Volume-Oriented Uncertainty for Inverse Problems

**Omer Belhasin, Yaniv Romano*, Daniel Freedman,
Ehud Rivlin, and Michael Elad**

{omerbe,danielfreedman,ehud,melad}@verily.com
yromano@technion.ac.il

Verily Life Sciences - Research, AI Team
Haifa, Israel

Abstract

Uncertainty quantification for imaging-related inverse problems is drawing much attention lately. Existing approaches towards this task define uncertainty regions per pixel while ignoring spatial correlations. In this paper we propose PUQ (Principal Uncertainty Quantification) – a novel definition of uncertainty that takes into account spatial relationships within the image, thus providing reduced uncertainty volume. Leveraging diffusion models, we derive uncertainty intervals around principal components of the empirical posterior distribution, accompanied by probabilistic guarantees. The proposed approach can operate globally on the entire image, or locally on patches, resulting in informative and interpretable uncertainty regions. We verify our approach on several inverse problems, showing a significantly tighter uncertainty regions compared to baseline methods.

1 Introduction

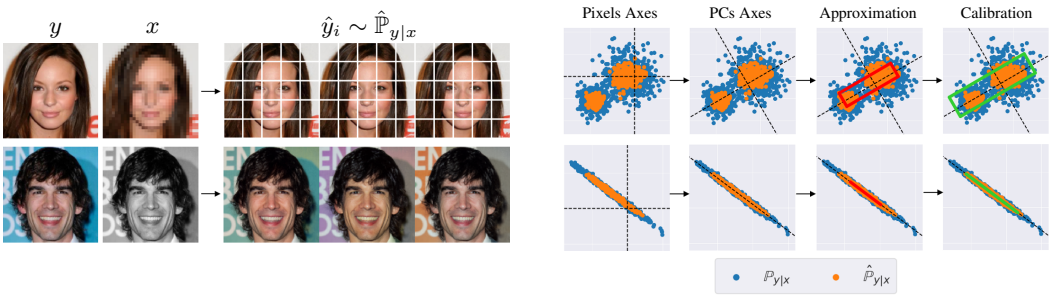
Inverse problems in imaging (e.g. colorization) are typically ill-posed, implying that such problems may have multiple solutions given the measurements. Uncertainty quantification aims to characterize the range of possible solutions, their spread, and variability. This may have an important role in various applications (e.g. astronomy and medical imaging) in which it is necessary to establish statistical boundaries for possible deviations.

Prior work [1, 2] addressed uncertainty assessment by forming intervals for each pixel. While this approach is simple, it disregards spatial correlations within the image, thus leading to an exaggerated uncertainty range. The work in [3] has improved over the above by quantifying the uncertainty in a latent space. However, by relying on a rigid and uncertainty-oblivious transformation, this method suffers from severe limitations. Appendix A gives a detailed discussion on related work.

In this paper we propose a novel approach that accounts for spatial relationships while operating in the image domain, thus enabling a full and clear interpretation of the quantified uncertainty region. The proposed technique, termed *Principal Uncertainty Quantification* (PUQ), uses the principal components of the empirical posterior probability density function (PDF), for describing the spread of possible solutions. This approach reduces the uncertainty volume, as demonstrated hereafter.

Our work leverages recent advancements in Diffusion models, serving as stochastic solvers for inverse problems. Diffusion models offer a systematic and well-motivated algorithmic path towards the task of sampling from a prior PDF, $\mathbb{P}_y$, through the repeated application of a trained image-denoiser [4, 5, 6]. An important extension of the above allows the sampler to become conditional, drawing from the posterior PDF, $\mathbb{P}_{y|x}$, where x represents the observed measurements. This approach has recently gained significant attention [6, 7, 8, 9], as it recovers a variety of candidate solutions while targeting high perceptual quality.

*Y. Romano is affiliated with the Department of Computer Science, Technion - Israel Institute of Technology.



(a) (b)

Figure 1: (a) Illustration of the sampling procedure for the approximation phase, both locally on patches (first row) and globally on the entire image (second row). (b) Illustration of our PUQ procedure in 2D ($d = 2$) for a single $x \in \mathcal{X}$. The top row corresponds to the case when $K = d = 2$ (as in E-PUQ), while the bottom row depicts the case when $K = 1 < d = 2$ (as in DA-PUQ).

Our goal is to generalize the pixelwise uncertainty assessment so as to incorporate spatial correlations between pixels. This is obtained by considering an instance-adaptive basis for a linear space that replaces the standard one in the pixelwise approach. To optimize the volume of the uncertainty region, we propose to compute and leverage the principal components of the candidate restorations, as obtained by the diffusion-based sampler. Our method may be applied both globally (on the entire image) or locally (on selected portions or patches), yielding a tighter and more accurate encapsulation of statistically valid uncertainty regions.

Our method offers two *conformal prediction* [10, 11, 12] calibration options using the Learn-then-Test [13] scheme, with a trade-off between precision and complexity. The former uses the entire set of principal components, and the later extracts a predetermined subset of them. These calibration procedures ensure the validity of the uncertainty region to contain the unknown true values with a user-specified confidence probability, while also ensuring the recovery of the unknown true values using the selected principal components. This approach allows for efficient navigation within the uncertainty region of highly probable solutions.

In summary, the contributions of this paper are the following: (1) We introduce a generalized definition of uncertainty that takes into account spatial correlations, thus providing tight uncertainty regions and better posterior coverage; (2) We present two novel calibration procedures that accompany the above by statistical guarantees; and (3) We provide a comprehensive empirical study to demonstrate the effectiveness of the proposed approach.

2 Principal Uncertainty Quantification (PUQ)

Let $\mathbb{P}_{x,y}$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y} \subset [0,1]^d$ represent the input and the output spaces, respectively, for the inverse problem at hand. E.g., for image colorization, $\mathcal{Y}$ could represent full-color high-quality images, while $\mathcal{X}$ represents their colorless versions to operate on.

Given an input measurement $x \in \mathbb{R}^d$, we aim to quantify the uncertainty of the possible solutions to the inverse problem, as manifested by the estimated d-dimensional posterior distribution, $\hat{\mathbb{P}}_{y|x}$. The idea is to improve upon the definition of pixelwise uncertainty intervals, as in [1], by considering the spatial correlation between pixels. This is achieved by constructing uncertainty intervals using a designated collection of orthonormal basis vectors for $\mathbb{R}^d$ instead of intervals over individual pixels. We denote this collection by $\hat{B}(x) = \{\hat{v}_1(x), \hat{v}_2(x) \dots \hat{v}_d(x)\}$, where $\hat{v}_i(x) \in \mathbb{R}^d$. Note that these vectors are instance-dependent, thus best adapted to their task. In our work we use the principal components (PC) of samples from $\hat{\mathbb{P}}_{y|x}$ in order to incorporate semantic factors within the image.

Similar to previous work [1, 2], we use intervals centered around the conditional mean image, $\hat{\mu}(x) = \mathbb{E}[y|x] \in \mathbb{R}^d$. The following function constructs prediction intervals along each basis-vector:

$$\mathcal{T}(x; \hat{B}(x))_i := \left[\hat{v}_i(x)^T \hat{\mu}(x) - \hat{l}(x)_i, \hat{v}_i(x)^T \hat{\mu}(x) + \hat{u}(x)_i\right], \tag{1}$$

where $i \in \{1, 2 \dots d\}$ is a basis-vector index, and $\{\hat{l}(x)_i, \hat{u}(x)_i\} \in \mathbb{R}^+$ are the lower and upper interval boundaries. If $\hat{y} \sim \hat{\mathbb{P}}_{y|x}$ is a candidate solution, $\hat{v}_i(x)^T \hat{y}$ is its i-th projection, and this value should fall within $\mathcal{T}(x; \hat{B}(x))_i$ with high probability. The uncertainty intervals using these basis-vectors form a d-dimensional hyper-rectangle, referred to as the *uncertainty region*. in order

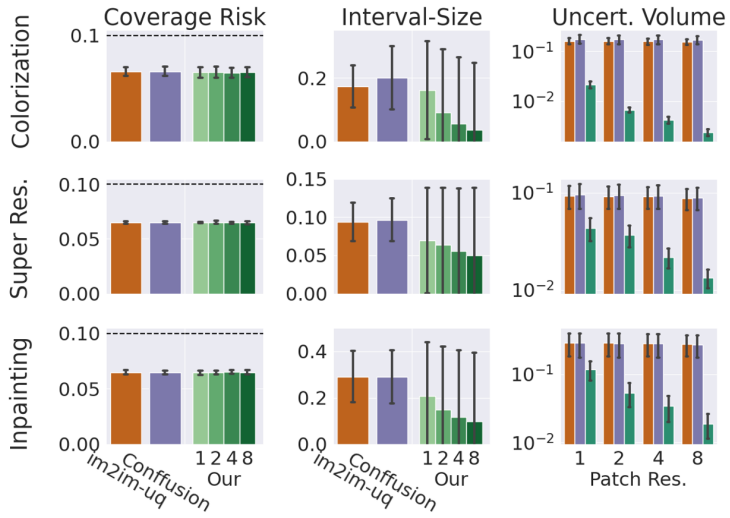


Figure 2: **Local Experiments:** A comparison of E-PUQ (applied on patches $\alpha = \delta = 0.1$) with [1, 2].

to quantify the obtained uncertainty across different regions, we introduce a new metric called the *uncertainty volume*, $\mathcal{V}(x; \mathcal{T}(x; \hat{B}(x)))$, given by

$$\mathcal{V}(x; \mathcal{T}(x; \quad \hat{B}(x))) := \sqrt[d]{\prod_{i=1}^{d} \left[\hat{u}(x)_i + \hat{l}(x)_i\right]}. \tag{2}$$

Naturally, our goal is to obtain the smallest possible volume while providing solid statistical guarantees. Our proposed method thus consists of two phases, as illustrated in Figure 1. In the first, referred to as the *approximation phase*, we use a conditional diffusion-model [8] to estimate the various terms in Equation (1): the PCs $\hat{B}(x) = \{\hat{v}_1(x), \hat{v}_2(x), \ldots, \hat{v}_K(x)\}$ ($K \leq d$), the interval boundaries $\{\hat{l}(x)_i, \hat{u}(x)_i\}_{i=1}^{K}$, and $\hat{\mu}(x)$. These estimations are used to construct heuristic uncertainty intervals in the K-dimensional space. The above-described approximation phase is merely an estimation, as the corresponding intervals may not contain the projected ground-truth values with a desired confidence. Additionally, the basis vectors may not be able to recover the ground-truth pixel values within an acceptable threshold if $K < d$. Therefore, a calibration-phase on an held-out set of data is in order.

The *calibration phase* we propose has two candidate procedures – a direct and complete one, and an approximated and thus faster alternative. The choice between the two depends on the user, taking into account the trade-off between precision and complexity. The two calibration strategies are the following (the detailed steps of these calibration procedures are brought in Appendix B):

(1) Exact PUQ (E-PUQ): Assuming that d PCs can be constructed and maintained in full, this calibration procedure is straightforward, involving only scaling of the intervals until they contain the appropriate proportion of the projected ground-truth values, similar to previous work [1, 2].

(2) Dimension-Adaptive PUQ (DA-PUQ): We may choose the leading $K < d$ PC's while omitting the rest. In this case, we should choose K while verifying a small guaranteed reconstruction error for unseen data. K is dynamically determined per input image, to accommodate the inner complexity of $\mathbb{P}_{y|x}$. Once K has been chosen, the calibration itself for the K intervals is done as in E-PUQ.

3 Experiments

We turn to present an empirical study of PUQ with three challenging tasks: image colorization, super-resolution, and inpainting, using the CelebA-HQ dataset [14]. Our approximation phase samples from the posterior, applied in our work by the SR3 conditional diffusion model [8]. The experiments we present verify that our method satisfies both the reconstruction and coverage guarantees and demonstrate that PUQ provides more confined uncertainty regions compared to prior work, including im2im-uq [1] and Conffusion [2]. Appendix C brings more details on these experiments.

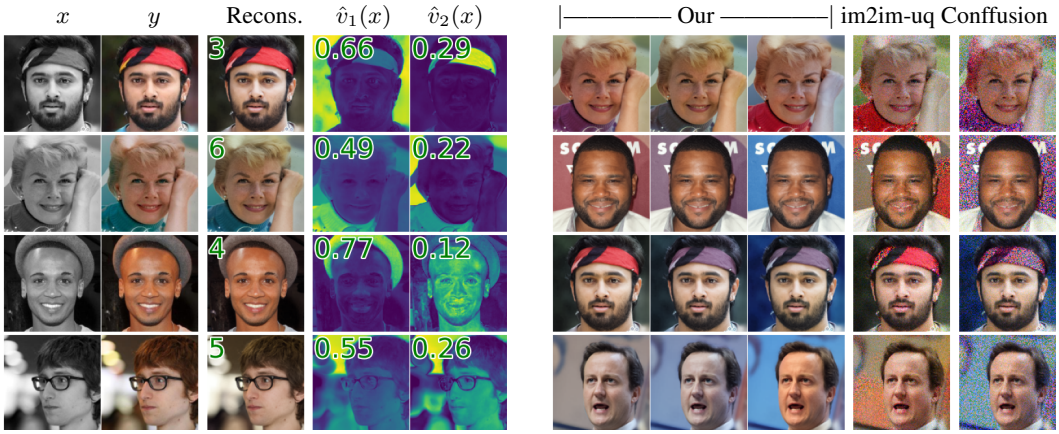


Figure 3: **Global Experiments:** Left: Uncertainty regions provided by DA-PUQ. Each row shows the measurement x, and original image y, the recovered image and how many PCs are used, and the first two PCs along with their importance. Right: Images sampled uniformly from the estimated global uncertainty regions - As can be seen, our images are realistic while the others are noisy and perceptually flawed.

3.1 Local Experiments on Patches

We apply E-PUQ on RGB patches of increasing size — 1x1, 2x2, 4x4, and 8x8 — for image colorization, super-resolution, and inpainting tasks. The obtained results are illustrated and compared to baseline methods in Figure 2. As can be seen, our method provides smaller uncertainty volumes, and thus more confined uncertainty regions, when compared to previous work in all tasks and for all patch resolutions, and while satisfying the same statistical guarantees in all cases. E-PUQ leads to an improvement factor of $\sim \times 100$ in the uncertainty volumes in colorization and $\sim \times 10$ in super-resolution and inpainting, when applied to the highest resolution of 8x8. Interestingly, as the patch resolution increases, we observe a desired trend of uncertainty volume reduction, indicating that our method takes into account spatial correlation to reduce uncertainty. Note that the standard deviation of the interval-size of our approach is higher than that of the baseline methods. This happens because the intervals of the first few PCs are wider than the remaining ones. However, the majority of the interval sizes are significantly smaller, resulting in a much smaller uncertainty volume.

3.2 Global Experiments on Images

We turn to examine the effectiveness and validity of DA-PUQ when applied to complete images at a resolution of 128×128, and consider only the colorization task, as this task incorporates strong spatial correlation. Under this setting, most of the image variability could be represented via the first few PCs while maintaining a small reconstruction risk. Note that the E-PUQ procedure does not apply to the global approach, as it requires computing and maintaining $d = 128 \times 128 \times 3$ PCs. The results are brought in Figure 3. Our method can operate with as few as 2-6 PCs, emerging from as few as 30 posterior samples, and yet maintain a guaranteed and small reconstruction risk, while providing significantly smaller uncertainty volumes compared to our local results and previous work [1, 2]. To illustrate of the tightness of the uncertainty region proposed, we sample from it, resulting is valid and colorful solutions, whereas such sampling with [1, 2] results in regions that look like noise.

4 Concluding Remarks

We present "Principal Uncertainty Quantification" (PUQ), a novel and effective approach for quantifying uncertainty for ill-posed image-to-image tasks. PUQ takes into account the spatial dependencies between pixels in order to achieve significantly tighter uncertainty regions. This method can be operated both locally or globally, with merits to each per the task being pursued. PUQ is shown to outperform existing methods in image colorization, super-resolution and inpainting, by improving substantially the uncertainty volume. Additionally, by allowing for a small reconstruction error, PUQ produces tight uncertainty regions with a few axes and thus improves computational complexity and interpretability at inference time.

References

[1] Anastasios N Angelopoulos, Amit Pal Kohli, Stephen Bates, Michael Jordan, Jitendra Malik, Thayer Alshaabi, Srigokul Upadhyayula, and Yaniv Romano. Image-to-image regression with distribution-free uncertainty quantification and applications in imaging. In *International Conference on Machine Learning*, pages 717–730. PMLR, 2022.

[2] Eliahu Horwitz and Yedid Hoshen. Conffusion: Confidence intervals for diffusion models. *arXiv preprint arXiv:2211.09795*, 2022.

[3] Swami Sankaranarayanan, Anastasios N Angelopoulos, Stephen Bates, Yaniv Romano, and Phillip Isola. Semantic uncertainty intervals for disentangled latent spaces. *arXiv preprint arXiv:2207.10074*, 3, 2022.

[4] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[6] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

[7] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23(47):1–33, 2022.

[8] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[9] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022.

[10] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

[11] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

[12] Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

[13] Anastasios N Angelopoulos, Stephen Bates, Emmanuel J Candès, Michael I Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control. *arXiv preprint arXiv:2110.01052*, 2021.

[14] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[15] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[16] Zahra Kadkhodaie and Eero P Simoncelli. Solving linear inverse problems using the prior implicit in a denoiser. *arXiv preprint arXiv:2007.13640*, 2020.

[17] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[18] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Stochastic image denoising by sampling from the posterior distribution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1866–1875, 2021.

[19] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

[20] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022.

[21] Zahra Kadkhodaie, Florentin Guth, Stéphane Mallat, and Eero P Simoncelli. Learning multi-scale local conditional probability models of images. *arXiv preprint arXiv:2303.02984*, 2023.

[22] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

[23] Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48):e2107794118, 2021.

[24] Matteo Sesia and Yaniv Romano. Conformal prediction using conditional histograms. *Advances in Neural Information Processing Systems*, 34:6304–6315, 2021.

[25] Chirag Gupta, Arun K Kuchibhotla, and Aaditya Ramdas. Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognition*, 127:108496, 2022.

[26] Danijel Kivaranovic, Kory D Johnson, and Hannes Leeb. Adaptive, distribution-free prediction intervals for deep networks. In *International Conference on Artificial Intelligence and Statistics*, pages 4346–4356. PMLR, 2020.

[27] Stephen Bates, Anastasios Angelopoulos, Lihua Lei, Jitendra Malik, and Michael Jordan. Distribution-free, risk-controlling prediction sets. *Journal of the ACM (JACM)*, 68(6):1–34, 2021.

[28] Anastasios N Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. *arXiv preprint arXiv:2208.02814*, 2022.

[29] Jacopo Teneggi, Matthew Tivnan, Web Stayman, and Jeremias Sulam. How to trust your diffusion model: A convex optimization approach to conformal risk control. In *International Conference on Machine Learning*, pages 33940–33960. PMLR, 2023.

[30] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[31] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

A Related Work

A promising recent approach towards solving inverse problems in imaging relies on generative AI techniques. These new tools enable to sample from the posterior distribution of the output image given the measurements. *Denoising Diffusion Probabilistic Models* (DDPM) [4, 5] is such a new paradigm for regular and conditional image synthesis [15, 16, 17, 18, 19, 6, 7, 8, 9, 20, 21]. SR3 [8] is the diffusion-based method we use in this work.

The field of uncertainty quantification has seen much progress in recent years [22, 23, 24, 25, 26]. A key paradigm is *conformal prediction* (CP) [10, 11, 12] and *risk-controlling* methods [27, 13, 28]. These allow to rigorously quantify the prediction uncertainty of a machine learning model with a user-specified probability guarantee. Only a few of these contributions have focused on uncertainty assessment for image restoration problems, and these include im2im-uq [1] and Conffusion [2]. The work reported in [29] is closely related as well, as it introduced a generalized, and thus improved, calibration scheme for Conffusion [2]. These works employed a risk-controlling paradigm [27] to provide prediction intervals over the pixel domain, ensuring the inclusion of ground-truth solutions. Both these approaches share the same limitation of operating in the pixel domain and disregarding spatial correlations within the image, thus leading to exaggerated volume of uncertainty.

An exception to the above is [3], which quantifies uncertainty in the latent space of GANs. Their migration from the image domain to the latent space is a rigid, global, non-linear, non-invertible and uncertainty-oblivious transformation. Therefore, quantification of the uncertainty in this domain is quite limited. More specifically, rigidity implies that this approach cannot adapt to the complexity of the problem by adjusting the latent space dimension; Globality suggests that it cannot be operated locally on patches in order to better localize the uncertainty assessments; Being non-linear implies that an evaluation of the uncertainty volume in the image domain is hard and next to impossible; Non-invertability of this transformation means that some energy is lost from the image in the analysis and not accounted for, thus hampering the validity of the statistical guarantees; Finally, note that the latent space is, at best, associated with the image content, but does not represent the prime axes of the uncertainty behavior. All these shortcomings are converted into advantages in our proposed approach advocated in this paper. Note, however, that due to the above, and especially the inability to provide certified volumes of uncertainty, a comparison of our method to [3] is impossible.

Inspired by the above contributions, this work proposes a novel alternative uncertainty quantification approach that takes spatial relationships into account. Our work provides tight uncertainty regions, compared to prior work, with user-defined statistical guarantees through the use of a CP-based paradigm. Specifically, we adopted the Learn then Test [13] procedure that provides statistical guarantees for controlling multiple risks in a general setting.

B Description of E-PUQ and DA-PUQ

Below is a formal and detailed description of the approximation phase and the two calibration procedures: E-PUQ and DA-PUQ.

B.1 Approximation Phase

This phase estimates the necessary terms for the uncertainty intervals, as defined in Equation (1). Leveraging recent advances in diffusion models, we generate high-quality samples from $\hat{\mathbb{P}}_{y|x}$. Formally, we define $f_\theta : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ as a stochastic regression solver for an inverse problem, where $\mathcal{Z}$ is the noise seed space. Given an input instance x, we propose to generate K samples, denoted by $\{f_\theta(x, z_i)\}_{i=1}^K$, where, $f_\theta(x, z_i) \sim \hat{\mathbb{P}}_{y|x}$. These are used to estimate the PCs of possible solutions and their singular values using the SVD decomposition of the generated samples around the mean image. The singular values serve as importance weights, denoted by $\hat{w}(x) \in \mathbb{R}^K$. Additionally, the samples are utilized to estimate the conditional mean, $\hat{\mu}(x)$, and the lower and upper bounds, $\tilde{l}(x)$ and $\tilde{u}(x)$, necessary for Equation (1). $\tilde{l}(x)$ and $\tilde{u}(x)$ are obtained by calculating quantiles of the projected samples onto each PC, with a user-specified miscoverage ratio $\alpha \in (0, 1)$. To capture the full spread and variability of $\hat{\mathbb{P}}_{y|x}$, it is necessary to generate at least $K = d$ samples to feed to the SVD procedure. Alternatively, we suggest working locally on patches, where d is small and fully

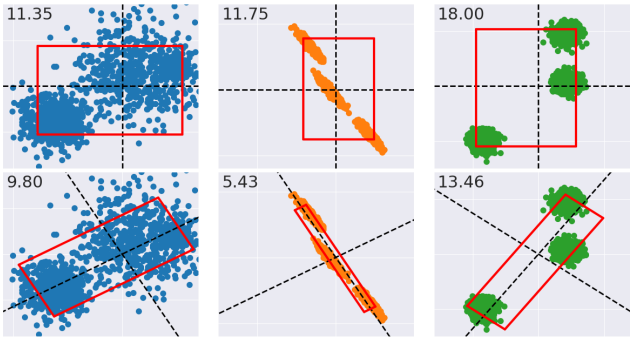


Figure 4: An illustration of uncertainty regions (in red) of 2d posterior distributions and considering three different PDF behaviors, shown in blue, orange, and green. These regions are defined by two intervals, where $\hat{l}(x)$ and $\hat{u}(x)$ represent the 0.05 and 0.95 quantiles over the dashed black axes. The top row presents regions in the pixel domain using standard basis vectors that ignores spatial correlations. The lower row presents the regions using the principal components as the basis. The uncertainty volume is indicated in the top left corner of each plot. The 90% coverage guarantee is satisfied by all regions. However, the lower row regions that take spatial dependencies into account, are significantly smaller than the pixelwise corresponding regions in the upper row.

controlled by the user by specifying the patch size to work on. Algorithm 1 formally summarizes the approximation phase.

Algorithm 1 Approximation Phase

Require: Instance $x \in \mathcal{X}$. Conditional stochastic generative model $f_\theta : \mathcal{X} \to \mathcal{Y}$ or $f_\theta : \mathcal{X} \to \mathcal{Y}_{\text{patch}}$. Maximal PCs / samples number $K \leq d$. Misscoverage ratio $\alpha \in (0, 1)$.

▷ Generate samples drawn from $\hat{\mathbb{P}}_{y|x}$

for $i = 1$ to K **do**
  $\hat{y}_i(x) \leftarrow f_\theta(x, z_i)$
end for

▷ Compute conditional mean

$\hat{\mu}(x) \leftarrow \frac{1}{K} \sum_{i=1}^{K} \hat{y}_i(x)$

▷ Apply SVD decomposition and extract the PCs and weights

$\hat{Y}(x) \leftarrow [\hat{y}_1(x), \hat{y}_2(x) \dots \hat{y}_K(x)] \in \mathbb{R}^{d \times K}$
$\hat{Y}(x) - \hat{\mu}(x) \cdot \mathbf{1}_K^T = \hat{V}(x)\hat{\Sigma}(x)\hat{U}(x)^T$
$\hat{B}(x) \leftarrow \{\hat{v}_1(x), \hat{v}_2(x) \dots \hat{v}_K(x)\}$, where $\hat{v}_i(x) = [\hat{V}(x)]_i$
$\hat{w}(x) \leftarrow \left[\hat{\sigma}_1(x)^2, \dots, \hat{\sigma}_K(x)^2\right] / c \in \mathbb{R}^K$, where $\hat{\sigma}_i(x) = [\hat{\Sigma}(x)]_i$ and $c = \sum_{j=1}^{K} \hat{\sigma}_j(x)^2$.

▷ Compute $\alpha/2$ and $1 - \alpha/2$ empirical quantiles of projected samples onto each PC

for $i = 1$ to K **do**
  $\tilde{l}(x)_i \leftarrow \hat{\mathcal{Q}}_{\alpha/2}(\{\hat{v}_i(x)^T(\hat{y}_j(x) - \hat{\mu}(x))\}_{j=1}^K)$
  $\tilde{u}(x)_i \leftarrow \hat{\mathcal{Q}}_{1-\alpha/2}(\{\hat{v}_i(x)^T(\hat{y}_j(x) - \hat{\mu}(x))\}_{j=1}^K)$
end for

Ensure: K PCs $\hat{B}(x)$, importance weights $\hat{w}(x)$, conditional mean $\hat{\mu}(x)$, lower and upper bounds $\tilde{l}(x)$ and $\tilde{u}(x)$.

B.2 Calibration Goals and Notations

The interval-valued function $\mathcal{T}$, defined in Equation (1), should produce valid intervals that contain a user-specified fraction of the projected ground-truth values, defined as $\{\hat{v}_i(x)^T y\}_{i=1}^K$, where x is an unseen test instance, $y \in \mathbb{R}^d$ is its unknown ground-truth, and $\hat{B}(x) := \{\hat{v}_1, \dots, \hat{v}_K\}$ are the selected axes of the uncertainty region. We define a user-specified risk level, $\alpha \in (0, 1)$, on the coverage of $\{\hat{v}_i(x)^T y\}_{i=1}^K$ in the uncertainty intervals in Equation (1). This is achieved by using the following holistic expression that aggregates the effect of all the intervals:

$$\mathbb{E}\left[\sum_{i=1}^{d} \hat{w}_i(x) \cdot \mathbf{1}\left\{\hat{v}_i(x)^T y \in \mathcal{T}(x; B(x))_i\right\}\right] > 1 - \alpha, \tag{3}$$

where $\hat{w}_i(x) \in [0,1]$ s.t. $\sum_{i=1}^{d} \hat{w}_i(x) = 1$ are weight factors that set the importance of covering the projected ground-truth values along each interval.

As demonstrated in Figure 4, if the orthonormal basis in Equation (1) is chosen to be the standard one, we get the pixel-based intervals that disregard spatial correlations within the image, thus leading to an exaggerated uncertainty region. In this work we address this limitation by transitioning to an instance-adapted orthonormal basis of $\mathbb{R}^d$ that allows the description of uncertainty using axes that are not necessarily pixel-independent, thereby providing tighter uncertainty regions. Our choice of a linear and orthonormal representation comes as a natural extension of the pixelwise approach, retaining much of the simplicity and efficiency of treating each axis separately. Note that the orthogonality enables the decomposition of y around $\hat{\mu}(x)$ via its projected values, $y = \hat{\mu}(x) + \sum_{i=1}^{d} \left[\hat{v}_i(x)^T (y - \hat{\mu}(x))\right] \hat{v}_i(x)$, which we refer to as the *exact reconstruction property*.

When operating in high dimensions (e.g. on the full image), providing uncertainty intervals for all the d-dimensions poses severe challenges, both in complexity and interpretability. In this case, constructing and maintaining the basis vectors becomes infeasible. Moreover, the uncertainty quantification using these intervals may be less intuitive compared to the conventional pixelwise approach because of the pixel-dependency between the basis vectors, which makes it difficult to communicate the uncertainty to the user. To mitigate these challenges, we propose an option of using $K \ll d$ basis vectors that capture the essence of the uncertainty. We later discuss how to dynamically adjust K to provide fewer uncertainty axes. While reducing the number of basis vectors benefits in interpretability and complexity, this option does not fulfill the exact reconstruction property that was discussed earlier. Therefore, we propose an extension to the conventional coverage validity of Equation (3) that takes into account the reconstruction error of the decomposed ground-truth images. Specifically, the user sets a ratio of pixels, $q \in \mathbb{R}$, and a maximum acceptable reconstruction error over this ratio, $\beta \in (0,1)$. This approximation allows us to reduce the number of basis vectors used to formulate $\hat{B}(x)$, such that the reconstruction will be valid according to the following condition:

$$\mathbb{E}\left[\hat{\mathcal{Q}}_q\left(\left\{\left|\sum_{j=1}^{K} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c\right|_i\right\}_{i=1}^{d}\right)\right] \leq \beta\,, \tag{4}$$

where $y_c := y - \hat{\mu}(x)$ is the ground-truth image centered around $\hat{\mu}(x)$, and $\hat{\mathcal{Q}}_q(\cdot)$ is the empirical quantile function defined by the smallest z satisfying $\frac{1}{d}\sum_{i=1}^{d} \mathbf{1}\{z_i \leq \hat{\mathcal{Q}}_q(z)\} \geq q$. As an example, setting $q = 0.9$ and $\beta = 0.05$ would mean that the maximal reconstruction error of 90% of the ground-truth pixels is no more than 5% of the $[0,1]$ dynamic range.

In order to refine the approximation phase and obtain valid uncertainty axes and intervals that satisfy the guarantees of Equation (3) and Equation (4) with high probability, it is necessary to apply a calibration phase over an held-out set of calibration data, denoted by $\mathcal{S}_{\text{cal}} := \{(x_i, y_i)\}_{i=1}^{n}$. We propose two different options based on particular conditions on the number of PCs to be constructed and maintained during the calibration procedure or during inference, when applied either globally or locally. Below we outline each of these options in more details.

B.3 Exact PUQ

The *Exact PUQ* (E-PUQ) procedure provides the complete uncertainty of the d-dimensional posterior distribution, $\mathbb{P}_{y|x}$. In this case, the exact reconstruction property discussed earlier is satisfied, and Equation (4) is fulfilled with 0% error ($\beta = 0$) across 100% ($q = 1.0$) of the pixels. Therefore, the calibration is simple, involving only a scaling of intervals to ensure Equation (3) is satisfied with high probability, similar to previous work [1, 2].

Formally, for each input instance x and its corresponding ground-truth $y \in \mathbb{R}^d$ in $\mathcal{S}_{\text{cal}}$, we use the estimators obtained in the approximation phase to get d PCs of possible solutions, $\hat{B}(x)$, their corresponding importance weights $\hat{w}(x)$, the conditional mean $\hat{\mu}(x)$, and the lower and upper bounds, denoted by $\tilde{l}(x)$ and $\tilde{u}(x)$. We then define the scaled intervals to be those specified in Equation (1), with the upper and lower bounds defined as $\hat{u}(x) := \lambda\tilde{u}(x)$ and $\hat{l}(x) := \lambda\tilde{l}(x)$, where $\lambda \in \mathbb{R}^+$ is a tunable parameter that controls the scaling. Notably, the size of the uncertainty intervals decreases as λ decreases. We denote the scaled uncertainty intervals by $\mathcal{T}_\lambda(x; \hat{B}(x))$.

The following weighted coverage loss function is used to guide our design of λ:

$$\mathcal{L}(x, y; \lambda) := \sum_{i=1}^{d} \hat{w}_i(x) \cdot \mathbf{1}\left\{\hat{v}_i(x)^T y \notin \mathcal{T}_\lambda(x; \hat{B}(x))_i\right\}. \tag{5}$$

This loss is closely related to the expression in Equation (3), and while it may seem arbitrary at first, this choice is a direct extension to the one practiced in [1, 2].

Our goal is to ensure that the expectation of $\mathcal{L}(x, y; \lambda)$ is below a pre-specified threshold, $\alpha \in (0, 1)$, with high probability over the calibration data. This is accomplished by a conformal prediction based calibration scheme, and in our paper we use the LTT [13] procedure, which guarantees the following:

$$\mathbb{P}\left(\mathbb{E}[\mathcal{L}(x, y; \hat{\lambda})] \leq \alpha\right) \geq 1 - \delta\,, \tag{6}$$

for a set of candidate values of λ, given as the set $\hat{\Lambda}$. $\delta \in (0, 1)$ is an error level on the calibration set and $\hat{\lambda}$ is the smallest value within $\hat{\Lambda}$ satisfying the above condition, so as to provide the smallest uncertainty volume over the scaled intervals, as defined in Equation (2), which we denote by $\mathcal{V}_{\hat{\lambda}}$.

Put simply, the above guarantees that more than $1 - \alpha$ of the ground-truth values projected onto the full d PCs of $\hat{\mathbb{P}}_{y|x}$ are contained in the uncertainty intervals with probability at least $1 - \delta$, where the latter probability is over the randomness of the calibration set. The scaling factor takes into account the weights to ensure that uncertainty intervals with high variability contain a higher proportion of projected ground-truth values than those with low variability. This is particularly important for tasks with strong pixel correlations, where the first few PCs capture most of the variability in possible solutions. We summarize the E-PUQ procedure in Algorithm 2.

Algorithm 2 Exact PUQ Procedure

Require: Calibration set $\mathcal{S}_{\text{cal}} := \{x_i, y_i\}_{i=1}^n$. Scanned calibration parameter values $\Lambda = [1 \ldots \lambda_{\text{max}}]$. Approximation phase estimations $\hat{B}, \hat{w}, \hat{\mu}, \tilde{u}, \tilde{l}$. Misscoverage ratio $\alpha \in (0, 1)$. Calibration error level $\delta \in (0, 1)$.

for $(x, y) \in \mathcal{S}_{\text{cal}}$ **do**
  $\hat{B}(x), \hat{w}(x), \hat{\mu}(x), \tilde{u}(x), \tilde{l}(x) \leftarrow$ Apply Algorithm 1 to x, with the choice of $K = d$ samples
  for $\lambda \in \Lambda$ **do**
    ▷ Scale uncertainty intervals
    $\hat{u}(x) \leftarrow \lambda \tilde{u}(x)$ and $\hat{l}(x) \leftarrow \lambda \tilde{l}(x)$
    $\mathcal{T}_\lambda(x; \hat{B}(x)) \leftarrow$ Equation (1) using $\hat{\mu}(x), \hat{u}(x), \hat{l}(x)$
    ▷ Compute weighted coverage loss, Equation (5)
    $\mathcal{L}(x, y; \lambda) \leftarrow$
      $\sum_{i=1}^{d} \hat{w}_i(x) \cdot \mathbf{1}\left\{\hat{v}_i(x)^T y \notin \mathcal{T}_\lambda(x; \hat{B}(x))_i\right\}$
  end for
end for
$\hat{\Lambda} \leftarrow$ Extract valid λs from LTT [13] applied on $\{\mathcal{L}(x, y; \lambda) : (x, y) \in S_{\text{cal}}, \lambda \in \Lambda\}$ at risk level α and error level δ, referring to Equation (6).
▷ Compute the minimizer for the uncert. volume, Equation (2)
$\hat{\lambda} \leftarrow \arg\min_{\lambda \in \hat{\Lambda}} \left\{\frac{1}{n} \sum_{i=1}^{n} \mathcal{V}_\lambda(x_i; \hat{B}(x))\right\}$

Ensure: Given a new instance $x \in \mathcal{X}$, obtain valid uncertainty intervals for it, $\mathcal{T}_{\hat{\lambda}}(x; \hat{B}(x))$.

B.4 Dimension-Adaptive PUQ

The E-PUQ procedure assumes the ability to construct and maintain d PCs, which can be computationally challenging both locally and globally. Furthermore, an uncertainty quantification over these axes may be less intuitive, due to the many axes involved, thus harming the method's interpretability. To address these, we propose the *Dimension-Adaptive PUQ* (DA-PUQ) procedure, which describes the uncertainty region with fewer axes, $K \leq d$. The use of only a few leading dimensions, e.g., $K = 3$, can lead to a more interpretable uncertainty region, enabling an effective visual navigation within the obtained uncertainty range.

While this approach does not satisfy the exact reconstruction property, the decomposed ground-truth values can still be recovered through the K PCs with a small user-defined error in addition to the

coverage guarantee. By doing so, we can achieve both the guarantees outlined in Equation (3) and Equation (4) with high probability.

To satisfy both the coverage and reconstruction guarantees while enhancing interpretability, we use a dynamic function, $\hat{k}(x) : \mathcal{X} \to \mathbb{N}$, and a scaling factor to control the reconstruction and coverage risks. The function $\hat{k}(x)$ determines the number of top PCs (out of K) to include in the uncertainty region, focusing on the smallest number that can satisfy both Equation (3) and (4), so as to increase interpretability.

Formally, for each input instance x and its corresponding ground-truth value $y \in \mathbb{R}^d$ in $\mathcal{S}_{\text{cal}}$, we use the estimators obtained in the approximation phase to estimate $K \leq d$ PCs of possible solutions, denoted by $\hat{B}(x)$, their corresponding importance weights, denoted by $\hat{w}(x)$, the conditional mean denoted by $\hat{\mu}(x)$, and the lower and upper bounds denoted by $\tilde{l}(x)$ and $\tilde{u}(x)$, respectively. We then introduce a threshold $\lambda_1 \in (0, 1)$ for the decay of the importance weights over the PCs of solutions to x. The adaptive number of PCs to be used is defined as follows:

$$\hat{k}(x; \lambda_1) := \min_{1 \leq k \leq K} \left\{ k \quad \text{s.t.} \quad \sum_{i=1}^{k} \hat{w}_i(x) \geq \lambda_1 \right\}. \tag{7}$$

Obviously, the importance weights are arranged in a descending order, starting from the most significant axis and ending with the least significant one. Furthermore, let $q \in (0, 1)$ be a specified ratio of pixels, and $\beta \in (0, 1)$ be a maximum allowable reconstruction error over this ratio. The reconstruction loss function to be controlled is defined as:

$$\mathcal{L}_1(x, y; \lambda_1) := \hat{\mathcal{Q}}_q \left(\left\{ \left| \sum_{j=1}^{\hat{k}(x;\lambda_1)} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c \right|_i \right\}_{i=1}^{d} \right), \tag{8}$$

where $\hat{\mathcal{Q}}_q(\cdot)$ selects the empirical q-quantile of the reconstruction errors, and $y_c = y - \hat{\mu}(x)$ is the ground-truth image centered around $\hat{\mu}(x)$.

At the same time, we also control the coverage risk over the $\hat{k}(x)$ PCs, with $\alpha \in (0, 1)$ representing a user-specified acceptable misscoverage rate and $\lambda_2 \in \mathbb{R}^+$ representing the calibration factor parameter. To control this coverage risk, we define the coverage loss function to be the same as in Equation (5), but limited to the $\hat{k}(x)$ PCs, that is:

$$\mathcal{L}_2(x, y; \lambda_1, \lambda_2) := \sum_{i=1}^{\hat{k}(x;\lambda_1)} \hat{w}_i(x) \cdot \mathbf{1} \left\{ \hat{v}_i(x)^T y \notin \mathcal{T}_{\lambda_2}(x; \hat{B}(x))_i \right\}. \tag{9}$$

Finally, using the reconstruction loss function of Equation (8) and the coverage loss function of Equation (9), we seek to minimize the uncertainty volume, defined in Equation (2), for the scaled intervals where any unused axes (out of d) are fixed to zero. We denote this uncertainty volume as $\mathcal{V}_{\lambda_1,\lambda_2}$. The minimization of $\mathcal{V}_{\lambda_1,\lambda_2}$ is achieved by minimizing λ_1 and λ_2, while ensuring that the guarantees of Equation (3) and Equation (4) hold with high probability over the calibration data. This can be provided, for example, through the LTT [13] calibration scheme, which guarantees the following:

$$\mathbb{P} \left(\begin{array}{c} \mathbb{E}[\mathcal{L}_1(x, y; \hat{\lambda}_1)] \leq \beta \\ \mathbb{E}[\mathcal{L}_2(x, y; \hat{\lambda}_1, \hat{\lambda}_2)] \leq \alpha \end{array} \right) \geq 1 - \delta \,, \tag{10}$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the minimizers for the uncertainty volume among valid calibration parameter results, $\hat{\Lambda}$, obtained through the LTT procedure. In other words, we can reconstruct a fraction q of the ground-truth pixel values with an error no greater than β, and a fraction of more than $1 - \alpha$ of the projected ground-truth values onto the first $\hat{k}(x; \hat{\lambda}_1)$ PCs of $\mathbb{P}_{y|x}$ are contained in the uncertainty intervals, with a probability of at least $1 - \delta$. A detailed description of the DA-PUQ procedure is given in Algorithm 3.

Algorithm 3 Dimension-Adaptive PUQ Procedure

Require: Calibration set $\mathcal{S}_{\text{cal}} := \{x_i, y_i\}_{i=1}^n$. Scanned calibration parameter values $\Lambda^1 \leftarrow [1 \dots \lambda_{1\,\text{max}}]$ and $\Lambda^2 \leftarrow [1 \dots \lambda_{2\text{max}}]$. Maximal PCs number $K \leq d$. Approximation phase estimators $\hat{B}, \hat{w}, \hat{\mu}, \tilde{u}, \tilde{l}$. Recovered pixels ratio $q \in (0,1)$. Reconstruction error $\beta \in (0,1)$. Misscoverage ratio $\alpha \in (0,1)$. Calibration error level $\delta \in (0,1)$. For an effective calibration, α, β, δ should be close to 0 while q should be close to 1.

for $(x, y) \in \mathcal{S}_{\text{cal}}$ **do**
  $\hat{B}(x), \hat{w}(x), \hat{\mu}(x), \tilde{u}(x), \tilde{l}(x) \leftarrow$ Apply Algorithm 1 to x, with the choice of K samples
  for $\lambda_1 \in \Lambda_1$ **do**
    ▷ Compute adaptive dimensionality, Equation (7)
    $\hat{k}(x; \lambda_1) \leftarrow \min_k \left\{ k : \sum_{i=1}^{K} \hat{w}_i(x) \geq \lambda_1 \right\}$
    ▷ Compute reconstruction loss, Equation (8)
    $y_c \leftarrow y - \hat{\mu}(x)$
    $\mathcal{L}_1(x, y; \lambda_1) \leftarrow \hat{\mathcal{Q}}_q\left(\left\{ \left| \sum_{j=1}^{\hat{k}(x;\lambda_1)} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c \right|_i \right\}_{i=1}^{d} \right)$
    for $\lambda_2 \in \Lambda_2$ **do**
      ▷ Scale uncertainty intervals
      $\hat{u}(x_i) \leftarrow \lambda_2 \tilde{u}(x)$ and $\hat{l}(x) \leftarrow \lambda_2 \tilde{l}(x)$
      $\mathcal{T}_{\lambda_2}(x; \hat{B}(x)) \leftarrow$ Eq. (1) using $\hat{\mu}(x), \hat{u}(x), \hat{l}(x)$
      ▷ Compute weighted coverage loss, Equation (5)
      $\mathcal{L}_2(x, y; \lambda_1, \lambda_2) \leftarrow \sum_{i=1}^{\hat{k}(x;\lambda_1)} \hat{w}_i(x) \cdot \mathbf{1}\left\{ \hat{v}_i(x)^T y \notin \mathcal{T}_{\lambda_2}(x; \hat{B}(x))_i \right\}$
    end for
  end for
end for
$\hat{\Lambda} \leftarrow$ Extract valid λs from LTT [13] applied on $\{(\mathcal{L}_1(x, y; \lambda_1), \mathcal{L}_2(x, y; \lambda_1, \lambda_2)) : (x, y) \in S_{\text{cal}}, \lambda_1 \in \Lambda^1, \lambda_2 \in \Lambda^2\}$ at risk levels (β, α) and error level δ, referring to Equation (10)
▷ Compute the minimizers for the uncer. volume, Equation (2)
$\hat{\lambda}_1, \hat{\lambda}_2 \leftarrow \arg\min_{\lambda_1, \lambda_2 \in \hat{\Lambda}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathcal{V}_{\lambda_1, \lambda_2}(x_i; \hat{B}(x_i)) \right\}$

Ensure: Given a new instance $x \in \mathcal{X}$, obtain valid uncertainty intervals for it, $\mathcal{T}_{\hat{\lambda}_2}(x; \hat{B}(x))$ over $\hat{k}(x; \hat{\lambda}_1) \leq K$ PCs.

C Experimental Details

This section provides details of the experimental methodology employed in the experiments of this work, including the datasets used, architectures implemented, and the procedural details and hyperparameters of our method.

C.1 Datasets and Preprocessing

Our machine learning system was trained using the Flickr-Faces-HQ (FFHQ) dataset [30], which includes 70,000 face images at a resolution of 128x128. We conducted calibration and testing on the CelebA-HQ (CelebA) dataset [14], which also consists of face images and was resized to match the resolution of our training data. To this end, we randomly selected 2,000 instances from CelebA, of which 1,000 were used for calibration and 1,000 for testing. For the colorization experiments, a grey-scale transformation was applied to the input images. For the super-resolution experiments, patches at a resolution of 32x32 were averaged to reduce the input image resolution by a factor of 4 in each dimension. For the inpainting experiments, we randomly cropped pixels from the input images during the training phase, either in squares or irregular shapes; while for the calibration and testing data, we cropped patches at a resolution of 64x64 at the center of the image.

C.2 Architecture and Training

In all our experiments, we applied the approximation phase using recent advancements in conditional image generation through diffusion-based models, while our proposed approximation phase in

Algorithm 1 can accommodate any stochastic regression solvers for inverse problems, such as conditional GANs [31]. In all tasks, we utilized the framework for conditional diffusion-based models proposed in the SR3 work [8], using a U-Net architecture. For each of the three tasks, we trained a diffusion model separately and followed the training regimen outlined in the code of [8]. To ensure a valid comparison with the baseline methods, we implemented them using the same architecture and applied the same training regimen. All experiments, including the baseline methods, were trained for 10,000 epochs with a batch size of 1,024 input images.

C.3 PUQ Procedures and Hyperparameters

Our experimental approach consists of 2 sets of experiments: local experiments on patches and global experiments on entire images. For the local experiments, we conducted 4 experiments of the E-PUQ procedure on RGB patch resolutions of 1x1, 2x2, 4x4, and 8x8. We used $K = 3$, $K = 12$, $K = 48$, and $K = 192$ PCs for each resolution, respectively. We set $\alpha = \delta = 0.1$ to be the user-specified parameters of the guarantee, defined in Equation (6). For the global experiments, we used entire images at a resolution of 128x128, in which we applied the DA-PUQ procedure. As global working is suitable for tasks that exhibit strong pixel correlation, we applied this experiment only on the task of image colorization. We set $q = 0.95$, $\beta = \alpha = \delta = 0.1$, to be the user-specified parameters of the guarantee of Equation (10), and used $K = 100$ PCs. In all experiments, we employed a computationally stable version of the uncertainty volume, defined in Equation (2), which is expressed as follows:

$$\mathcal{V}(x; \mathcal{T}(x; \hat{B}(x))) \approx \exp\left(\frac{1}{d}\sum_{i=1}^{d} \log\left(\hat{u}(x)_i + \hat{l}(x)_i + \epsilon\right)\right) - \epsilon\,,$$

where we have selected $\epsilon = 1e - 10$.