Single Image Blind Parameter Recovery

Stephen Tierney
Business School, The University of Sydney
Camperdown, 2050, Australia
stephen.tierney@sydney.edu.au

Abstract

Historically, image formation parameters and hyper-parameters have been tuned by hand for image restoration due to the lack of a broadly applicable selection framework. This has substantially limited the application of state-of-the-art image processing methods. We show that recasting the problem of image restoration to include the additional task of inpainting pixels permits an automatic hyperparameter selection procedure for image restoration through a pixel-wise hold-out cross-validation. This procedure is algorithm-agnostic and works on single images without any prior information. We show that the estimator in the procedure is affinely unbiased under very general settings and can be used for model selection. We provide a reference implementation to demonstrate how to extend existing methods to support inpainting and present empirical results. We hope that this eliminates manual tuning in image restoration pipelines.

1. Introduction

Digital imaging is subject to defects, such as blurring and noise, at each step of the imaging pipeline. This is due to imperfections in lenses, subject motion, vibration, lighting conditions, sensor damage, quantisation error and sensor read noise. As a result, the problem of recovering the original image from a corrupted source finds wide applications and has been well studied.

In this paper, we assume a general image–formation model and problem

$$X_{\text{obs}} = F_{\phi}\big(X_{\text{true}}\big) + n, \tag{1}$$

where F_{ϕ} is a forward operator parametrised by $\phi \in \Phi$ (e.g. a convolution with PSF parameters), and n is any zero-mean, finite-variance noise. The problem of recovering the true image, X_{true}, from an observation, X_{obs}, without knowledge of the parameters is known as "blind" image restoration. Both images are assumed to be $\in \mathbb{R}^{m \times n}$.

Attempts to solve (1) fall into two broad classes. Variational (model-based) methods add an explicit penalty to stabilise the inversion [9] e.g. total variation (TV) [24], or non-local means priors [10]. Their performance is controlled by a small set of hyper-parameters, such as regularisation weights.

Data-driven methods learn a direct mapping from corrupted to clean images from paired clean/corrupted images. These have evolved over time from relatively simple multi-layer perceptron architectures, through to Convolutional Neural Networks and recently Diffusion models [8]. While they achieve state-of-the-art accuracy, they demand substantial training compute. Moreover, they may not generalise to imaging domains that are not part of the training set distribution. Nor do they directly reveal image formation parameters since they are encoded in the layers of the network used.

Therefore unsupervised single-image restoration is still of interest and remains an open problem. Unfortunately we are left with uncertainty around selecting optimal image formation parameters and hyper-parameters. Historically both kinds of parameters have been tuned by hand due to the lack of a broadly applicable selection framework. We remark that if a solver never sees a subset of pixels in an image, then the error that it makes is an estimate of the generalisation error. We show that it is possible to exploit this "self-consistency" observation to devise an automatic parameter selection procedure for image restoration through a pixel-wise hold-out cross-validation. Our contributions and differences from prior works are:

- A model-agnostic, single-image procedure that recovers plausible image formation and optimal hyper-parameters,
- A proof that the estimator is affinely unbiased under arbitrary zero-mean finite-variance noise,
- A derivation of the estimator variance, confidence bounds and guidance on practical use of the estimator,

2. Related Work

2.1. Variational Image Restoration

Classical variational image restoration formulates the forward operator and a prior as an objective that is minimised iteratively. Total-Variation (TV) regularisation [24] remains the canonical prior for edge–preserving recovery and spawned a long line of inquiry into optimisation schemes. The iterative soft-thresholding method (ISTA) [7] and its accelerated variant FISTA [3] are still the template for proximal algorithms. Numerous primal–dual methods (e.g. PDHG [22]) and Split-Bregman variants [11] further refined these ideas. The Generalised Accelerated Proximal Gradient (GAPG) approach of Zuo and Lin [34] offers improved runtime performance over the aforementioned ideas by removing the need for an expensive inner total variation solver. However none provide a mechanism for selecting optimal blur kernel or regularisation weight and in fact this problem is usually ignored.

2.2. Hyper-Parameter Selection

Selecting hyper-parameters is a challenging problem by itself. Classical point-estimators include the L-curve [13], Generalised Cross-Validation (GCV) [12] and Stein's Unbiased Risk Estimate (SURE) [26]. Unfortunately, all three make strict assumptions that break down once the forward model contains unknown parameters.

2.3. Non-Negative Matrix Factorisation

We draw inspiration from work in matrix factorisation for selection of the factorisation rank. This work dates back to Wold's work on PCA [30] which used the idea of masking a random sub-matrix, performed a fit on the remaining data and predicted the masked values. This work was more extended to Non-Negative Matrix Factorisation (NMF) [15, 21]. At the time Wold remarked that:

> The present cross-validatory procedure also rests on the property of the NIPALS procedure to work with incomplete data

which highlights the need for complimentary image restoration methods.

2.4. Image Inpainting

Inpainting is the task of restoring an image where parts of the image are marked as missing or corrupt. These missing images are denoted by a mask provided to the inpainting solver. Early work propagated colour along image structures with differential equations. The first of these were based on fluid flow [4], then curvature-driven diffusion [5]. Simple harmonic and biharmonic interpolation [17, 14, 6] remain popular because they run in milliseconds.

The rise of deep learning has made the field explode. CNN-based approaches such as Contextual Attention [31], Partial Convolutions [18] and Gated Convolutions [32] can plausibly fill large irregular holes while respecting texture. Subsequent two-stage pipelines first hallucinate structure (edges or semantics) and then colour it in. Influential examples include EdgeConnect [20] and CoMod-GAN [33]. More recently, LaMa [27] shows that large receptive fields and fast Fourier convolutions can match GAN quality and diffusion can be used to solve inpainting tasks e.g. RePaint [19] and Palette [25].

2.5. Self-Consistency in Deep Learning

A recent wave of self-supervised methods exploits the independence between held-out and predicted pixels. Deep Image Prior (DIP) fits the weights of an untrained CNN to a single degraded image and uses the network's architecture as a prior [28]. Noise2Self extends the idea to denoising with arbitrary "J-invariant" predictors [2]. The Noise-Tolerant Self-Supervised Inversion framework (NT-SSI) applies the same principle to deconvolution σ^2 [16].

2.6. Summary

Our work is closest in spirit to Noise2Self and NT-SSI yet differs in three key aspects. First we tackle the problem of simultaneously recovering both plausible image formation parameters and hyper-parameters rather than fitting a black-box deep neural network. Second we clarify and expand on their theoretical results by investigating the variance of the estimator and derive variance bounds that link the mask density to image size so that we can analytically choose a hold-out rate. Third and finally, our procedure is more general in that it uses any available convex solver, such as a TV solver or even a deep neural network, enabling the user to make a tradeoff between computational time and restoration quality.

During final proofing, we discovered a relatively unknown body of work by Reeves [23] from the early 90s that overlaps our work. It was both frustrating and relieving to discover this work since it takes priority (credit) of this idea, yet at the same time shows that the we are not alone in this line of thinking. However that body of work does not focus on pixel-wise holdout as Reeves quickly move onto generalized cross validation, which we believe is due to their choice of L2 regularisation and the lack of solvers for other penalties at that time. The authors wish to draw the readers attention to the fact that the body of work by Reeves does not provide theoretical justification for pixel-wise holdout as is presented in this paper. We also wish to point out that in the discussion we mention many further extensions to this idea that have not been proposed, to the best of our knowledge.

3. Pixel-Wise Cross-Validation

The core idea of pixel-wise cross-validation is: if a restoration algorithm never "sees" a chosen subset of pixels, then the reconstruction error on those withheld pixels serves as an unbiased proxy for its out-of-sample performance. We call this concept "self-consistency".

Concretely, we randomly mask out a small fraction of pixels, run any black-box solver on the remaining data, and then re-evaluate the solver's predictions only on the held-out set. By repeating this for each candidate pair of forward-model parameters and solver hyper-parameters, we obtain a score whose minimizer identifies the combination that best generalizes to unseen pixels.

We assume access to an image restoration solver:

$$X^*(\phi,\theta) = \textsc{Solver}\big(X_{\text{obs}},\, M,\, \phi,\, \theta\big), \tag{2}$$

which in practice solves

$$X^*(\phi,\theta) \;\approx\; \arg\min_X \underbrace{\big\| M \odot \big(F_\phi(X) - X_{\text{obs}}\big)\big\|_2^2}_{\text{masked data fidelity}} + \underbrace{\text{Reg}_\theta(X)}_{\text{solver-specific prior}}, \tag{3}$$

where $M \in \{0,1\}^{m\times n}$ is a binary mask with held-out pixels set to zero or one otherwise, Reg_θ denotes any priors of the solver such as regularisation, filter, or network weights controlled by hyperparameters $\theta \in \Theta$.

Since (3) makes no assumption on the form of SOLVER, one can plug in simple filters, variational methods or deep neural networks - so long as they support masking pixels.

3.1. Procedure

The procedure selects the best (ϕ,θ) by measuring prediction error on held-out pixels. The procedure is as follows:

1. **Masking.** Draw $M_{ij} \sim \text{Bernoulli}(1-p)$, where p is the proportion of held-out pixels, the training set is $\Omega_t = \{M_{ij} = 1\}$, the validation set is $\Omega_v = \{M_{ij} = 0\}$ and $V = |\Omega_v|$.

2. **Masked reconstruction.** For each $(\phi,\theta) \in \Phi \times \Theta$ solve

$$X^*(\phi,\theta) = \textsc{Solver}\big(X_{\text{obs}},\, M,\, \phi,\, \theta\big), \tag{4}$$

3. **Hold-out scoring.** Compute the mean-squared error on Ω_v:

$$\widehat{R}(\phi,\theta) = \frac{1}{V}\sum_{(i,j)\in\Omega_v}\Big[F_\phi\big(X^*(\phi,\theta)\big)_{ij} - X_{\text{obs},ij}\Big]^2. \tag{5}$$

Algorithm 1: Pixel-Wise Cross-Validation

Input: X_{obs}, candidate parameters Φ, Θ, hold-out rate p

Sample $M_{ij} \sim \text{Bernoulli}(1-p)$;
Define $\Omega_v = \{M_{ij} = 0\}$, $V = |\Omega_v|$;
foreach $\phi \in \Phi,\ \theta \in \Theta$ **do**
  $X^*(\phi,\theta) \leftarrow \textsc{Solver}(X_{\text{obs}}, M, \phi, \theta)$;
  $E \leftarrow \frac{1}{V}\sum_{(i,j)\in\Omega_v}\big[F_\phi(X^*(\phi,\theta))_{ij} - X_{\text{obs},ij}\big]^2$;
  Store $\widehat{R}(\phi,\theta) = E$;
$(\phi^\star,\theta^\star) \leftarrow \arg\min \widehat{R}(\phi,\theta)$;
return $(\phi^\star,\theta^\star)$

4. **Selection.** Choose the minimiser

$$(\phi^\star,\theta^\star) = \arg\min_{(\phi,\theta)\in\Phi\times\Theta} \widehat{R}(\phi,\theta), \tag{6}$$

then optionally recompute $X^*(\phi^\star,\theta^\star)$ on the full image ($M = \mathbf{1}$).

This procedure is summarised in Algorithm 1.

3.2. Affinely Unbiased Estimator

If $\widehat{R}$ systematically over- or under-estimates the true predictive risk $R_{\text{pred}}(\phi,\theta)$, it could lead us to choose suboptimal (ϕ,θ). Therefore one may be interested in knowing how faithfully $\widehat{R}$ estimates the true predictive risk.

Lemma 3.1. *Under any zero-mean, finite-variance noise the pixel-wise CV score* (5)

$$\mathbb{E}\big[\widehat{R}\big] = \underbrace{\frac{1}{V}\sum_{(i,j)\in\Omega_v}\big[F_\phi(X^*)_{ij} - F_{\phi_\star}(X_{\text{true}})_{ij}\big]^2}_{R_{\text{pred}}(\phi,\theta)} + \sigma^2. \tag{7}$$

where we use $\widehat{R} = \widehat{R}(\phi,\theta)$ *for brevity.*

Therefore $\widehat{R}$ is an affinely unbiased estimator of the true predictive risk R_{pred}. Since σ is independent of (ϕ,θ), minimising $\widehat{R}$ is equivalent to minimising R_{pred} and can be used to select optimal (ϕ,θ).

Proof. $X^*(\phi,\theta)$ depends only on the training pixels, so for each held-out (i,j) the deterministic bias

$$d_{ij} = F_\phi(X^*)_{ij} - F_{\phi_\star}(X_{\text{true}})_{ij} \tag{8}$$

is independent of the noise term n_{ij}. Meanwhile

$$X_{\text{obs},ij} = F_{\phi_\star}(X_{\text{true}})_{ij} + n_{ij}, \tag{9}$$

so the squared residual can be written

$$\begin{aligned}\big[F_\phi(X^*)_{ij} - X_{\text{obs},ij}\big]^2 &= (d_{ij} - n_{ij})^2 \\ &= d_{ij}^2 - 2\, d_{ij}\, n_{ij} + n_{ij}^2. \end{aligned} \tag{10}$$

Taking expectations, the cross-term vanishes and $\mathbb{E}[n_{ij}^2] = \sigma^2$, giving

$$\mathbb{E}\big[(d_{ij} - n_{ij})^2\big] = d_{ij}^2 + \sigma^2. \tag{11}$$

Finally, averaging over all V held-out pixels completes the proof:

$$\begin{aligned}\mathbb{E}[\widehat{R}] =& \frac{1}{V} \sum_{(i,j)\in\Omega_v} \mathbb{E}\big[(d_{ij} - n_{ij})^2\big] \\ =& \frac{1}{V} \sum_{(i,j)\in\Omega_v} (F_\phi(X^*)_{ij} - F_{\phi_\star}(X_{\text{true}})_{ij})^2 + \sigma^2. \end{aligned} \tag{12}$$

□

3.3. Variance

Using a random mask introduces randomness into the procedure. At the same time there is a bias-variance trade-off to be made as increasing p will decrease variance of the estimator but increase bias. Therefore one may be interested in the variance of our estimator for different values of p. We show that in practice, for images from modern consumer devices the standard deviation is very tight for modest values of p.

Lemma 3.2. *Recall Ω_v as the set of V held-out pixels and $\widehat{R}(\phi,\theta)$ defined as (5) then*

$$\operatorname{Var}\big[\widehat{R}(\phi,\theta)\big] = \frac{M_4 - \sigma^4}{V} + \frac{4\sigma^2}{V^2} \sum_{(i,j)\in\Omega_v} d_{ij}^2 \tag{13}$$

where M_4 is the fourth moment, for symmetric noise with zero mean.

Proof. For each $(i,j) \in \Omega_v$, write the prediction residual as

$$Y_{ij} \;=\; \big[F_\phi(X^*)_{ij} - X_{\text{obs},ij}\big]^2 = \big(d_{ij} - n_{ij}\big)^2, \tag{14}$$

where d_{ij} is the bias. Since d_{ij} is fixed (once X^* is computed) and n_{ij} is independent across pixels,

$$\operatorname{Var}\big[\widehat{R}\big] = \operatorname{Var}\Big(\tfrac{1}{V}\sum Y_{ij}\Big) = \frac{1}{V^2}\sum_{\Omega_v} \operatorname{Var}[Y_{ij}]. \tag{15}$$

Expanding and since d_{ij} is constant

$$\operatorname{Var}[Y_{ij}] = \operatorname{Var}[-2\, d_{ij}\, n_{ij} + n_{ij}^2]. \tag{16}$$

Using the variance of the sum of two random variables we have

$$\operatorname{Var}[Y_{ij}] = \operatorname{Var}(n_{ij}^2) + 4d_{ij}^2 \operatorname{Var}(n_{ij}) - 4d_{ij}\operatorname{Cov}(n_{ij}^2, n_{ij}). \tag{17}$$

Under symmetric noise with zero mean, we have $\operatorname{Cov}(n_{ij}^2, n_{ij}) = \mathbb{E}[n_{ij}^3] = 0$ therefore

$$\operatorname{Var}[Y_{ij}] = (M_4 - \sigma^4) \;+\; 4d_{ij}^2\sigma^2. \tag{18}$$

Substituting back yields

$$\begin{aligned}\operatorname{Var}[\widehat{R}] =& \frac{1}{V^2} \cdot V\,(M_4 - \sigma^4) + \frac{4\sigma^2}{V^2} \sum_{(i,j)\in\Omega_v} d_{ij}^2 \\ =& \frac{M_4 - \sigma^4}{V} + \frac{4\sigma^2}{V^2} \sum_{(i,j)\in\Omega_v} d_{ij}^2 \end{aligned} \tag{19}$$

as claimed.

□

Example As of writing, consumer devices have on average 54MP sensors [1], although these are often binned or down-sampled to 12-24MP final images. For illustrative purposes we assume: a typical final image will be roughly 4000×6000 pixels, Gaussian noise with fourth moment $M_4 = 3\sigma^4$, a hold out rate of $p = 0.05$ and a high quality reconstruction permitting us to drop the deterministic bias term. This yields:

$$\operatorname{Var}[\widehat{R}] = \frac{3\sigma^4 - \sigma^4}{V} + \frac{4\sigma^2}{V^2} = \frac{2\sigma^4}{V} \tag{20}$$

Plugging in the image size, and a reasonable approximation of the image noise $\sigma = 0.01$ we have

$$\text{Std} \approx 1.29 \times 10^{-7}, \tag{21}$$

which is usually negligible compared with typical differences between competing (ϕ,θ) pairs.

3.4. Distribution-Free Confidence Bounds

Throughout this section let

$$\widetilde{R}(\phi,\theta) \;:=\; \widehat{R}(\phi,\theta) - \sigma^2, \tag{22}$$

so that $\mathbb{E}[\widetilde{R}] = R_{\text{pred}}$. Lemma 3.2 shows that, for symmetric i.i.d. noise with variance σ^2 and fourth moment M_4,

$$\operatorname{Var}[\widetilde{R}] = \frac{M_4 - \sigma^4}{V} + \frac{4\sigma^2}{V}\,\bar{d}^2, \quad \bar{d}^2 := \frac{1}{V}\sum_{\Omega_v} d_{ij}^2. \tag{13}$$

Chebyshev tail. Because (13) holds for all noise distributions with finite M_4, Chebyshev's inequality gives a distribution–free deviation:

$$\begin{aligned}\Pr\big(|\widetilde{R} - R_{\text{pred}}| \geq \varepsilon\big) &\leq \frac{\operatorname{Var}[\widetilde{R}]}{\varepsilon^2} \\ &\leq \frac{M_4 - \sigma^4 + 4\sigma^2\bar{d}^2}{V\,\varepsilon^2}. \end{aligned} \tag{23}$$

Required mask size. To achieve half-width ε with confidence $1-\delta$ it suffices to choose

$$V = \frac{M_4 - \sigma^4 + 4\sigma^2\bar{d}^2}{\delta\,\varepsilon^2}, \qquad p = \frac{V}{mn}. \tag{24}$$

Multiple hyper-parameter trials. When D candidate pairs (ϕ, θ) are scored on the same mask, a union (Bonferroni) bound on (23) yields

$$\begin{aligned}&\Pr\Big(\max_{t\leq D}|\widetilde{R}(t) - R_{\text{pred}}(t)| \geq \varepsilon\Big) \\ &\qquad \leq \min\Big\{1,\ \frac{D\,[M_4 - \sigma^4 + 4\sigma^2\bar{d}^2]}{V\,\varepsilon^2}\Big\},\end{aligned} \tag{25}$$

and therefore

$$V = \frac{D\,[M_4 - \sigma^4 + 4\sigma^2\bar{d}^2]}{\delta\,\varepsilon^2} \tag{26}$$

guarantees the same (ε, δ) band simultaneously for all D configurations.

Practical impact. At first sight (26) suggests an explosion in the number of held-out pixels when a fine (ϕ, θ) grid is explored. In practice the bound is overly conservative because:

1. the D risk estimates are highly correlated,
2. modern sensors exhibit $\sigma \approx 10^{-4}$ in $[0,1]$-normalised units, so the leading factor $M_4 - \sigma^4 = 2\sigma^4$ is $< 2 \times 10^{-8}$,
3. averaging over K independent random masks divides the variance by K, and a coarse-to-fine grid search reduces D by orders of magnitude before the tight mask is needed.

4. Join Restoration and Inpainting

To demonstrate how one can jointly inpaint and restore an image we extend the GAPG [34] total-variation-based deblurring algorithm. We note that while [34] discusses deblurring and inpainting, they were considered separate tasks with separate optimisation schemes. We show how to unify them.

4.1. GAPG-Inpaint

We consider the optimization problem

$$\begin{aligned}\min_{l\leq X\leq u}\ f(X) + g(X), \qquad (27)\\ f(X) = \frac{1}{2}\big\| M \odot \big(P * X - X_{\text{obs}}\big)\big\|_F^2, \\ g(X) = 2\lambda\,\mathrm{TV}(X),\end{aligned}$$

where $f(X)$ is the data-fidelity term, $P * X$ denotes convolution with a fixed PSF P, $g(X)$ is either isotropic or anisotropic total variation, λ is the TV regularization weight and $[l, u]$ enforces magnitude constraints on X to generate a valid pixel intensities.

We summarize the GAPG-Inpaint algorithm in Algorithm 2. The only change is that the gradient is computed only on the observed pixels. Concretely, we replace $\nabla f(Y) = P^\top(P * Y - B_{\text{obs}})$ by the masked version

$$\nabla f_M(Y) = P^\top\big(M \odot (P * Y - B_{\text{obs}})\big), \tag{28}$$

so that any contribution from a held-out pixel (where $M_{ij} = 0$) is zeroed out.

Note that in real computation, the matrix-vector products in Algorithm (2) must be replaced by operators acting on images. This means that $P * Y$ and $P^\top(\cdot)$ are implemented via 2D convolution, not matrix–vector products.

Algorithm 2: GAPG-Inpaint: Accelerated Proximal-Gradient for Joint Deblurring+Inpainting

Input: Observed image B_{obs}, PSF kernel P, mask M,
TV weight λ, bounds l, u, iterations K
Output: Recovered image X_K

Estimate Lipschitz constant L of ∇f;
Initialize $X_0 = B_{\text{obs}},\ Y_0 = X_0,\ t_0 = 1$;
for $k = 1$ **to** K **do**
  $t_k \leftarrow \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$
  $\alpha \leftarrow (t_{k-1} - 1)/t_k$
  $Y \leftarrow X_{k-1} + \alpha\,(X_{k-1} - X_{k-2})$
  $\nabla f(Y) \leftarrow P^\top\big(M \odot (P * Y - B_{\text{obs}})\big)$
  $Z \leftarrow \mathrm{clip}\big(Y - \frac{1}{L}\nabla f(Y),\, l, u\big)$
  $X_k \leftarrow \mathrm{prox}_{g/L}(Z)$
return X_K

4.2. Practical Considerations

In practice we warm start the GAPG algorithm by estimating missing pixels using Biharmonic Inpainting [6] as implemented by [29]. This substantially reduces the number of iterations of GAPG-Inpaint required for convergence and ultimately the running time of the algorithm.

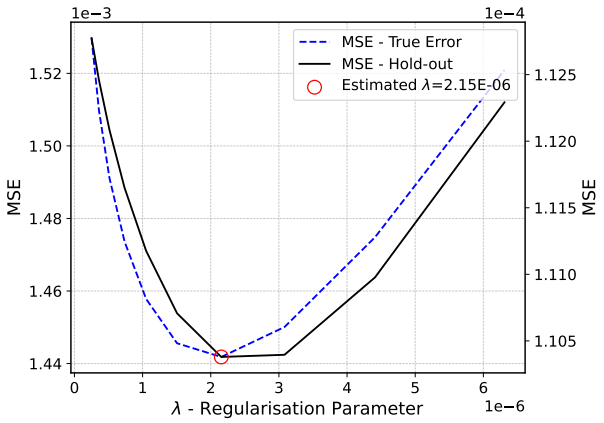


Figure 1: Comparison of $\widehat{R}$ and R_{pred} over a range of λ with known Gaussian blur kernel.

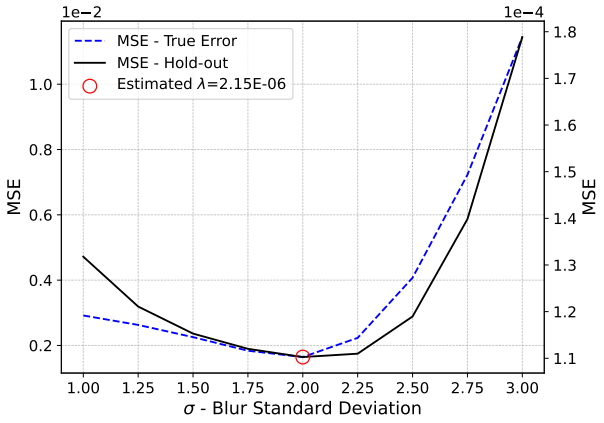


Figure 2: Comparison of $\widehat{R}$ and R_{pred} over a range of blur standard deviation for a Gaussian blur kernel with fixed and optimal λ from Figure 1.

5. Experiments

In this section we use the GAPG-Inpaint algorithm with anisotropic total variation for all image restoration tasks.

5.1. Proof of Concept

To demonstrate the concept we perform a toy experiment of recovering an image from a blurred and noisy image i.e.

$$X_{\text{obs}} = P * X_{\text{true}} + n, \tag{29}$$

where P is a Gaussian blurring kernel, $*$ is the convolution operator and n is Gaussian noise.

For this experiment we use the astronaut image from [29], see Figure 3, which is an excellent test image due to the varied textures and feature size. We use a 19×19 Gaussian kernel with mean zero and standard deviation $\sigma_P = 2$. To this we add Gaussian noise with mean zero and standard deviation 10^{-2}. For this experiment we set the hold-out rate p to 20%.

We first search for an optimal λ under this setting by calculating $\widehat{R}$ using the true blur kernel. We report both $\widehat{R}$ and the R_{pred}. Results are presented in Figure 1. We find that the minimum of $\widehat{R}$ aligns with the minimum of the true predictive risk.

Next we fix λ to the value corresponding to the minimum $\widehat{R}$ error and then search for the blur kernel standard deviation. Again we report $\widehat{R}$ and the R_{pred}. Results are presented in Figure 2. We find that the minimum of $\widehat{R}$ aligns with the minimum of the true predictive risk and closely matches the true $\sigma_P = 2$.

This toy experiment shows that the self-consistency concept and the pixel-wise cross validation error $\widehat{R}$ can inform and guide selection of image formation and hyper-parameters. This can be observed from the re-blurred best reconstruction almost exactly matching the observed image, see Figure 3. We note that this a single run i.e. a single random mask was used. We expect that repeated evaluation with different masks would yield an even tighter estimate on σ_P.

5.2. Hold Out Sensitivity

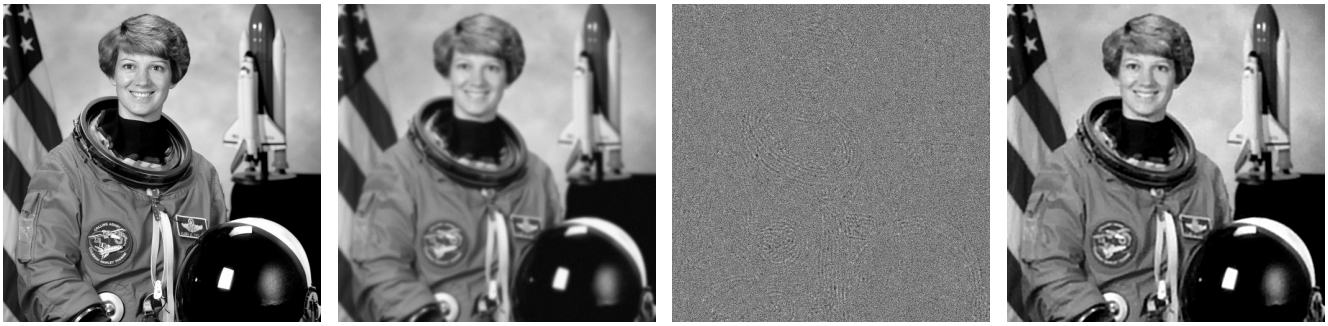
Next we investigate the impact of the hold-out fraction p on the ability to identify parameters. Under the same settings as the previous subsection we sweep over p between 0.1 and 0.9 with a correctly specified Gaussian blur kernel and optimal λ. We observe as expected that the reconstruction quality degrades with decreasing p, shown in Figure 4. What is less predictable, however, is that the degradation is non-linear with a sharp loss of quality below $p = 0.2$.

The next question is whether this relative degradation impacts on estimating parameters. To investigate this we sweep over p while varying the blur kernel standard deviation in a similar fashion to the previous subsection. We repeated this process with 5[1] replicates at each setting and reported the mean standard deviation. Results are visualised in Figures 5 and 6. We removed the $p = 0.1$ case since with so few replicates the results were considered unreliable. We observe a general trend that the correct parameters are recovered on average, even for low values of p.

6. Discussion and Future Work

We introduced pixel–wise cross-validation, a simple procedure that automatically recovers plausible image formation and optimal hyper-parameters by minimising a hold-out error on masked pixels. The method is entirely algorithm-agnostic - it wraps any restoration solver that

[1] We would prefer this to be much greater but we're trying to meet the paper deadline!



(a) (b) (c) (d)

Figure 3: Images used in the proof of concept experiment. (a) The original astronaut image, (b) the observed blurred and noisy image, (c) the error between the observed image and the re-blurred best reconstruction, and (d) the best reconstruction.

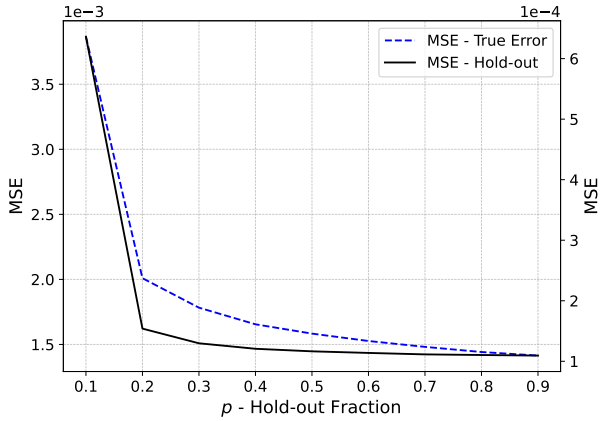


Figure 4: Comparison of $\widehat{R}$ and R_{pred} over a range of p using the correct Gaussian blur kernel and optimal λ.

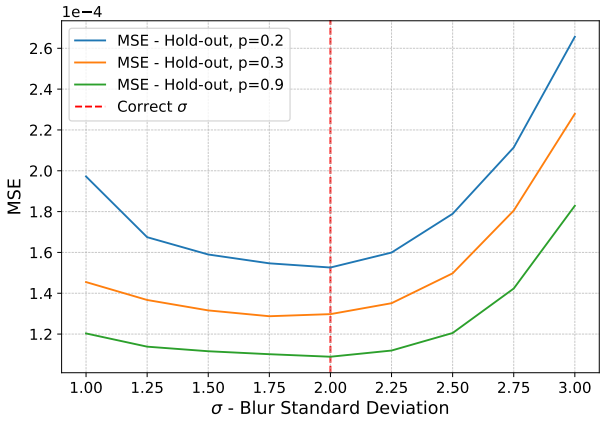


Figure 5: Average $\widehat{R}$ for given values of p over a range of blur standard deviation for a Gaussian blur kernel with fixed but potentially sub-optimal λ.

accepts a binary mask. Theoretical analysis shows that the score is an affinely unbiased estimate of the predictive risk and provided the first finite-sample variance formula and distribution-free confidence interval for pixel-mask cross-validation.

The theoretical analysis opens many statistical and algorithmic extensions for future work such as:

- tighter variance bounds and parameter error bounds,
- improved search strategies by using the variance results, bayesian optimisation or smoothness of the parameter space,
- proving that a unique minimiser exists for the hold out error.

7. Code

Code for this paper can be found at https://github.com/sjtrny/sibpr-dicta25.

References

[1] Average Primary Smartphone Camera Resolution Reaches 54MP. https://www.counterpointresearch.com/insight/average-primary-smartphone-camera-resolution-reaches-54mp [Accessed 15-07-2025]. 4

[2] J. Batson and L. Royer. Noise2Self: Blind denoising by self-supervision. In *Proc. ICML*, pages 524–533, 2019. 2

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. 2

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Navier-stokes, fluid dynamics, and image and video inpainting. In *Proc. IEEE CVPR*, pages 355–362, 2001. 2

[5] T. F. Chan and J. Shen. Mathematical models for local non-texture inpainting. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043, 2002. 2

[6] S. B. Damelin and N. S. Hoang. On surface completion

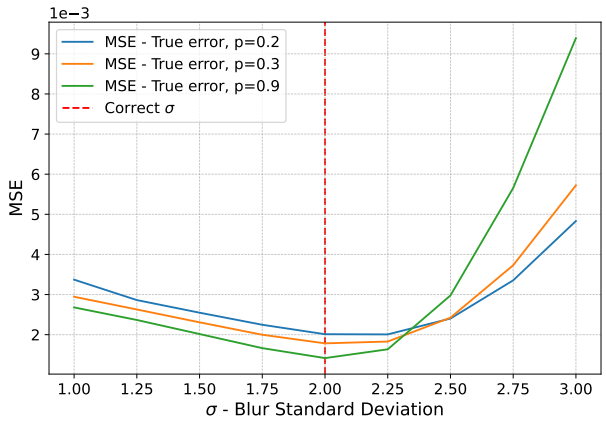


Figure 6: Average R_{pred} for given values of p over a range of blur standard deviation for a Gaussian blur kernel with fixed but potentially sub-optimal λ.

and image inpainting by biharmonic functions: Numerical aspects. *International Journal of Mathematics and Mathematical Sciences*, 2018(1):3950312, 2018. 2, 5

[7] I. Daubechies, M. Defrise, and C. D. Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. 2

[8] M. Elad, B. Kawar, and G. Vaksman. Image denoising: The deep learning revolution and beyond—a survey paper. *SIAM Journal on Imaging Sciences*, 16(3):1594–1654, 2023. 1

[9] L. Fan, F. Zhang, H. Fan, and C. Zhang. Brief review of image denoising techniques. *Visual computing for industry, biomedicine, and art*, 2(1):7, 2019. 1

[10] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2009. 1

[11] T. Goldstein and S. Osher. The split bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. 2

[12] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979. 2

[13] P. C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34(4):561–580, 1992. 2

[14] S. Hoffmann, G. Plonka, and J. Weickert. Discrete green's functions for harmonic and biharmonic inpainting with sparse atoms. In *International workshop on energy minimization methods in computer vision and pattern recognition*, pages 169–182. Springer, 2015. 2

[15] B. Kanagal and V. Sindhwani. Rank selection in low-rank matrix approximations: A study of cross-validation for nmfs. In *Proc Conf Adv Neural Inf Process*, volume 1, pages 10–15, 2010. 2

[16] H. Kobayashi, A. Solak, J. Batson, and L. A. Royer. Image deconvolution via noise-tolerant self-supervised inversion. *arXiv:2006.06156*, 2020. 2

[17] J. P. Lewis and S. Sylwan. Edge-aware inpainting with oriented biharmonic interpolation. In *ACM SIGGRAPH Sketches*, 2007. 2

[18] G. Liu, F. A. Reda, K. J. Shih, T. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *ECCV*, 2018. 2

[19] A. Lugmayr, M. Danelljan, R. Timofte, and L. Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022. 2

[20] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *ICCV Workshops*, 2019. 2

[21] A. B. Owen and J. Wang. Bi-Cross-Validation for Factor Analysis. *Statistical Science*, 31(1):119 – 139, 2016. 2

[22] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *ICCV*, 2011. 2

[23] S. J. Reeves. *A Cross-Validation Approach to Image Restoration and Blur Identification*. PhD thesis, Citeseer, 1990. 2

[24] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 1, 2

[25] C. Saharia, W. Chan, H. Chang, C. Lee, J. Cao, Jonathan Ho, T. Salimans, D. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *SIGGRAPH*, 2022. 2

[26] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981. 2

[27] R. Suvorov, E. Logacheva, A. Mashikhin, and all. Lama: Resolution-independent image inpainting with fast fourier convolutions. *arXiv:2109.07161*, 2022. 2

[28] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. *CVPR*, pages 9446–9454, 2018. 2

[29] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014. 5, 6

[30] S. Wold. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, 20(4):397–405, 1978. 2

[31] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang. Generative image inpainting with contextual attention. In *CVPR*, 2018. 2

[32] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang. Free-form image inpainting with gated convolution. In *ICCV*, 2019. 2

[33] B. Zhao, E. Weng, A. Torralba, and P. Isola. Large mask inpainting with cascade co-modulated generative adversarial networks. In *CVPR*, 2021. 2

[34] W. Zuo and Z. Lin. A generalized accelerated proximal gradient approach for total-variation-based image restoration. *IEEE Transactions on Image Processing*, 20(10):2748–2759, 2011. 2, 5