USIM-DAL: Uncertainty-aware Statistical Image Modeling-based Dense Active Learning for Super-resolution

Abstract

Dense regression is a widely used approach in computer vision for tasks such as image super-resolution, enhancement, depth estimation, etc. However, the high cost of annotation and labeling makes it challenging to achieve accurate results. We propose incorporating active learning into dense regression models to address this problem. Active learning allows models to select the most informative samples for labeling, reducing the overall annotation cost while improving performance. Despite its potential, active learning has not been widely explored in high-dimensional computer vision regression tasks like super-resolution. We address this research gap and propose a new framework called *USIM-DAL* that leverages the statistical properties of colour images to learn informative priors using probabilistic deep neural networks that model the heteroscedastic predictive distribution allowing uncertainty quantification. Moreover, the aleatoric uncertainty from the network serves as a proxy for error that is used for active learning. Our experiments on a wide variety of datasets spanning applications in natural images (visual genome, BSD100), medical imaging (histopathology slides), and remote sensing (satellite images) demonstrate the efficacy of the newly proposed *USIM-DAL* and superiority over several dense regression active learning methods.

1 INTRODUCTION

The paradigm of dense prediction is very important in computer vision, given that pixel-level regression tasks like super-resolution, restoration, depth estimation etc., help in holistic scene understanding. A common example of a pixel-level (i.e., dense) regression task is *Image super-resolution* (SR) is the process of recovering high-resolution (HR) images from their low-resolution (LR) versions. It is an important class of image processing techniques in computer vision, deep learning, and image processing and offers a wide range of real-world applications, such as medical imaging [Li et al., 2021], satellite imaging [Verpoorter et al., 2014], surveillance [Caner et al., 2003] and security [Gohshi, 2015], and remote sensing [Yang et al., 2015a], to name a few. The well-performing techniques for super-resolution often rely on deep learning-based methods that are trained in a supervised fashion, requiring high-resolution data as groundtruth. However, the acquisition of high-resolution imaging data (to be served as labels) for many real-world applications may be infeasible. Consider the example of histopathology microscopy from medical imaging, where the typical digital microscope takes significantly longer to acquire the high-resolution scans (i.e., at high magnification) image of the slide than low-magnification [Aeffner et al., 2018, Hamilton et al., 2014]. Moreover, the acquired high-resolution scans also have a significantly larger memory footprint leading to an increase in storage resources [Bertram and Klopfleisch, 2017]. Similarly, acquiring high spatial resolution images from satellites for remote sensing requires expensive sensors and hardware and has significantly higher operating costs [Cornebise et al., 2018, 2022]. In such scenarios, generating a large volume of training samples is infeasible.

As a remedy, concepts like zero-shot SR or single-image SR have been proposed. Nevertheless, zero-shot SR still requires ample supervision from the test image patches [Shocher et al., 2018] to learn the transferrable model for novel scenarios with divergent distributions [Soh et al., 2020], and the performance of the single-image SR models is still affected by the lack of sufficient labeled data [Lim et al., 2017]. Notwithstanding these discussions, there are situations where there are restrictions on dealing with training samples within a pre-defined budget. For example, in histopathology microscopy, the constraint on available resources may allow high-resolution acquisition for only a limited number of patients/microscopy slides.

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

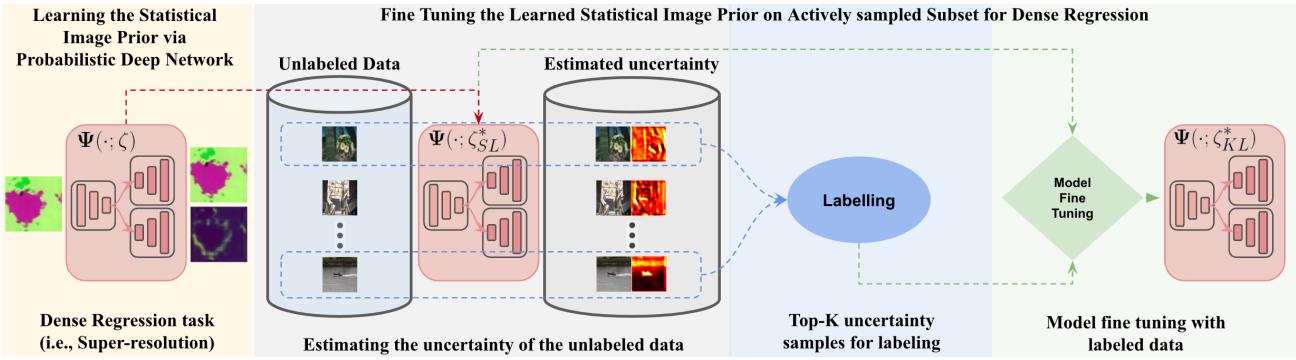

Figure 1: The proposed framework *USIM-DAL*. (Left-to-right) We train a probabilistic deep network for a dense regression task (e.g., super-resolution) on synthetic samples obtained from statistical image models as described in Section 3. The pre-trained model is used to identify the high-uncertainty samples from the domain-specific unlabeled set. Top-K highly uncertain samples are chosen for labeling on which the pre-trained network is further fine-tuned.

One of the viable solutions in this regard is to select a subset of highly representative training samples from the available training set while respecting the budget and deploying them to train the SR model. This corresponds to the notion of active learning for subset selection. However, selecting the subset is challenging considering the fact that we need a quantitative measurement for the eligibility of a given training LR-HR pair to be selected.

Many works have explored different *query functions* to select a subset to label from a larger dataset [Beluch et al., 2018, Gorriz et al., 2017, Roy and McCallum, 2001]. However, most of them have been applied to classification or low-dimensional regression problems [Jain and Grauman, 2016], and there still exists a gap on how to address this for dense regression tasks (e.g., super-resolution). Active learning technique to label those points for which the current model is least certain has been studied well in the context of classification [Yang et al., 2015b]. While there are recent advances in uncertainty estimation using deep neural networks for dense regression tasks in computer vision [Kendall and Gal, 2017], it is yet to be studied if they can be leveraged in active learning for dense regression tasks.

In summary, our contributions are as follows: (i) We show how statistical image models can help alleviate the need for a large volume of high-resolution imaging data. (ii) We show that probabilistic deep networks, along with the statistical image models, can be used to learn informative prior about niche domain datasets that may allow limited access to high-resolution data. (iii) Our probabilistic deep network trained with the statistical image models allows us to estimate the uncertainty for the sample in a niche domain that can be leveraged for active learning as illustrated in Figure 1.

2 RELATED WORK

Active Learning. These are a set of techniques that involve selecting a minimal data subset to be annotated, representing the entire dataset, and providing maximum performance gains. Querying strategies for active learning can be broadly categorized into three categories: heterogeneity-based, performance-based, and representativeness-based models. Uncertainty sampling [Beluch et al., 2018, Gorriz et al., 2017, Wang et al., 2016, Roy and McCallum, 2001, Ebrahimi et al., 2019], a type of heterogeneity-based model, is a standard active learning strategy where the learner aims to label those samples which have the most uncertain labelings. Non-Bayesian approaches[Brinker, 2003, Wang and Ye, 2015] dealing with entropy, distance from decision boundary, etc., also exist but are not scalable for deep learning Sener and Savarese [2017]. Representation-based methods that aim at increasing the diversity in a batch[Jain and Grauman, 2016] have also been studied. However, most of these works have been studied in the context of classification or low-dimensional regression problems, and the literature on dense regression is still sparse.

Statistical Image models. The $n \times n$ RGB images occupy the space of $\mathbb{R}^{3n^2}$. However, the structured images occupy a small region in that space. The statistical properties of the samples in this small structured space can be leveraged to generate synthetic data that have similar statistics to real-world structured images. For instance, the observation that natural images follow a power law with respect to the magnitude of their Fourier Transform (FT) formed the basis for Wiener image denoising[Simoncelli, 2005], Dead Leaves models [Lee et al., 2001] and fractals as image models [Redies et al., 2008]. Similarly, works like [Field, 1987, Simoncelli, 2005, Kretzmer, 1952] showed that outputs of zero mean wavelets to natural images are sparse and follow a generalized Laplacian distribution. Works like [Heeger and Bergen, 1995, Portilla and Simoncelli, 2000] showed statistical models capable of producing realistic-looking textures. The recent work [Baradad Jurjo et al., 2021] takes this research a step closer to realistic image generation by learning from procedural noise processes and using the generated samples for pre-training the neural networks. However, they only apply these methods to classification settings.

Super-resolution. This consists of CNN-based methods to enhance the resolution of the image [Ledig et al., 2017, Wang et al., 2018]. Attention mechanism has proven to be ubiquitous, with [Woo et al., 2018] introducing channel and spatial attention modules for adaptive feature refinement. Transformers-based endeavors such as [Liang et al., 2021], achieve state-of-the-art results using multi-head self-attention for SR. [Saharia et al., 2022] uses a probabilistic diffusion model and performs SR through an iterative denoising process. Works like [Shocher et al., 2018, Bose et al., 2022] use internal and external recurrence of information to get superior SR performance during inference. However, these works do not consider the problem of super-resolution in the active learning context, leaving a gap in the literature.

Uncertainty Estimation. Quantifying uncertainty in machine learning models is crucial for safety-critical applications [Nair et al., 2020, Sudarshan et al., 2021]. Uncertainty can be broadly categorized into two classes: (i) Epistemic uncertainty (i.e., uncertainty in model weights [Blundell et al., 2015, Daxberger et al., 2021, Graves, 2011, Kendall and Gal, 2017]). (ii) Aleatoric uncertainty (i.e., noise inherent in the observations) [Bae et al., 2021, Wang et al., 2019]. The dense predictive uncertainty may be considered as a proxy for error and can be used for active learning purposes [Laves et al., 2020, Upadhyay et al., 2022].

3 METHOD

We first formulate the problem in Section 3.1, and present preliminaries on active learning, statistical image models, and uncertainty estimation in Section 3.2. In Section 3.3, we describe the construction of *USIM-DAL* that learns a prior via statistical image modeling, which is later used to select the most informative samples from the unlabeled set for labeling and further improving the model.

3.1 PROBLEM FORMULATION

Let $\mathcal{D}_U = \{\mathbf{x}_i\}_{i=1}^N$ be the unlabeled set of input images from domain $\mathbf{X}$ (i.e., $\mathbf{x}_i \in \mathbf{X} \forall i$). We consider the task where images ($\mathbf{x}$) are to be mapped to another set of dense continuous labels ($\mathbf{y}$, e.g., other images, such that $\mathbf{y}_i \in \mathbf{Y} \forall i$). We want to learn a mapping $\mathbf{\Psi}$ for the same, i.e., $\mathbf{\Psi} : \mathbf{X} \rightarrow \mathbf{Y}$. However, we want to learn it under the constraint that we do not have sufficient *budget* to "label" all the N samples in $\mathcal{D}_U$ (i.e., acquire all the corresponding $\mathbf{y}$), but we do have a budget to label a significantly smaller subset of $\mathcal{D}_U$ with $K << N$ samples, say $\mathcal{D}_U^K$. This is a real-world constraint, as discussed in Section 2. In this work, we focus on the problem of super-resolution where the domain $\mathbf{Y}$ consists of high-resolution images (corresponding to the low-resolution images in domain $\mathbf{X}$).

We tackle the problem of choosing the set of $K << N$ samples ($\mathcal{D}_U^K$) that are highly representative of the entire unlabeled training set $\mathcal{D}_U$, such that the learned mapping $\mathbf{\Psi}$ on unseen data from a similar domain performs well.

3.2 PRELIMINARIES

Active Learning. As discussed above, given a set of N unlabeled images $\mathcal{D}_U$, we want to choose a set of $K << N$ samples ($\mathcal{D}_U^K$) that are highly representative of the entire unlabeled training set $\mathcal{D}_U$. This is the problem of active learning, which consists of *query strategies* that maps the entire unlabeled set $\mathcal{D}_U$ to its subset. That is, the query strategy (constrained to choose K samples and parameterized by ϕ) is given by, $\mathcal{Q}_{K,\phi} : \mathcal{D}_U \rightarrow \mathcal{D}_U^K$. Many works explore designing the query strategy $\mathcal{Q}_{K,\phi}$ [Beluch et al., 2018, Gorriz et al., 2017, Wang et al., 2016]. However, they seldom attempt to design such a strategy for dense regression.

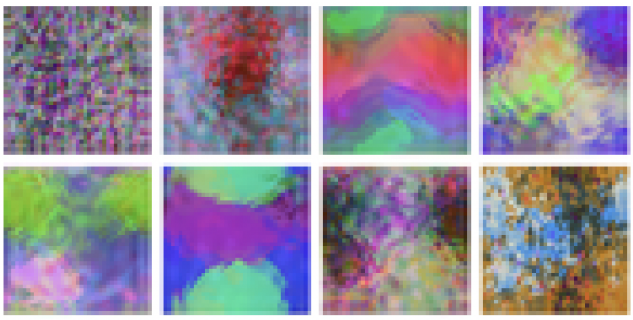

Figure 2: Samples generated from Statistical Image Models (combination of Spectrum + WMM + Color histogram).

Statistical Image Models (SIM). As discussed in [Baradad Jurjo et al., 2021], the statistical properties of RGB images can be exploited to generate synthetic images that can serve as an excellent pre-training learning signal. The generative model (based on statistical properties of RGB images) is described as $\mathcal{G}(\cdot;\theta_G) : \mathbf{z} \rightarrow \mathbf{x}$ where $\mathbf{z}$ is a stochastic latent variable and $\mathbf{x}$ is an image. The image generation is modelled as a hierarchical process in which, first, the parameters of a model are sampled. Then the image is sampled given these parameters and stochastic noise. Previous works [Baradad Jurjo et al., 2021] highlight the following statistical models. (i) **Spectrum:** based on the magnitude of the Fourier transform (FT). The FT of many natural images follows a power law, i.e., $\frac{1}{|f|^\alpha}$, where $|f|$ is the magnitude of frequency f, and α is a constant close to 1. For generative models, the sampled images are constrained to be random noise images that have FT magnitude following $\frac{1}{|f_x|^a+|f_y|^b}$ with a and b being two random numbers uniformly sampled as detailed in [Baradad Jurjo et al., 2021]. (ii) **Wavelet-marginal model (WMM):** Generates the texture by modeling their histograms of wavelet coefficient as discussed in [Simoncelli, 2005, Kretzmer, 1952]. (iii) **Color histograms:** As discussed in [Baradad Jurjo et al., 2021], this generative model follows the color distribution of the dead-leaves

model [Baradad Jurjo et al., 2021]. Combining all these different models allows for capturing colour distributions, spectral components, and wavelet distributions that mimic those typical for natural images. Figure 2 shows examples of generated samples from such models.

Uncertainty Estimation. Various works [Lakshminarayanan et al., 2016, Kendall and Gal, 2017] have proposed different methods to model the uncertainty estimates in the predictions made by DNNs for different tasks. Interestingly recent works [Kendall and Gal, 2017, Upadhyay et al., 2022] have shown that for many real-world vision applications, modeling the aleatoric uncertainty allows for capturing erroneous predictions that may happen with out-of-distribution samples. To estimate the uncertainty for the regression tasks using deep network (say $\mathbf{\Psi}(\cdot;\zeta) : \mathbf{X} \rightarrow \mathbf{Y}$), the model must capture the output distribution $\mathcal{P}_{Y|X}$. This is often done by estimating $\mathcal{P}_{Y|X}$ with a parametric distribution and learning the parameters of the said distribution using the deep network, which is then used to maximize the likelihood function. That is, for an input $\mathbf{x}_i$, the model produces a set of parameters representing the output given by, $\{\hat{\mathbf{y}}_i, \hat{\nu}_i \dots \hat{\rho}_i\} := \mathbf{\Psi}(\mathbf{x}_i;\zeta)$, that characterizes the distribution $\mathcal{P}_{Y|X}(\mathbf{y};\{\hat{\mathbf{y}}_i, \hat{\nu}_i \dots \hat{\rho}_i\})$, such that $\mathbf{y}_i \sim \mathcal{P}_{Y|X}(\mathbf{y};\{\hat{\mathbf{y}}_i, \hat{\nu}_i \dots \hat{\rho}_i\})$. The likelihood $\mathscr{L}(\zeta;\mathcal{D}) := \prod_{i=1}^{N} \mathcal{P}_{Y|X}(\mathbf{y}_i;\{\hat{\mathbf{y}}_i, \hat{\nu}_i \dots \hat{\rho}_i\})$ is then maximized to estimate the optimal parameters of the network. Typically, the parameterized distribution is chosen to be *heteroscedastic* Gaussian distribution, in which case $\mathbf{\Psi}(\cdot;\zeta)$ is designed to predict the *mean* and *variance* of the Gaussian distribution, i.e., $\{\hat{\mathbf{y}}_i, \hat{\sigma}_i^2\} := \mathbf{\Psi}(\mathbf{x}_i;\zeta)$. The optimization problem becomes,

$$\zeta^* = \underset{\zeta}{\text{argmin}} \sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|^2}{2\hat{\sigma}_i^2} + \frac{\log(\hat{\sigma}_i^2)}{2} \tag{1}$$

With Uncertainty$(\hat{\mathbf{y}}_i) = \hat{\sigma}_i^2$. An important observation from Equation 1 is that, ignoring the dependence through ζ, the solution to Equation 1 decouples estimation of $\hat{\mathbf{y}}_i$ and $\hat{\sigma}_i$. That is, for minimizing with respect to $\hat{\mathbf{y}}_i$ we need,

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|^2}{2\hat{\sigma}_i^2} + \frac{\log(\hat{\sigma}_i^2)}{2} \right)}{\partial \hat{\mathbf{y}}_i} = 0 \tag{2}$$

$$\frac{\partial^2 \left(\sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|^2}{2\hat{\sigma}_i^2} + \frac{\log(\hat{\sigma}_i^2)}{2} \right)}{\partial \hat{\mathbf{y}}_i^2} > 0 \tag{3}$$

Equation 2 & 3 lead to $\hat{\mathbf{y}}_i = \mathbf{y}_i \; \forall i$. Similarly for minimizing with respect to $\hat{\sigma}_i$ we need,

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|^2}{2\hat{\sigma}_i^2} + \frac{\log(\hat{\sigma}_i^2)}{2} \right)}{\partial \hat{\sigma}_i} = 0 \tag{4}$$

$$\frac{\partial^2 \left(\sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|^2}{2\hat{\sigma}_i^2} + \frac{\log(\hat{\sigma}_i^2)}{2} \right)}{\partial \hat{\sigma}_i^2} > 0 \tag{5}$$

Equation 4 & 5 lead to $\hat{\sigma}_i^2 = |\hat{\mathbf{y}}_i - \mathbf{y}_i|^2 \; \forall i$. That is, the estimation $\hat{\sigma}_i^2$ should perfectly reflect the squared error. Therefore, a higher $\hat{\sigma}_i^2$ indicates higher error. We leverage this observation to design our dense active learning framework as described in Section 3.3.

3.3 CONSTRUCTING *USIM-DAL*

To tackle the problem mentioned in Section 3.1 (i.e., choosing a small subset), we leverage the fact that even before training the model with the labelled set, we can train a model based on the samples that we get from statistical image model as described above, which can then be used to make inference on the unlabeled domain-specific dataset identifying the high-uncertainty samples. The high-uncertainty samples can then be labelled and used to fine-tune the model.

We constraint the generative process for statistical image models as, Similar to [Baradad Jurjo et al., 2021], we treat image generation as a hierarchical process in which first the parameters of a model, θ_G, are sampled. Then the image is sampled given these parameters and stochastic noise, i.e.,

$$\theta_G \sim prior(\theta_G) \text{ and } \mathbf{z} \sim prior(\mathbf{z}) \tag{6}$$

$$\mathbf{x} = \mathcal{G}(\mathbf{z};\theta_G) \tag{7}$$

In particular, for super-resolution, we create a large (synthetic) labelled dataset using the samples from the statistical image models, say $\mathcal{D}_{SL} = \{(\texttt{low}(\mathbf{x}_{s,i}), \mathbf{x}_{s,i})\}_{i=1}^{M}$. Where $\mathbf{x}_{s,i}$ are generated samples from statistical image model and $\texttt{low}(\cdot)$, is the 4× down-sampling operation. We then train the network $\mathbf{\Psi}(\cdot;\zeta)$ on $\mathcal{D}_{SL}$ using Equation 1, leading to the optimal parameter ζ_{SL}^*, as shown in Figure 1. The trained model $\mathbf{\Psi}(\cdot;\zeta_{SL}^*)$ is then run in inference mode on all the samples of the unlabeled set $\mathcal{D}_U$ and gather the top uncertain samples for labeling, that is,

$$\{\hat{\mathbf{y}}_i, \hat{\sigma}_i\} := \mathbf{\Psi}(\mathbf{x}_i;\zeta_{SL}^*) \; \forall \mathbf{x}_i \in \mathcal{D}_U \tag{8}$$

$$\mathcal{D}_U^K := \{\mathbf{x}_j\} \forall j \in \texttt{topK}\left(\{\langle\hat{\sigma}_i\rangle\}_{i=1}^{N}\right) \tag{9}$$

Where, $\langle\cdot\rangle$ represents the mean operation, and $\texttt{topK}\left(\{\langle\hat{\sigma}_i\rangle\}_{i=1}^{N}\right)$ returns the indices of "top-K" most uncertain samples (i.e., mean uncertainty is high). We then acquire the labels for the samples in $\mathcal{D}_U^K$, giving us, $\mathcal{D}_{UL}^K = \{(\mathbf{x}_j, \mathbf{y}_j)\}$. As discussed in Section 3.2, the input samples in $\mathbf{D}_{UL}^K$ serve as a proxy to the set of K samples that would have the highest error between the prediction made by the model $\mathbf{\Psi}(\cdot;\zeta_{SL}^*)$ and the ground truth. That leads to better fine-tuning. The model $\mathbf{\Psi}(\cdot;\zeta_{SL}^*)$ is then fine-tuned on $\mathcal{D}_{UL}^K$ via Equation 1, leading to the final state of the model $\mathbf{\Psi}(\cdot;\zeta_{KL}^*)$ (shown in Figure 1) that can be used for inferring on the new sample.

4 EXPERIMENTS AND RESULTS

We provide an overview of the experiments performed and the results obtained. In Section 4.1, we describe the task and various methods used for comparison. Section 4.3 analyzes

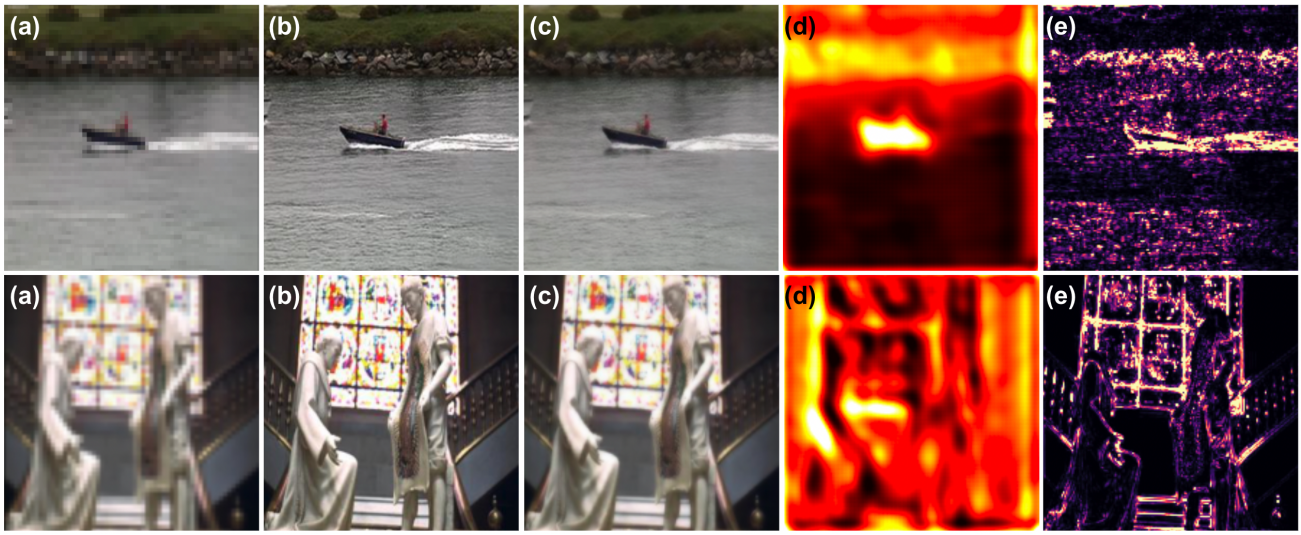

Figure 3: Output of the pre-trained probabilistic deep network (which is trained using synthetic images sampled from statistical image models) on samples from *unseen* natural image datasets. (a) LR input, (b) HR groundtruth, (c) Predicted output, SR, from the network, (d) Predicted uncertainty from the network, (e) Error between SR and groundtruth.

the performance of various dense active learning algorithms for super-resolution and shows that our proposed method *USIM-DAL* can help greatly improve the performance when constrained with a limited budget.

4.1 TASKS, DATASETS, AND METHODS

We present the results of all our experiments on the super-resolution task. We demonstrate our proposed framework using a probabilistic SRGAN (which is the adaptation of SRGAN [Ledig et al., 2017] that estimates pixel-wise uncertainty as described in [Kendall and Gal, 2017]) model. We evaluate the performance of various models on a wide variety of domains like (i) Natural Images (with Set5, Set14, BSD100, and Visual Genome dataset [Ledig et al., 2017, Martin et al., 2001, Krishna et al., 2017]). (ii) Satellite Images (with PatternNet dataset [Zhou et al., 2018]). (ii) Histopathology Medical Images (with Camelyon dataset [Litjens et al., 2018]). The evaluation protocol is designed to constraint all the training domain datasets to be restricted by a small fixed number of images (also called *training budget*). We used different training budgets of 500, 1000, 2000, 3000 and 5000 images for natural and satellite domains. For both natural and satellite images, the input image resolution was set to 64×64. For natural images the training dataset was obtained from Visual Genome (separate from the test-set). Similarly, for the histopathology medical images, the input image resolution was set to 32×32 and we used training budgets of 4000, 8000, 12000, and 16000.

We compare the super-resolution performance in terms of metrics MSE, MAE, PSNR, and SSIM [Wang et al., 2004] for the following methods on respective test sets: (i) SRGAN model trained from scratch with a randomly chosen subset satisfying the training budget from the entire training data (called *Random*). (ii) SRGAN model trained from scratch on a large synthetically generated dataset via statistical image modeling (as described in Section 3.2). This model is called *SIM*. (iii) SRGAN model trained from scratch on a large synthetically generated dataset via statistical image modeling and then fine-tuned on a randomly chosen subset satisfying the training budget from the entire training data, called *SIM+Random*. (iv) SRGAN model trained from scratch on a large synthetically generated dataset via statistical image modeling and then fine-tuned on a subset chosen using uncertainty estimates, satisfying the training budget from the entire training data, called *USIM-DAL*.

4.2 DENSE ACTIVE LEARNING VIA UNCERTAINTY ESTIMATION

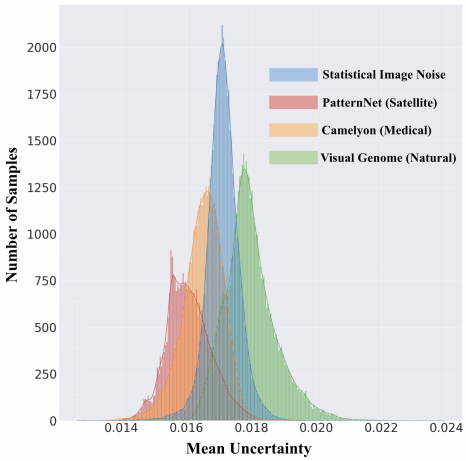

Figure 4: Distribution of mean uncertainty for samples in Statistical Image Noise, PatternNet (satellite), Camelyon (medical), Visual Genome (natural) datasets.

Our method proposes to utilize a probabilistic network that is learned from synthetic images sampled from statistical

D	Methods	Budgets (Number of images)				
		500	1000	2000	3000	5000
		MSE / MAE / PSNR / SSIM $\times 10^3$ $\times 10^2$ $\times 10^0$ $\times 10^2$	MSE / MAE / PSNR / SSIM $\times 10^3$ $\times 10^2$ $\times 10^0$ $\times 10^2$	MSE / MAE / PSNR / SSIM $\times 10^3$ $\times 10^2$ $\times 10^0$ $\times 10^2$	MSE / MAE / PSNR / SSIM $\times 10^3$ $\times 10^2$ $\times 10^0$ $\times 10^2$	MSE / MAE / PSNR / SSIM $\times 10^3$ $\times 10^2$ $\times 10^0$ $\times 10^2$
Set5	Random	4.129 / 3.854 / 24.784 / 7.232	3.898 / 3.720 / 24.957 / 7.319	3.660 / 3.588 / 25.271 / 7.422	3.586 / 3.529 / 25.334 / 7.465	3.500 / 3.420 / 25.514 / 7.539
	SIM	3.431 / 3.524 / 25.641 / 7.541	3.431 / 3.524 / 25.641 / 7.541	3.431 / 3.524 / 25.641 / 7.541	3.431 / 3.524 / 25.641 / 7.541	3.431 / 3.524 / 25.641 / 7.541
	SIM + Random	2.976 / 3.139 / 26.283 / 7.839	2.958 / 3.099 / 26.377 / 7.872	2.941 / 3.081 / 26.435 / 7.896	2.934 / 3.088 / 26.436 / 7.910	2.912 / 3.056 / 26.546 / 7.935
	USIM-DAL	**2.926 / 3.088 / 26.484 / 7.869**	**2.884 / 3.069 / 26.550 / 7.894**	**2.848 / 3.027 / 26.619 / 7.931**	**2.843 / 3.029 / 26.644 / 7.944**	**2.831 / 3.025 / 26.699 / 7.943**
Set14	Random	6.254 / 4.750 / 22.535 / 6.333	6.111 / 4.669 / 22.576 / 6.382	5.942 / 4.564 / 22.701 / 6.468	5.862 / 4.539 / 22.616 / 6.488	5.800 / 4.450 / 22.886 / 5.594
	SIM	4.852 / 4.303 / 22.897 / 6.383	4.852 / 4.303 / 22.897 / 6.383	4.852 / 4.303 / 22.897 / 6.383	4.852 / 4.303 / 22.897 / 6.383	4.852 / 4.303 / 22.897 / 6.383
	SIM + Random	4.488 / 3.907 / 23.748 / **7.016**	4.485 / 3.871 / 23.787 / **7.082**	4.444 / 3.828 / 24.106 / 7.159	4.426 / 3.828 / 24.162 / 7.179	4.396 / 3.798 / 24.090 / 7.198
	USIM-DAL	**4.376 / 3.836 / 23.810** / 6.984	**4.366 / 3.816 / 23.818** / 7.000	**4.331 / 3.767 / 24.288 / 7.177**	**4.317 / 3.749 / 24.422 / 7.208**	**4.292 / 3.728 / 24.553 / 7.227**
BSD100	Random	4.857 / 4.338 / 23.357 / 6.072	4.778 / 4.294 / 23.427 / 6.098	4.670 / 4.226 / 23.583 / 6.160	4.630 / 4.207 / 23.598 / 6.187	4.600 / 4.160 / 23.703 / 6.214
	SIM	3.526 / 3.738 / 24.805 / 6.713	3.526 / 3.738 / 24.805 / 6.713	3.526 / 3.738 / 24.805 / 6.713	3.526 / 3.738 / 24.805 / 6.713	3.526 / 3.738 / 24.805 / 6.713
	SIM + Random	3.362 / 3.578 / 25.007 / 6.786	3.352 / 3.559 / 25.043 / 6.794	3.328 / 3.539 / 25.092 / 6.812	3.323 / 3.540 / 25.085 / 6.816	3.305 / 3.519 / 25.137 / 6.834
	USIM-DAL	**3.299 / 3.520 / 25.174 / 6.826**	**3.293 / 3.520 / 25.191 / 6.830**	**3.282 / 3.504 / 25.207 / 6.838**	**3.277 / 3.496 / 25.212 / 6.844**	**3.262 / 3.486 / 25.263 / 6.854**
Visual Genome	Random	4.442 / 3.946 / 23.935 / 6.853	4.346 / 3.892 / 24.033 / 6.889	4.231 / 3.818 / 24.200 / 6.954	4.182 / 3.797 / 24.216 / 6.983	4.120 / 3.718 / 24.353 / 7.032
	SIM	4.310 / 3.963 / 24.055 / 6.826	4.310 / 3.963 / 24.055 / 6.826	4.310 / 3.963 / 24.055 / 6.826	4.310 / 3.963 / 24.055 / 6.826	4.310 / 3.963 / 24.055 / 6.826
	SIM + Random	4.038 / 3.721 / 24.396 / 7.036	4.026 / 3.690 / 24.423 / 7.056	3.993 / 3.663 / 24.496 / 7.088	3.977 / 3.661 / 24.515 / 7.101	3.943 / 3.631 / 24.563 / 7.126
	USIM-DAL	**3.966 / 3.668 / 24.543 / 7.056**	**3.949 / 3.657 / 24.570 / 7.069**	**3.925 / 3.623 / 24.624 / 7.109**	**3.908 / 3.608 / 24.656 / 7.126**	**3.880 / 3.593 / 24.721 / 7.143**

Table 1: Evaluating different methods on natural image datasets (Set5, Set14, BSD100, Visual Genome) using MSE, MAE, PSNR, SSIM. Lower MSE/MAE is better. Higher PSNR/SSIM is better. "D": Datasets. Best results are in **bold**.

image models (i.e., $\mathbf{\Psi}(\cdot;\zeta^*_{SL})$ mentioned in Section 3.3). Figure 3 shows the output of probabilistic SRGAN trained on synthetic images evaluated on samples from natural images. We observe that (i) The predicted super-resolved images (Figure 3-(c)) are still reasonable. (ii) The uncertainty estimates (Figure 3-(d)) still resemble the structures from the images and are a reasonable proxy to the error maps (Figure 3-(e)) between the predictions and the ground truth, even though the model has never seen the natural images.

We use the predicted uncertainty from this model to identify the samples from the real-world domain that would lead to high errors. Figure 4 shows the distribution of mean uncertainty values for samples in (i) Statistical Noise (ii) Natural (ii) Satellite (iii) Medical image datasets. We notice that the model trained on synthetic images leads to a gaussian distribution for the mean uncertainty values on the synthetic image datasets. We obtain similar distributions for other datasets from different domains. This further emphasizes that uncertainty estimates obtained from $\mathbf{\Psi}(\cdot;\zeta^*_{SL})$ can be used as a proxy to identify the highly uncertain (therefore erroneous) samples from different domains (i.e., the samples close to the right tail of the distributions).

4.3 *USIM-DAL* FOR SUPER-RESOLUTION

Table 1 shows the performance of different methods on multiple natural image datasets, including Set5, Set14, BSD100, and Visual Genome (VG). We observe that with the smallest training budget of 500 images, *USIM-DAL* performs the best with a PSNR/MAE of 25.174/0.035 (Table 1 shows the results with a scaling factor for better accommodation) compared to *SIM+Random* with PSNR/MAE of 25/0.039 and *SIM* with PSNR/MAE of 24.8/0.037. We also notice that at this budget, choosing the random subset of the training dataset to train the model from scratch performs the worst with PSNR/MAE of 23.36/0.043. As the budget increases (left to right in Tabel 1), the performances of all the methods also improve. However, a similar trend is observed where the *USIM-DAL* performs better than *SIM+Random*, *SIM*, and *Random*. We observe a similar trend for other natural image datasets. This allows us to make the following observations:

(i) Using a synthetic training image dataset (sampled from the statistical image model, discussed in Section 3.2) leads to better performance than using a small random subset of training images from the original domain (i.e., *SIM* better than *Random*). (ii) Using the above synthetic training image dataset to train a model and later fine-tuning it with domain-specific samples lead to further improvements (i.e., both *USIM-DAL* and *SIM+Random* better than *SIM*). (iii) With a limited budget, fine-tuning a model (pre-trained on synthetic training image dataset) using high-uncertainty samples from the training set (as decided by the *USIM-DAL*) is better than using the random samples from the training set (i.e., *USIM-DAL* better than *SIM+Random*).

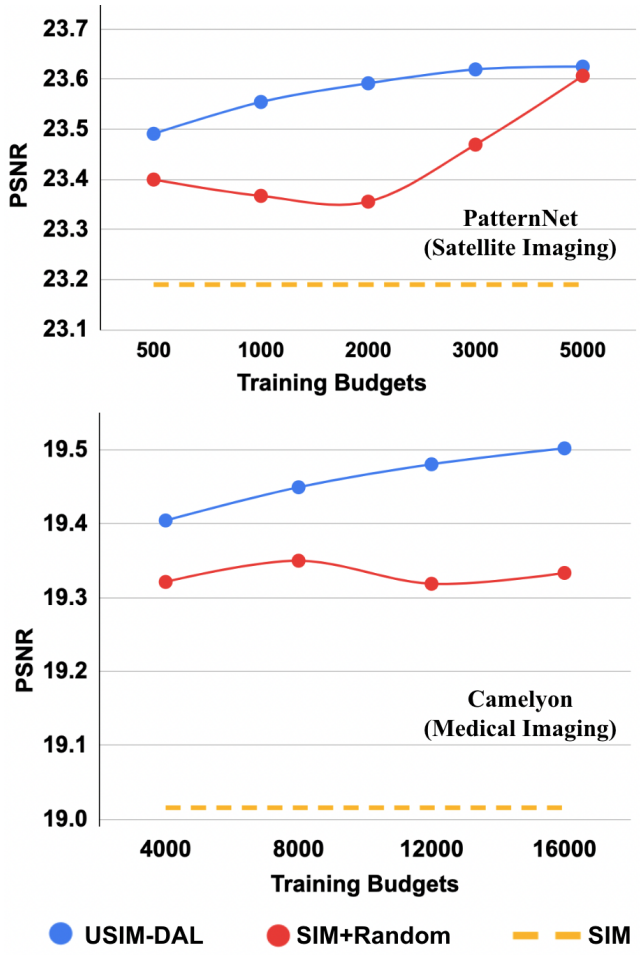

Figure 5: Evaluation of various methods on histopathology medical domain (i.e., Camelyon dataset) and satellite imaging domain (i.e., PatternNet dataset) at various fine-tuning budgets. The yellow curve is the *SIM* baseline. The red curve is the SIM model fine-tuned with random samples (i.e., *SIM+Random*). The blue curve is the SIM model fine-tuned with the highest uncertain samples (i.e., *USIM-DAL*).

We perform a similar set of experiments with other imaging domains, namely, (i) Satellite imaging (using PatternNet dataset) and (ii) Medical imaging (using Camelyon histopathology dataset). We observe a similar (to natural images) trend in these domains. Figure 5 shows the performance (measured using PSNR) for different methods on these two domains, with varying training budgets. For satellite imaging, at the lowest training budget of 500 images, *USIM-DAL* with PSNR of 23.5 performs better than *SIM+Random* with PSNR of 23.4 and *SIM* with a PSNR of 23.2. We observe that as the training budget increases to 2000 images, *USIM-DAL* (with PSNR of 23.6) outperforms *SIM+Random* (with PSNR of 23.35) with an even higher margin. As we increase the training budget further, the *SIM+Random* model starts performing similarly to *USIM-DAL*. With a budget of 5000 samples, *USIM-DAL* has a performance of 23.62, and *SIM+Random* has a performance of 23.60. Given a domain with large (specific to datasets) training budgets, the performance achieved from random sampling and active learning strategies will converge.

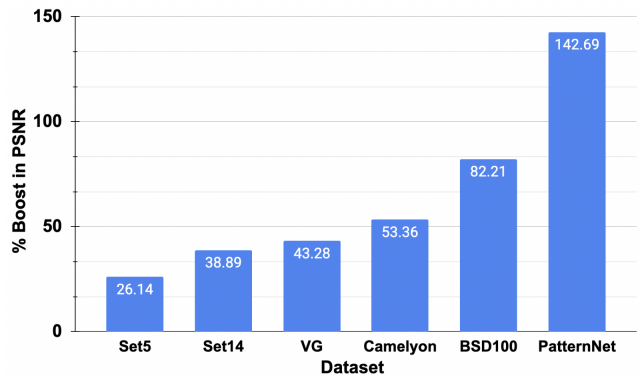

Figure 6: Relative % boost in PSNR of *USIM-DAL* relative to *SIM+Random* over *SIM* baseline (Equation 10) at optimal budget for six datasets across three domains.

For Camelyon dataset, we use the input image resolution of 32×32. We observe that *USIM-DAL* performs the best across all budgets when compared to *SIM+Random* and *SIM*. We also note that high-frequency features that are typically present in high-resolution scans (i.e., obtained at 20× or 40× magnification from the histopathology microscope) make the super-resolution problem harder and require more data to achieve good performance.

Figure 6 summarizes the performance gain (in terms of PSNR) by using *USIM-DAL* (i.e., uncertainty-based active learning strategy for dense regression) compared to *SIM+Random* (i.e., no active learning, randomly choosing a subset from real training domain), relative to *SIM* (i.e., no real samples used from the domain) at best performing limited budgets. That is, the relative percentage boost in performance is reported as:

$$\frac{(\text{PSNR}_{\text{USIM-DAL}} - \text{PSNR}_{\text{SIM+Random}}) * 100}{\text{PSNR}_{\text{SIM+Random}} - \text{PSNR}_{\text{SIM}}} \quad (10)$$

We note that *USIM-DAL* consistently performs better than *SIM+Random*, with the relative percentage boost in PSNR

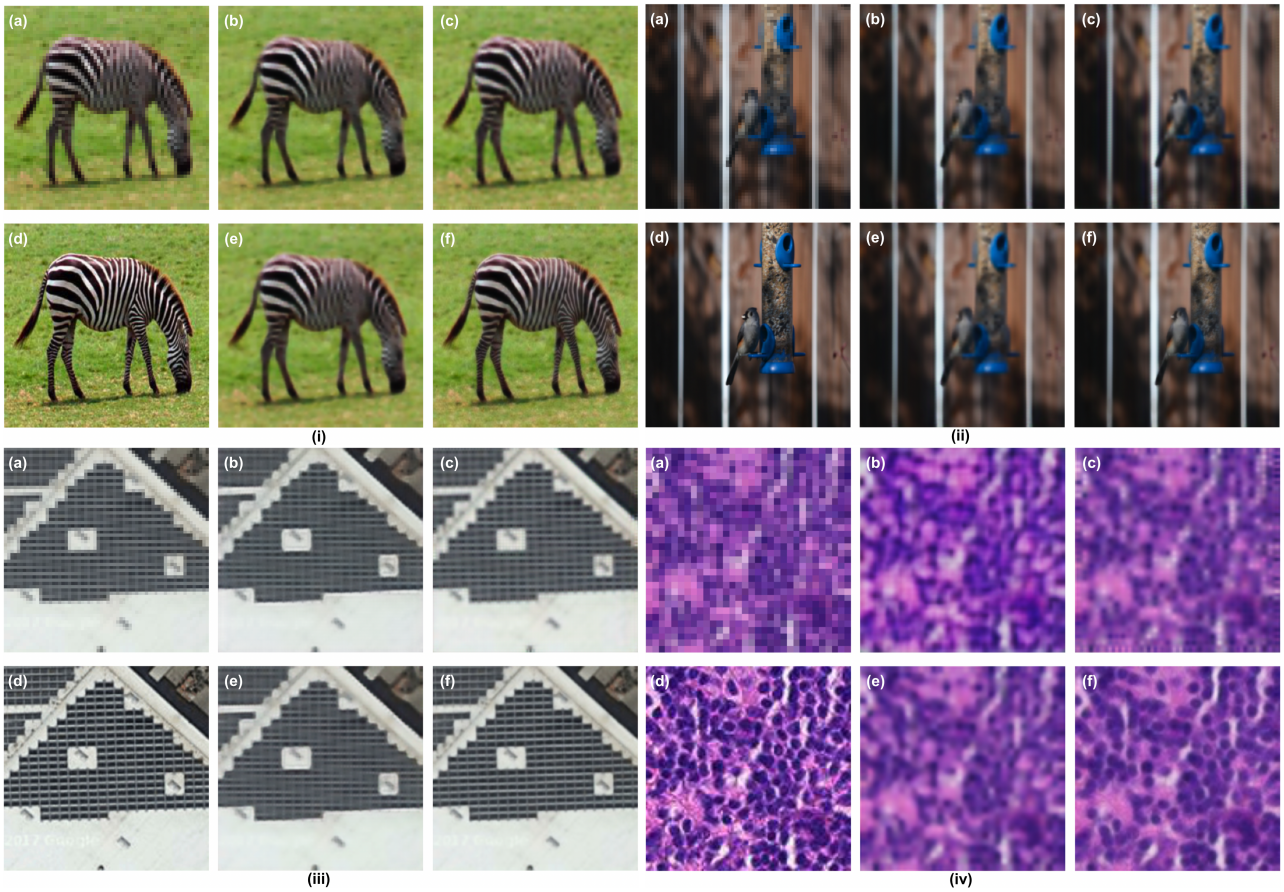

Figure 7: Qualitative results from different methods (performing $4\times$ super-resolution) including (b) *Random*, (c) *SIM*, (e) *SIM+Random*, (f) *USIM-DAL* on (i) BSD100, (ii) Visual Genome, (iii) PatternNet, and (iv) Camelyon datasets. (a) LR input, and (d) HR groundtruth. Input resolution for BSD100, Visual Genome, and PatternNet is 64×64, and for Camelyon is 32×32. (f) *USIM-DAL* produces the most visually appealing outputs.

of 26.14% for Set5 to 142.69% for PatternNet. Figure 7 shows the qualitative outputs of different models on multiple datasets. On all the datasets, we notice that the output obtained by *USIM-DAL* is better than the output of *SIM+Random* that is better than *SIM* and *Random*.

5 DISCUSSION AND CONCLUSION

In this work, we presented a novel framework called *USIM-DAL* that is designed to perform active learning for dense-regression tasks, such as image super-resolution. Dense-regression tasks, such as super-resolution, are an important class of problem for which deep learning offers a wide range of solutions applicable to medical imaging, security, and remote sensing. However, most of these solutions often rely on supervision signals derived from high-resolution images. Due to the time-consuming acquisition of high-resolution images or expensive sensors, hardware, and operational costs involved, it is not always feasible to generate large volumes of high-resolution imaging data. But in real-world scenarios, a limited budget for acquiring high-resolution data is often available. This calls for active learning that chooses a subset from large unlabeled set to perform labeling to train the models. While multiple querying strategies (in the context of active learning) exist for the classification tasks, the same for dense regression tasks are seldom discussed. Our work paves the way for using modern uncertainty estimation techniques for active learning in dense regression tasks. We show that a large synthetic dataset acquired using statistical image models can be used to learn informative priors for various domains, including natural images, medical images, satellite images, and more. The learned prior can then be used to choose the subset consisting of high-uncertainty samples that can then be labeled and used to fine-tune the prior further. Through extensive experimentation, we show that our approach generalizes well to a wide variety of domains, including medical and satellite imaging. we show that active learning performed by proposed querying strategy (i.e., *USIM-DAL*) leads to gains of upto 140% / 53% with respect to a random selection strategy (i.e., SIM+Random) relative to no dataset-specific fine-tuning (i.e., *SIM*) on satellite/medical imaging.

References

Famke Aeffner, Hibret A Adissu, Michael C Boyle, Robert D Cardiff, Erik Hagendorn, Mark J Hoenerhoff, Robert Klopfleisch, Susan Newbigging, Dirk Schaudien, Oliver Turner, et al. Digital microscopy, image analysis, and virtual slide repository. *ILAR journal*, 2018.

Gwangbin Bae, Ignas Budvytis, and Roberto Cipolla. Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In *ICCV*, 2021.

Manel Baradad Jurjo, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. *NeurIPS*, 2021.

William H. Beluch, Tim Genewein, Andreas Nürnberger, and Jan M. Köhler. The power of ensembles for active learning in image classification. In *CVPR*, 2018.

Christof A Bertram and Robert Klopfleisch. The pathologist 2.0: an update on digital pathology in veterinary medicine. *Veterinary pathology*, 2017.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *ICML*, 2015.

Rupak Bose, Vikrant Rangnekar, Biplab Banerjee, and Subhasis Chaudhuri. Zero-shot remote sensing image super-resolution based on image continuity and self tessellations. In *GCPR*, 2022.

Klaus Brinker. Incorporating diversity in active learning with support vector machines. In *ICML*, 2003.

G. Caner, A.M. Tekalp, and W. Heinzelman. Super resolution recovery for multi-camera surveillance imaging. In *International Conference on Multimedia and Expo*, 2003.

Julien Cornebise, Daniel Worrall, Micah Farfour, and Milena Marin. Witnessing atrocities: quantifying villages destruction in darfur with crowdsourcing and transfer learning. In *Proc. AI for Social Good NeurIPS2018 Workshop, NeurIPS'18*, 2018.

Julien Cornebise, Ivan Oršolić, and Freddie Kalaitzis. Open high-resolution satellite imagery: The worldstrat dataset–with application to super-resolution. *arXiv preprint arXiv:2207.06418*, 2022.

Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux-effortless bayesian deep learning. *NeurIPS*, 2021.

Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, and Marcus Rohrbach. Uncertainty-guided continual learning with bayesian neural networks. *arXiv preprint arXiv:1906.02425*, 2019.

David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 1987.

Seiichi Gohshi. Real-time super resolution algorithm for security cameras. In *International Joint Conference on e-Business and Telecommunications (ICETE)*, 2015.

Marc Gorriz, Axel Carlier, Emmanuel Faure, and Xavier Giro-i Nieto. Cost-effective active learning for melanoma segmentation. *arXiv preprint arXiv:1711.09168*, 2017.

Alex Graves. Practical variational inference for neural networks. *NeurIPS*, 2011.

Peter W Hamilton, Peter Bankhead, Yinhai Wang, Ryan Hutchinson, Declan Kieran, Darragh G McArt, Jacqueline James, and Manuel Salto-Tellez. Digital pathology and image analysis in tissue biomarker research. *Methods*, 2014.

David J Heeger and James R Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 1995.

Suyog Dutt Jain and Kristen Grauman. Active image segmentation propagation. In *CVPR*, 2016.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*, 2017.

Ernest R Kretzmer. Statistics of television signals. *The bell system technical journal*, 1952.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *IJCV*, 2017.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*, 2016.

Max-Heinrich Laves, Sontje Ihler, Jacob F Fast, Lüder A Kahrs, and Tobias Ortmaier. Well-calibrated regression uncertainty in medical imaging with deep learning. In *Medical Imaging with Deep Learning*, 2020.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017.

Ann B Lee, David Mumford, and Jinggang Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *IJCV*, 2001.

Y Li, Bruno Sixou, and F Peyrin. A review of the deep learning methods for medical images super resolution problems. *Irbm*, 2021.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCVw*, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRw*, 2017.

Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi, Oscar Geessink, Maschenka Balkenhol, Peter Bult, Altuna Halilovic, Meyke Hermsen, Rob van de Loo, Rob Vogels, et al. 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience*, 2018.

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.

Tanya Nair, Doina Precup, Douglas L Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Medical image analysis*, 2020.

Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 2000.

Christoph Redies, Jens Hasenstein, and Joachim Denzler. Fractal-like image statistics in visual art: similarity to natural scenes. *Spatial vision*, 2008.

Nicholas Roy and Andrew McCallum. Toward optimal active learning through monte carlo estimation of error reduction. *ICML, Williamstown*, 2001.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 2022.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *CVPR*, 2018.

Eero P Simoncelli. 4.7 statistical modeling of photographic images. *Handbook of Video and Image Processing*, 2005.

Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, 2020.

Viswanath P Sudarshan, Uddeshya Upadhyay, Gary F Egan, Zhaolin Chen, and Suyash P Awate. Towards lower-dose pet using physics-based uncertainty-aware multimodal learning with robustness to out-of-distribution data. *Medical Image Analysis*, 2021.

Uddeshya Upadhyay, Shyamgopal Karthik, Yanbei Chen, Massimiliano Mancini, and Zeynep Akata. BayesCap: Bayesian identity cap for calibrated uncertainty in frozen neural networks. In *European Conference on Computer Vision*, pages 299–317. Springer, 2022.

Charles Verpoorter, Tiit Kutser, David A Seekell, and Lars J Tranvik. A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 2014.

Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 2019.

Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCVw*, 2018.

Zheng Wang and Jieping Ye. Querying discriminative and representative samples for batch mode active learning. *ACM TKDD*, 2015.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 2004.

Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, 2018.

Daiqin Yang, Zimeng Li, Yatong Xia, and Zhenzhong Chen. Remote sensing image super-resolution: Challenges and approaches. In *IEEE DSP*, 2015a.

Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G Hauptmann. Multi-class active learning by uncertainty sampling with diversity maximization. *IJCV*, 2015b.

Weixun Zhou, Shawn Newsam, Congmin Li, and Zhenfeng Shao. Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS journal of photogrammetry and remote sensing*, 2018.