Principal Uncertainty Quantification with Spatial Correlation for Image Restoration Problems

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

Abstract—Uncertainty quantification for inverse problems in imaging has drawn much attention lately. Existing approaches towards this task define uncertainty regions based on probable values per pixel, while ignoring spatial correlations within the image, resulting in an exaggerated volume of uncertainty. In this paper, we propose PUQ (Principal Uncertainty Quantification) – a novel definition and corresponding analysis of uncertainty regions that takes into account spatial relationships within the image, thus providing reduced volume regions. Using recent advancements in stochastic generative models, we derive uncertainty intervals around principal components of the empirical posterior distribution, forming an ambiguity region that guarantees the inclusion of true unseen values with a user confidence probability. To improve computational efficiency and interpretability, we also guarantee the recovery of true unseen values using only a few principal directions, resulting in ultimately more informative uncertainty regions. Our approach is verified through experiments on image colorization, super-resolution, and inpainting; its effectiveness is shown through comparison to baseline methods, demonstrating significantly tighter uncertainty regions.

Index Terms—Uncertainty and probabilistic reasoning, Probability and Statistics, Restoration, Inverse problems, Stochastic processes, Correlation and regression analysis.

I. Introduction

RESTORATION problems in imaging are widely encountered in various disciplines, including cellular cameras, surveillance, experimental physics, medical imaging, and more. Inverse problems, such as colorization, super-resolution, and inpainting, are typically ill-posed, implying that multiple solutions can explain the unknown target image given the measurements. In this context, uncertainty quantification aims to characterize the range of possible solutions, their spread, and variability. This has an especially important role in applications such as astronomy and medical diagnosis, where it is necessary to establish statistical boundaries for possible gray-value deviations. The ability to characterize the range of permissible solutions with accompanying statistical guarantees has thus become an important and useful challenge, which we address in this paper.

Prior work on this topic [1], [2] has addressed the uncertainty assessment by constructing intervals of possible values for each pixel via quantile regression [3], or other heuristics such as estimations of per-pixel residuals. While this line of thinking is appealing due to its simplicity, it disregards

O. Belhasin (omerbe@verily.com), D. Freedman (danielfreedman@verily.com), Ehud Rivlin (ehud@verily.com) and M. Elad (melad@verily.com) are with Verily Life Sciences - Research, Haifa, Israel.

O. Belhasin (omer.be@cs.technion.ac.il) and Y. Romano (yromano@technion.ac.il) are with the Department of Computer Science, Technion - Israel Institute of Technology, Haifa, Israel.
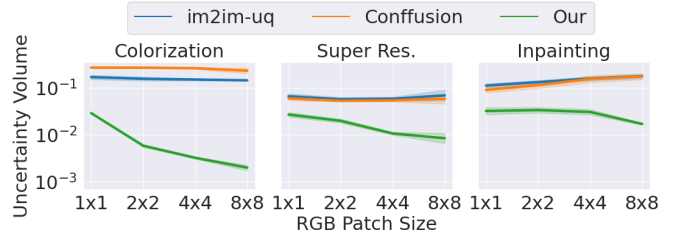

Fig. 1. Comparison of PUQ's performance on the CelebA-HQ dataset in image colorization, super-resolution, and inpainting tasks using the Exact PUQ procedure (Section IV-B1) applied on RGB image patches of varying size. As demonstrated, our method provides tighter uncertainty regions with significantly smaller uncertainty volumes ($\sim$x10 in super-resolution and inpainting, and $\sim$x100 in colorization). The compared methods are im2im-uq [1] and Conffusion [2].

spatial correlations within the image, and thus provides an exaggerated uncertainty range. The study presented in [4] has improved over the above by quantifying the uncertainty in a latent space, thus taking spatial dependencies into account. However, by relying on a rigid, global, non-linear, non-invertible and uncertainty-oblivious transformation, this method suffers from severe interpretability limitations.

In this paper, we propose *Principal Uncertainty Quantification* (PUQ) – 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. PUQ uses the principal components of the empirical posterior probability density function, which describe the spread of possible solutions. This approach reduces the uncertainty volume[1], as demonstrated in Figure 1. This figure presents a comparison between our proposed Exact PUQ procedure (see Section IV-B1) and previous work [1], [2], showing a much desired trend of reduced *uncertainty volume* that further decreases as the size of the patch under consideration grows.

Our work aims to improve the quantification of the uncertainty volume by leveraging recent advancements in generative models serving as stochastic solvers for inverse problems. While our proposed approach is applicable using any such solver (e.g., conditional GAN [5]), we focus in this work on diffusion-based techniques, which have recently emerged as the leading image synthesis approach, surpassing GANs and other alternative generators [6]. Diffusion models offer a systematic and well-motivated algorithmic path towards the task of sampling from a prior probability density function (PDF), $\mathbb{P}_y$, through the repeated application of a trained image-

---

[1]The definition of this volume, which plays a critical part in this work, is further discussed in later sections and given in Equation (3).

denoiser [7], [8]. An important extension of these models allows the sampler to become conditional, drawing samples from the posterior PDF, $\mathbb{P}_{y|x}$, where x represents the observed measurements. This approach has recently gained significant attention [6], [9], [10], [11] and yields a new and fascinating viewpoint to inverse problems, recovering a variety of candidate solutions to such problems while targeting high perceptual quality.

In this work, we generalize the pixelwise uncertainty assessment, as developed in [1], [2], so as to incorporate spatial correlations between pixels. This generalization is obtained by considering an image-adaptive basis for a linear space that replaces the standard basis in the pixelwise approach. To optimize the volume of the output uncertainty region, we propose a statistical analysis of the posterior obtained from a diffusion-based sampler (e.g., [10], [11]), considering a series of candidate restorations. 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. For the purpose of adapting the basis, we compute and leverage the principal components of the candidate restorations. As illustrated in Figure 2 for a simple 2-dimensional PDFs, the pixelwise regions are less efficient and may contain vast empty areas, and especially so in cases where pixels exhibit strong correlation. Clearly, as the dimension increases, the gap between the standard and the adapted uncertainty quantifications is further amplified.

Our proposed method offers two *conformal prediction* [12], [13], [14] based calibration options (specifically, using the Learn then Test [15] scheme) for users to choose from, with a trade-off between precision and complexity. These include (i) using the entire set of principal components, (ii) using a predetermined subset of them[2]. The proposed 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 when only a subset is used. Applying these approaches allows for efficient navigation within the uncertainty region of highly probable solutions.

We conduct various local and global experiments to verify our method, considering three challenging tasks: image colorization, super-resolution, and inpainting, all described in Section V, and all demonstrating the advantages of the proposed approach. For example, when applied locally on $8{\times}8{\times}3$ patches, our experiments show a reduction in the guaranteed uncertainty volume by a factor of $\sim$10-100 compared to previous approaches, as demonstrated in Figure 1. Moreover, this local approach can have a substantially reduced computational complexity while retaining the statistical guarantees, by drawing far fewer posterior samples and using a small subset of the principal components. As another example, the global tests on the colorization task provide an unprecedented tightness in uncertainty volumes. This is accessible via a

---

[2]We also propose a variation of this option that enables controlling the number of necessary principal components to be used, thus resulting in a reduced complexity at calibration and test time.
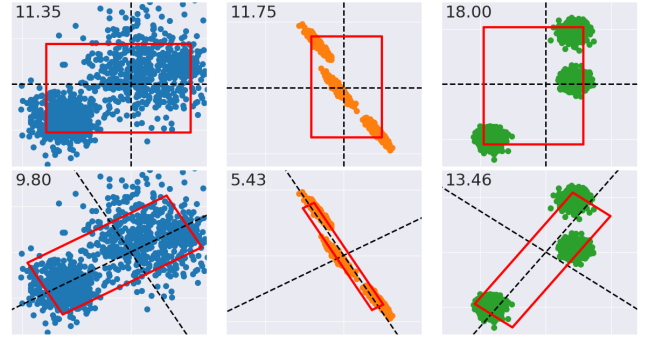

Fig. 2. An illustration of uncertainty regions (in red) of 2d posterior distributions and considering three different PDF behaviors, shown in blue, orange, and green. The uncertainty regions are formed from intervals, as defined in Equation (1), where $\hat{l}(x)$ and $\hat{u}(x)$ represent the 0.05 and 0.95 quantiles over the dashed black axes. The top row presents the uncertainty region in the pixel domain using standard basis vectors that ignores the spatial correlations, while the lower row presents the regions using the principal components as the basis. The uncertainty volume, defined in Equation (3), is indicated in the top left corner of each plot. The 90% coverage guarantee, outlined in Equation (2) with $w_i := 1/2$, 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.

reduced set of drawn samples, while also allowing for efficient navigation within the solution set.

In summary, the contributions of this paper are the following:

1) We introduce a novel generalized definition of uncertainty region that leverages an adapted linear-space basis for better posterior coverage.
2) We propose a new method for quantifying the uncertainty of inverse problems that considers spatial correlation, thus providing tight uncertainty regions.
3) We present two novel calibration procedures for the uncertainty quantification that provide statistical guarantees for unknown data to be included in the uncertainty region with a desired coverage ratio while being recovered with a small error by the selected linear axes.
4) We provide a comprehensive empirical study of three challenging image-to-image translation tasks: colorization, super-resolution, and inpainting, demonstrating the effectiveness of the proposed approach in all modes.

II. RELATED WORK

Inverse problems in imaging have been extensively studied over the years; this research domain has been deeply influenced by the AI revolution [16], [17], [18], [19], [5]. A promising recent approach towards image-to-image translation problems relies on the massive progress made on learned generative techniques. These new tools enable to model the conditional distribution of the output images given the input, offering a fair sampling from this PDF. Generative-based solvers of this sort create a new and exciting opportunity for getting high perceptual quality solutions for the problem in hand, while also accessing a diverse set of such candidate solutions.

Recently, *Denoising Diffusion Probabilistic Models* (DDPM) [7], [8] have emerged as a new paradigm for image generation, surpassing the state-of-the-art results achieved by GANs [20], [6]. Consequently, several *conditional* diffusion methods have been explored [6], [9], [10], [11], including SR3 [10] – a diffusion-based method for image super-resolution, Palette [11] – a diffusion-based unified framework for image-to-image translation tasks, and more (e.g. [21], [22], [23], [24], [25], [26], [27]). Note, however, that most current conditional algorithms for inverse problems do not offer statistical guarantees that protect against model deviations and hallucinations.

Moving to uncertainty quantification, the field of machine learning has been seeing rich work on the derivation of statistically rigorous confidence intervals for predictions [28], [29], [30], [31], [32]. One key paradigm in this context is *conformal prediction* (CP) [12], [13], [14] and *risk-controlling* methods [33], [15], [34], which allow to rigorously quantify the prediction uncertainty of a machine learning model with a user-specified probability guarantee. Despite many proposed methods, only a few have focused on mitigating uncertainty assessment in image restoration problems, including im2im-uq [1] and Conffusion [2]. These works have employed a risk-controlling paradigm [33] to provide statistically valid prediction intervals over the pixel domain, ensuring the inclusion of ground-truth solutions in the output intervals. However, both these approaches share the same limitation of operating in the pixel domain while disregarding spatial correlations within the image or the color layers. This leads to an unnecessarily exaggerated volume of uncertainty.

An exception to the above is [4], 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 (discussed in Section III) 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 [4] is impossible.

Inspired by the above contributions, we propose 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 [15] procedure that

provides statistical guarantees for controlling multiple risks in a general setting.

III. PROBLEM FORMULATION

Let $\mathbb{P}_{x,y}$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ represent the input and the output space, respectively, for the inverse problem at hand. E.g., for the task of image colorization, $\mathcal{Y}$ could represent full-color high-quality images, while $\mathcal{X}$ represents their colorless versions to operate on. We assume that $\mathcal{X}, \mathcal{Y} \subset [0,1]^d \subset \mathbb{R}^d$, where, without loss of generality, d is assumed to be the dimension of both spaces. In the context of examining patches within output images, we define $\mathcal{Y}_{\text{patch}}$ as the patch space of the output images. For simplicity, we use the same notation, d, for $\mathcal{Y}$ and $\mathcal{Y}_{\text{patch}}$, while it is clear that the dimension of $\mathcal{Y}_{\text{patch}}$ is smaller and controlled by the user through the patch size to work on.

Given an input measurement $x \in \mathbb{R}^d$, our objective is 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 enhance the definition of pixelwise uncertainty intervals as introduced in [1] by integrating the spatial correlations between pixels to yield a better structured uncertainty region. To achieve this, we propose to construct 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 $\hat{B}(x) = \{\hat{v}_1(x), \hat{v}_2(x) \ldots \hat{v}_d(x)\}$, where $\hat{v}_i(x) \in \mathbb{R}^d$. Note that these vectors are instance-dependent, thus best adapted to their task. One intuitive example of such a basis is the standard one, $\hat{B}(x) = \{e_1, e_2 \ldots e_d\}$, where $e_i \in \mathbb{R}^d$ is the one-hot vector with value 1 in the ith entry. In our work, we use a set of principal components of $\hat{\mathbb{P}}_{y|x}$, which will be discussed in detail in Section IV.

Similar to previous work [1], [2], we use an interval-based method centered around the conditional mean image, i.e., an estimate of $\mathbb{E}[y|x] \in \mathbb{R}^d$, denoted by $\hat{\mu}(x)$. Formally, we utilize the following interval-valued function that constructs prediction intervals along each basis vector around the estimated conditional mean:

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

In the above, $i \in \{1, 2 \ldots d\}$ is a basis vector index, and $\hat{l}(x)_i \in \mathbb{R}^+$ and $\hat{u}(x)_i \in \mathbb{R}^+$ are the lower and upper interval boundaries for the projected values of candidate solutions emerging from $\hat{\mathbb{P}}_{y|x}$. That is, if $\hat{y} \sim \hat{\mathbb{P}}_{y|x}$ is such a 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. Returning to the example of the standard basis, the above equation is nothing but pixel-wise prediction intervals, which is precisely the approach taken in [1], [2]. By leveraging this generalization, the uncertainty intervals using these basis vectors form a d-dimensional hyper-rectangle, referred to as the *uncertainty region*.

Importantly, we propose that the interval-valued function, $\mathcal{T}$, should produce valid intervals that contain a user-specified fraction of the projected ground-truth values within a risk level of $\alpha \in (0, 1)$. In other words, more than $1 - \alpha$ of the projected ground-truth values should be contained within the intervals, similar to the approach taken in previous work in the pixel domain. To achieve this, we propose a holistic expression that

aggregates the effect of all the intervals, $\mathcal{T}(x; \hat{B}(x))$. This expression leads to the following condition that should be satisfied:

$$\mathbb{E}\left[\sum_{i=1}^{d} \hat{w}_i(x) \cdot \mathbf{1}\left\{ \hat{v}_i(x)^T y \in \mathcal{T}(x; B(x))_i \right\} \right] > 1 - \alpha, \quad (2)$$

where $y \in \mathbb{R}^d$ is the unknown ground-truth and $\hat{w}_i(x) \in [0, 1]$ s.t. $\sum_{i=1}^{d} \hat{w}_i(x) = 1$ are the weight factors that set the importance of covering the projected ground-truth values along each interval. In Section IV we discuss the proposed holistic expression and a specific choice of these weights. As an example, we could set $\alpha = 0.1$ and $\hat{w}_i(x) := 1/d$, indicating that more than 90% of the projected ground-truth values onto the basis vectors are contained in the intervals, as illustrated in a 2d example in Figure 2 for different kinds of $\hat{\mathbb{P}}_{y|x}$.

As discussed above and demonstrated in Figure 2, 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. While such a basis could have been defined analytically using, for example, orthonormal wavelets [35], we suggest a learned and thus a better-tuned one. The choice to use a linear and orthonormal representation for the uncertainty quantification 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 hereafter as the *exact reconstruction property*.

To evaluate the uncertainty across different uncertainty regions, we introduce a new metric called the *uncertainty volume*, $\mathcal{V}(x; \mathcal{T}(x; \hat{B}(x)))$, which represents the dth root of the uncertainty volume with respect to intervals $\mathcal{T}(x; \hat{B}(x))$, defined in the following equation:

$$\mathcal{V}(x; \mathcal{T}(x; \hat{B}(x))) := \sqrt[d]{\prod_{i=1}^{d} \left[\hat{u}(x)_i + \hat{l}(x)_i \right]} \quad (3)$$

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

where $\epsilon > 0$ is a small hyperparameter used for numerical stability. In Section V we demonstrate that our approach results in a significantly reduction in these uncertainty volumes when compared to previous methods.

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
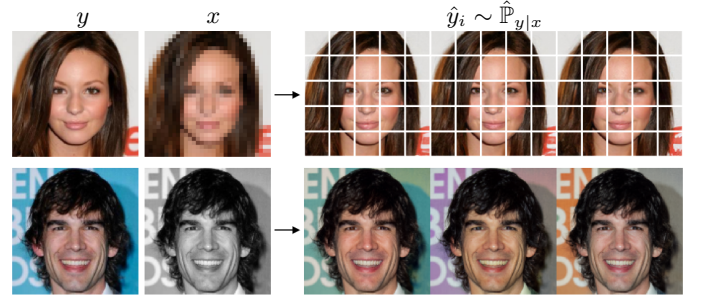

Fig. 3. The sampling procedure for two image restoration problems using a conditional stochastic generator. The top row corresponds to super-resolution in local mode with patches, while the bottom row shows colorization in global mode. The implementation details are described in Section IV-A.

challenges, we propose an option of using $K \ll d$ basis vectors that capture the essence of the uncertainty. In Section IV, we 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. Therefore, we propose an extension to the conventional coverage validity of Equation (2) 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, $\beta^3 \in (0, 1)$. 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:

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

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 $\frac{1}{d} \sum_{i=1}^{d} \mathbf{1}\{z_i \leq \hat{\mathcal{Q}}_q(z)\} \geq q$. In Section IV, we discuss this expression for assessing the validity of the basis vectors. 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.

IV. PUQ: PRINCIPAL UNCERTAINTY QUANTIFICATION

In this section, we present *Principal Uncertainty Quantification* (PUQ), our method for quantifying the uncertainty in inverse problems while taking into account spatial dependencies between pixels. PUQ uses the *principal components* (PCs) of the solutions to the inverse problem for achieving its goal. In Appendix A, we provide an intuition behind the choice of the PCs as the basis. Our approach can be used either globally across the entire image, referred to as the *global mode*; or locally within predefined patches or segments of interest, referred to as the *local mode*. Local uncertainty quantification can be applied to any task, where the dimensionality of the target space is fully controlled by the user. In contrast,

---

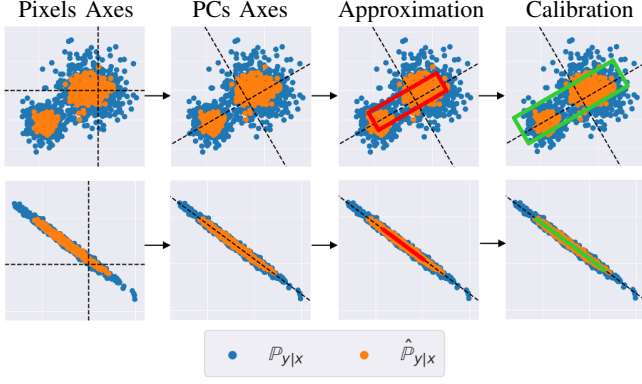[3]This should not be confused with the acceptable coverage level parameter in Equation (2).

Fig. 4. Illustration of our PUQ procedure in 2D ($d = 2$) for a single instance $x \in \mathcal{X}$. 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 and RDA-PUQ). The procedure begins by drawing samples $\hat{y}_i \sim \hat{\mathbb{P}}_{y|x}$. Next, these samples are projected onto the PCs domain: $\hat{V}^T \hat{y}_i$, where $\hat{V} := [\hat{v}_1, \ldots, \hat{v}_K] \in \mathbb{R}^{d \times K}$. Then, we compute bounds along the PCs to contain the samples at the correct ratio, forming the intervals specified in Equation (1). Finally, the intervals are scaled to statistically guarantee Equation (2) and contain the correct ratio over solutions for unseen input instances. In the bottom row, the procedure also statistically guarantees Equation (4) by ensuring a small recovery error of solutions to unseen input instances, as demonstrated by the small variance around the single PC.

global quantification is particularly advantageous for tasks that exhibit strong spatial correlations between pixels.

Our proposed method consists of two phases. In the first, referred to as the *approximation phase*, a machine learning system is trained to predict the PCs of possible solutions, denoted by $\hat{B}(x) = \{\hat{v}_1(x), \hat{v}_2(x), \ldots, \hat{v}_K(x)\}$ (where $K \le d$), as well as a set of importance weights, $\hat{w}(x) \in \mathbb{R}^K$, referring to the vectors in $\hat{B}(x)$. In addition, the system estimates the necessary terms in Equation (1), which include the conditional mean, $\hat{\mu}(x) \in \mathbb{R}^d$, and the lower and upper bounds, $\tilde{l}(x) \in \mathbb{R}^K$ and $\tilde{u}(x) \in \mathbb{R}^{K\ 4}$, for the spread of projected solutions over $\hat{B}(x)$. These estimations are then used to construct heuristic uncertainty intervals in the K-dimensional space. In Section IV-A, we discuss a method for deriving these terms using recent advances in stochastic regression solvers for inverse problems, specifically using diffusion models. Our approach is based on a sampling technique, as illustrated in Figure 3.

The above-described approximation phase is merely an estimation, as the corresponding heuristic intervals of Equation (1) may not contain the projected ground-truth values with a desired ratio. Additionally, the basis vectors may not be able to recover the ground-truth pixel values within an acceptable threshold when $K < d$, or the basis set may contain insignificant axes in terms of variability. Therefore, in the second, *calibration phase*, we offer two calibration procedures on an held-out set of calibration data, denoted by $\mathcal{S}_{\mathrm{cal}} := \{(x_i, y_i)\}_{i=1}^n$. These assess the validity of our proposed uncertainty region over unseen data, which is composed by the intervals defined in Equation (1). The choice between the two

calibration procedures depends on the user, taking into account the trade-off between precision and complexity. The steps of our proposed method are summarized in Algorithm 1, and the two calibration strategies are as follows:

(1) Exact PUQ (E-PUQ - Section IV-B1): In the setting of an exact uncertainty assessment, while assuming that d PCs can be constructed and maintained in full, the exact reconstruction property is satisfied. Consequently, the 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 over the pixel domain.

(2) Dimension-Adaptive PUQ (DA-PUQ - Section IV-B2, RDA-PUQ - Appendix D): In the setting of an approximate uncertainty assessment, while allowing for a small recovery error of projected ground-truth instances to full-dimensional instances, either due to complexity or interpretability reasons (see Section III), the exact reconstruction property is no longer satisfied. Hence, in addition to the scaling procedure outlined above, we must verify that the K PCs can decompose the ground-truth pixel values with a small error. In this calibration process, we also control the minimum number of the first $\hat{k}(x)$ PCs out of the K PCs, such that a small reconstruction error can be guaranteed for unseen data. This number is dynamically determined per input image, so that instances with greater pixel correlations are assigned more PCs than those with weaker correlations. As manually determining K might be challenging, we introduce the Reduced Dimension-Adaptive PUQ (RDA-PUQ) procedure that also controls that value as part of the calibration - see Appendix D.

Algorithm 1 Generating PUQ Axes and Intervals

Input: Training set. Calibration set. Number of PCs $K \in \mathbb{N}$. An unseen input instance $x \in \mathbb{R}^d$.
Output: Statistically valid uncertainty axes and intervals for x.

> ▷ *Approximation phase*
1: Train a machine learning system (e.g., Section IV-A) to estimate the following:
    K PCs of $\hat{\mathbb{P}}_{y|x}$
    Importance weights of PCs
    The conditional mean
    Lower and upper bounds on the PCs

> ▷ *Calibration phase*
2: **if** Exact uncertainty (accurate) **then**
3:     Apply E-PUQ using the calibration data
4: **else if** Approximate uncertainty (reduced complexity) **then**
5:     Apply DA-PUQ / RDA-PUQ using the calibration data
6: **end if**

> ▷ *Inference*
7: Provide statistically valid uncertainty axes and intervals in terms of Equation (2) and Equation (4), applied to an unseen input instance x

---

In Section V, we demonstrate a significant decrease in the uncertainty volume, as defined in Equation (3) for each procedure, whether applied globally or locally, compared to prior work. The E-PUQ procedure is the simplest and can be applied locally to any task, while the DA-PUQ and RDA-PUQ procedures are more complex and particularly effective

---

[4]Note that these bounds are meant for $\hat{\mathbb{P}}_{y|x}$ and not for $\mathbb{P}_{y|x}$, and thus marked with tilde. The $\hat{l}(x), \hat{u}(x)$ bounds that are related to Equation (1) are defined later in the calibration schemes (Section IV-B).

in tasks in which pixels exhibit strong correlations, such as in the image colorization task. Our method is visually illustrated in Figure 4, showing a sampling methodology and a calibration scheme using the full PCs or only a subset of them.

A. Diffusion Models for the Approximation Phase

The approximation phase, summarized in Algorithm 1 in RED, can be achieved in various ways. In this section, we describe the implementation we used to obtain the results detailed in Section V. While we aim to construct the uncertainty axes and intervals in the most straightforward way, further exploration of more advanced methods to achieve the PCs is left for future work.

In our implementation, we leverage the recent advances in stochastic regression solvers for inverse problems based on diffusion models, which enable to train a machine learning model to 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 in global mode, where $\mathcal{Z}$ is the noise seed space. Similarly, in local mode, we consider $f_\theta : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}_{\text{patch}}$. Given an input instance $x \in \mathbb{R}^d$, 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 samples are used to estimate the PCs of possible solutions and their importance weights using the SVD decomposition of the generated samples. The importance weights assign high values to axes with large variance among projected samples, and low ones to those with small variance. In Section IV-B, we elaborate on how these weights are used in the calibration phase. 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 miss-coverage 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, which is computationally challenging for high-dimensional data. As a way out, we suggest working locally on patches, where d is small and fully controlled by the user by specifying the patch size to work on. However, for tasks with strong pixel correlation, such as image colorization, a few PCs can describe the variability of $\hat{\mathbb{P}}_{y|x}$ with a very small error. Therefore, only a few samples (i.e., $K \ll d$) are required for the SVD procedure to construct meaningful PCs for the entire image, while capturing most of the richness in $\hat{\mathbb{P}}_{y|x}$. We formally summarize our sampling-based methodology, in either global or local modes, in Algorithm 2.

B. Calibration Phase

In order to refine the approximation phase and obtain valid uncertainty axes and intervals that satisfy the guarantees of Equation (2) and Equation (4), it is necessary to apply a calibration phase, as summarized in Algorithm 1 in BLUE. This phase includes 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.

Algorithm 2 Approximation Phase via Sampling

Input: 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 $\hat{\mathbb{P}}_{y|x}$*
1: **for** $i = 1$ to K **do**
2:     $\hat{y}_i(x) \leftarrow f_\theta(x, z_i)$
3: **end for**
   ▷ *Compute conditional mean*
4: $\hat{\mu}(x) \leftarrow \frac{1}{K} \sum_{i=1}^K \hat{y}_i(x)$
   ▷ *Apply SVD decomposition and extract the PCs and weights*
5: $\hat{Y}(x) \leftarrow [\hat{y}_1(x), \hat{y}_2(x) \dots \hat{y}_K(x)] \in \mathbb{R}^{d \times K}$
6: $\hat{Y}(x) - \hat{\mu}(x) \cdot \mathbf{1}_K^T = \hat{V}(x)\hat{\Sigma}(x)\hat{U}(x)^T$
7: $\hat{B}(x) \leftarrow \{\hat{v}_1(x), \hat{v}_2(x) \dots \hat{v}_K(x)\}$, where $\hat{v}_i(x) = [\hat{V}(x)]_i$
8: $\hat{w}(x) \leftarrow [\hat{\sigma}_1(x)^2, \dots, \hat{\sigma}_K(x)^2]/c \in \mathbb{R}^K$, 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*
9: **for** $i = 1$ to K **do**
10:     $\tilde{l}(x)_i \leftarrow \hat{\mathcal{Q}}_{\alpha/2}(\{\hat{v}_i(x)^T(\hat{y}_j(x) - \hat{\mu}(x))\}_{j=1}^K)$
11:     $\tilde{u}(x)_i \leftarrow \hat{\mathcal{Q}}_{1-\alpha/2}(\{\hat{v}_i(x)^T(\hat{y}_j(x) - \hat{\mu}(x))\}_{j=1}^K)$
12: **end for**
Output: K PCs $\hat{B}(x)$, importance weights $\hat{w}(x)$, conditional mean $\hat{\mu}(x)$, lower and upper bounds $\tilde{l}(x)$ and $\tilde{u}(x)$.

---

1) 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 in Section III 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 (2) is satisfied with high probability, similar to previous work [1], [2].

Formally, for each input instance x and its corresponding ground-truth value $y \in \mathbb{R}^d$ in the calibration data, 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:

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

This loss is closely related to the expression in Equation (2), and while it may seem arbitrary at first, this choice is a direct extension to the one practiced in [1], [2]. In Appendix B we provide an additional justification for it, more tuned to the realm discussed in this paper.

Our goal is to ensure that the expectation of $\mathcal{L}(x, y; \lambda)$ is below a pre-specified threshold, α, 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 [15] procedure, which guarantees the following:

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

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 (3), 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 describe in detail the E-PUQ procedure in Algorithm 3.

Algorithm 3 Exact PUQ Procedure

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

1: **for** $(x, y) \in \mathcal{S}_{\text{cal}}$ **do**
2:     $\hat{B}(x), \hat{w}(x), \hat{\mu}(x), \tilde{u}(x), \tilde{l}(x) \leftarrow$ Apply Algorithm 2 to x, with the choice of $K = d$ samples
3:     **for** $\lambda \in \Lambda$ **do**
       ▷ *Scale uncertainty intervals*
4:        $\hat{u}(x) \leftarrow \lambda \tilde{u}(x)$ and $\hat{l}(x) \leftarrow \lambda \tilde{l}(x)$
5:        $\mathcal{T}_\lambda(x; \hat{B}(x)) \leftarrow$ Equation (1) using $\hat{\mu}(x), \hat{u}(x), \hat{l}(x)$
       ▷ *Compute weighted coverage loss, Equation (5)*
6:        $\mathcal{L}(x, y; \lambda) \leftarrow$
7:        $\sum_{i=1}^d \hat{w}_i(x) \cdot \mathbf{1} \left\{ \hat{v}_i(x)^T y \notin \mathcal{T}_\lambda(x; \hat{B}(x))_i \right\}$
8:     **end for**
9: **end for**
10: $\hat{\Lambda} \leftarrow$ Extract valid λs from LTT [15] 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 (3)*
11: $\hat{\lambda} \leftarrow \arg\min_{\lambda \in \hat{\Lambda}} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{V}_\lambda(x_i; \hat{B}(x)) \right\}$
Output: Given a new instance $x \in \mathcal{X}$, obtain valid uncertainty intervals for it, $\mathcal{T}_{\hat{\lambda}}(x; \hat{B}(x))$.

---

2) 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 (see discussion in Section III). 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 (see Section III), 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 (2) 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 (2) 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 the calibration data, 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:

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

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:

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

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 $\hat{\mu}(x)$. In Appendix C, we discuss further this specific loss function for controlling the capability of the linear subspace to capture the richness of the complete d-dimensional posterior distribution.

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:

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

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 (3), 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 (2) and Equation (4) hold with high probability over the calibration data. This can be provided, for example, through the LTT [15] calibration scheme, which guarantees the following:

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

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 4.

The above-described DA-PUQ procedure reduces the number of PCs to be constructed to $K \leq d$ while using $\hat{k}(x; \hat{\lambda}_1) \leq K$ PCs, leading to increased efficiency in both time and space during inference. However, determining manually the smallest K value that can guarantee both Equation (2) and Equation (4) can be challenging. To address this, we propose an expansion of the DA-PUQ procedure; the *Reduced Dimension-Adaptive PUQ* (RDA-PUQ) procedure that also controls the maximum number of PCs, K, required for the uncertainty assessment. This approach is advantageous for inference as it reduces the number of samples required to construct the PCs using Algorithm 2, while ensuring both the coverage and reconstruction guarantees of Equation (2) and Equation (4) with high probability. The RDA-PUQ procedure is fully described in Appendix D.

Algorithm 4 Dimension-Adaptive PUQ Procedure

Input: Calibration set $\mathcal{S}_{\text{cal}} := \{x_i, y_i\}_{i=1}^n$. Scanned calibration parameter values $\Lambda^1 \leftarrow [1 \ldots \lambda_{1\max}]$ and $\Lambda^2 \leftarrow [1 \ldots \lambda_{2\max}]$. 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.

1: **for** $(x, y) \in \mathcal{S}_{\text{cal}}$ **do**
2:      $\hat{B}(x), \hat{w}(x), \hat{\mu}(x), \tilde{u}(x), \tilde{l}(x) \leftarrow$ Apply Algorithm 2 to x, with the choice of K samples
3:      **for** $\lambda_1 \in \Lambda_1$ **do**
       ▷ *Compute adaptive dimensionality, Equation* (7)
4:          $\hat{k}(x; \lambda_1) \leftarrow \min_k \left\{ k : \sum_{i=1}^K \hat{w}_i(x) \geq \lambda_1 \right\}$
       ▷ *Compute reconstruction loss, Equation* (8)
5:          $y_c \leftarrow y - \hat{\mu}(x)$
6:          $\mathcal{L}_1(x, y; \lambda_1) \leftarrow$
         $\hat{\mathcal{Q}}_q \left(\left\{ \left| \sum_{j=1}^{\hat{k}(x;\lambda_1)} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c \right|_i \right\}_{i=1}^d \right)$
7:        **for** $\lambda_2 \in \Lambda_2$ **do**
         ▷ *Scale uncertainty intervals*
8:            $\hat{u}(x_i) \leftarrow \lambda_2 \tilde{u}(x)$ and $\hat{l}(x) \leftarrow \lambda_2 \tilde{l}(x)$
9:            $\mathcal{T}_{\lambda_2}(x; \hat{B}(x)) \leftarrow$ Eq. (1) using $\hat{\mu}(x), \hat{u}(x), \hat{l}(x)$
         ▷ *Compute weighted coverage loss, Equation* (5)
10:           $\mathcal{L}_2(x, y; \lambda_1, \lambda_2) \leftarrow \sum_{i=1}^{\hat{k}(x;\lambda_1)} \hat{w}_i(x) \cdot$
            $\mathbf{1}\left\{ \hat{v}_i(x)^T y \notin \mathcal{T}_{\lambda_2}(x; \hat{B}(x))_i \right\}$
11:        **end for**
12:      **end for**
13: **end for**
14: $\hat{\Lambda} \leftarrow$ Extract valid λs from LTT [15] applied on $\left\{ (\mathcal{L}_1(x, y; \lambda_1), \mathcal{L}_2(x, y; \lambda_1, \lambda_2)) : (x, y) \in S_{\text{cal}}, \lambda_1 \in \Lambda^1, \lambda_2 \in \Lambda^2 \right\}$ at risk levels (β, α) and error level δ, referring to Equation (10)
    ▷ *Compute the minimizers for the uncer. volume, Equation* (3)
15: $\hat{\lambda}_1, \hat{\lambda}_2 \leftarrow \arg\min_{\lambda_1, \lambda_2 \in \hat{\Lambda}} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{V}_{\lambda_1, \lambda_2}(x_i; \hat{B}(x_i)) \right\}$

Output: 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.

V. EMPIRICAL STUDY

This section presents a comprehensive empirical study of our proposed method PUQ, applied to three challenging tasks: image colorization, super-resolution, and inpainting, over the CelebA-HQ dataset [36]. Our approximation phase starts with a sampling from the posterior, applied in our work by the SR3 conditional diffusion model [10]. Figure 5 presents typical sampling results for these three tasks, showing the expected diversity in the images obtained.

The experiments we present herein 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]. Through the experiments, we present superiority in uncertainty volume, as defined in Equation (3), and in interpretability through the use of only a few PCs to assess the uncertainty of either a patch or a complete image. All the experiments were conducted over 100 calibration-test splits . For in-depth additional details of our experiments, we refer the reaxxxder to Appendix E.

A. Evaluation Metrics

Before presenting the results, we pause to discuss the metrics used to evaluate the performance of the different methods. Although our approach is proved to guarantee Equation (6) for E-PUQ and Equation (10) for DA-PUQ, (through LTT [15] procedure), we assess the validity and tightness of these guarantees as well.

Empirical coverage risk. We measure the risk associated with the inclusion of projected unseen ground-truth values in the uncertainty intervals. In E-PUQ, we report the average coverage loss, defined in Equation (5). In the case of DA-PUQ and RDA-PUQ, we report the value defined by Equation (9).

Empirical reconstruction risk. We measure the risk in recovering unseen ground-truth pixel values using the selected PCs. In the case of E-PUQ, this risk is zero by definition. However, for DA-PUQ and RDA-PUQ, we report the average reconstruction loss, defined by Equation (8).

Interval-Size. We report the calibrated uncertainty intervals' sizes of Equation (1), and compare them with baseline methods. For E-PUQ, we compare intervals over the full basis set of PCs with the intervals in the pixel domain used in previous work. In the DA-PUQ and RDA-PUQ procedures, we apply dimensionality reduction to $K \ll d$ dimensions. To validly compare the intervals' sizes of these methods to those methods over the full d dimensions, we pad the remaining $d - K$ dimensions with zeros as we assume that the error in reconstructing the ground-truth from the dimensionally reduced samples is negligible.

Uncertainty Volume. We report the uncertainty volumes, defined in Equation (3), for the calibrated uncertainty regions and compare them with previous work. A smaller uncertainty volume implies a higher level of certainty in probable solutions to $\mathbb{P}_{y|x}$. In E-PUQ, we compare volumes over the full basis set of PCs to the pixel domain used in earlier studies. In the DA-PUQ and RDA-PUQ procedures, we pad the remaining dimensions with zeros.
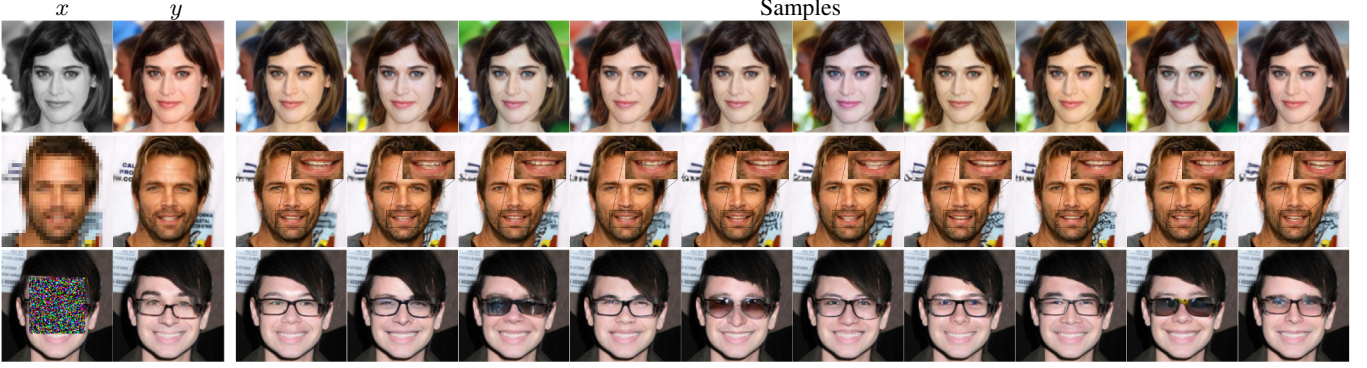
Fig. 5. The three image recovery tasks, colorization (top), super-resolution (middle) and inpainting (bottom). For each we present a given measurement x, the ground-truth y, and 10 candidate samples from the (approximated) posterior distribution. These samples fuel the approximation phase in our work.

TABLE I
LOCAL EXPERIMENTS: QUANTITATIVE COMPARISON OF THE MEANS AND STANDARD DEVIATIONS OF OUR LOCALLY APPLIED PUQ METHOD ON RGB PATCH RESOLUTION OF 8x8, UTILIZING THE TWO PROPOSED PROCEDURES. NOTE THAT IN THIS EXPERIMENT $d = 8 \times 8 \times 3 = 192$ AND $\epsilon = 1e - 10$ FOR THE VOLUME COMPUTATION.

		Recons. Risk	Dim. $\hat{k}(x)$ / K	Uncert. Volume
Colorization	im2im-uq [1]	0	192 / 192	$1.6e-1 \pm 7.2e-2$
	Conffusion [2]	0	192 / 192	$1.7e-1 \pm 1.3e-1$
	E-PUQ	0	192 / 192	$2.3e-3 \pm 9.8e-4$
	DA-PUQ	$2.5e-2 \pm 5.3e-4$	1.6 ± 0.77 / 100	$2.4e-11 \pm 1.3e-11$
	RDA-PUQ	$1.7e-2 \pm 9.7e-4$	3.8 ± 1.8 / 11.8 ± 6.4	$6.8e-11 \pm 3.9e-11$
Super-Resolution	im2im-uq [1]	0	192 / 192	$8.8e-2 \pm 5.7e-2$
	Conffusion [2]	0	192 / 192	$8.8e-2 \pm 6.9e-2$
	E-PUQ	0	192 / 192	$1.3e-2 \pm 7.5e-3$
	DA-PUQ	$2.5e-2 \pm 3.4e-4$	11.1 ± 5.5 / 192	$3.4e-10 \pm 3.5e-10$
	RDA-PUQ	$2.0e-2 \pm 0.0e-0$	22.8 ± 6.8 / 70.0 ± 0.0	$1.6e-9 \pm 1.3e-9$
Inpainting	im2im-uq [1]	0	192 / 192	$2.8e-1 \pm 1.4e-1$
	Conffusion [2]	0	192 / 192	$2.7e-1 \pm 1.6e-1$
	E-PUQ	0	192 / 192	$1.9e-2 \pm 1.0e-2$
	DA-PUQ	$1.8e-2 \pm 1.0e-4$	39.4 ± 11.8 / 192	$3.6e-8 \pm 7.3e-8$
	RDA-PUQ	$1.8e-2 \pm 1.6e-3$	55.6 ± 10.2 / 72.6 ± 30.3	$1.3e-7 \pm 8.8e-8$

Fig. 6. **Local Experiments:** A comparison of E-PUQ (see Section IV-B1) with previous work – im2im-uq [1] and Conffusion [2]. These methods are applied locally on patches with $\alpha = \delta = 0.1$. Each column corresponds to a relevant metric (see Section V-A), and each row corresponds to a specific task. The uncertainty volume was computed with $\epsilon = 1e - 10$. Results indicate that our approach achieves superior uncertainty volume.

B. Local Experiments on Patches

We apply our proposed methods on RGB patches of increasing size — 1x1, 2x2, 4x4, and 8x8 — for image colorization, super-resolution, and inpainting tasks. The obtained results are illustrated in Figure 6 and Figure 7, where Figure 6 compares our exact procedure, E-PUQ, to baseline methods, and Figure 7 examines our approximation procedures, DA-PUQ and RDA-PUQ. In Table I we present a numerical comparison of uncertainty volumes across tasks at 8x8 patch resolution. We also provide visual representations of the uncertainty volume maps for patches at varying resolutions in Figure 8.

The results shown in Figure 6 and Figure 7 demonstrate that our method provides smaller uncertainty volumes, and thus more confined uncertainty regions, when compared to previous work in all tasks and patch resolutions, and while satisfying the same statistical guarantees in all cases. More specifically, Figure 6 compares our exact procedure, E-PUQ, to baseline methods. Following this figure, one can see that using the E-PUQ procedure we obtained an improvement of $\sim \times 100$ in the uncertainty volumes in colorization and an improvement of $\sim \times 10$ in super-resolution and inpainting, when applied to the highest resolution of 8x8. Additionally, 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.
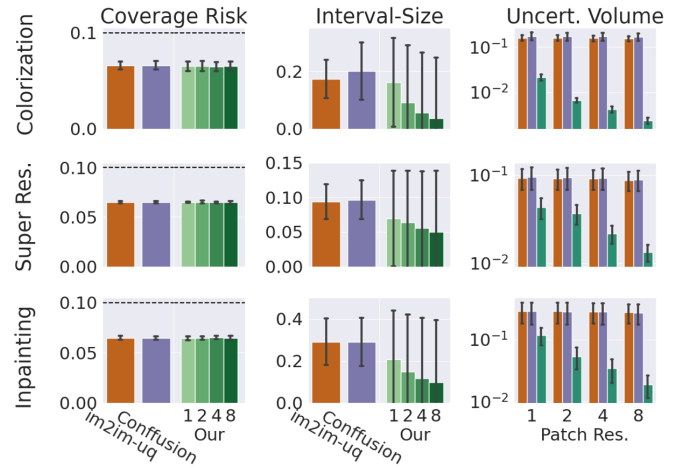
Note that even a patch size of 1×1 brings a benefit in the evaluated volume, due to the exploited correlation within the three color channels. E-PUQ reduces trivially to im2im-uq [1] and Conffusion [2] when applied to scalars ($1 \times 1 \times 1$ patches).

In Figure 7, we examine our approximation methods, DA-PUQ and RDA-PUQ, in which we set a relatively small reconstruction risk of $\beta = 0.05$. Observe the significantly smaller uncertainty volumes obtained; this effect is summarized in Table I as well. Figure 7 also portrays the dimensionality of the uncertainty region used with our method using two overlapping bars. The outer bar in yellow refers to the number of PCs that need to be constructed, denoted as K in DA-PUQ and $\hat{K}$ in RDA-PUQ. The smaller this number is, the lower the test time computational complexity. The inner bar in green refers to the average number of the adaptively selected PCs, denoted as $\hat{k}(x)$. A lower value of $\hat{k}(x)$ indicates better interpretability, as fewer PCs are used at inference than those that were constructed. For example, in the colorization task, it can be seen that the RDA-PUQ procedure is the
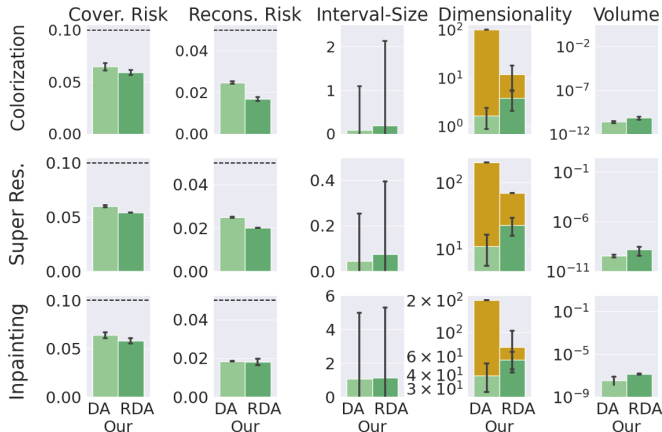
Fig. 7. **Local Experiments:** A comparison of DA-PUQ (see Section IV-B2) and RDA-PUQ (see Appendix D), when applied locally on 8x8 patches with $\alpha = \delta = 0.1$, $\beta = 0.05$ and $q = 0.9$. Each column corresponds to a relevant metric (see Section V-A), and each row corresponds to a specific task. The uncertainty volume was computed with $\epsilon = 1e - 10$. Here, the dimensionality is presented by two overlapping bars, where the yellow bars represent the distribution of K in DA-PUQ and $\hat{K}$ in RDA-PUQ, and the inner bars represent the distribution of $\hat{k}(x)$ in both cases.

TABLE II
GLOBAL EXPERIMENTS: QUANTITATIVE COMPARISON OF THE MEANS AND STANDARD DEVIATIONS OF OUR GLOBALLY APPLIED PUQ METHOD IN THE COLORIZATION TASK, UTILIZING THE PROPOSED DA-PUQ (SEE SECTION IV-B2) AND RDA-PUQ (SEE APPENDIX D) PROCEDURES.

	Recons. Risk	Dim. $\hat{k}(x)$ / K	Uncert. Volume
im2im-uq [1]	0	49152 / 49152	$1.4e{-}1 \pm 3.2e{-}2$
Conffusion [2]	0	49152 / 49152	$1.4e{-}1 \pm 5.5e{-}2$
DA-PUQ	$5.0e{-}2 \pm 1.1e{-}3$	2.2 ± 0.93 / 100	$1.2e{-}13 \pm 5.0e{-}14$
RDA-PUQ	$4.3e{-}2 \pm 2.8e{-}3$	5.5 ± 4.5 / 22.3 ± 10.9	$3.1e{-}13 \pm 2.4e{-}13$

task. While all PUQ procedures can be applied locally for any task, working globally is more realistic in tasks that exhibit strong pixel correlation. Under this setting, most of the image variability could be represented via DA-PUQ or RDA-PUQ while (i) maintaining a small reconstruction risk, and (ii) using only a few PCs to assess the uncertainty of the entire images. Note that the tasks of super-resolution and inpainting are ill-served by a global mode since their estimated PCs are local in nature, thus requiring huge number of them for an effective uncertainty representation. Also, the E-PUQ procedure does not apply to the global approach, as it requires computing and maintaining $d = 128 \times 128 \times 3$ PCs.

Figure 9 visually demonstrates the performance of our approximation methods, also summarized in Table II. These results demonstrate that our method provides significantly smaller uncertainty volumes compared to our local results in Figure 7 and previous works, but this comes at the cost of introducing a small reconstruction risk of up to $\beta = 0.1$. Observe how our approximation methods improve interpretability: the uncertainty regions consist of only 2-5 PCs in the full dimensional space of the images. The DA-PUQ procedure produces the tightest uncertainty regions; see the uncertainty volumes in Table II. In addition, the mean interval-size with our procedures is very small and almost equal to zero, indicating that the constructed uncertainty regions are tight and narrow due to strong correlation structure of pixels. However, similar to the previous results, the standard deviation of interval-size is spread across a wide range. This is because few of the first PCs have wide intervals. The RDA-PUQ procedure is the most computationally efficient as it required to construct only ~ 30 PCs during inference to ensure statistical validity.

Figure 10 presents selected uncertainty regions that were provided by our proposed RDA-PUQ procedure when applied globally. As can be seen, the projected ground-truth images using only $\hat{k}(x)$ PCs results in images that are very close to the originals. This indicates that the uncertainty region can describe the spread and variability among solutions with small reconstruction errors. The first two axes of our uncertainty regions exhibit semantic content, which is consistent with a method that accounts for spatial pixel correlation. The fact that these PCs capture foreground/background or full-object content highlights a unique strength of our approach. We provide the importance weights of the first two PCs, indicating impressive proportions of variability among projected samples onto these components (see Section IV-A). For example, in the third row, we observe that 77% of the variability in $\hat{\mathbb{P}}_{y|x}$ is captured by $\hat{v}_1(x)$, which mostly controls a linear color range

most computationally efficient methodology, requiring only $\hat{K} \approx 12$ PCs to be constructed at inference, while the DA-PUQ procedure is the most interpretable results, with uncertainty regions consisting of only $\hat{k}(x) \in \{1, 2, 3\}$ axes.

In all experiments demonstrated in Figure 6 and Figure 7, it is noticeable that the standard deviation of the interval-size of our approach is higher than that of the baseline methods. This effect happens because a few intervals along the first few PCs are wider than those along the remaining PCs. However, the majority of the interval sizes are significantly smaller, resulting in a much smaller uncertainty volume. Interestingly, the uncertainty intervals of the DA-PUQ and RDA-PUQ procedures in Figure 7 exhibit larger standard deviation compared to the E-PUQ procedure in Figure 6. We hypothesize that this is caused when only a few intervals (e.g., 2 intervals) are used for the calibration process while small miscoverage ratio is set by the user ($\alpha = 0.1$). As an example in the case of using 2 intervals with all samples of the calibration set, it is necessary to enlarge all the intervals to ensure the coverage guarantee, resulting in wider intervals over the first few PCs.

The heat maps presented in Figure 8 compare the uncertainty volumes of our patch-based E-PUQ procedure to baseline methods. Each pixel in the presented heat maps corresponds to the value of Equation (3) evaluated on its corresponding patch. The results show that as the patch resolution increases, pixels with strong correlation structure, such as pixels of the background area, also exhibit lower uncertainty volume in their corresponding patches. This indicates that the proposed method indeed takes into account spatial correlation, leading to reduced uncertainty volume.

C. Global Experiments on Images

We turn to examine the effectiveness and validity of DA-PUQ and RDA-PUQ when applied to complete images at a resolution of 128×128, and consider only the colorization
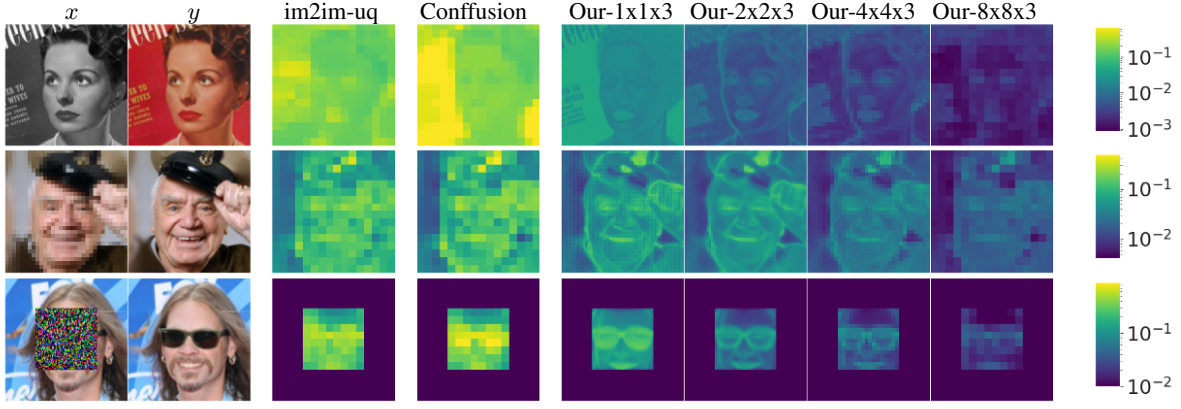
Fig. 8. **Local Experiments:** Uncertainty volume maps for patches applied in image colorization (top), super-resolution (middle), and inpainting (bottom) with E-PUQ, im2im-uq [1] and Conffusion [2]. Each pixel in the maps corresponds to the uncertainty volume, defined in Equation (3), of its corresponding patch. These results expose the effectiveness of our method that incorporates spatial correlations, resulting in a reduction of the uncertainty volume.

Fig. 9. **Global Experiments:** A comparison of DA-PUQ (see Section IV-B2) and RDA-PUQ (see Appendix D), when applied globally on the colorization task with $\alpha = \beta = \delta = 0.1$ and $q = 0.95$. The uncertainty volume was computed with $\epsilon = 1e - 10$.
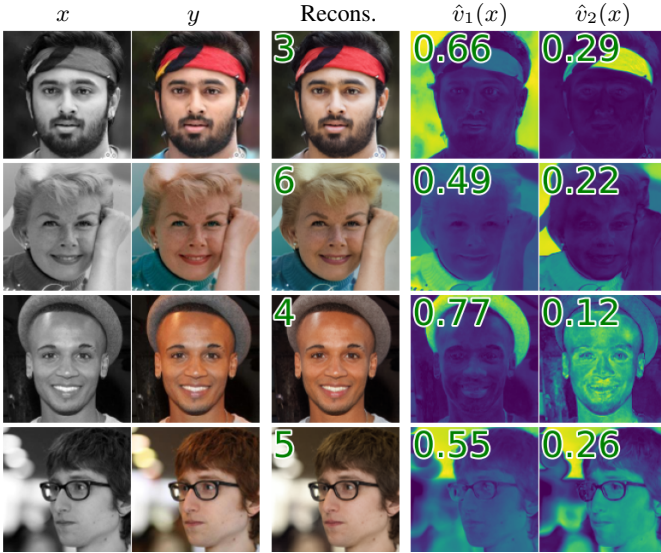

Fig. 10. **Global Experiments:** Visual presentation of uncertainty regions provided by RDA-PUQ (Appendix D) when applied globally for the colorization task. The reconstructed image is given by $\hat{\mu}(x) + \sum_{i=1}^{\hat{k}(x)} \hat{v}_i(x)^T y_c \hat{v}_i(x)$, where $y_c := y - \hat{\mu}(x)$. The values of $\hat{k}(x)$, $\hat{w}_1(x)$ and $\hat{w}_2(x)$ are shown in the top left corners of the corresponding columns.
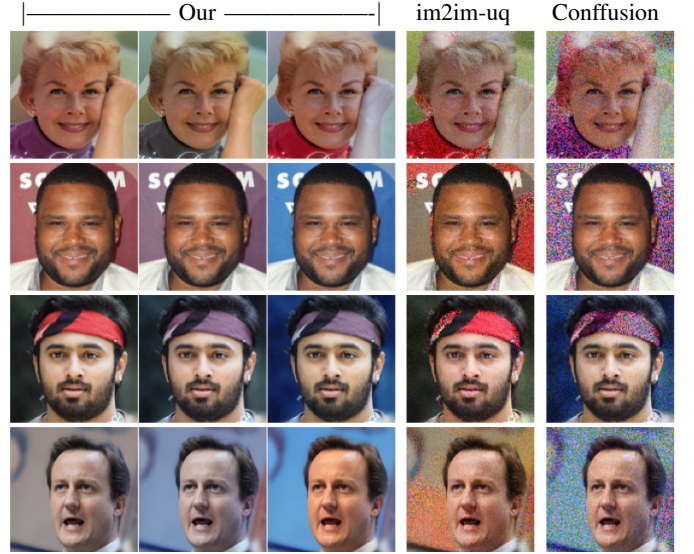

Fig. 11. **Global Experiments:** Images sampled uniformly from the estimated global uncertainty regions, referring to the colorization task. Using RDA-PUQ results with high-perceptual images, while im2im-uq [1] and Conffusion [2] produce unlikely images. These results indicate that our uncertainty regions are significantly more confined than those of previous works.

of the pixels associated with the hat in the image.

In Figure 11 we visually compare samples that were generated from the corresponding estimated uncertainty regions, by sampling uniformly a high dimensional point (i.e., an image) from the corresponding hyper-rectangle. As can be seen, the samples extracted from our uncertainty region are of high perceptual quality, whereas im2im-uq [1] and Conffusion [2] produce highly improbable images. This testifies to the fact that our method provides much tighter uncertainty regions, whereas previous work results in exaggerated uncertainty regions that contain unlikely images.

VI. CONCLUDING REMARKS

This paper presents "Principal Uncertainty Quantification" (PUQ), a novel and effective approach for quantifying uncertainty in any image-to-image task. PUQ takes into account the spatial dependencies between pixels in order to achieve significantly tighter uncertainty regions. We introduced two different procedures for uncertainty assessment, which are selected based on the user's preference for balancing precision and complexity. The experimental results demonstrate that PUQ outperforms existing methods in image colorization,

super-resolution and inpainting, by improving the uncertainty volume. Additionally, by allowing for a small reconstruction error when recovering ground-truth images, PUQ produces tight uncertainty regions with a few axes and thus improves computational complexity and interpretability at inference. As a result, PUQ achieves state-of-the-art performance in uncertainty quantification for image-to-image problems.

Referring to future research, more sophisticated choices that rely on recent advancements in stochastic image regression models could be explored, so as to improve the complexity of our proposed approximation phase. Additionally, investigating alternative geometries for uncertainty regions could be interesting in order to further reduce the gap between the provided region of uncertainty and the high-density areas of the true posterior distribution. This includes an option to divide the spatial domain into meaningful segments, while minimizing the uncertainty volume, or consider a mixture of Gaussians modeling of the samples of the estimated posterior distribution.

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] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[4] 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.

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

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

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

[16] Stéphane Lathuilière, Pablo Mesejo, Xavier Alameda-Pineda, and Radu Horaud. A comprehensive analysis of deep regression. *IEEE transactions on pattern analysis and machine intelligence*, 42(9):2065–2081, 2019.

[17] Venkataraman Santhanam, Vlad I Morariu, and Larry S Davis. Generalized deep image to image regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5609–5619, 2017.

[18] Xinchen Yan, Jimei Yang, Kihyuk Sohn, and Honglak Lee. Attribute2image: Conditional image generation from visual attributes. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 776–791. Springer, 2016.

[19] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. In *International conference on machine learning*, pages 1462–1471. PMLR, 2015.

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

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

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

[23] 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.

[24] 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.

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

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

[27] 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.

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

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

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

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

[32] 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.

[33] 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.

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

[35] Yves Meyer. Orthonormal wavelets. In *Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987*, pages 21–37. Springer, 1990.

[36] 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.

[37] 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.
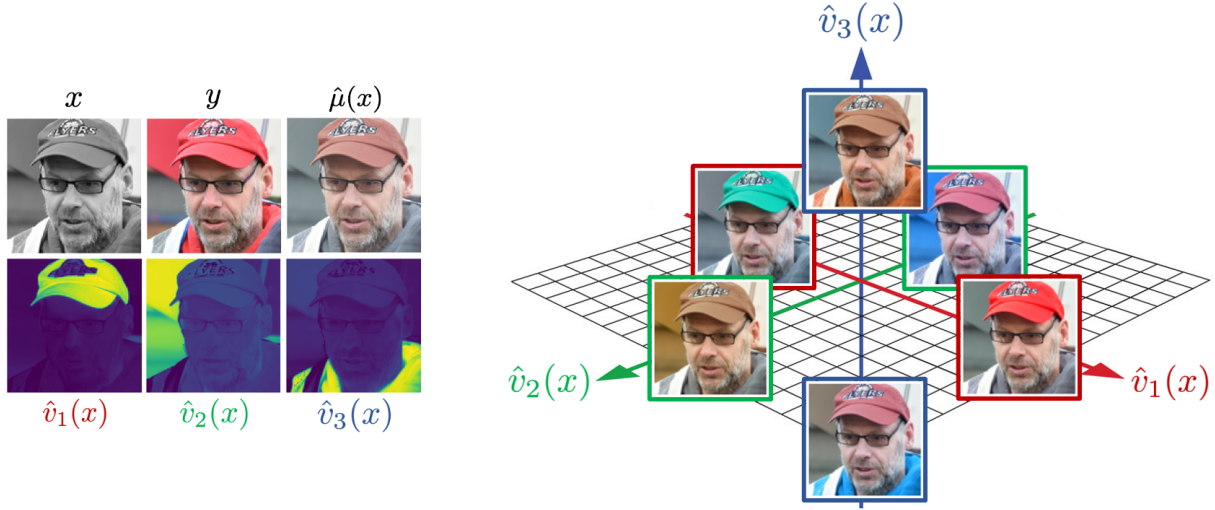
Fig. 12. A visual representation demonstrating the intuition behind utilizing principal components (PCs) as the basis, $\hat{B}(x)$, in Equation (1) for the colorization task. The left part illustrates that the PCs incorporate spatial correlation, with $\hat{v}_1(x)$ primarily controlling the hat color, $\hat{v}_2(x)$ governing the background color, and $\hat{v}_3(x)$ influencing the clothing color. On the right side, an illustration of the uncertainty region is presented, composed of these axes, where the origin is $\hat{\mu}(x)$, and each image is defined by $\hat{\mu}(x) + \hat{v}_i(x)^T y_c + a$, where $y_c := y - \hat{\mu}(x)$, and $a \in \mathbb{R}$ is a controllable parameter that moves along the axis.

APPENDIX

A. Visualizing the Principal Component Vectors

Figure 12 depicts the role of the Principal Component (PCs) vectors in the context of the image colorization task. This figure provides an intuition behind employing these vectors for the uncertainty quantification. We show the estimation of the first three PCs using our globally applied PUQ and visualize the uncertainty region formed by these axes. Our approach facilitates efficient exploration within the uncertainty region, thanks to the linear axes that incorporate spatial correlation, as illustrated by the visualization of $\hat{v}_1(x)$, $\hat{v}_2(x)$, and $\hat{v}_3(x)$.

B. Coverage Loss Justification

This section aims to justify our choice for the loss-function for tuning λ in Equation (5), and the weights used in it, $\hat{w}_i(x)$. Recall, this expression is given as:

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

Our starting point is the given d-dimensional hyper-rectangle obtained from the approximation phase, oriented along the d PC directions. This shape serves as our initially estimated uncertainty region. Given the calibration data, $\mathcal{S}_{\text{cal}} := \{(x_i, y_i)\}_{i=1}^{n}$, our goal is to inflate (or deflate, if this body proves to be exaggerated) this shape uniformly across all axes so that it contains the majority of the ground truth examples.

Focusing on a single pair from this dataset, (x, y), the degraded image x is used to ignite the whole approximation phase, while the ground truth y serves for assessing the obtained hyper-rectangle, by considering the projected coordinates $\{\hat{v}_i(x)^T y_c\}_{i=1}^{d}$, where $y_c := y - \hat{\mu}(x)$. The following function measures a potential deviation in the i-th axis,

$$h_i(x, y) := \max\{\hat{v}_i(x)^T y_c - \hat{u}(x)_i, 0\} \tag{11}$$
$$+ \max\{-\hat{v}_i(x)^T y_c + \hat{l}(x)_i, 0\}.$$

Written differently, this expression is also given by

$$h_i(x, y) := \begin{cases} \hat{v}_i(x)^T y_c - \hat{u}(x)_i & \text{if } \hat{v}_i(x)^T y_c > \hat{u}(x)_i > 0 \\ \hat{l}(x)_i - \hat{v}_i(x)^T y_c & \text{if } \hat{v}_i(x)^T y_c < -\hat{l}(x)_i < 0 \\ 0 & \text{otherwise.} \end{cases}$$

If positive, this implies that in this axis the example spills outside the range of the rectangle, and the value itself is the distance from it's border.

The following expression quantifies the weighted amount of energy that should be invested in projecting back the $\{\hat{v}_i(x)^T y\}_{i=1}^{d}$ coordinates to the closest border point:

$$\text{Energy}(x, y) = \sum_{i=1}^{d} \hat{\sigma}_i(x)^2 h_i(x, y)^2. \tag{12}$$

Note that in our weighting we prioritize high-variance axes, in which deviation from the boundaries is of greater impact. Naturally, we should tune λ, which scales $\hat{u}(x)_i$ and $\hat{l}(x)_i$, so as to reduce this energy below a pre-chosen threshold, thus guaranteeing that the majority of ground truth images fall within the hyper-rectangle. While this expression is workable, it suffers from two shortcomings: (1) It is somewhat involved to compute; and (2) The threshold to use with it is hard to interpret and thus to choose. Therefore, similar to previous approaches [1], [2], we opted in this work for a binary version of Equation (11) of the form

$$b_i(x, y) := \begin{cases} 1 & \text{if } h_i(x, y) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In addition, we divide the energy expression, defined in Equation (12), by the sum of squares of all the singular values, and this way obtain exactly $\mathcal{L}(x, y; \lambda)$ as in Equation (5). Observe that, by definition, we get that $0 \leq \mathcal{L}(x, y; \lambda) \leq 1$, where the bottom bound corresponds to a point fully within the rectangle, and the upper bound for the case where the point is fully outside in all axes. Therefore, thresholding the expectation of this value with $\alpha \ll 1$ is intuitive and meaningful.

C. Reconstruction Loss Justification

This section aims to discuss our choice for the loss function for tuning λ_1 in Equation (8), given by

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

Recall the process: We begin with $K \leq d$ PCs obtained from the approximation phase, and then choose $\hat{k}(x;\lambda_1) \leq K$ of them as instance-specific number of PCs for the evaluation of the uncertainty.

Given a calibration pair (x,y), x is used to derive $\hat{k}(x;\lambda_1)$, defining a low-dimensional subspace $\hat{V}(x) := [\hat{v}_1(x),\ldots,\hat{v}_{\hat{k}(x;\lambda_1)}(x)] \in \mathbb{R}^{\hat{k}(x;\lambda_1)\times d}$. This, along with the conditional-mean, $\hat{\mu}(x)$, represent $\mathbb{P}_{y|x}$ as an affine subspace. The ground-truth image y is then projected onto this slab via:

$$\text{Projection}(y) := \hat{\mu}(x) + \hat{V}(x)\hat{V}(x)^T y_c \qquad (14)$$

$$= \hat{\mu}(x) + \sum_{j=1}^{\hat{k}(x;\lambda_1)} \hat{v}_j(x)^T y_c \hat{v}_j(x) ,$$

where $y_c := y - \hat{\mu}(x)$.

The parameter λ_1 should be tuned so as to guarantee that this projection entails a bounded error, $dist(y,\text{Projection}(y))$ in expectation. A natural distance measure to use here is the L_2-norm of the difference, which aligns well with our choice to use SVD in the approximation phase. However, L_2 accumulates the error over the whole support, thus losing local interpretability. An alternative is using L_∞ which quantifies the worst possible pixelwise error induced by the low-dimensional projection,

$$dist(y,\text{Projection}(y)) := \left\|\sum_{j=1}^{\hat{k}(x;\lambda_1)} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c\right\|_\infty .$$

While this measure is applicable in many tasks, there are cases (e.g., inpainting) in which controlling a small maximum error requires the use of a large number of PCs, $\hat{k}(x;\lambda_1)$. To address this, we propose a modification by considering the maximum error over a user-defined ratio of pixels, $q \in (0,1)$, a value close to 1. This is equivalent to determining the q-th empirical quantile, $\hat{\mathcal{Q}}_q$, of the error values among the pixels, providing a more flexible and adaptive approach, which also aligns well with the rationale of uncertainty quantification, in which the statistical guarantees are given with probabilistic restrictions.

D. Reduced Dimension-Adaptive PUQ

The DA-PUQ procedure (see Section IV-B2) reduces the number of PCs to be constructed to $K \leq d$ while using $\hat{k}(x;\hat{\lambda}_1) \leq K$ PCs, leading to increased efficiency in both time and space during inference. However, determining manually the smallest K value that can guarantee both Equation (2) and Equation (4) with high probability can be challenging. To address this, we propose an expansion of the DA-PUQ procedure; the *Reduced Dimension-Adaptive PUQ* (RDA-PUQ) procedure that also controls the maximum number of PCs required for the uncertainty assessment. While this approach is computationally intensive during calibration, it is advantageous for inference as it reduces the number of samples required to construct the PCs using Algorithm 2.

Specifically, for each input instance x and its corresponding ground-truth value y in the calibration data, we use the estimators obtained in the approximation phase, to estimate $\hat{K}_{\lambda_3}$ 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. Note that these estimates are now depend on λ_3, we omit the additional notation for simplicity. Then, for each choice of λ_3, we use these $\hat{K}_{\lambda_3}$-dimensional estimates exactly as in the DA-PUQ procedure to achieve both the coverage and reconstruction guarantees of Equation (2) and Equation (4) with high probability.

Similar to previous approaches, we aim to minimize the uncertainty volume, defined in Equation (3), for the scaled $\hat{K}_{\lambda_3}$-dimensional intervals where any additional axis ($d - \hat{K}_{\lambda_3}$ axes) is fixed to zero. We denote the uncertainty volume in this setting as $\mathcal{V}_{\lambda_1,\lambda_2,\lambda_3}$. The minimization of $\mathcal{V}_{\lambda_1,\lambda_2,\lambda_3}$ is achieved by minimizing λ_1, λ_2 and λ_3, while ensuring that the guarantees of Equation (2) and Equation (4) are satisfied with high probability. This can be provided using a conformal prediction scheme, for example, through the LTT [15] calibration scheme, which ensures that the following holds:

$$\mathbb{P}\left(\begin{array}{c}\mathbb{E}[\mathcal{L}_1(x,y;\hat{\lambda}_1,\hat{\lambda}_3)] \leq \beta \\ \mathbb{E}[\mathcal{L}_2(x,y;\hat{\lambda}_1,\hat{\lambda}_2,\hat{\lambda}_3)] \leq \alpha\end{array}\right) \geq 1 - \delta , \qquad (15)$$

where $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ are the minimizers for the uncertainty volume among valid calibration parameter results, $\hat{\Lambda}$, obtained through the LTT procedure. Note that the loss functions, $\mathcal{L}_1$ and $\mathcal{L}_2$, in the above are exactly those of the DA-PUQ procedure, defined in Equation (8) and Equation (9), while replacing K with $\hat{K}_{\lambda_3}$.

Intuitively, Equation (15) guarantees that a fraction q of the ground-truth pixel values is recovered with an error no greater than β using no more than $\hat{K}_{\hat{\lambda}_3}$ principal components, and a fraction of more than $1-\alpha$ of the projected ground-truth values onto the first $\hat{k}(x;\hat{\lambda}_1)$ principal components (out of $\hat{K}_{\hat{\lambda}_3}$) are contained in the uncertainty intervals, with a probability of at least $1-\delta$. The RDA-PUQ procedure is formally described in Algorithm 5.

E. Experimental Details

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

1) Datasets and Preprocessing: Our machine learning system was trained using the Flickr-Faces-HQ (FFHQ) dataset [37], which includes 70,000 face images at a resolution of 128x128. We conducted calibration and testing on the CelebA-HQ (CelebA) dataset [36], 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

Algorithm 5 Reduced Dimension-Adaptive PUQ Proc.

Input: Calibration set $\mathcal{S}_{\text{cal}} := \{x_i, y_i\}_{i=1}^n$. Scanned calibration parameter values $\Lambda^1 \leftarrow [1 \ldots \lambda_{1\max}]$, $\Lambda^2 \leftarrow [1 \ldots \lambda_{2\max}]$ and $\Lambda^3 \leftarrow [1 \ldots \lambda_{3\max}]$. 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 $\delta \in (0,1)$.

1: **for** $(x,y) \in \mathcal{S}_{\text{cal}}$ **do**
2:    **for** $\lambda_3 \in \Lambda_3$ **do**
    ▷ *Reduce dimensionality*
3:       $\hat{K}_{\lambda_3} \leftarrow \lfloor K \cdot \lambda_3 \rfloor$
4:       $\hat{B}(x), \hat{w}(x), \hat{\mu}(x), \tilde{u}(x), \tilde{l}(x) \leftarrow$ Apply Algorithm 2
         to x, with the choice of $\hat{K}_{\lambda_3}$ samples
5:       **for** $\lambda_1 \in \Lambda_1$ **do**
    ▷ *Compute adaptive dimensionality, Equation* (7)
6:         $\hat{k}(x; \lambda_1, \lambda_3) \leftarrow$
            $\min_k \left\{ k : \sum_{i=1}^{\hat{K}_{\lambda_3}} \hat{w}_i(x) \geq \lambda_1 \right\}$
    ▷ *Compute reconstruction loss, Equation* (8)
7:         $y_c \leftarrow y - \hat{\mu}(x)$
8:         $\mathcal{L}_1(x,y; \lambda_1, \lambda_3) \leftarrow$
     $\hat{\mathcal{Q}}_q \left(\left\{ \left| \left| \sum_{j=1}^{\hat{k}(x;\lambda_1,\lambda_3)} \hat{v}_j(x)^T y_c \hat{v}_j(x) - y_c \right| \right|_i \right\}_{i=1}^d \right)$
9:         **for** $\lambda_2 \in \Lambda_2$ **do**
    ▷ *Scale uncertainty intervals*
10:           $\hat{u}(x) \leftarrow \lambda_2 \tilde{u}(x)$ and $\hat{l}(x) \leftarrow \lambda_2 \tilde{l}(x)$
11:           $\mathcal{T}_{\lambda_2}(x; \hat{B}(x)) \leftarrow$
           Equation (1) using $\hat{\mu}(x), \hat{u}(x), \hat{l}(x)$
    ▷ *Compute weighted coverage loss, Equation* (5)
12:           $\mathcal{L}_2(x,y; \lambda_1, \lambda_2, \lambda_3) \leftarrow \sum_{i=1}^{\hat{k}(x;\lambda_1,\lambda_3)} \hat{w}_i(x) \cdot$
           $\mathbf{1} \left\{ \hat{v}_i(x)^T y \notin \mathcal{T}_{\lambda_2}(x; \hat{B}(x))_i \right\}$
13:         **end for**
14:       **end for**
15:    **end for**
16: **end for**
17: $\hat{\Lambda} \leftarrow$ Extract valid λs from LTT [15] applied on $\{(\mathcal{L}_1(x,y; \lambda_1, \lambda_3), \mathcal{L}_2(x,y; \lambda_1, \lambda_2, \lambda_3)) : (x,y) \in S_{\text{cal}}, \lambda_1 \in \Lambda^1, \lambda_2 \in \Lambda^2, \lambda_3 \in \Lambda^3\}$ at risk levels (β, α).
   ▷ *Compute the minimizers for the uncer. volume, Equation* (3)
18: $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3 \leftarrow \arg\min_{\lambda_1, \lambda_2, \lambda_3 \in \hat{\Lambda}} \left\{ \frac{1}{n} \sum_{i=1}^n \mathcal{V}_{\lambda_1, \lambda_2}(x_i; \hat{B}(x_i)) \right\}$
Output: Given a new instance $x \in \mathcal{X}$, obtain valid uncertainty intervals for it, $\mathcal{T}_{\hat{\lambda}_2}(x; \hat{B}(x))$ over $\hat{k}(x; \hat{\lambda}_1) \leq \hat{K}_{\hat{\lambda}_3}$ PCs.

---

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.

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 general scheme in Algorithm 1 can accommodate any stochastic regression solvers for inverse problems, such as conditional GANs [5]. In all tasks, we utilized the framework for conditional diffusion-based models proposed in the SR3 work [10], 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 [10]. 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.

3) PUQ Procedures and Hyperparameters: Our experimental approach follows the general scheme presented in Algorithm 1 and 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 (detailed in Section IV-B1) 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). In addition, we conducted another 2 experiments of the DA-PUQ (detailed in Section IV-B2), and RDA-PUQ (detailed in Appendix D) procedures on RGB patch resolution of 8x8. We set $q = 0.9$, $\beta = 0.05$ and $\alpha = \delta = 0.1$, to be the user-specified parameters of the guarantees of both Equation (10) and Equation (15). In total, we conducted 18 local experiments across three tasks. For the global experiments, we used entire images at a resolution of 128x128, in which we applied the DA-PUQ and the RDA-PUQ procedures. As global working is suitable for tasks that exhibit strong pixel correlation, we applied these 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 guarantees of both Equation (10) and Equation (15). Both locally and globally, for the DA-PUQ and RDA-PUQ experiments, we used $K = 100$ PCs in the colorization task and $K = 200$ PCs in super-resolution and inpainting. We note that in the RDA-PUQ experiments, we used $\hat{K}$ PCs during inference, as discussed in Appendix D. In all experiments we used $\epsilon = 1e - 10$ for the computation of the uncertainty volume, defined in Equation (3).