PANOM: Automatic Hyper-parameter Tuning for Inverse Problems

Tianci Liu[1], Quan Zhang[2], and Qi Lei[3]

[1]Purdue University, [2]Michigan State University, [3]Princeton University

Inverse Problems as a Bilevel optimization task

Many existing methods for inverse problems have objectives of the form

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmin}}\, L_{\text{recon}}(G(\boldsymbol{z}), \boldsymbol{y}) + \beta L_{\text{reg}}(\boldsymbol{z}), \quad \hat{\boldsymbol{x}} = G(\hat{\boldsymbol{z}}), \tag{1}$$

for some type of loss L_{recon} and penalty L_{reg}. Recent works used generative priors as regularizers and achieved state-of-the-art performances using normalizing flows [1, 2].

Empirically, the choice of hyper-parameter β is critical, but without validation data, automatic hyper-parameter tuning methods cannot be applied. We propose a new criterion by formulating it as a bilevel optimization task. Let $\mathcal{R}(\beta, \hat{\boldsymbol{x}}) := \alpha Q_1(\text{PSNR}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^{\text{MLE}})) + (1-\alpha)Q_2(\|\hat{\boldsymbol{z}}\| \mid \|\hat{\boldsymbol{z}}^{\text{MLE}}\|)$ where Q_1 and Q_2 are two probability density functions estimated by training samples of prior G. We define

$$\max_{\beta:\beta\geq 0} \;\mathcal{R}(\beta, \hat{\boldsymbol{x}}), \text{ s.t., } \hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, L_{\text{recon}}(G(\boldsymbol{z}), \boldsymbol{y}) + \beta L_{\text{reg}}(\boldsymbol{z}). \tag{2}$$

We solve Eqn (2) through a zeroth optimization by using maximum likelihood estimate given by $\beta = 0$. Define

$$\gamma_1 := \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}(\text{PSNR}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}^{\text{MLE}}) > \text{PSNR}(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}_n^{\text{MLE}})) - 0.5,\;\; \gamma_2 := \|\hat{\boldsymbol{z}}\|/h(\|\hat{\boldsymbol{z}}^{\text{MLE}}\|) - 1, \tag{3}$$

we propose *PSNR And NOrm Matching* (PANOM) shown in Algorithm 1. Intuitively, γ_1 and γ_2 encourage the recovery to have a reasonable PSNR and ratio $\|\boldsymbol{z}\|_2/\|\hat{\boldsymbol{z}}^{\text{MLE}}\|$ with high probability estimated by empirical Q_1 and Q_2.

Algorithm 1 PANOM based automated hyper-parameter tuning for inverse problems

Input: data $\boldsymbol{y}$, prior Q_1, Q_2, h, and G, β_{init}, learning rate η, and maximum number of iterations T.
Initialize $\hat{\boldsymbol{z}} = \mathbf{0}, \beta = \beta_{\text{init}}$; estimate $\hat{\boldsymbol{z}}^{\text{MLE}}(\hat{\boldsymbol{x}}^{\text{MLE}})$ with maximum $\frac{T}{3}$ steps.
while iteration number not reaching T **do**
  Inner-loop steps: update $\hat{\boldsymbol{z}}$ by K steps of gradient ascent of $\log p(\boldsymbol{y} \mid G(\boldsymbol{z})) + \beta \log p_G(G(\boldsymbol{z}))$
  Outer-loop step: update β with learning rate η by
  $$\beta \leftarrow \beta \times (1 + \eta\Delta), \quad \text{where } \Delta = \alpha \operatorname{sign}(\gamma_1)\gamma_1^2 + (1-\alpha)\operatorname{sign}(\gamma_2)\gamma_2^2$$
end while

Experimental Results: Quantitative Comparison

We use Grid Search as baseline to evaluate the efficiency of PANOM on Denoising and Noisy Compressed Sensing tasks. We compare how many grids are needed to achieve comparable recovery performances of PANOM.

Task	Noise	Examples	Ours, $\alpha =$ .5	Ours, $\alpha =$ 0	Ours, $\alpha =$ 1	Grid Search, #(grids) = 5(9)	Grid Search, #(grids) = 10(14)	Grid Search, #(grids) = 15(19)
Denoise	Gaussian	Test set	29.8 ± 0.1	$\mathbf{30.1 \pm 0.1}$	29.0 ± 0.1	28.1 ± 0.1	29.9 ± 0.1	29.8 ± 0.1
		OOD	$\mathbf{29.0 \pm 0.5}$	27.6 ± 0.5	28.5 ± 0.4	26.5 ± 0.4	28.9 ± 0.4	28.7 ± 0.5
	Poisson	Test set	30.6 ± 0.3	$\mathbf{31.0 \pm 0.2}$	29.8 ± 0.3	29.5 ± 0.2	30.7 ± 0.1	30.8 ± 0.2
		OOD	29.1 ± 0.6	29.0 ± 0.6	$\mathbf{29.2 \pm 0.6}$	27.2 ± 0.6	28.8 ± 0.6	28.5 ± 0.6
NCS	Gaussian	Test set	32.2 ± 0.2	33.9 ± 0.2	29.7 ± 0.2	34.2 ± 0.2	34.3 ± 0.2	$\mathbf{34.9 \pm 0.3}$
		OOD	31.6 ± 1.1	31.9 ± 1.0	30.9 ± 1.1	$\mathbf{32.3 \pm 0.9}$	32.1 ± 1.0	32.2 ± 0.9

Fig. 2: Averages and standard errors of recovered PSNR$(\hat{\boldsymbol{x}}, \boldsymbol{x})$ on test and OOD examples, best results are in bold. Numbers $9, 14, 19$ in parentheses in $\#$(grids) are total grid numbers on NCS task. Computational cost of Grid Search is $\frac{3}{4}\#$(grids) times of our method.

Experimental Results: Qualitative Comparison

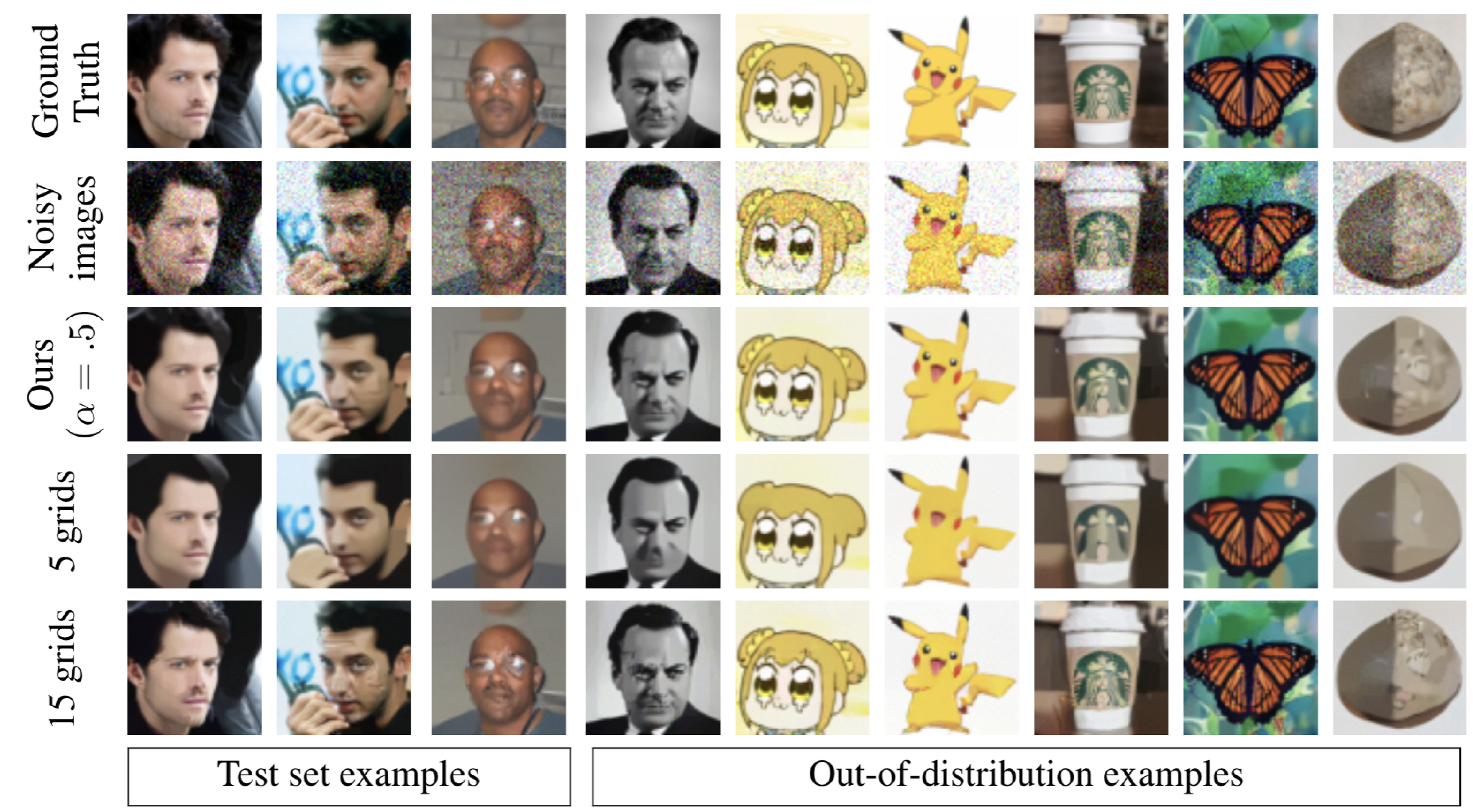

Fig. 3: Result of denoising Gaussian noise on CelebA-HQ faces and out-of-distribution images. For grid search we show recovered images achieving the highest PSNR$(\hat{\boldsymbol{x}}, \boldsymbol{x})$.

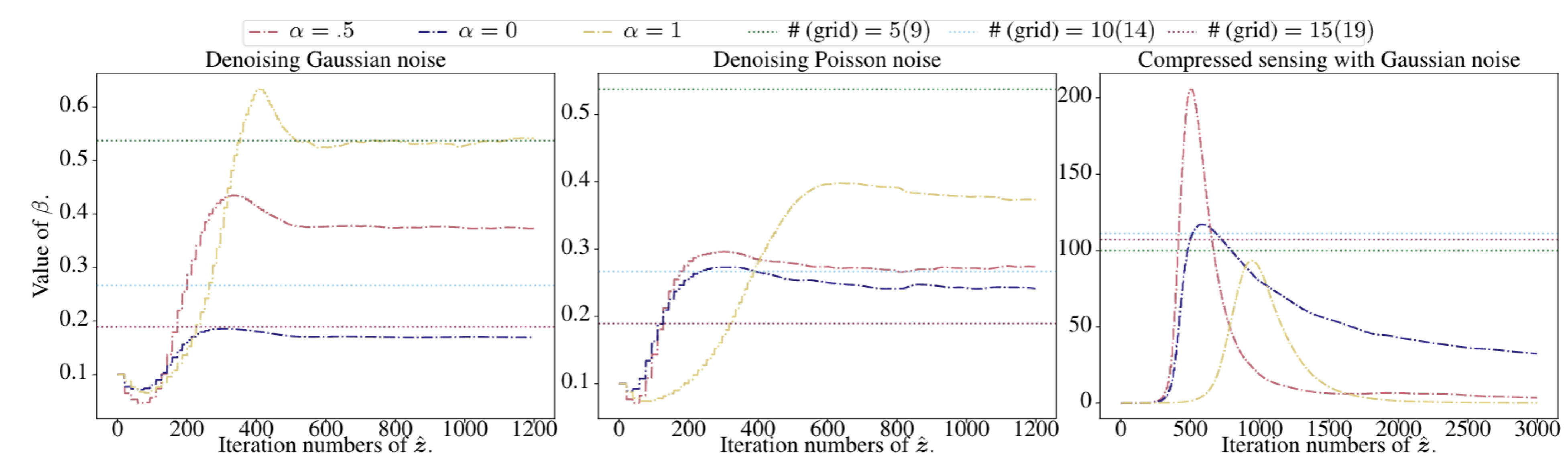

Fig. 4: Trajectories of β from different tasks on OOD examples. Grid Search dotted lines are best choices we found; numbers 9, 14, 19 in parentheses are total grid numbers on NCS task.

Conclusion

In this work, we propose PANOM, an automated hyper-parameter tuning strategy for inverse problems. Our method shows comparable recovery performance with manual grid search, but requires no human guidance and is much more efficient and convenient. We also conjecture that the proposed bilevel optimization starting from small β will facilitate the optimization procedure, and we leave the theoretical understanding to future studies.

References

[1] Muhammad Asim, Ali Ahmed, and Paul Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *CoRR* abs/1905.11672 (2019). arXiv: 1905.11672. URL: http://arxiv.org/abs/1905.11672.

[2] Jay Whang, Qi Lei, and Alexandros G Dimakis. "Compressed sensing with invertible generative models and dependent noise". In: *arXiv preprint arXiv:2003.08089* (2020).