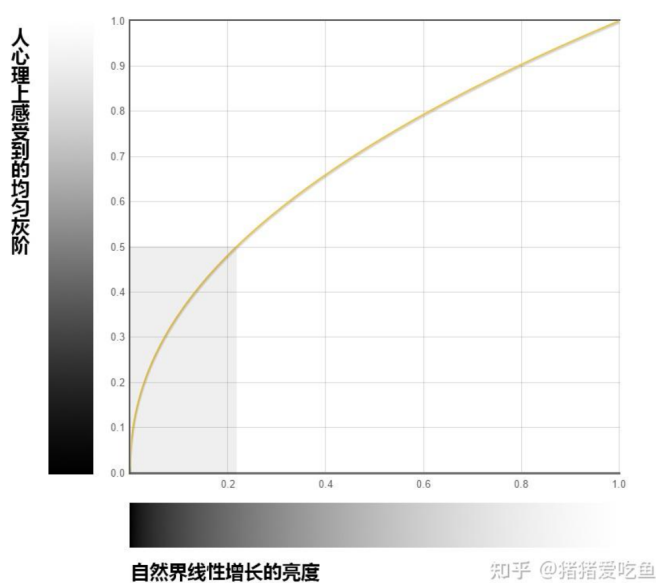


空域时域降噪 (STNR)

STNR是一种2D+3D降噪方法，它通过一套算法判别一个像素是属于前景还是背景，被判决为背景的像素将会参与时域平滑，被判决为前景的像素将会参与空域平滑，而判决条件则每一帧都在动态更新，以尽可能保证判决准确性。



7. Gamma:



人眼感知到的不同亮度和自然界的亮度并非线性对应

更多优质电子书 PPT 报告 课程资料 详询VX: 727266980

传感器接收到的光信号往往是线性的，这实际上与人眼感觉并不相同，人眼感知到的不同亮度和自然界的亮度并非线性对应，而是如上图曲线所示，在暗的时候感知到的亮度比实际的亮度要高，因此为了让最终图片更符合人眼所见，会对图片进行gamma矫正。

Gamma矫正前后结果如下图所示，左图为矫正前，右图为矫正后。



8. 色彩校正矩阵（CCM）

由于相机的颜色匹配特性通常不满足卢瑟条件（也就是说sensor的RGB响应通常不是线性无关的），即相机的颜色匹配特性与CIE标准观察者之间并不存在线性关系。因此，我们需要某种方法来校正相机的特性，使其接近标准观察者。在实际的ISP处理中，这种方法通常以3x3矩阵的形式出现，称为色彩校正矩阵（colour correction matrix）

$$\begin{bmatrix} R_{out} \\ G_{out} \\ B_{out} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} R_{in} \\ G_{in} \\ B_{in} \end{bmatrix}$$

下图是色彩矫正前后的结果，左图是矫正前，右图是矫正后



9. 锐化（Sharpening）

锐化（sharpen）就是使用一定的算法对图像进行处理以检测图像中存在的边缘，然后对符合条件的边缘采取一定的滤波措施以达到提高边缘锐度的目的。

那么，一幅图像是不是锐度越高越好呢？当然不是。从信息的角度看，通过滤波算法提高图像的锐度并不会增加图像中所包含的信息，它只是对原有信息进行一些加工处理，使画面更符合人类视觉的响应特性，可以减轻人类大脑的负担。但是如果算法处理不当，则会将一些原本并不重要甚至并不存在的细节（包括图像噪声）也一起放大，这反而会违背人类的正常认知，会加重大脑的负担，起到反作用。



因此，锐化的主要目标应该是弥补成像过程中图像的锐度损失，从而恢复场景本应具有的自然锐度，同时需要注意避免过度锐化。下面将讨论造成图像锐度损失的一些原因。

2.2 AI ISP

AI ISP是这两年才出来的一个新的技术概念。传统的ISP有三种组织形式，整个pipeline中有不同的模块负责不同的功能。AI ISP已经有很多公司进行研究，比如英特尔实验室的《[Learning to see in the dark](#)》论文中提出，可以通过整个神经网络来实现ISP的全部功能。在这篇论文中，通过一个raw图进入，然后输出RGB或者YUV图像。但是在之前的研究中，没有一家公司能够真正把它进行产品化，现在能实际落地的仅有几个模块：

- 超分辨率
- 白平衡
- 解马赛克
- 暗光降噪
- HDR。

这其中讲一下降噪和HDR，因为这两者是传统ISP中难以进一步优化的点。

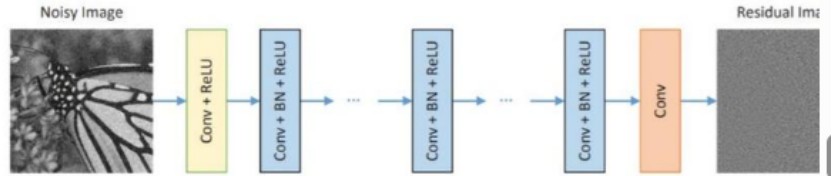
Zhang et al. TIP'17

Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang

Abstract—Discriminative model learning for image denoising has been recently attracting considerable attention due to its favorable denoising performance. In this paper, we take one step forward by investigating the construction of feed-forward denoising convolutional neural networks (DnCNNs) to enhance the progress in very deep architectures, learning algorithms, and regularization method into image denoising. Specifically, residual learning and batch normalization are utilized to speed up the training process as well as boost the denoising performance. Different from the existing discriminative denoising models which usually train a specific model for additive white Gaussian noise (AWGN) at a certain noise level, our DnCNN model is able to handle Gaussian denoising with unknown noise level (i.e., blind Gaussian denoising). With the residual learning strategy, DnCNN implicitly removes the latent clean image in the hidden layers. This property motivates us to train a single DnCNN model to tackle with several general image denoising tasks such as Gaussian denoising, single image super-resolution and JPEG image deblocking. Our extensive experiments demonstrate that our DnCNN model can not only exhibit high effectiveness in several general image denoising tasks, but also be efficiently

random field (MRF) models [10], [11], [12]. In particular, the NSS models are popular in state-of-the-art methods such as BM3D [1], LSSC [1], NCSR [1] and WNSM [13]. Despite their high denoising quality, most of the image prior-based methods typically suffer from two major drawbacks. First, these methods generally involve a complex optimization problem in the testing stage, making the denoising process time-consuming [1], [13]. Thus, most of the prior-based methods can hardly achieve high performance without sacrificing computational efficiency. Second, the models in general are non-convex and involve several manually chosen parameters, providing some leeway to boost denoising performance. To overcome the limitations of prior-based approaches, several discriminative learning methods have been recently developed to learn image prior models in the context of truncated inference procedure. The resulting models are able to get rid of the iterative optimization procedure in the test phase.

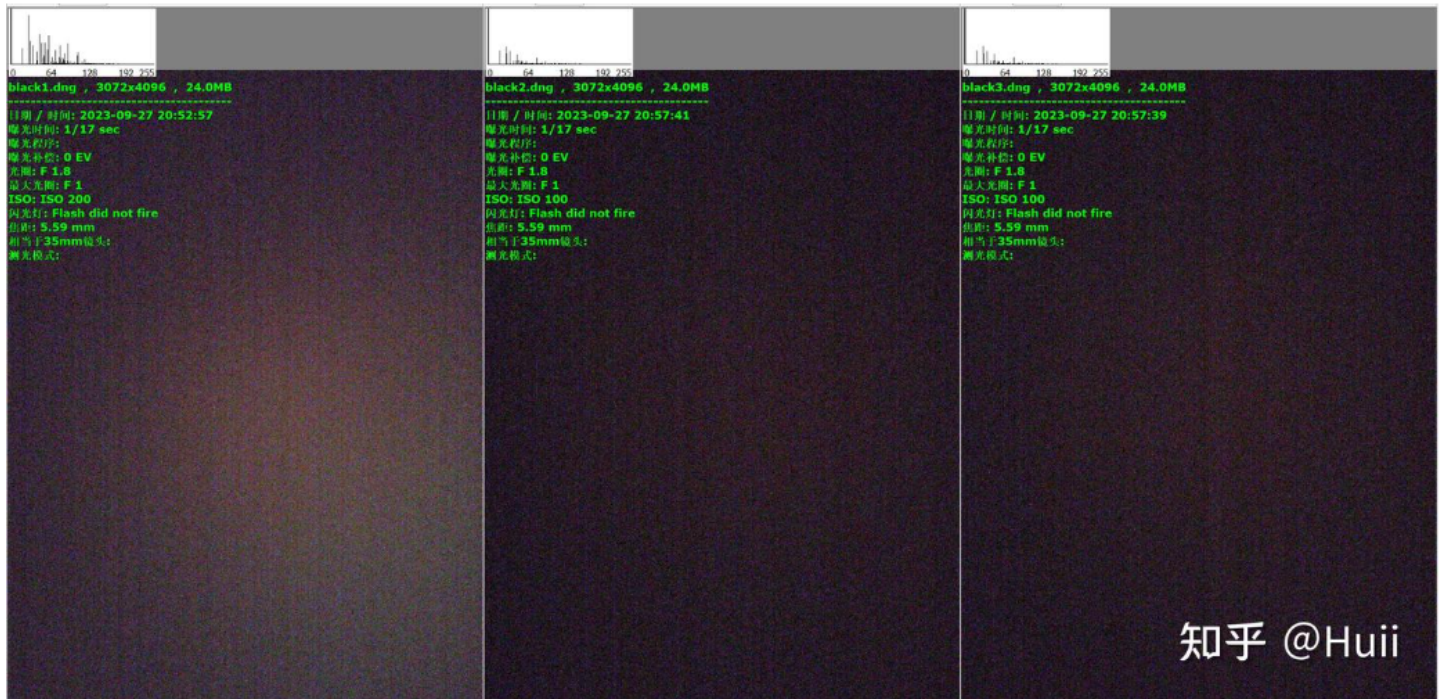


- Straight-forward network based on deep residual learning (Kim SR-ResNet).
- Introduced batch normalization to the network.
- Predicts the residual noise layer.

知乎 @VegDo

降噪

AI降噪并非是预想中的给定噪声图和干净图进行训练，然后可以输入噪声图得到干净图。目前更成熟更稳定的做法是用网络预测噪声，然后把噪声减掉从而实现降噪。如果要取得比较好的效果，则这一步需要在raw上去做，rgb图像的噪声建模很复杂，没有清晰的分布，就不太好造训练数据；尤其在有sensor数据的情况下，对sensor采集基准帧，计算均值方差等数据属性来拟合噪声分布，从而得到目标camera噪声模型的参数。这样预测全场景的噪声效果会更好。

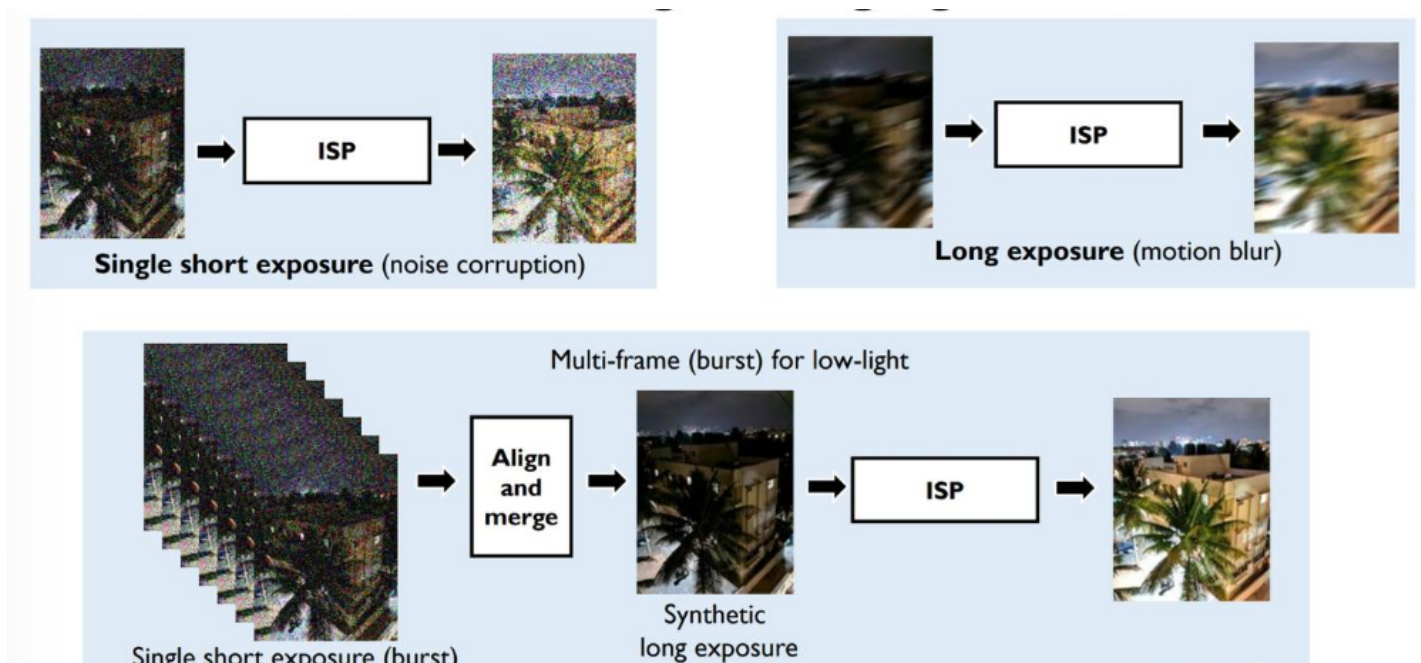


知乎 @Huii

最终效果如下，左图是降噪前，右图是降噪后

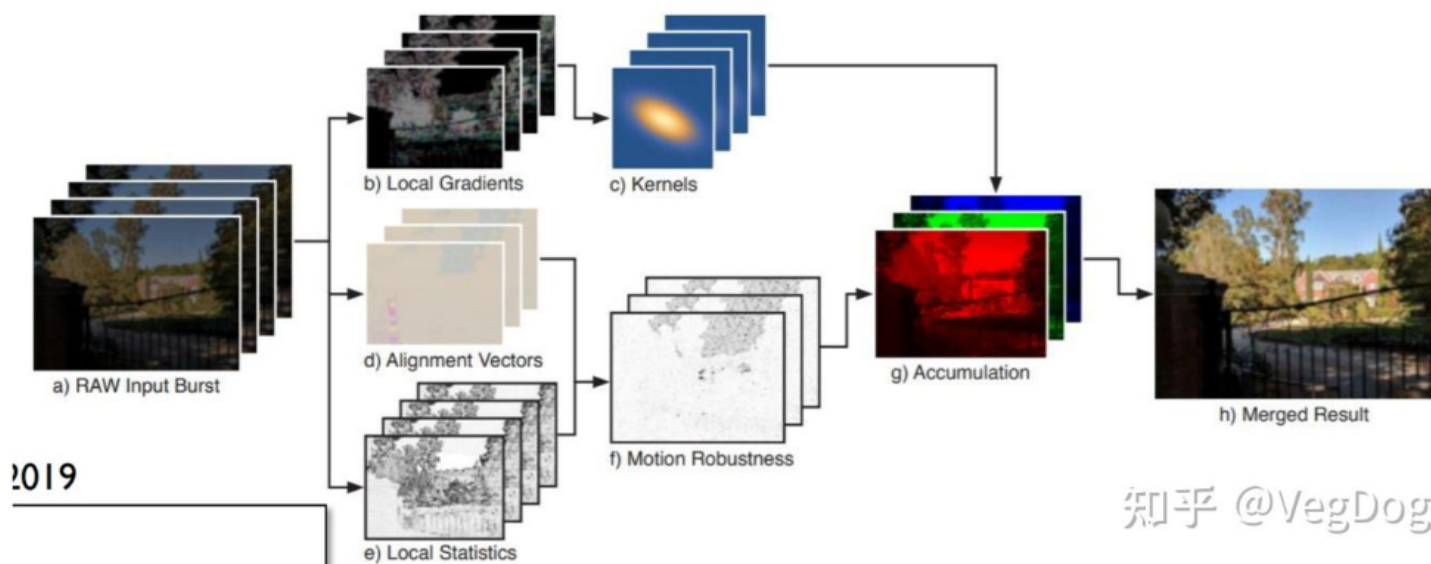


再进一步的，在暗光情况下最常用的方法是通过多帧进行降噪。其实靠提高ISO本身就可以放大采集到的信号，只不过噪声也会随之增大，如果解决了这个过程噪声，就能解决传统ISP降噪中最为头疼的夜景问题。

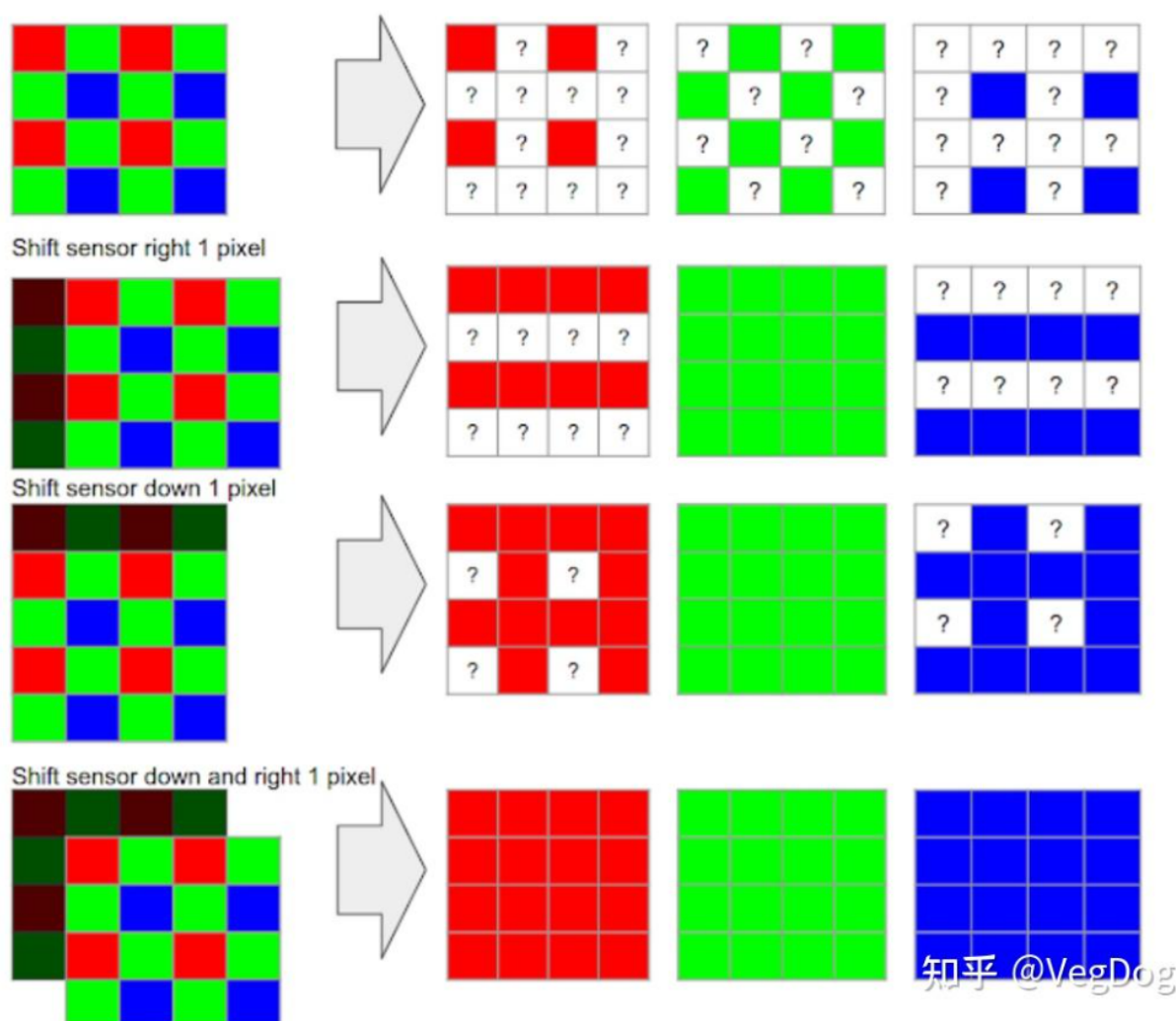


使用的方法是先对多帧图片对齐后进行融合，然后再经过ISP,多帧融合效果相当于长曝光，既得到了更多的光信号，又不会产生通过长曝光拍摄单帧时出现的抖动模糊现象。

大名鼎鼎的Google pixel,同样也针对Raw格式图像，结合多帧图像超分



同时通过分析抖动数据，可以恢复拜尔阵列，也就是demosaic



利用抖动进行demosaic

HDR

在利用多帧进行降噪的同时，通过多帧不同曝光的照片同样也能进行融合，恢复更多高光和暗部的信息，给图片更高的动态范围。



得到不同曝光图像的拉普拉斯金字塔，然后以对比度和饱和度作为指导信息去确定融合的权重，因为在过曝和死黑的区域的像素值接近0和255，梯度很小比较平滑，而饱和度同样是接近于0的。

3. 总结与展望

传统ISP在图像处理领域奠定了坚实的基础，但其固定算法和分步处理模式已难以满足日益复杂的图像处理需求。AI ISP通过引入深度学习和神经网络技术，不仅提升了图像质量，还赋予了设备更强的智能化能力。从数码相机到智能手机，再到AI眼镜等新型终端，AI ISP的应用正在不断扩展。

随着AI智能眼镜中视觉理解模型的进步，对图像处理的要求也日益严格，将会进一步驱动ISP向高端迭代，高端ISP将朝着高精度、高性能及多功能的方向发展。具体来看，包括更高要求的图像处理性能、实时性与低延迟、支持边缘计算、适应多样化光照条件等。

ISP向高端迭代主要有以下两大需求，一是能高效地处理原始数据，二是内置足够的计算资源。

在高效地处理原始数据方面，随着视觉理解模型的能力不断增强，要求ISP芯片高效地处理来自摄像头传感器的原始数据，通过色彩校正、自动曝光控制等技术，为视觉理解模型提供准确的图像；同时，还要支持更高的像素密度和更快的数据吞吐量，以确保提供给AI模型的输入是准确且详细的。

在内置足够的计算资源方面，具备拍摄功能的要求ISP需要内置足够的计算资源来运行轻量级的AI算法或者预处理任务。当前，有不少AI眼镜的AI任务在边缘侧（设备端）进行，特别是一些语音交互功能，而不是依赖于云端服务器。这样的设计可以减少延迟、提高隐私保护，并且能够在没有稳定互联网连接的情况下工作。例如百度推出的小度AI眼镜具有拍照录像功能，支持基于百度文心大模型构建的语音交互功能，这意味着它可以在本地完成语音识别和响应。

最终AI眼镜要实际把带有ISP的芯片放进去必然离不开小型化，而ISP在往高效处理原始数据的过程中，也需要去不断减少中间过程同时不断优化其中的功耗，从而为硬件减负，减少需要的体积。最理想的方法莫过于一个模型完成端到端的ISP，即ISP的过程都可以用一个模型包揽，raw经过模型后直接得到jpg，只是现在这种方法的效果和稳定性都不如AI和传统ISP融合的方法。但我相信，这些问题都会随着时代的进步和行业内每个人的努力，全都会一一解决。

全网第一商业资料社群：

- 每日分享50+行业报告、思维导图、行业资讯、社群课程等

- 全行业覆盖：新零售、AR、房地产、人工智能、新基建、生鲜、物联网、母婴、机器人、新能源汽车工业互联网、直播短视频等 460+个行业

- 全网唯一终身制知识社群
长按识别右侧二维码，立即加入



长按二维码加入