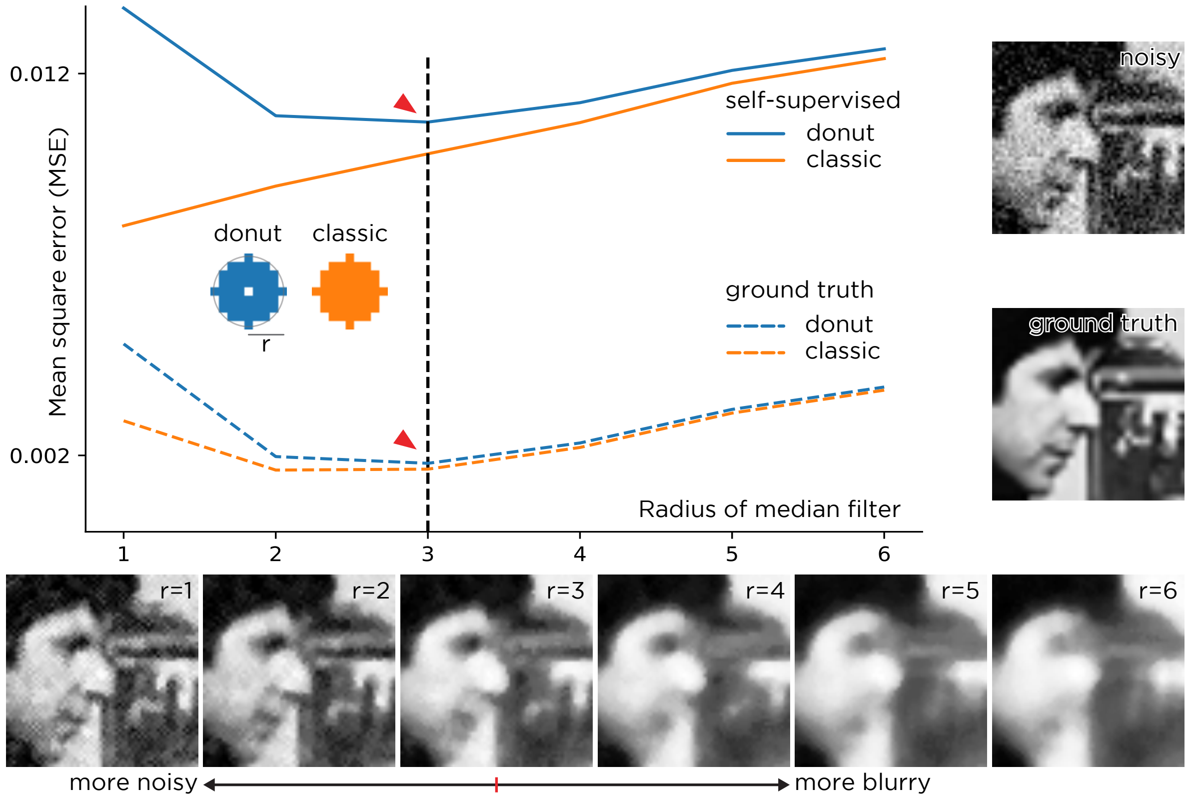

Figure 2. Calibrating a median filter without ground truth. Different median filters may be obtained by varying the filter's radius. Which is optimal for a given image? The optimal parameter for $\mathcal{J}$-invariant functions such as the donut median can be read off (red arrows) from the self-supervised loss.

compute the self-supervised loss $\|f_\theta(x) - x\|^2$. For general f_θ, it is unrelated to the ground truth loss, but if f_θ is $\mathcal{J}$-invariant, then it is equal to the ground truth loss plus the noise variance (Eqn. 2), and will have the same minimizer.

In Figure 2, we compare both losses for the median filter g_r, which replaces each pixel with the median over a disk of radius r surrounding it, and the "donut" median filter f_r, which replaces each pixel with the median over the same disk *excluding the center*, on an image with i.i.d. Gaussian noise. For $\mathcal{J} = \{\{1\}, \ldots, \{m\}\}$ the partition into single pixels, the donut median is $\mathcal{J}$-invariant. For the donut median, the minimum of the self-supervised loss $\|f_r(x) - x\|^2$ (solid blue) sits directly above the minimum of the ground truth loss $\|f_r(x) - y\|^2$ (dashed blue), and selects the optimal radius $r = 3$. The vertical displacement is equal to the variance of the noise. In contrast, the self-supervised loss $\|g_r(x) - x\|^2$ (solid orange) is strictly increasing and tells us nothing about the ground truth loss $\|g_r(x) - y\|^2$ (dashed orange). Note that the median and donut median are genuinely different functions with slightly different performance, but while the former can only be tuned by inspecting the output images, the latter can be tuned using a principled loss.

More generally, let g_θ be any classical denoiser, and let $\mathcal{J}$ be any partition of the pixels such that neighboring pixels are in different subsets. Let $s(x)$ be the function replacing each pixel with the average of its neighbors. Then the function f_θ defined by

$$f_\theta(x)_J := g_\theta(\mathbf{1}_J \cdot s(x) + \mathbf{1}_{J^c} \cdot x)_J, \tag{3}$$

for each $J \in \mathcal{J}$, is a $\mathcal{J}$-invariant version of g_θ. Indeed, since the pixels of x in J are replaced before applying g_θ, the output cannot depend on x_J.

In Supp. Figure 1, we show the corresponding loss curves for $\mathcal{J}$-invariant versions of a wavelet filter, where we tune the threshold σ, and NL-means, where we tune a cut-off distance h (Buades et al., 2005a; Chang et al., 2000; van der Walt et al., 2014). The partition $\mathcal{J}$ used is a 4x4 grid. Note that in all these examples, the function f_θ is genuinely different than g_θ, and, because the simple interpolation procedure may itself be helpful, it sometimes performs better.

In Table 1, we compare all three $\mathcal{J}$-invariant denoisers on a single image. As expected, the denoiser with the best self-supervised loss also has the best performance as measured by Peak Signal to Noise Ratio (PSNR).