Representation Learning of Distribution Flows using Ensemble Control Systems

Wei Zhang* Lin Tang* Jr-Shin Li*†
*Department of Electrical & Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA
†Division of Computational & Data Sciences, Washington University in St. Louis, St. Louis, MO, USA
{wei.zhang, l.tang1, jsli}@wustl.edu

***Abstract*—Constructing interpretable representation learning models for flows of probability distributions has become a thriving research area in machine learning and artificial intelligence. This effort not only sparks new research avenues but also provides distinctive insights into established fields such as image processing. For example, the recently developed flow matching (FM) model has demonstrated its effectiveness in addressing image restoration problems. However, a unified, comprehensive, and interpretable framework for learning probability flows from data in a general setting remains underexplored. To fill this gap, we develop an ensemble control system (ECS) model for learning probability flows. Our model is represented as an ensemble of heterogeneous control systems, with the control inputs acting as time-dependent trainable parameters. The heterogeneous dynamics and time-dependent parameters significantly enhance the model's capabilities, making it exceptionally powerful. To further capitalize on these strengths, we introduce a moment kernel transform that creates a reduced kernel representation of the ECS model over a reproducing kernel Hilbert space. This approach enables efficient training without compromising learning performance. We demonstrate the significant advantages of the ECS model through various image restoration tasks and provide a detailed comparison with baseline FM-based image processing models.**

***Index Terms*—ensemble control systems, kernel methods, distribution flow learning, image restoration, reproducing kernel Hilbert space**

I. Introduction

Constructing representation learning models for the generation of desired flows of probability distributions forms the foundation of flow-based generative models [1]–[3]. The emergence of such models has opened new research avenues in generative artificial intelligence (AI) and has provided novel insights and tools for well-established fields, notably image processing and generation [1], [4]–[8]. The recently developed flow matching (FM) technique falls into this category of generative models [8], [9]. When integrated with state-of-the-art algorithms for linear inverse problems, pre-trained FM models have demonstrated remarkable capabilities in various image restoration tasks, including denoising, deblurring, inpainting, and super-resolution [10]–[14].

The flow-based characteristic of FM elucidates its connection with dynamical systems. From this perspective, the learning objective of FM is the vector field of an ordinary differential equation (ODE) system whose phase space distribution, given by the solution of the associated Liouville equation (continuity equation), is the desired probability flow [8], [15]–[17].

a) Our contributions: Inspired by the dynamical system characterization of FM, we propose a control system-embodied representation learning model for flows of probability distributions. Our model comprises an ensemble of structurally identical control systems with distinct dynamic characteristics, referred to as an *ensemble control system (ECS)*, in contrast to a single unforced dynamical system used in FM. By utilizing the control inputs for targeted manipulation of the system dynamics, they naturally stand as trainable parameters in the ECS model. The time-dependence of the trainable control inputs also distinguishes ECS from most machine learning models, in which trainable parameters are typically constant. This feature, combined with the heterogeneity in ECS dynamics, not only enhances model capability but also renders ECS a high-dimensional model, potentially infinite-dimensional, in both the system state and training parameter spaces. To effectively leverage this infinite-dimensionality, we develop a moment kernel transform, which generates a kernel representation of ECS over a reproducing kernel Hilbert space (RKHS). This ensures efficient training of the ECS model while maintaining satisfactory model capability and learning performance. We apply ECS to various image restoration tasks to demonstrate its efficiency and performance, showcasing a new dynamic image processing paradigm. Additionally, we conducted a detailed comparison with state-of-the-art FM-based image processing models. The main contributions of our work are summarized as follows:

- Construction of ECS representing models for learning and generation of probability flows.
- Development of the moment kernelization framework to enable efficient training of the ECS model over an RKHS without compromising its model capability.
- Application of ECS to various image restoration tasks to establish a new dynamic image processing paradigm.

b) Related works: The fundamental objective of this work is the targeted transportation of probability distributions using ECS's. The problem of transporting probability distributions was initially formulated by G. Monge in 1781 [18], laying the foundation for optimal transport theory [19]–

This work was supported in part by the Air Force Office of Scientific Research under award FA955021-1-0335 and the National Science Foundation under award 2508439.

[21]. The connection between optimal transport and dynamical systems was discovered by J.-D. Benamou and Y. Brenier, where probability distributions are transported along flows of dynamical systems [22]. Recently, the significance of this viewpoint was recognized in the realm of generative AI, catalyzing the development of FM [3], [8], [9].

This emerging FM model particularly provides new insights and tools for image processing. The core idea in this thread of research centers around integrating a pre-trained FM model with linear inverse algorithms to enhance image restoration performance. State-of-the-art FM-based image processing models include OT-ODE, D-Flow, Flow Priors, and PnP-Flow [10]–[13]. In OT-ODE, a gradient-based adaptation term is added to the pretrained unconditional flow model, converting an unconditional vector field to a conditional one to improve the perceptual quality of restored images [10]. In D-Flow, controlled generation is added to an FM or diffusion model, imposing a prior on the optimization process in an image inverse problem, resulting in better model generalizability [11]. In Flow Priors, the Iterative Corrupted Trajectory Matching (ICTM) algorithm is applied to approximate the maximum-a-posteriori (MAP) estimator in an FM model, improving image restoration performance [12]. In the latest PnP-Flow, the Plug-and-Play framework is combined with a pre-trained FM, yielding a time-dependent image denoiser with improved computational efficiency and memory consumption [13].

In the proposed ECS probability flow learning model, the technical tool for addressing infinite dimensionality is the moment kernel transform, which is motivated by the method of moments in functional analysis and probability theory [23]–[25]. This method was developed by P. L. Chebyshev in 1887 and later formalized by his student A. Markov [26]. It soon gained considerable attention and was studied under different settings, notably, the Hausdorff, Hamburger, and Stieltjes moment problems [27]–[31]. In our work, we adopt the modern mathematical formulation of the moment method [24] and extend it to the dynamical systems setting to establish the ECS learning model.

II. Representation of Distribution Flows using Dynamic Ensemble Systems

Building an interpretable and trustworthy model for learning probability flows is an emerging research area in contemporary AI, particularly in generative models. The time-dependence characteristic of flows inherently highlights the close relationship between such models and dynamical systems. In this section, we introduce a new representation learning model for probability flows, described by an ensemble of control systems with distinct dynamic characteristics. This heterogeneity endows the ECS model with enhanced model capability compared to the existing flow-based generative models inspired by dynamical systems, such as FM.

A. Ensemble systems representation of distribution flows

An *ensemble control system (ECS)* is a parameterized family of control systems evolving on a common manifold $M \subseteq \mathbb{R}^n$, described by

$$\frac{d}{dt}x(t,\beta) = f\big(t,\beta,x(t,\beta),u(t)\big), \tag{1}$$

where the system parameter β varies over $\Omega \subseteq \mathbb{R}^d$, $x(t,\beta) \in M$ holds for all $t \in [0,T]$ and $\beta \in \Omega$ with $T > 0$, $f(t,\beta,\cdot,u(t))$ is a vector field on M, and $u(t) \in \mathbb{R}^p$ is the control input. Systems of this sort are natural models for describing the dynamics of large-scale population systems arising from numerous domains, including robot swarms, quantum spin ensembles, and spiking neurons [32]–[34].

In practice, the data of the states of individual systems within the ensemble are distributed differently at each time t. Namely, at each $t \in [0,T]$, these data, denoted $X_t = \{x(t,\beta) \in M : \beta \in \Omega\}$, have an underlying distribution μ_t over M, representing how these systems are allocated in the state-space M. For example, they may be concentrated at one point, making μ_t a point mass, or they may be uniformly spread out over a region $D \subseteq M$, making μ_t the uniform distribution on D. Therefore, this *ensemble distribution* μ_t is time-varying and governed by the vector field f. In this distributional interpretation of the ECS in (1), the control input $u(t)$ serves as the training parameter for learning a desired distribution flow μ_t^*. In this context, the learning task admits the following optimal tracking control formulation:

$$\begin{aligned} \min_{u:[0,T]\to\mathbb{R}^p} \quad & \mathbb{E}\int_0^T |f(t,\beta,x(t,\beta),u(t)) - f^*(t,x(t,\beta))|^2 dt, \\ \text{s.t.} \quad & \frac{d}{dt}x(t,\beta) = f\big(t,\beta,x(t,\beta),u(t)\big) \end{aligned} \tag{2}$$

with the initial ensemble distribution μ_0 induced by data X_0, where $\mathbb{E}$ denotes the expectation with respect to a predetermined probability measure λ on the parameter space Ω, and f^* is the velocity field of μ_t^*.

B. Comparison with flow matching-based models

The fundamental difference of the proposed ECS model from FM lies in its characteristics of having training parameters as functions and the richness of its system dynamics generated by these functional parameters. In most existing machine learning models, including FM, training parameters are finite-dimensional constant vectors; however, in the ECS model described in (1), they are $\mathbb{R}^p$-valued functions defined on $[0,T]$. This expansion significantly enhances the learning capability of the ECS model.

In addition, the heterogeneity introduced by the varying parameter β provides rich diversity to the model. Specifically, the ECS in (1) evolves on the space of M-valued functions over Ω, which is infinite-dimensional. In contrast, the underlying dynamical system in an FM model is a homogeneous ODE system defined on a finite-dimensional space. The expansion in the dimension of the system's state-space empowers the ECS model with the full capability to represent richer classes of distributions, thereby enabling ECS to tackle a wider range of learning tasks. For instance, FM is incapable of handling a classification task in which the desired distribution flow necessarily converges to a totally discrete probability distribution consisting of point masses. This requires the sample paths of the distribution flow, composed of the solution trajectories of the associated ODE with different initial conditions, to intersect at the mass points, which violates the uniqueness of solutions to the ODE [35]. This limitation in FM is overcome by the proposed ECS model, where every sample path is generated by a distinct control system characterized by a different value of β.

Accompanied by the enhanced model capability through the enlargement of the training parameter and state spaces comes potentially increased computational complexity in training and operating the ECS model. This presents challenges to both the theoretical and computational aspects of the model. To address this issue, we develop a new learning representation for probability flows by mapping the ECS model to a reproducing kernel Hilbert space (RKHS), which is the main focus of the next section.

III. Moment Kernelization for Distribution Flow Learning

In this section, we develop a moment kernel representation of the ECS model for learning probability flows. This kernelization maps the ECS model to a reduced moment representation, enabling rigorous and efficient probability flow learning on an RKHS.

A. Moment kernel transform

The core idea of the moment kernel transform is to identify a Borel measure with an infinite sequence. In this work, we focus on Radon probability measures defined on the parameter space Ω and denote the space of such measures by $\mathcal{P}(\Omega)$. Because every positive linear functional on $C_c(\Omega)$, the space of compactly supported continuous real-valued functions defined on Ω, is uniquely representable as a measure on Ω by the Riesz–Markov–Kakutani representation theorem [36], this allows us to introduce the notion of *moments* for $\mu_t \in \mathcal{P}(\Omega)$ by using the dual-primal pairing $\langle\cdot,\cdot\rangle : C_c(\Omega)\times\mathcal{P}(\Omega)\to\mathbb{R}$ as

$$m_k(t) = \langle \varphi_k, \mu_t\rangle = \int_\Omega \varphi_k d\mu_t, \quad k\in\mathbb{N}. \tag{3}$$

Here, $\{\varphi_k\}_{k\in\mathbb{N}}$ is chosen to be a basis for $C_c(\Omega)$, which guarantees a one-to-one correspondence between the measure μ_t and its *moment sequence* $m(t) = \big(m_k(t)\big)_{k\in\mathbb{N}}$ [24].

To illustrate how to incorporate the moment kernelization into the ECS model, we focus on Radon measures μ_t on Ω that are absolutely continuous with respect to a predetermined probability measure $\lambda \in \mathcal{P}(\Omega)$ with the Radon-Nikodym derivative given by the ensemble state x_t, i.e., $d\mu_t = x_t d\lambda$. In this case, the dual-primal pairing can be identified with the L^2-inner product of the functions φ_k and x_t with respect to the measure λ as $m_k(t) = \langle\varphi_k,\mu_t\rangle = \int_\Omega \varphi_k x_t d\lambda = \langle\varphi_k, x_t\rangle_{L^2}$, provided $\|x_t\|^2 = \int_\Omega |x_t|^2 d\lambda < \infty$, i.e., $x_t \in L^2(\Omega,\lambda)$. This identification further allows the choice of $\{\varphi_k\}_{k\in\mathbb{N}}$ as an orthonormal basis of $L^2(\Omega,\lambda)$. In this case, the moment sequence $m(t)$ satisfies $\|m(t)\|^2_{\ell^2} = \sum_{k\in\mathbb{N}} |m_k(t)|^2 = \|x_t\|^2_{L^2} < \infty$, indicating that $m(t)$ is an element in ℓ^2, the space of square-summable sequences [36]. Since the ℓ^2-space is an RKHS with the reproducing kernel induced by the ℓ^2-inner product $\langle m(t_1), m(t_2)\rangle_{\ell^2} = \sum_{k\in\mathbb{N}} m_k(t_1)m_k(t_2)$ [37], this demonstrates that $m(t)$ is essentially a kernel representation of μ_t. We will leverage this property to carry out ensemble control-based probability flow learning.

B. Moment kernel representation of ensemble distribution flow

The essential step in kernelizing the probability flow learning problem in (2) is to derive the moment kernel representation of the ensemble system. Taking the time-derivative of $m_k(t)$ defined in (3) yields

$$\begin{aligned}\frac{d}{dt}m_k(t) &= \frac{d}{dt}\langle\varphi_k,\mu_t\rangle = \frac{d}{dt}\int_\Omega \varphi_k x_t d\lambda \\ &= \int_\Omega \varphi_k \frac{d}{dt}x_t d\lambda = \int_\Omega \varphi_k f(t,\cdot,x_t,u(t))d\lambda,\end{aligned}$$

where the interchange of integration and differentiation follows from the dominated convergence theorem [36]. Note that the last term above is essentially the k^{th} moment of the x_t-dependent measure whose Radon-Nikodym derivative with respect to λ is $f(t,\beta,x_t(\beta),u(t))$ as a function of β. Representing the moment sequence of f as $\bar f = \big(\bar f_k\big)_{k\in\mathbb{N}}$, the moment kernel representation of the ensemble system follows $\frac{d}{dt}m(t) = \bar f\big(t,m(t),u(t)\big)$. On the other hand, the training loss in (2) satisfies $\mathbb{E}\int_0^T |f(t,\beta,x(t,\beta),u(t)) - f^*(t,x(t,\beta))|^2 dt = \int_\Omega\int_0^T |f(t,\ \beta,x(t,\beta),u(t)) - f^*(t,x(t,\beta))|^2 dt d\lambda(\beta) = \int_0^T\int_\Omega |f(t,\ \beta,x(t,\beta),u(t)) - f^*(t,x(t,\beta))|^2 d\lambda(\beta)dt = \int_0^T \|f(t,\cdot,x_t,u(t)) - f^*(t,x_t)\|^2_{L^2}dt = \int_0^T \|\bar f(t,m(t),u(t)) - \bar f^*(t,m(t))\|^2_{\ell^2}dt$, where the change in the order of integration follows from Tonelli's theorem [36]. We have now reached the moment kernelized ensemble distribution flow learning problem in (2), given by

$$\begin{aligned}\min_{u:[0,T]\to\mathbb{R}^p} \quad & \int_0^T \|\bar f(t,m(t),u(t)) - \bar f^*(t,m(t))\|^2 dt, \\ \text{s.t.} \quad & \frac{d}{dt}m(t) = \bar f\big(t,m(t),u(t)\big), \\ & m(0) = \Big(\langle\varphi_k,\mu_0\rangle\Big)_{k\in\mathbb{N}}.\end{aligned} \tag{4}$$

The structure of this moment kernelized problem defined on ℓ^2 suggests an effective training approach for the model in (2) by using a finite truncation of the moment sequence $m(t)$ in the moment kernelized ensemble FM model (4). Let $\hat m_q = (m_0, m_1, \ldots, m_q)$ and $\hat f_q = (\bar f_0, \bar f_1, \ldots, \bar f_q)$ denote the respective order-q truncated moment sequences for m and $\bar f$. We can then construct a finite-dimensional approximation of model (4) as

$$\begin{aligned}\min_{u:[0,T]\to\mathbb{R}^p} \quad & \int_0^T \|\hat f_q(t,\hat m_q(t),u(t)) - \hat f^*_q(t,\hat m_q(t))\|^2 dt, \\ \text{s.t.} \quad & \frac{d}{dt}\hat m_q(t) = \hat f_q\big(t,\hat m_q(t),u(t)\big), \\ & \hat m_q(0) = \Big(\langle\varphi_k,\mu_0\rangle\Big)_{k=0}^q.\end{aligned} \tag{5}$$

Theorem 1: Let $m^*(t)$ and $\hat{m}_q^*(t)$ be the moment kernelized probability flows learned by training the infinite-dimensional and truncated models in (4) and (5), respectively. Then, $\hat{m}_q^*(t) \to m^*(t)$ uniformly in $t \in [0,1]$ as the truncation order $q \to \infty$.

Proof. See Appendix A. □

In addition to the effective training through moment truncation, in practice the moments can be easily estimated from the data. Taking advantage of the fact that μ_t is a probability measure in Ω, the k^{th} moment defined in (3) can be viewed as the expectation of the "random variable" φ_k with respect to μ_t. Therefore, $m_k(t)$ can be approximated using the sample moment of the data sampled from μ_t, according to the law of large numbers [23]. Similarly, the moment sequence of f, $\bar{f}$, also admits the sample moment approximation.

IV. Dynamic Image Processing through Ensemble Distribution Flow Learning

One of the primary applications of distribution flow learning is image processing. In this section, we apply the developed ECS model to address various image processing tasks, including denoising, undistortion, and undegradation, using benchmark datasets. Furthermore, we conduct a detailed comparison with existing models to assess performance differences.

A. Experimental evaluation

a) Datasets: We validate the proposed ECS model on two benchmark datasets, CelebA and AFHQ-Cat, which contain RGB images of resolutions 128×128 and 256×256, respectively [38], [39]. For CelebA, images are decomposed into three color channels, each of which is reshaped into a 128^2-dimensional vector and normalized to a (discretized) Lebesgue probability density function on $[-1, 1]$. Images in the AFHQ-Cat dataset are processed in the same manner, resulting in 3-tuples of 256^2-dimensional vectors. Each vector is similarly normalized to a (discretized) probability density function on ([-1,1]).

b) ECS model setup: Images are typically processed using convolution operations, which are linear transforms [40], [41]. This inspires the use of a linear ensemble system to process the "image-distributions." Specifically, we build the ensemble system model with linear variation in its natural dynamics, controlled by multiple parameter-dependent inputs, in the form

$$\frac{d}{dt}x(t,\beta) = \beta x(t,\beta) + \sum_{i=0}^{p} \varphi_i(\beta)u_i(t), \tag{6}$$

where $\beta \in [-1, 1]$ and φ_i is the order-i Legendre polynomial for $i = 0, \dots, p$. The initial and target distributions, μ_0 and μ_1, correspond to the degraded and ground truth images, respectively. The probability flow to be learned is chosen as the linear interpolation $\mu_t = (1-t)\mu_0 + t\mu_1$, $t \in [0, 1]$. We employ Legendre polynomials to moment-kernelize the linear ensemble and choose the reference measure on $[-1, 1]$ to be the normalized Lebesgue measure, expressed as $d\lambda(\beta) = d\beta/2$. This approach yields the k^{th} moment as $m_k(t) = \langle \varphi_k, \mu_t \rangle = \int_{-1}^{1} \varphi_k(\beta) d\mu_t(\beta) = \frac{1}{2}\int_{-1}^{1} \varphi_k(\beta) x_t(\beta) d\beta$ for all $k \in \mathbb{N}$. The orthonormality of Legendre polynomials provides an explicit reconstruction of the probability flow to be learned using moments, given by $d\mu_t = x_t d\lambda = \sum_{k\in\mathbb{N}} m_k(t)\varphi_k d\lambda$. To improve numerical stability, image vectors are normalized to discretized Lebesgue density functions using the Legendre-Gauss-Lobatto (LGL) rule [42].

c) ECS model hyperparameter selection: This linear ECS model has two hyperparameters: the number p of control inputs and the truncation order q of the moment kernelization. To ensure sufficient model capability, we set $p = q$ in all the experiments. Specifically, p and q are chosen to be 150 for images in the CelebA dataset and 250 for images in the AFHQ-Cat dataset.

d) Baseline models: We compare the proposed ECS model with four state-of-the-art FM-based image processing models, including OT-ODE [10], D-Flow [11], Flow-Priors [12], and PnP-Flow [13]. The performance measures used for comparison are the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM), with higher values indicating better performance.

B. Experiment results

To showcase the performance of the proposed ECS model, we consider four different image restoration tasks: denoising, deblurring, super-resolution, and inpainting. (1) Denoising: The images are perturbed by additive Gaussian noise with a standard deviation of 0.2. (2) Deblurring: two types of blur are applied to the images: (i) Gaussian blur: using 61×61 Gaussian kernels with standard deviations of 1 and 3 for CelebA and AFHQ-Cat, respectively; (ii) Motion blur: with a blur length of 45 pixels and an angle of $45°$. (3) Super-resolution: The images are downsampled by a factor of 2 for CelebA and 4 for AFHQ-Cat. (4) Inpainting: two inpainting methods are considered: (i) Box inpainting: centered square masks of size 40×40 and 80×80 pixels are applied to CelebA and AFHQ-Cat images, respectively; (ii) Random inpainting: 70% of the pixels are randomly masked.

a) ECS model evaluation: Figures 1 and 2 illustrate the restoration performance of the ECS model on the CelebA and AFHQ-Cat datasets. In each figure, one representative image is selected to demonstrate the model performance across different restoration tasks. Specifically, the first and second columns represent the ground truth and degraded images, respectively; the third to sixth columns display the outputs of the ECS model at time points 0.25, 0.5, 0.75, and 1. It can be observed that the restored images at the final time point closely match the ground truth images. This high level of fidelity is also indicated by the high PSNR values.

b) Comparison with baseline models.: To make a fair comparison with the baseline models, additional Gaussian noise is added to the images. The standard deviation is chosen to be 0.05 for deblurring, super-resolution, and box inpainting tasks, and 0.01 for random inpainting tasks. For each image restoration task, the performance of the ECS model is evaluated using the average PSNR and SSIM over 100 selected

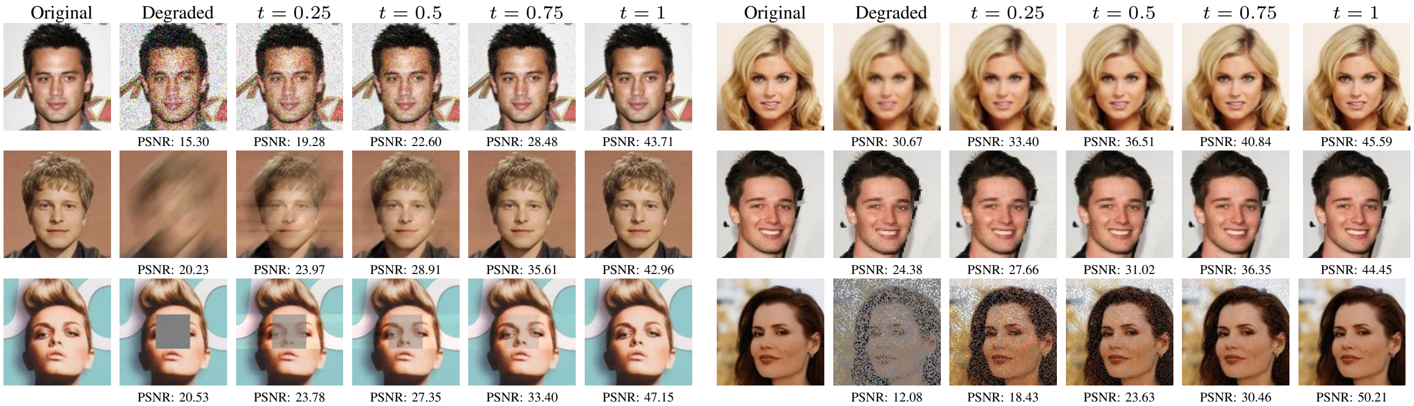

Fig. 1. Illustration of image restoration processes generated by the ECS model on the CelebA dataset. Row 1 shows the results of denoising and deblurring with Gaussian blur; Row 2 shows the results of deblurring with motion blur and super-resolution; Row 3 shows the results of single-mask and multi-mask inpainting. Columns 1 and 2 are the ground-truth and degraded images; columns 3 to 6 display the ECS outputs at time points $t = 0.25$, 0.5, 0.75, and 1, along with their corresponding PSNR values.

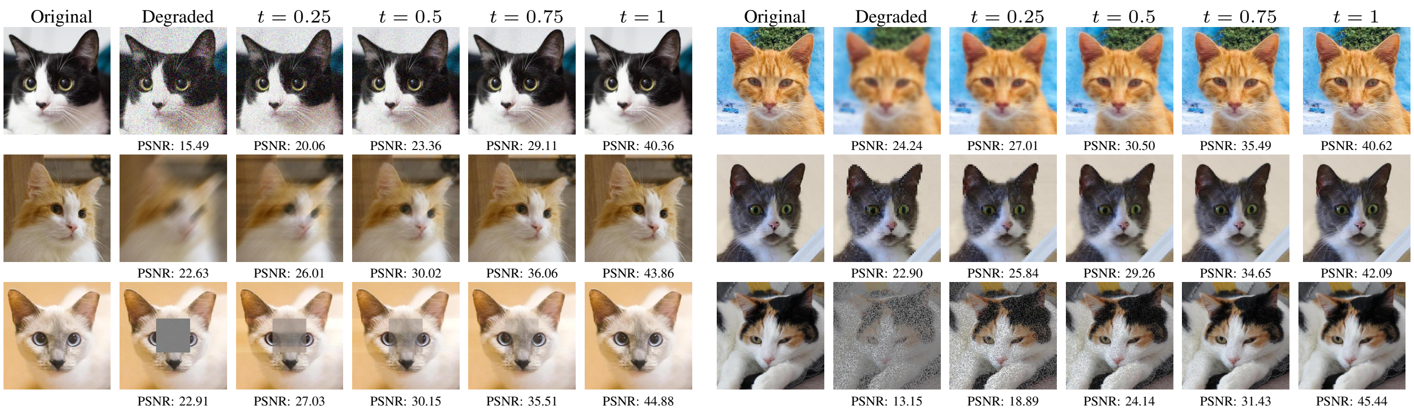

Fig. 2. Illustration of image restoration processes generated by the ECS model on the AFHQ-Cat dataset. Row 1 shows the results of denoising and deblurring with Gaussian blur; Row 2 shows the results of deblurring with motion blur and super-resolution; Row 3 shows the results of single-mask and multi-mask inpainting. Columns 1 and 2 are the ground-truth and degraded images; columns 3 to 6 display the ECS outputs at time points $t = 0.25$, 0.5, 0.75, and 1, along with their corresponding PSNR values.

images. The comparison results for the CelebA and AFHQ-Cat datasets are shown in Tables I and II, respectively. Figures 3 and 4 display one representative image restored by each model for each task. Overall, the proposed ECS model consistently outperforms the baseline models, as indicated by the PSNR values. Moreover, the restoration performance remains superb across different tasks, with PSNR variations on the order of 10^{-3} to 10^{-4}, demonstrating the robustness of the ECS model.

C. Discussion on the ECS Model

Interpretability and high training efficiency also stand out as major advantages of the ECS model, in addition to its model capability discussed in Section II-A. As a control system-embodied representation learning model, the trainable parameters of the ECS model are interpreted as control inputs, shaping the system dynamics to create the desired behavior. Regarding training efficiency, since the linear image processing ECS model in (6) is linear in the control inputs, its training process is essentially a linear inverse problem, which eliminates the need for time-consuming pretraining of deep neural networks. In our experiments, the average time required for training and processing the CelebA image deblurring task is 8.73 seconds per image on an Apple M4 Max chip with 128GB RAM. This is comparable to the processing times of baseline models solely, measured on a more powerful NVIDIA RTX 6000 Ada Generation with 48GB RAM, including 1.50 seconds for OT-ODE, 16.01 seconds for Flow-Priors, 32.19 seconds for D-Flow, and 3.40 seconds for PnP-Flow [13].

Although the ECS model has achieved superior image restoration performance, the learning accuracy can be further improved by tuning the hyperparameters of the model. In the current model setup as in (6), x_t is an real-valued function. We now increase the dimension of the model to $\frac{d}{dt}x(t,\beta) = \beta I x(t,\beta) + \sum_{i=0}^{p} \varphi_i(\beta) I u_i(t)$, where I denotes the n-by-n identity matrix, such that $x(t,\beta), u_i(t) \in \mathbb{R}^n$. Correspondingly, we also uniformly partition each image vector into n^2 vectors so that μ_t becomes a probability flow on $\mathbb{R}^{n^2}$. The restoration performance, measured by the average PSNR over 20 images, of this modified ECS model is shown in Table III, where n is varied from 1 to 4. It can be observed that the PSNR values are significantly increased as n increases.

TABLE I
COMPARISON BETWEEN THE PROPOSED ECS MODEL AND BASELINE MODELS USING THE CELEBA DATABASE FOR IMAGE DENOISING, DEBLURRING, SUPER-RESOLUTION, AND INPAINTING WITH BOX AND RANDOM MASKS.

Method	**Denoising**		**Deblurring**		**Super-resolution**		**Box inpainting**		**Random inpainting**	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Degraded	20.00	0.348	27.67	0.740	10.17	0.182	22.12	0.742	11.82	0.197
OT-ODE	30.50	0.867	32.63	0.915	31.05	0.902	28.84	0.914	28.36	0.865
D-Flow	26.42	0.651	31.07	0.877	30.75	0.866	29.70	0.893	33.07	0.938
Flow-Priors	29.26	0.766	31.40	0.856	28.35	0.717	29.40	0.858	32.33	0.945
PnP-Flow	32.45	0.911	34.51	0.940	31.49	0.907	30.59	0.943	33.54	0.953
Ensemble Control (ours)	46.1347	0.9979	46.1363	0.9979	46.1353	0.9979	46.1346	0.9979	46.1346	0.9979

TABLE II
COMPARISON BETWEEN THE PROPOSED ECS MODEL AND BASELINE MODELS USING THE AFHQ-CAT DATABASE FOR IMAGE DENOISING, DEBLURRING, SUPER-RESOLUTION, AND INPAINTING WITH BOX AND RANDOM MASKS.

Method	**Denoising**		**Deblurring**		**Super-resolution**		**Box inpainting**		**Random inpainting**	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Degraded	20.00	0.319	23.77	0.514	11.59	0.216	21.50	0.744	13.35	0.234
OT-ODE	29.90	0.831	26.43	0.709	25.17	0.711	23.88	0.874	28.84	0.838
D-Flow	26.22	0.620	27.49	0.740	24.10	0.595	26.69	0.833	31.37	0.888
Flow-Priors	29.32	0.768	25.78	0.692	23.34	0.573	25.85	0.822	31.76	0.909
PnP-Flow	31.65	0.876	27.62	0.763	26.75	0.774	26.87	0.904	32.98	0.930
Ensemble Control (ours)	40.4245	0.9891	40.4261	0.9891	40.4257	0.9891	40.4255	0.9891	40.4255	0.9891

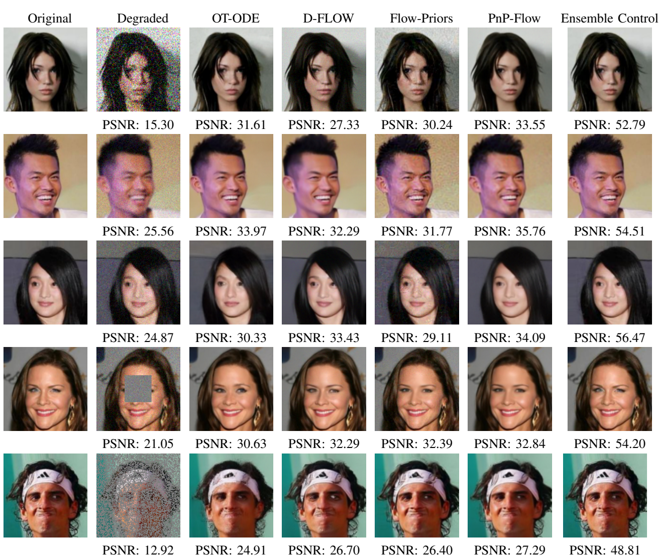

Fig. 3. Illustration of the comparison results between the proposed ECS model and baseline models for imaging denoising (row 1), deblurring (row 2), super-resolution (row 3), inpainting of single masks (row 4) and a multi mask (row 5) on CelebA dataset. Columns 1 and 2 are the ground truth and degraded images; columns 3 to 6 are the restored images output from the baseline models, including OT-ODE, D-Flow, Flow-Priors, and PnP-Flow; column 7 display the images restored by the ECS model.

Fig. 4. Illustration of the comparison results between the proposed ECS model and baseline models for imaging denoising (row 1), deblurring (row 2), super-resolution (row 3), inpainting of single masks (row 4) and a multi mask (row 5) on AFHQ dataset. Columns 1 and 2 are the ground truth and degraded images; columns 3 to 6 are the restored images output from the baseline models, including OT-ODE, D-Flow, Flow-Priors, and PnP-Flow; column 7 display the images restored by the ECS model.

V. CONCLUSION

In this paper, we introduced an ECS representation model for learning flows of probability distributions. Unlike existing models, such as FM, which use a single unforced system with constant trainable parameters, the ECS model generates distribution flows using an ensemble of heterogeneous control systems. The control inputs serve as time-varying trainable parameters, and coupled with heterogeneous dynamics, significantly enhance the model's capabilities. To

TABLE III
IMAGE RESTORATION PERFORMANCE OF THE MODIFIED ESC MODEL USING THE AFHQ-CAT DATASET

n	Denoising	Deblurring	Super-res.	Box inpaint.	Rand. inpaint.
1	41.1986	41.1992	41.1991	41.1988	41.1983
2	57.2110	57.2216	57.2182	57.2150	57.1982
3	71.9343	72.0013	71.9867	72.0128	71.6403
4	77.5737	77.9042	77.8156	77.7761	76.8013

further capitalize on these strengths, we developed a moment kernel transform that creates a reduced kernel representation of the ECS model over an RKHS. This approach facilitates efficient training without compromising learning performance. We demonstrated these significant advantages of the ECS model through various image restoration tasks and provided a detailed comparison with baseline FM-based image processing models.

APPENDIX A
PROOF OF THEOREM 1

Assumption. The vector field $f(t,\beta,x_t,u(t))$, as a function in β, is an L^2-function on Ω, which is Lipschitz continuous in (t,x_t) uniformly for $u(t)$. Meaning, there exists $L<\infty$, independent of u, such that $\|f(t_1,\cdot,x_{t_1},u(t_1))-f(t_2,\cdot,x_{t_2},u(t_2))\|_{L^2}\leq L\big(|t_1-t_2|+\|x_{t_1}-x_{t_2}\|_{L^2}\big)$.

Guaranteed by this condition, driven by any control input $u(t)$, the ECS model has a unique solution x_t [43]. Because the moment kernel transform $L^2(\Omega)\to\ell^2$ given by $x_t\mapsto m(t)$, defined in Section III-A, is an isometry, the vector field $\bar{f}$ in the moment kernelized ECS model also satisfies the Lipschitz continuity condition with the same Lipschitz constant L. Moreover, as a projection onto the first q components of $\bar{f}$, this assumption also holds for the vector field $\hat{f}_q$ in the truncated moment kernelized ECS model for all q.

Proof of Theorem 1. It suffices to prove that, with a fixed trainable control function $u(t)$, the output $\hat{m}_q(t)$ of the truncated moment kernelized ECS model in (5) converges to that $m(t)$ of the entire moment kernelized ECS model in (4).

We first show $\hat{f}_q\to\bar{f}$ as $q\to\infty$. As pointed out in Section III-B, $\bar{f}(t,m(t),u(t))\in\ell^2$ is the moment sequence of $f(t,\cdot,x_t,u(t))\in L^2(\Omega)$ and $\hat{f}_q(t,m(t),u(t))$ is the order-q truncation of $\bar{f}(t,m(t),u(t))$, yielding $\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|_{\ell^2}\to 0$ as $q\to\infty$, where to simplify the notation we use $\|\cdot\|$ to denote the ℓ^2-norm. The Lipschitz continuity of $\bar{f}$ and $\hat{f}_q$ particularly implies their continuity, and hence we also have for any fixed $t\in[0,T]$, $\|\bar{f}(t,m(t),u(t))-\bar{f}(t,\hat{m}_q(t),u(t))\|\to 0$ and $\|\hat{f}_q(t,m(t),u(t))-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\to 0$ as a result of $\|m(t)-\hat{m}_q(t)\|\to 0$ as $q\to\infty$ [44]. This vector field convergence then follows from the triangle inequality as

$$\begin{aligned}\|\bar{f}(t,m(t),u(t))&-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\\&\leq\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|\\&\quad+\|\hat{f}_q(t,m(t),u(t))-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\to 0\end{aligned}$$

as $q\to\infty$.

The convergence of the vector fields will be now adopted to prove the desired convergence for the trajectories of the moment and truncated moment kernelized ECS models. To avoid possible confusion about the moment order and truncation order, in the following, we use superscript to denote the moment order, e.g., $m(t)=\big(m^k(t)\big)_{k\in\mathbb{N}}$. Then, we have both of

$$\begin{aligned}\frac{d}{dt}\Big(m^k(t)-\hat{m}_q^k(t)\Big)&=\bar{f}^k(t,m(t),u(t))-\hat{f}_q^k(t,\hat{m}_q(t),u(t))\\&\leq\big|\bar{f}^k(t,m(t),u(t))-\hat{f}_q^k(t,\hat{m}_q(t),u(t))\big|\end{aligned}$$

and

$$\begin{aligned}\frac{d}{dt}\Big(\hat{m}_q^k(t)-m^k(t)\Big)&=\hat{f}_q^k(t,\hat{m}_q(t),u(t))-\bar{f}^k(t,m(t),u(t))\\&\leq\big|\bar{f}^k(t,m(t),u(t))-\hat{f}_q^k(t,\hat{m}_q(t),u(t))\big|,\end{aligned}$$

which yield

$$\frac{d}{dt}\Big|m^k(t)-\hat{m}_q^k(t)\Big|\leq\Big|\bar{f}^k(t,m(t),u(t))-\hat{f}_q^k(t,\hat{m}_q(t),u(t))\Big|.$$

Following from monotone convergence theorem, we have

$$\begin{aligned}&\frac{d}{dt}\|m(t)-\hat{m}_q(t)\|^2=\frac{d}{dt}\sum_{k=0}^{\infty}\Big|m^k(t)-\hat{m}_q^k(t)\Big|^2\\&\leq 2\sum_{k=0}^{\infty}\Big|m^k(t)-\hat{m}_q^k(t)\Big|\\&\qquad\cdot\Big|\bar{f}^k(t,m(t),u(t))-\hat{f}_q^k(t,\hat{m}_q(t),u(t))\Big|\\&\leq 2\|m(t)-\hat{m}_q(t)\|\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\\&\leq 2\|m(t)-\hat{m}_q(t)\|\Big(\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|\\&\qquad+\|\hat{f}_q(t,m(t),u(t))-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\Big).\end{aligned}$$

Using $\frac{d}{dt}\|m(t)-\hat{m}_q(t)\|^2=2\|m(t)-\hat{m}_q(t)\|\frac{d}{dt}\|m(t)-\hat{m}_q(t)\|$, we obtain

$$\begin{aligned}&\frac{d}{dt}\|m(t)-\hat{m}_q(t)\|\leq\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|\\&\qquad+\|\hat{f}_q(t,m(t),u(t))-\hat{f}_q(t,\hat{m}_q(t),u(t))\|\\&\leq\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|+L\|m(t)-\hat{m}_q(t)\|\end{aligned}$$

where we use the Lipschitz continuity of $\hat{f}_q$. Because of the convergence of $\hat{f}_q$ to $\bar{f}$, for any $\varepsilon>0$, we can choose a large enough q such that $\|\bar{f}(t,m(t),u(t))-\hat{f}_q(t,m(t),u(t))\|<\varepsilon$, which gives

$$\frac{d}{dt}\|m(t)-\hat{m}_q(t)\|\leq L\|m(t)-\hat{m}_q(t)\|+\varepsilon.$$

By Grönwall's inequality [43], the above estimate yields an upper bound for the cumulative truncation error, given by,

$$\begin{aligned}\|m(t)-\hat{m}_q(t)\|&\leq e^{Lt}\Big(\|m(0)-\hat{m}_q(0)\|+\varepsilon t\Big)\\&\leq e^{LT}\Big(\|m(0)-\hat{m}_q(0)\|+\varepsilon T\Big)\end{aligned}$$

for $t\in[0,T]$. Because $\hat{m}_q(0)$ is chosen to be the order-q truncation of $m(0)$, $\|m(0)-\hat{m}_q(0)\|\to 0$ holds for $q\to\infty$.

Then, letting $\varepsilon \to 0$, we obtain $\sup_{t\in[0,T]} \|m(t) - \hat{m}_q(t)\| \to 0$ as $q \to \infty$ as desired. $\square$

References

[1] J. Ho, X. Chen, A. Srinivas, Y. Duan, and P. Abbeel, "Flow++: Improving flow-based generative models with variational dequantization and architecture design," in *International conference on machine learning*. PMLR, 2019, pp. 2722–2730.

[2] S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, "Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7327–7347, 2022.

[3] J. Tomczak, *Deep Generative Modeling*, ser. Computer Science. Springer International Publishing, 2024.

[4] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 8780–8794.

[5] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, "Photorealistic text-to-image diffusion models with deep language understanding," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 36 479–36 494.

[6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 10 684–10 695.

[7] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "SDXL: Improving latent diffusion models for high-resolution image synthesis," in *The Twelfth International Conference on Learning Representations*, 2024.

[8] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *The Eleventh International Conference on Learning Representations*, 2023.

[9] Y. Lipman, M. Havasi, P. Holderrieth, N. Shaul, M. Le, B. Karrer, R. T. Q. Chen, D. Lopez-Paz, H. Ben-Hamu, and I. Gat, "Flow matching guide and code," 2024.

[10] A. Pokle, M. J. Muckley, R. T. Q. Chen, and B. Karrer, "Training-free linear image inverses via flows," *Transactions on Machine Learning Research*, 2024.

[11] H. Ben-Hamu, O. Puny, I. Gat, B. Karrer, U. Singer, and Y. Lipman, "D-flow: differentiating through flows for controlled generation," in *Proceedings of the 41st International Conference on Machine Learning*, ser. ICML'24. JMLR.org, 2024.

[12] Y. Zhang, P. Yu, Y. Zhu, Y. Chang, F. Gao, Y. N. Wu, and O. Leong, "Flow priors for linear inverse problems via iterative corrupted trajectory matching," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[13] S. T. Martin, A. Gagneux, P. Hagemann, and G. Steidl, "Pnp-flow: Plug-and-play image restoration with flow matching," in *The Thirteenth International Conference on Learning Representations*, 2025.

[14] G. Ohayon, T. Michaeli, and M. Elad, "Posterior-mean rectified flow: Towards minimum MSE photo-realistic image restoration," in *The Thirteenth International Conference on Learning Representations*, 2025.

[15] Y. Mei, M. Al-Jarrah, A. Taghvaei, and Y. Chen, "Flow matching for stochastic linear control systems," in *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, ser. Proceedings of Machine Learning Research, N. Ozay, L. Balzano, D. Panagou, and A. Abate, Eds., vol. 283. PMLR, 04–06 Jun 2025, pp. 484–496.

[16] R. Brockett, "Optimal control of the liouville equation," *Studies in Advanced Mathematics*, vol. 39, pp. 23–35, 01 2007.

[17] ——, *Notes on the Control of the Liouville Equation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 101–129.

[18] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, p. 666–704, 1781.

[19] C. Villani, *Topics in Optimal Transportation*, ser. Graduate studies in mathematics. American Mathematical Society, 2003, vol. 58.

[20] ——, *Optimal Transport: Old and New*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 2009, vol. 338.

[21] L. Ambrosio, N. Gigli, and G. Savare, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, ser. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2008.

[22] J.-D. Benamou and Y. Brenier, "A numerical method for the optimal time-continuous mass transport problem and related problems," in *Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)*, ser. Contemp. Math. Amer. Math. Soc., Providence, RI, 1999, vol. 226, pp. 1–11.

[23] P. Billingsley, *Probability and Measure*, ser. Wiley Series in Probability and Statistics. Wiley, 1995.

[24] K. Yoshida, *Functional Analysis*, ser. Classics in mathematics / Springer. World Publishing Company, 1980.

[25] P. Lax, *Functional Analysis*, ser. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2002.

[26] G. W. Mackey, "Harmonic analysis as the exploitation of symmetry - a historical survey," *Bulletin (New Series) of the American Mathematical Society.*, vol. 3, no. 1, pp. 543–698, 1980.

[27] F. Hausdorff, "Momentprobleme für ein endliches intervall." *Mathematische Zeitschrift*, vol. 16, no. 1, pp. 220–248, 1923.

[28] H. L. Hamburger, "Über eine erweiterung des stieltjesschen momentenproblems," *Mathematische Annalen*, vol. 82, pp. 120–164, 1920.

[29] ——, "Über eine erweiterung des stieltjesschen momentenproblems," *Mathematische Annalen*, vol. 82, pp. 168–187, 1921.

[30] ——, "Über eine erweiterung des stieltjesschen momentenproblems," *Mathematische Annalen*, vol. 81, pp. 235–319, 1921.

[31] T. J. Stieltjes, *Œuvres Complètes II - Collected Papers II*, ser. Springer Collected Works in Mathematics, G. van Dijk, Ed. Springer Berlin, Heidelberg, 1993.

[32] A. Becker, G. Habibi, J. Werfel, M. Rubenstein, and J. McLurkin, "Massive uniform manipulation: Controlling large populations of simple robots with a common input signal," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 520–527.

[33] W. Zhang and J.-S. Li, "Uniform and selective excitations of spin ensembles with rf inhomogeneity," in *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 5766–5771.

[34] J.-S. Li, I. Dasanayake, and J. Ruths, "Control and synchronization of neuron ensembles," *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 1919–1930, 2013.

[35] V. Arnold and R. Silverman, *Ordinary Differential Equations*. MIT Press, 1978.

[36] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, ser. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 1999.

[37] V. I. Paulsen and M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016, vol. 152.

[38] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. USA: IEEE Computer Society, 2015, p. 3676–3684.

[39] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8185–8194.

[40] T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1845–1866, 2018.

[41] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006.

[42] C. Canuto, M. Hussaini, A. Quarteroni, and T. Zang, *Spectral Methods: Evolution to Complex Geometries and Applications to Fluid Dynamics*, ser. Scientific Computation. Springer Berlin Heidelberg, 2007.

[43] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, ser. Modern Birkhäuser Classics. Birkhäuser Boston, 2009.

[44] T. Apostol, *Mathematical Analysis*, ser. Addison-Wesley series in mathematics. Addison-Wesley, 1974.