

# Noise2Void - Learning Denoising from Single Noisy Images

Dezeming Family

2023 年 07 月 20 日

DezemingFamily 系列文章和电子书**全部都有免费公开的电子版**，可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列电子书，可以从我们的网站 [<https://dezeming.top/>] 找到最新的版本。对文章的内容建议和出现的错误也欢迎在网站留言。

## 目录

<b>一 Introduction</b>	<b>1</b>
<b>二 相关工作</b>	<b>1</b>
<b>三 方法</b>	<b>1</b>
3.1 图像建立	1
3.2 有监督学习	1
3.3 Noise2Void	2
3.4 实现方案	2
<b>参考文献</b>	<b>3</b>

# 摘要

最近已经表明，去噪方法可以在没有 clean target 的情况下进行训练。相反，可以使用独立的噪声图像对，这种方法被称为 NOISE2NOISE (N2N)。在这里，我们介绍了 NOISE2VOID (N2V)，这是一种将这一想法进一步发扬光大的训练方案。它不需要有噪声的图像对，也不需要干净的目标图像。因此，N2V 允许我们直接在要去噪的数据体上进行训练，因此可以在其他方法无法应用的情况下应用。特别令人感兴趣的是生物医学图像数据的应用，在生物医学图像数据中，通常不可能获得干净或有噪声的训练目标。我们将 N2V 的性能与具有干净目标图像和/或噪声图像对的方法进行了比较。直观地说，不能指望 N2V 优于在训练过程中获得更多信息的方法。尽管如此，我们观察到 Noise2Void 的去噪性能适度下降，并且与无训练的去噪方法相比是有利的。

## 一 Introduction

我们假设噪声图像是原始图像也噪声图像的和。而且每个像素的值都是邻域相关的，并不是独立存在的。

N2N 方法虽然比较好，不需要无噪声参考图，但是它需要有噪图像对，这就意味着在获取有噪图像对时，两个图像相对静止，这在很多时候都不是很容易达到的。

我们提出的 Noise2Void 方法不需要有噪图像对，而是直接对有噪图像做自监督训练，我们对图像做了两点假设：(1) 信号  $s$  之间（各像素之间）颜色不是独立的；(2) 给定信号  $s$ ，噪声  $n$  是条件独立的。

## 二 相关工作

一些不错的去噪网络架构，以及其对应的解释。

## 三 方法

### 3.1 图像建立

设一幅图像  $x = s + n$ ，表示为分布：

$$p(s, n) = p(s)p(n|s) \quad (三.1)$$

假设信号  $s$  各个像素之间不是完全独立的，有一定关系，但是噪声  $n$  是条件独立的，也就是说：

$$p(n|s) = \prod_i p(n_i|s_i) \quad (三.2)$$

给定了信号，其噪声在各个像素之间并没有什么关系。

此外，还需要满足噪声的期望是 0，也就是说  $x$  的期望是  $s$ 。就好比蒙特卡洛渲染的图像，一大堆相同场景下估计得到的有噪图像取平均，就能得到无噪声图像。

### 3.2 有监督学习

设  $j$  表示第几个图像对的序号， $i$  是每个图像第几个像素的序号。

由于使用卷积网络训练是存在感受野的，因此，对于每个预测的像素输出，都是对应一个感受野的输入，用  $x_{RF(i)}^j$  表示有噪图像  $x^j$  的像素  $i$  在网络中的感受野，由此得到：

$$f(x_{RF(i)}^j; \theta) = \hat{s}_i^j \quad (三.3)$$

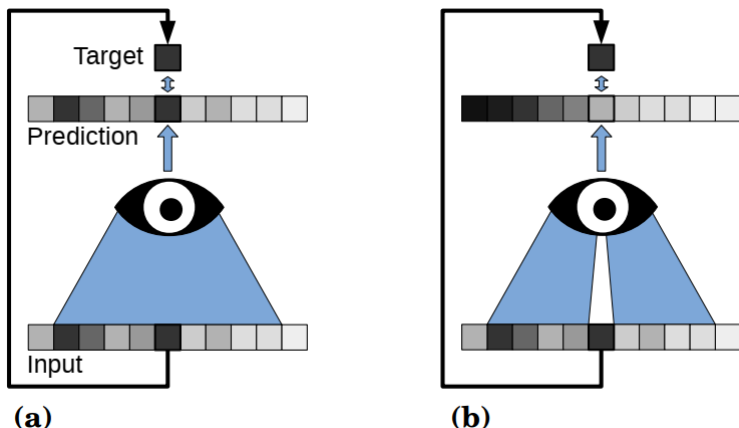
去噪目标如下：

$$\arg \min_{\theta} \sum_j \sum_i L(f(x_{RF(i)}^j; \theta) = \hat{s}_i^j, s_i^j) \quad (三.4)$$

### 3.3 Noise2Void

这里的构思非常巧妙。

如果我们用以前的卷积网络，把一个有噪图像作为输入，去预测有噪输出，那么网络就会训练成把中心像素输出的网络，如下图左：



现在换种思路，把网络输入设置为用周围像素预测中央像素，如右图。如果以前知道 Noise2Noise 的原理，那么应该会冥冥之中对这种方法有一丝“恍然大悟”的感受。

盲点网络可用于预测的信息略少，与正常网络相比，我们可以预期其准确性会略有下降。然而，考虑到整个感受野中只有一个像素被去除，我们可以假设它仍然表现得相当好。

盲点网络的优点在于不能学习到对等映射关系，且无法预测中间像素的噪声值，这是因为假设噪声之间都是独立的。然而由于像素之间值不独立，所以网络仍然能学习到像素值的统计特性。在大量统计下，就可以预测出理想的结果了。

### 3.4 实现方案

虽然盲点网络原则上可以仅使用单独的噪声训练图像来训练。然而，实现这样一个仍然可以高效运行的网络并非易事。我们提出了一种 mask 方案来避免这个问题，并实现与任何标准 CNN 相同的特性。方法是用从周围区域随机选择的值来替换每个输入 patch 中心的值（详见补充材料）。这有效地擦除了像素的信息，并防止网络学习恒定映射。

论文的实现方案也是一大亮点，因为需要跟现有的卷积模块兼容来训练。

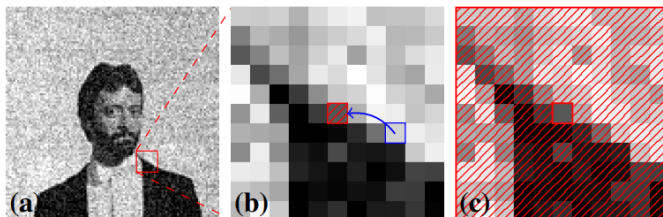


Figure 3: Blind-spot masking scheme used during NOISE2VOID training. (a) A noisy training image. (b) A magnified image patch from (a). During N2V training, a randomly selected pixel is chosen (blue rectangle) and its intensity copied over to create a blind-spot (red and striped square). This modified image is then used as input image during training. (c) The target patch corresponding to (b). We use the original input with unmodified values also as target. The loss is only calculated for the blind-spot pixels we masked in (b).

第一，先准备有噪图，然后把有噪图分为  $64 \times 64$  pixels 的 patches。这个 patches 要比感受野大。

然后，在每个 patch 中用分层抽样法（防止团簇）来选择  $N$  个像素，并将这些像素 masked。然后我们使用原始的有噪输入作为 target。不过要注意的是，不是说这  $N$  个 mask 的像素都 mask 到一张输入 patch 上，而是把输入复制  $N$  份，每份都选择不同的中心像素来 mask。

在训练时，只计算被 make 的像素的预测 loss，不考虑其他像素。这样相当于一次对  $N$  个像素的预测进行梯度下降优化，也是节省训练时间的方案。

实际训练时，batch size 是 128，且  $N = 64$ 。

## 参考文献

- [1] Krull A, Buchholz T O, Jug F. Noise2void-learning denoising from single noisy images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2129-2137.
- [2] <https://zhuanlan.zhihu.com/p/563746026>