Invertible Convolutional Networks

Abstract:

This paper addresses recent advances in developing invertible neural network architectures, with a specific focus on convolutional layers. The authors propose a method for making standard convolutional layers invertible by leveraging the Fourier transform to allow exact inversion, as well as providing a tractable method for calculating log determinants using matrix operations over real numbers. Additionally, they introduce bijective activation functions and downsampling techniques to make non-residual convolutional networks invertible without significantly affecting their classification performance.

By applying these adjustments, the authors demonstrate that traditional CNN architectures can be adapted for use in normalizing flow models, enhancing their potential for generative modeling alongside discriminative tasks. This approach maintains the strengths of CNNs for spatial and translational tasks while introducing invertibility, which is beneficial for feature visualization, low-memory training, and expressive normalizing flows.

Why Normalizing Flows Fail to Detect Out-of-Distribution Data:

Introduction:

This paper examines why normalizing flows, despite their potential as generative models, often struggle with out-of-distribution (OOD) detection. Specifically, flows tend to assign higher likelihoods to OOD data than to in-distribution data, such as a flow model trained on clothing images assigning higher likelihoods to handwritten digits. The authors attribute this issue to normalizing flows learning local pixel correlations and general image-to-latent-space mappings that are not specific to the target dataset's semantic content.

Focusing on flows with coupling layers, they show that these models often transform in-distribution and OOD images similarly, due to inductive biases in their architectures. This tendency undermines OOD detection by prioritizing local correlations over the data's semantic structure. The authors propose architectural modifications to the coupling layers to bias flows toward learning representations more specific to the target data, thereby enhancing OOD detection. Their findings suggest that the same properties allowing flows to generate high-fidelity images can impede their effectiveness in detecting OOD samples.

Conclusion:

Inductive biases in normalizing flows prioritize local pixel features over semantic structure, limiting their effectiveness in out-of-distribution (OOD) detection. While flows can generate samples resembling training data, they struggle with capturing specific semantics. Further research into their sampling processes is needed to address potential overfitting and improve OOD performance.

Reversible Vision Transformers:

Introduction:

This paper presents Reversible Vision Transformers (Rev-ViT), a memory-efficient adaptation of Vision Transformer (ViT) and Multiscale Vision Transformers (MViT) architectures. By decoupling memory footprint from model depth, Rev-ViTs allow for deeper scaling without proportional increases in GPU memory use, achieved through on-the-fly recomputation of activations rather than caching. This design is particularly useful in memory-limited settings like video recognition, where models are often constrained to small batch sizes due to memory demands.

The authors address training instabilities in deep reversible models by reconfiguring residual connections and adding lateral paths, resulting in stable, memory-efficient networks. They also introduce lighter data augmentation strategies, leveraging reversible transformers' inherent regularization. Extensive benchmarking across tasks such as image classification, object detection, and video classification shows that Rev-ViTs match non-reversible models' accuracy while significantly reducing memory usage—up to 15.5× in ViT-Large—and achieving 3.9× higher throughput, making them ideal for resource-constrained applications.

Conclusion:

Reversible Vision Transformers (Rev-ViT, Rev-MViT) achieve competitive accuracy with up to 15.5× memory savings and improved throughput, offering scalable, memory-efficient solutions for deep visual recognition tasks

Semi-Supervised Learning with Normalizing Flows

Introduction:

FlowGMM is a generative semi-supervised learning model that leverages normalizing flows to create a flexible, interpretable approach for handling both labeled and unlabeled data. The model integrates a latent Gaussian mixture structure with normalizing flows to jointly model data density, achieving exact likelihoods for both data types. FlowGMM uses the advantages of normalizing flows—such as invertibility, controlled latent representations, and efficient sampling—to unify the treatment of labeled and unlabeled data while providing high-quality, interpretable structure. Its applications extend beyond image data to text and tabular formats, showing effectiveness on tasks like AG-News and Yahoo Answers classification, tabular data analysis, and semi-supervised image classification. FlowGMM also offers real-time, optimization-free feature visualizations, interpretable latent space structures, and well-calibrated predictions. This approach advances the potential of normalizing flows in semi-supervised settings, offering both predictive accuracy and valuable interpretability in diverse data domains.

Conclusion:

FlowGMM offers a simple, interpretable approach for semi-supervised learning using normalizing flows, showing promise especially in tabular data and text classification. While it lags behind top discriminative models for image tasks, FlowGMM's interpretability, latent space access, and sampling potential make it valuable for further development in generative semi-supervised learning.

ButterflyFlow: Building Invertible Layers with Butterfly Matrices:

Introduction:

ButterflyFlow introduces a new class of normalizing flow models utilizing invertible butterfly layers, which capture complex structures like permutations and periodicity, commonly found in real-world data. These butterfly layers are efficient and maintain exact likelihood estimation, making them ideal for applications with structured data. ButterflyFlow combines these layers with traditional coupling layers and Glow-like architecture, enhancing representational power and computational efficiency. This model excels not only in standard generative tasks on image datasets (such as MNIST, CIFAR-10, and ImageNet-32×32) but also shows significant improvements on structured datasets, notably outperforming baselines by around 200% on the MIMIC-III patient dataset. ButterflyFlow's capability to model intricate data patterns and reduce memory usage makes it an effective choice for a range of generative tasks, particularly those involving structured datasets in domains like healthcare and scientific imaging.

Conclusion:

ButterflyFlow introduces a new generative model using invertible butterfly layers, enhancing flow-based architecture for density estimation. It excels in handling structured real-world data while maintaining strong performance on standard image datasets. Future work includes automating input partitioning and exploring broader applications beyond density estimation.

The Convolution Exponential and Generalized Sylvester Flows

Introduction:

This paper presents a novel method for constructing linear flows by taking the exponential of a linear transformation, which does not need to be invertible. This approach guarantees invertibility, enables straightforward inverse computation, and simplifies the log Jacobian determinant calculation to the trace of the transformation. The authors introduce new invertible transformations, including convolution exponentials and graph convolution exponentials, that preserve equivariance. They also propose Convolutional Sylvester Flows, generalizing Sylvester Flows and utilizing convolution exponentials as a basis change. Empirical results demonstrate that convolution exponentials outperform existing linear transformations on CIFAR10, while graph convolution exponentials enhance graph normalizing flows. Additionally, Convolutional Sylvester Flows show improved performance over residual flows in generative modeling, as measured by log-likelihood.

Conclusion:

In this paper we introduced a new simple method to construct invertible transformations, by taking the exponential of any linear transformation. Unlike prior work, we observe that the exponential can be computed implicitly. Using this we developed new invertible transformations named convolution exponentials and graph convolution exponentials, and showed that they retain their equivariance properties under exponentiation. In addition, we generalize Sylvester Flows and propose Convolutional Sylvester Flows.

Task-agnostic Continual Learning with Hybrid Probabilistic Models

Introduction:

This paper introduces HCL, a Hybrid generative-discriminative approach for Continual Learning (CL) in classification tasks. HCL utilizes normalizing flows to model the distribution of each task and class, enabling effective data distribution learning, classification, task change detection, and mitigation of catastrophic forgetting through generative replay and functional regularization techniques. The model employs state-of-the-art anomaly detection methods to identify task changes by assessing the typicality of the model's statistics. HCL demonstrates strong performance across various continual learning benchmarks, including split-MNIST, split-CIFAR, and SVHN-MNIST. The study addresses the challenge of domain-incremental continual learning, where the input distribution evolves while maintaining the same target space, all without knowledge of task boundaries, which is a significant hurdle in existing methods.

Conclusion:

HCL is a hybrid model for continual learning utilizing normalizing flows, demonstrating strong performance in image classification and automatic detection of tasks through flow statistics. Its simplicity and extensibility enable adaptations for class-incremental learning, multi-task scenarios, and semi-supervised continual learning, facilitating easy derivation of training objectives.

Woodbury Transformations for Deep Generative Flows

Introduction:

This paper introduces Woodbury transformations as a novel approach for deep generative models using normalizing flows. These transformations leverage the Woodbury matrix identity for efficient invertibility and Sylvester's determinant identity for effective likelihood calculations. Unlike existing flow operations that often compromise between interaction, sampling efficiency, and likelihood evaluation, Woodbury transformations achieve all three, enabling high-dimensional interactions while maintaining computational efficiency. Traditional methods, such as affine coupling layers and 1×1 convolutions, fall short in capturing complex dependencies among dimensions. In contrast, Woodbury transformations model dependencies along both spatial and channel axes, ensuring that both training and sampling time complexities remain linear relative to the input size. Additionally, a memory-efficient variant is introduced, providing similar advantages with reduced memory requirements for high-dimensional variables. Empirical evaluations demonstrate that models employing Woodbury transformations can achieve high likelihoods on various image datasets while preserving efficiency, positioning them as a strong alternative to current flow architectures.

Conclusion:

Woodbury transformations efficiently compute inverses and Jacobian determinants, capturing correlations among dimensions with linear complexity. They outperform state-of-the-art methods on image datasets but lack parameter sharing like convolutional layers.

MixerFlow: MLP-Mixer meets Normalising Flows

Introduction:

This paper introduces MixerFlow, a novel normalizing flow architecture based on the MLP-Mixer design, which aims to unify generative and discriminative modeling. Unlike the predominant Glow-based architectures, MixerFlow efficiently implements weight sharing, demonstrating competitive or superior density estimation on image datasets. It scales well with increasing image resolutions and provides more informative embeddings. MixerFlow also allows integration of structured transformations like splines or Kolmogorov-Arnold Networks, enhancing its expressiveness without the excessive parameter count typical of Glow models. The architecture emphasizes the dual purpose of normalizing flows as both density estimators and generative models.

Conclusion:

MixerFlow, inspired by the MLP-Mixer architecture, consistently outperforms existing models in negative log-likelihood on standard datasets. It demonstrates strong scalability for larger image sizes and effectively integrates MAF layers, enhancing its adaptability and versatility beyond traditional coupling layers in normalizing flow architectures.

ORTHONORMAL CONVOLUTIONS FOR THE ROTATION BASED ITERATIVE GAUSSIANIZATION

Introduction:

This paper presents Convolutional RBIG, an extension of the rotation-based iterative Gaussianization (RBIG) method designed to handle high-dimensional data, such as images. Traditional RBIG relies on principal or independent component analysis for rotation operations, limiting its application to mediumdimensional data and small image patches. Convolutional RBIG addresses this by using convolutional rotations (orthonormal convolutions), which are learned by optimizing reconstruction loss through transposed convolution. The authors introduce various regularizers, including one that promotes sparsity in activations, which extends convolutional independent component analysis to multilayer architectures. Additionally, the method facilitates the extraction of statistical properties, such as multivariate mutual information. The paper includes a simple texture synthesis example to illustrate the transform's behavior and visualizes stimuli that maximize responses in specific features and layers. Overall, Convolutional RBIG provides a more scalable approach for density estimation and information theory applications in image processing.

Conclusion:

This work introduces orthonormal convolutions within the RBIG framework, enhancing its ability to efficiently process images while maintaining advantages like easy training, invertibility, and the computation of statistical properties.

Parallel Backpropagation for Inverse of a Convolution with Application to Normalizing Flows

Introduction:

This paper presents a fast parallel backpropagation algorithm for computing the inverse of convolutions, crucial for applications like Normalizing Flows and image deblurring. The traditional Gaussian elimination approach has a computational complexity of O(n3)O(n^3)O(n3), but the proposed algorithm reduces this to O(n)O(\sqrt{n})O(n) for square images, significantly improving efficiency. By implementing inverse convolutions in the forward pass of Normalizing Flows and using standard convolutions for the sampling pass, the InverseFlow models achieve faster sampling times without sacrificing quality. This approach utilizes GPU optimization to enhance performance, allowing for rapid, high-quality sample generation necessary for large-scale data tasks. Experimental results demonstrate that InverseFlow models maintain competitive bits per dimension while significantly improving sampling speed compared to previous architectures, thereby addressing the efficiency challenges typically associated with convolutions in deep learning frameworks.

Conclusion:

This paper introduces Inverse-Flow, a flow-based model featuring a fast backpropagation algorithm for inverse convolutions, enhancing learning and sampling efficiency. It significantly reduces sampling time, advancing generative modeling and expanding the practical applicability of flow-based approaches.

By Tying Embeddings You Are Assuming the Distributional Hypothesis:

Introduction:

This work explores the impact of tied input-output embeddings, a technique that reduces model size and can enhance training. Our analysis reveals a connection to Harris's distributional hypothesis, suggesting that semantically similar words share input embeddings, while words in similar contexts share output embeddings. This organization of embeddings in foundational language models reflects semantic relationships, supporting the idea that tying embeddings is beneficial when the distributional hypothesis holds true. We introduce the concept of semantic equivalence for symbols, akin to programming languages, to formalize why these embeddings form a semantic space. Our findings indicate that optimal input and output embeddings convey the same semantic information, elucidating the effectiveness of weight tying in natural language processing. We complement our theoretical insights with experiments that validate these claims, shedding light on the structure of embeddings in foundation models.

Conclusion:

This work analyzed weight tying in embeddings, linking it to the distributional hypothesis. We found semantically equivalent symbols share input embeddings, suggesting weight tying should be considered under specific conditions. Future research may explore alternative interpretations.

Invertible Convolution with Symmetric Paddings:

Introduction:

Convolutional Neural Networks (CNNs) have excelled in computer vision by transforming complex image data distributions into manageable latent feature domains for tasks like classification and generation. Recent research has identified invertible CNN operators, enabling bidirectional transformations between image and latent domains. This paper focuses on the invertibility of symmetric padded convolution, a less-explored area compared to conventional circular convolution, which can confuse training due to its violation of spatial locality. We demonstrate that multiple invertible convolution operators can be formed using various symmetric padding combinations. Our findings allow for the conversion of conventional CNNs into their invertible counterparts with minimal structural modifications. Additionally, we present efficient methods for analytical inversion using the Discrete Fourier Transform (DFT), enhancing the understanding of CNN processes and offering a unified approach for generative and discriminative tasks.

Conclusion:

We comprehensively analyze several different symmetric and anti-symmetric padding modes for the convolution operation and show that multiple cases exist where the convolution can be inverted. We consider this derivation as a contribution to the research of invertible networks and generative models.

The Convolution Exponential:

Introduction:

This paper presents a novel approach for constructing linear flows using the exponential of a linear transformation. Notably, this linear transformation does not need to be invertible, and the resulting exponential has several advantageous properties: it is guaranteed to be invertible, its inverse can be easily computed, and the log Jacobian determinant equals the trace of the linear transformation. A key insight is that the exponential can be computed implicitly, which facilitates the integration of convolutional layers into the flow. This leads to the development of new invertible transformations known as convolution exponentials and graph convolution exponentials, both of which preserve the equivariance of their underlying transformations. Empirical results demonstrate that convolution exponentials significantly outperform other linear transformations in generative flows on the CIFAR10 dataset, while graph convolution exponentials enhance the performance of graph normalizing flows. This work opens avenues for more efficient and effective generative modeling by leveraging the mathematical properties of matrix exponentials within the context of deep learning.

Conclusion:

In this paper we introduced a new simple method to construct invertible transformations, by taking the exponential of any linear transformation. Unlike prior work, we observe that the exponential can be computed implicitly. Using this we developed new invertible transformations named convolution exponentials and graph convolution exponentials, and showed that they retain their equivariance properties under exponentiation.

Inverting Toeplitz matrices for Convolutional Neural Networks:

Introduction:

This thesis explores the inversion of convolutional neural networks (CNNs) by investigating a method to translate convolution into matrix multiplication through special matrices related to tridiagonal Toeplitz structures. The primary focus is on developing efficient algorithms to compute the inverse of these matrices, which is essential for understanding the reversible nature of CNN operations. The work begins with decomposing the matrices to facilitate inversion and deriving the determinant of the tridiagonal Toeplitz matrix, as this determinant is fundamental to the expression of the matrices involved. The theoretical findings are categorized into two groups: the first deals with expressions for matrix inverses using linear algebra, while the second focuses on recursive coefficients linked to the determinant and strives to derive a clear formula for it using complex analysis. The effectiveness of these theoretical ideas is validated through Python implementations, revealing numeric errors in practical applications. The thesis also discusses experimental results related to matrix inversion in CNNs and highlights the limitations encountered. This work contributes to the understanding of CNN interpretability and reversibility, laying the groundwork for future research in this area.

Conclusion:

This thesis successfully explores the mathematical foundations of inverting convolutional neural networks, revealing significant theoretical results and potential improvements. Future work should focus on controlling determinants and inverses, addressing numerical errors, and refining implementations for better experimental outcomes.

i-REVNET: DEEP INVERTIBLE NETWORKS

Abstract:

This paper challenges the prevailing notion that successful deep convolutional networks achieve performance by discarding uninformative variability from inputs. We present the i-RevNet, an invertible architecture constructed from homeomorphic layers that retains all information up to the final classification stage. This design allows for effective learning without the necessary loss of information commonly seen in traditional architectures. Our findings reveal that while previous research suggested that reducing mutual information is crucial for generalization, our invertible approach demonstrates that information retention can occur without sacrificing discriminability.

The i-RevNet employs a progressive contraction strategy, effectively managing variability through one-to-one mappings in intermediate representations. This architecture performs comparably to existing non-invertible networks like RevNet and ResNet on the ImageNet dataset, validating the potential for invertible networks in practical applications. Our analysis of learned representations shows that the i-RevNet facilitates a gradual separation of signal classes, suggesting that effective learning arises from both contraction and the preservation of input information. Thus, our work not only provides a new perspective on the role of information in deep learning but also highlights a viable path for constructing networks that enhance interpretability and recoverability in image classification tasks.

Conclusion:

This work provides the first empirical evidence that invertible representations in deep learning can maintain input information while achieving high classification accuracy on large-scale datasets. The i-RevNet architecture demonstrates that effective generalization can occur without discarding information, relying instead on progressive separation and contraction of feature representations.

A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT:

Introduction:

In recent years, Artificial Intelligence Generated Content (AIGC) has garnered widespread attention beyond the computer science community, capturing societal interest in various content generation products from major tech companies like OpenAI's ChatGPT and DALL-E 2. AIGC refers to content produced using advanced Generative AI (GAI) techniques, allowing for rapid automation in content creation. ChatGPT, for example, is a language model designed for conversational AI, adept at understanding and responding to human language. DALL-E 2, another OpenAI innovation, generates unique, high-quality images from textual descriptions, exemplifying the capabilities of AIGC. The rapid advancements in AIGC signal the dawn of a new AI era with the potential for significant global impact.

Technically, AIGC involves two main steps: extracting intent information from human instructions and generating content that aligns with these intentions. While the dual-step paradigm is not entirely new, recent improvements stem from training more sophisticated generative models on larger datasets, utilizing more substantial foundational architectures, and leveraging increased computational resources. For instance, GPT-3 has expanded its training data and model size significantly compared to GPT-2, enhancing its generalization capabilities. Researchers are also integrating new technologies with GAI algorithms, such as reinforcement learning from human feedback in ChatGPT, to further improve reliability and accuracy.

Conclusion:

This survey offers a detailed overview of the history and recent advancements in Artificial Intelligence Generated Content (AIGC), focusing on both unimodal and multimodal generative models. It discusses applications, techniques, and concerns about trustworthiness and responsibility within the field. The survey aims to provide readers with a thorough understanding of current developments and future challenges in generative AI, distinguishing contemporary models from earlier ones. Ultimately, it serves as a foundation for further investigation into AIGC, fostering deeper insights and potential innovations.

Normalizing Flows for Probabilistic Modeling and Inference:

Introduction:

Normalizing flows offer a powerful framework for constructing complex probability distributions by applying a series of bijective transformations to a simple base distribution. This review aims to unify the understanding of normalizing flows within the context of probabilistic modeling and inference. By emphasizing core principles such as expressive power and computational trade-offs, it provides insights into the design and functionality of flows. The authors explore the flexibility of normalizing flows, demonstrating how even basic transformations can lead to highly complex distributions, making them suitable for various statistical tasks, including modeling, inference, and simulation.

This paper builds upon previous works, such as Papamakarios (2019) and Kobyzev et al. (2020), by providing a more comprehensive tutorial that covers foundational topics and addresses ongoing challenges, such as extending flows to discrete variables and Riemannian manifolds. The review serves as a valuable resource for researchers seeking to understand the latest developments in normalizing flows and their applications in generative modeling, approximate inference, and supervised learning. Through this holistic approach, the authors aim to clarify the relationships between recent advancements and the foundational principles of normalizing flows, promoting further exploration in this dynamic area of machine learning.

Conclusion:

This review outlines the principles and applications of normalizing flows in probabilistic modeling and inference, highlighting their expressive power and construction. It emphasizes key concepts, trade-offs, and implementation considerations, aiming to guide practitioners in effectively utilizing flows for various tasks in the evolving landscape of machine learning.

CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion:

Introduction:

Multi-modality (MM) image fusion is an essential field in image processing that aims to create informative fused images by combining critical information from various sources, such as infrared and visible images. This fusion enhances the visibility of functional highlights while preserving detailed textures, addressing the limitations of individual modalities. For instance, while visible images can lose detail in low-light conditions, infrared images may exhibit noise and lower resolution. Consequently, MM image fusion techniques are particularly valuable in applications like infrared-visible image fusion (IVF) and medical image fusion (MIF), where integrating diverse imaging modalities can significantly aid in diagnosis and analysis.

Existing approaches to multi-modality image fusion often rely on convolutional neural networks (CNNs) in an auto-encoder framework. However, these methods face significant challenges, including the difficulty in modeling cross-modality features and the risk of losing high-frequency information during forward propagation. This paper proposes a novel Correlation-Driven feature Decomposition Fusion (CDDFuse) network to tackle these challenges. The CDDFuse architecture utilizes Restormer blocks for extracting shallow cross-modality features and incorporates a dual-branch feature extractor, combining Lite Transformer (LT) blocks for global feature extraction and Invertible Neural Network (INN) blocks for capturing local features. This design not only enhances feature extraction but also improves interpretability and control over the fusion process. Through extensive experimentation, we demonstrate that CDDFuse outperforms existing methods in various fusion tasks, making it a promising advancement in the field of image fusion.

Conclusions:

This work introduces the Correlation-Driven feature Decomposition Fusion (CDDFuse) network, addressing challenges in multi-modality image fusion. By leveraging Transformer-CNN architectures and Invertible Neural Networks, CDDFuse effectively extracts and fuses cross-modality features. Extensive experiments confirm its effectiveness across various fusion tasks, enhancing downstream performance.

Normalizing Flows: An Introduction and Review of Current Methods:

Introduction:

A primary objective in statistics and machine learning is to model probability distributions based on samples drawn from those distributions, a process known as generative modeling. This task is particularly significant due to the abundance of unlabeled data compared to labeled data, making it vital for applications such as density estimation, outlier detection, prior construction, and dataset summarization.

Various methods for generative modeling exist, ranging from direct analytic approaches that approximate data with fixed distribution families to variational methods that introduce latent variables for added flexibility, albeit at the cost of increased complexity in learning and inference. Graphical models explicitly represent the conditional dependencies among random variables, while recent advancements include generative neural approaches like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which have shown remarkable performance in tasks like learning natural image distributions. However, these models face challenges, including the inability to evaluate the exact probability density of new points and difficulties in training.

Normalizing Flows (NF) emerge as a family of generative models that offer tractable distributions, enabling efficient and exact sampling and density evaluation. This article provides a comprehensive review of Normalizing Flows, covering their fundamentals, current literature, and identifying open questions and future research directions.

Conclusion:

Normalizing Flows represent a powerful approach to generative modeling, offering exact sampling and density evaluation. This review highlights their significance in various applications and emphasizes the need for ongoing research to address challenges and explore future directions, ultimately enhancing the field of distribution learning in machine learning.

Hamiltonian Neural Networks:

Introduction:

Neural networks excel in various tasks but often struggle to learn fundamental physical laws. This paper explores the potential of endowing neural networks with better inductive biases by drawing inspiration from Hamiltonian mechanics, which emphasizes conservation laws. We introduce Hamiltonian Neural Networks (HNNs) designed to learn and maintain exact conservation laws, specifically focusing on energy conservation in systems such as the two-body problem and pendulum observations.

Our approach involves parameterizing the Hamiltonian using a neural network, allowing the model to learn the conservation laws directly from data rather than relying on handcrafted equations. The results demonstrate that HNNs train faster and generalize better than traditional neural networks, showcasing their effectiveness in scenarios where conservation of energy is crucial. Additionally, an intriguing outcome of this method is that HNNs exhibit perfect time reversibility, reflecting the underlying physical principles.

By integrating Hamiltonian mechanics into the design of neural networks, we provide a promising framework for enhancing the ability of models to learn and respect physical laws, potentially leading to advancements in physics-informed machine learning and more accurate simulations of physical systems. This work opens avenues for further exploration of physics-based inductive biases in neural networks.

Conclusion:

Hamiltonian Neural Networks (HNNs) bridge the gap between established Hamiltonian mechanics and emerging deep learning techniques. By integrating first-principles physics with data-driven learning, HNNs enhance model accuracy and generalization. This synthesis of approaches offers a promising avenue for advancing our understanding of complex physical systems through machine learning.

Opening the black box of Deep Neural Networks via Information:

Introduction:

Deep Neural Networks (DNNs) have achieved remarkable success across various machine learning tasks, yet a comprehensive understanding of their optimization processes and internal organization remains elusive. This paper builds on the Information Plane framework proposed by Tishby and Zaslavsky, demonstrating that DNN training occurs in two distinct phases: empirical error minimization and representation compression. During the initial phase, significant gradients facilitate rapid increases in mutual information related to labels. In contrast, the compression phase involves stochastic fluctuations dominating the gradients, leading to slow representation adjustments under training error constraints. This phase accounts for the absence of overfitting in DNNs, revealing the existence of many randomized networks with similar performance. Our analysis indicates that optimized layers approach the Information Bottleneck (IB) bound, suggesting a self-consistent relationship between encoders and decoders. Additionally, we show that hidden layers significantly enhance computational efficiency, aligning with the observed diffusive dynamics of Stochastic Gradient Descent (SGD).

Conclusion:

Our findings highlight the Information Bottleneck framework's role in elucidating the inner workings of Deep Learning. By visualizing layers in the information plane, we uncover distinct SGD optimization phases—drift and diffusion—that facilitate efficient internal representation through noise-induced compression, aligning with recent insights into noise's significance in DNNs.

3D Packing for Self-Supervised Monocular Depth Estimation:

Introduction:

Accurate depth estimation is crucial for robotics tasks, yet traditional methods often rely on costly sensors. This paper introduces a self-supervised monocular depth estimation technique using a novel deep network architecture, PackNet, which learns from unlabeled monocular videos. PackNet employs innovative packing and unpacking blocks, utilizing 3D convolutions to create detail-preserving representations. It demonstrates superior performance on the KITTI benchmark compared to existing self-supervised, semi-supervised, and fully supervised methods, especially in longer-range scenarios, and generalizes well on out-of-domain datasets like NuScenes. The model's 3D inductive bias allows it to scale efficiently with input resolution and parameters without overfitting, and it operates in real-time without requiring extensive supervised pretraining on ImageNet. Additionally, the authors present the Dense Depth for Automated Driving (DDAD) dataset, which provides more accurate depth evaluation through high-density LiDAR data from self-driving cars, further enhancing the evaluation of depth estimation techniques.

Conclusion:

We introduce PackNet, a convolutional network for self-supervised monocular depth estimation that excels in compressing and decompressing high-resolution visual data. Trained on unlabeled videos, it outperforms existing methods and achieves real-time performance, demonstrating superior generalization and accuracy, particularly on longer depth ranges in the DDAD dataset.

Invertible Residual Networks:

Introduction:

This paper introduces invertible residual networks (i-ResNets), which adapt standard ResNet architectures to be invertible, allowing the same model to be used for classification, density estimation, and generation. By modifying the normalization scheme during training, i-ResNets maintain competitive performance on tasks like image classification across datasets such as MNIST and CIFAR. The authors present a novel approximation method for calculating the Jacobian log-determinant of residual blocks, facilitating the training of i-ResNets as maximum likelihood generative models using unlabeled data. The results demonstrate that i-ResNets can achieve performance comparable to state-of-the-art image classifiers and flow-based generative models, thus bridging the gap between discriminative and generative learning in a single architecture. This work highlights the potential of i-ResNets to unify various tasks within a common framework, moving closer to the goal of general-purpose neural network architectures.

Conclusion:

We presented i-ResNets, an architecture enabling flexible layer designs while providing accurate density estimates. Despite achieving notable performance, challenges remain, including a biased log-determinant estimator and the complexity of enforcing Lipschitz constraints. Future research could enhance performance through unbiased estimators and improved control over Lipschitz constants.

Video Super-Resolution Based on Deep Learning: A Comprehensive Survey:

Introduction:

This survey presents a comprehensive review of 37 state-of-the-art deep learning-based video super-resolution (VSR) methods, a crucial task for reconstructing high-resolution videos from low-resolution inputs. Recognizing the importance of inter-frame information, we propose a taxonomy categorizing these methods into seven sub-categories based on their utilization of this information. The paper details architectural designs and implementation specifics, along with performance comparisons on benchmark datasets. Additionally, we explore various applications of VSR algorithms and highlight ongoing challenges that researchers face in the field. Notably, this work is the first systematic review of VSR techniques, aiming to advance understanding and development in this area of deep learning. The structure of the paper includes sections on background, taxonomy, method classifications, performance analysis, challenges, and concluding remarks, providing a thorough insight into current VSR methodologies and future directions for research.

Conclusion:

This survey reviewed deep learning approaches for video super-resolution, classifying algorithms into seven categories based on inter-frame information utilization. We summarized key methods, compared performance on benchmark datasets, and highlighted various applications. Additionally, we identified eight open issues to guide future research and enhance VSR algorithm development.

Invertibility of Convolutional Generative Networks from Partial Measurements:

Abstract:

This work addresses the problem of inverting convolutional generative neural networks, aiming to recover input latent codes from network outputs, particularly in image inpainting contexts. While the network inversion problem is typically non-convex and computationally challenging, we provide theoretical proof that for a two-layer convolutional network with ReLU and Gaussian-distributed weights, the latent code can be efficiently recovered using gradient descent. This indicates a one-to-one mapping from the low-dimensional latent space to the high-dimensional image space, even when the network output is partially observed. Furthermore, we demonstrate that this finding extends to deeper networks, various activation functions, and weights trained on real datasets. Understanding the invertibility of these networks not only facilitates applications such as image reconstruction and manipulation but also enhances insights into the mappings within generative models, potentially addressing mode collapse issues common in training generative adversarial networks (GANs).

Conclusion:

This work establishes the invertibility of a two-layer ReLU convolutional generative neural network, even with partial outputs, providing a solution to the mode collapse problem in GAN training. Future research may extend these findings to other activation functions and deeper architectures, including scenarios with noisy measurements.

GAN Inversion: A Survey:

Introduction:

Generative Adversarial Networks (GANs) are deep generative models that use adversarial training to generate new data. Comprising two neural networks—a generator (G) and a discriminator (D)—GANs enable the generator to create fake data that mimics real data while the discriminator learns to distinguish between them. Recent advancements in GANs, such as PGGAN, BigGAN, and StyleGAN, have improved the synthesis of high-quality images by leveraging rich semantic information encoded in latent spaces. However, the challenge remains that manipulating latent codes is typically limited to images generated by the GAN and does not extend to real images. GAN inversion addresses this issue by enabling the conversion of real images into their corresponding latent codes within a pretrained GAN model, allowing for faithful reconstruction and manipulation of real images based on learned attributes. This evolving area integrates GANs with interpretable machine learning, facilitating flexible image editing and enhancing our understanding of deep generative models.

Conclusion:

Deep generative models such as GANs learn the underlying variation factors of the training data through the weak supervision of image generation. Discovering and steering the interpretable latent representations in image generation facilitate a wide range of image editing applications. This paper presents a comprehensive survey of GAN inversion methods with an emphasis on algorithms and applications

Interpreting the Latent Space of GANs for Semantic Face Editing:

Introduction:

Generative Adversarial Networks (GANs) have achieved remarkable success in high-fidelity image synthesis, yet the mechanisms behind their ability to map random latent codes to realistic images remain unclear. This work introduces InterFaceGAN, a framework for semantic face editing that interprets the latent semantics learned by GANs. Our research reveals that the latent codes in well-trained generative models exhibit disentangled representations after linear transformations, allowing for more precise control over facial attributes such as gender, age, expression, and even pose variations. Additionally, we explore the relationships among these semantics, enabling the decoupling of entangled attributes through subspace projection. InterFaceGAN facilitates effective semantic face editing without requiring retraining of the GAN model and extends to real image manipulation when integrated with GAN inversion methods. Our extensive experimental results demonstrate that the synthesis process inherently promotes a controllable representation of facial attributes, paving the way for advanced applications in image editing.

Conclusion:

InterFaceGAN provides a framework for interpreting the semantics in GAN latent spaces, enabling precise control of facial attributes using any fixed GAN model, effectively transforming unconditional GANs into controllable ones. Our extensive experiments also indicate its applicability in real image editing. This work is supported by the Early Career Scheme and SenseTime Collaborative Grant.

In-Domain GAN Inversion for Real Image Editing:

Introduction:

Recent research has revealed that Generative Adversarial Networks (GANs) learn diverse semantics in their latent space during image synthesis. However, applying this latent space manipulation to real images is challenging due to limitations in existing inversion methods, which often focus solely on pixel-wise reconstruction. These methods may not accurately position the inverted code within the original latent space, hindering effective semantic editing. To address this, we introduce an in-domain GAN inversion approach that not only reconstructs the input image faithfully but also generates semantically meaningful inverted codes. This is achieved through a novel domain-guided encoder that projects images into the native latent space of GANs, combined with domain-regularized optimization to fine-tune the encoded latent codes. Our extensive experiments demonstrate that this method not only ensures high-quality image reconstruction but also significantly enhances various image editing tasks, outperforming current state-of-the-art techniques.

Conclusion:

We present an in-domain GAN inversion method that enables effective semantic editing of real images by ensuring that inverted latent codes are semantically meaningful. Our approach combines a domain-guided encoder with domain-regularized optimization, leading to accurate reconstructions and improved image manipulation capabilities, surpassing existing methods.

Image Quality Assessment: Unifying Structure and Texture Similarity:

Introduction:

This paper introduces a novel full-reference image quality assessment (IQA) method that explicitly tolerates texture resampling, addressing the limitations of traditional measures like mean squared error (MSE) and structural similarity index (SSIM) which are sensitive to pixel alignment. Using a convolutional neural network, we create an injective function that transforms images into multi-scale overcomplete representations, allowing the capture of texture appearance through spatial averages of feature maps. Our approach combines "texture similarity" based on these averages with "structure similarity" from the feature maps, optimizing parameters to align with human quality ratings. The proposed measure effectively explains perceptual scores on both conventional and texture-specific databases, demonstrating competitive performance in tasks like texture classification and retrieval. Additionally, our method shows resilience to geometric transformations without requiring specialized training or data augmentation, making it a significant advancement in perceptual image quality assessment.

Conclusion:

We introduced DISTS, a full-reference image quality assessment method that accommodates texture resampling. By integrating structure and texture similarity, DISTS accurately predicts human quality ratings for both textures and natural images, demonstrating robustness to geometric distortions and strong performance in texture classification and retrieval. The approach leverages global mean convolution responses from a pre-trained VGG network, creating a low-dimensional yet rich representation of texture appearance.

Image Processing Using Multi-Code GAN Prior:

Introduction:

Generative Adversarial Networks (GANs) have shown remarkable success in image synthesis, but their application to real image processing tasks poses significant challenges. Traditional methods for inverting a target image back to the latent space often rely on back-propagation or the development of additional encoder models, which typically result in suboptimal reconstructions. In response to this limitation, we introduce a novel approach known as mGANprior, which leverages well-trained GANs as effective priors for various image processing tasks. Our method employs multiple latent codes to generate several feature maps from intermediate layers of the GAN generator, allowing for adaptive channel importance when reconstructing the input image. This strategy of over-parameterizing the latent space leads to a marked improvement in image reconstruction quality, surpassing existing methods. The high-fidelity reconstructions achieved through mGANprior facilitate the application of trained GAN models to a range of real-world tasks, including image colorization, super-resolution, image inpainting, and semantic manipulation. Additionally, we analyze the properties of the learned layer-wise representations within the GAN models, providing insights into the specific knowledge each layer can convey, thereby enhancing the understanding of GANs in practical applications.

Conclussion:

We present mGANprior that employs multiple latent codes for reconstructing real images with a pre-trained GAN model. It enables these GAN models as powerful prior to a variety of image processing tasks. Acknowledgement: This work is supported in part by the Early Career Scheme (ECS) through the Research Grants Council of Hong Kong under Grant No.24206219 and in part by SenseTime Collaborative Grant.

Unsupervised 3D Shape Completion through GAN Inversion:

Introduction:

3D shape completion aims to reconstruct complete geometries from partial shapes, a crucial task in applications such as robotics navigation and scene understanding. Traditional approaches predominantly rely on fully supervised learning with paired partial-complete shape data, which, while effective on in-domain datasets, struggle with generalization to out-of-domain data, including real-world scans and varying forms of partial shapes. This paper introduces ShapeInversion, a novel approach that integrates Generative Adversarial Network (GAN) inversion into the realm of 3D shape completion for the first time. ShapeInversion leverages a pre-trained GAN on complete shapes to search for a latent code that produces a complete shape best matching a given partial input. This method eliminates the need for paired training data, allowing for the incorporation of rich priors from well-trained generative models. Experimental results on the ShapeNet benchmark demonstrate that ShapeInversion outperforms state-of-the-art unsupervised methods and achieves performance comparable to supervised approaches. Furthermore, it exhibits remarkable generalization abilities, providing robust outcomes for various forms of partial inputs and real-world scans. The involvement of the pre-trained GAN also enables additional capabilities, such as generating multiple valid completions for ambiguous inputs and facilitating shape manipulation and interpolation.

Conclusion:

We present ShapeInversion for unsupervised point cloud completion, addressing domain gaps in partial scans using GAN inversion. As the first approach of its kind, it introduces uniform loss and a robust degradation function tailored for point clouds. Leveraging a well-trained GAN enhances generalization for diverse real-world inputs, enabling multiple reasonable completions and shape manipulation. Future work will focus on improving fidelity in multiclass models for cross-category shape completion using conditional GANs. Acknowledgments go to SenseTime and A*STAR for support through collaborative funding.

InOut : Diverse Image Outpainting via GAN Inversion:

Introduction:

Image outpainting aims to extend an input image semantically beyond its existing content, distinguishing it from inpainting, which fills in missing pixels based on neighboring data. This work presents a novel approach by leveraging generative adversarial networks (GANs) to tackle outpainting as an inversion problem. Instead of treating it as a conventional image-to-image translation task, we develop a method that generates micro-patches conditioned on their joint latent codes and positions, facilitating the synthesis of diverse outpainted regions. This strategy enhances visual richness and structure in the output, overcoming limitations of existing methods that often produce repetitive textures. Our formulation also allows for categorical conditioning, providing flexible user controls. Extensive experiments demonstrate that our method outperforms state-of-the-art techniques in terms of visual quality and diversity, as evaluated on datasets like Place365 and Flickr-Scenery, using metrics such as Frechet Inception Distance (FID) and user studies for realism assessment. Ultimately, this approach opens new avenues for applications in content creation and immersive experiences.

Conclusion:

In this work, we tackle the image outpainting task from the GAN inversion perspective. We first train a generator to synthesize micro-patches conditioned on their positions. Based on the trained generator, we propose an inversion process that seeks for multiple latent codes recovering available regions as well as predicting outpainting regions. The proposed framework can generate diverse samples and support categorical specific outpainting, enabling more flexible user controls. Qualitative and quantitative experiments demonstrate the effectiveness of the proposed framework in terms of visual quality and diversity.

Deep generative image priors for semantic face manipulation:

Introduction:

This paper presents a novel framework for semantic face manipulation using generative adversarial networks (GANs). While previous studies primarily focused on high-fidelity image synthesis, our approach leverages the knowledge learned by GANs to control facial attributes more effectively. We adapt latent codes through an attribute prediction model, which is pretrained by inverting synthesized face images back into the GAN latent space. This method explicitly captures the semantics encoded within the latent space, enabling precise editing of attributes such as eyeglasses, smiling, baldness, age, mustache, and gender in high-resolution face images. Our approach effectively unifies face attribute prediction and GANs, enhancing the accuracy of attribute estimation and manipulation. Moreover, we introduce a semantic-aware encoder that faithfully reconstructs input images while ensuring that the inverted codes retain semantic meaning. Extensive experiments demonstrate that our method surpasses state-of-the-art performance in both face attribute prediction and semantic manipulation, highlighting its potential for various high-level computer vision tasks.

Conclusion:

In this study, we propose to leverage the natural image prior learned by a deep generative model for face attribute prediction and semantic face manipulation. We show that the extended W space in StyleGAN is more flexible to be inverted by an encoder network, which provides an effective network initialization for high-level face attribute prediction. Furthermore, we propose an attribute-guided latent space manipulation approach to edit a single or multiple attributes by explicitly considering

Solving inverse problems using conditional invertible neural networks

Abstract:

In this work, we present a novel inverse surrogate model designed to estimate high-dimensional spatially-varying property fields from sparse and noisy observations. This approach addresses the inherent challenges of inverse modeling, particularly in systems described by complex multiscale partial differential equations (PDEs). By employing a combination of invertible and conditional neural networks trained end-to-end with limited data, we create a robust framework that maps observations to unknown input fields. Specifically, our method focuses on a multiphase flow problem where the goal is to recover a heterogeneous log-permeability field from pressure and saturation measurements. The results demonstrate that our two- and three-dimensional models yield diverse sample realizations of the non-Gaussian log-permeability field, with the predictive mean closely approximating the true values. This model effectively leverages the available information, showcasing its potential for practical applications in subsurface modeling and resource management.

Conclusion:

In conclusion, our inverse surrogate model effectively estimates high-dimensional spatially-varying property fields from sparse and noisy observations. By combining invertible and conditional neural networks, we achieve accurate recovery of heterogeneous log-permeability fields in multiphase flow problems, demonstrating robustness and potential for real-world applications.

Conditional Invertible Neural Networks for Medical Imaging:

Introduction:

The task of image reconstruction in computed tomography (CT) and magnetic resonance imaging (MRI) is fundamentally an inverse problem, where the goal is to retrieve the true image from noisy measurements. Traditional approaches often focus on obtaining stable reconstructions, but they typically yield only point estimates, which can overlook critical uncertainties, particularly in ill-posed problems. Recent advancements in deep learning have introduced data-driven methods that aim to address these challenges, including generative models like variational autoencoders (VAEs), generative adversarial networks (GANs), and flow-based models using invertible neural networks. These models offer a promising framework for estimating the entire conditional distribution of images given noisy data, enhancing our understanding of uncertainties inherent in the reconstruction process. This paper explores the application of generative flow-based models specifically for low-dose CT and accelerated MRI tasks, evaluating various architectures of invertible neural networks through extensive ablation studies. Notably, the findings indicate that employing a radial distribution as the base distribution in flow-based models can significantly enhance the quality of reconstructions compared to the standard Gaussian distribution, providing a more robust approach to medical image reconstruction.

Conclusion:

This study demonstrates that generative flow-based models, particularly those using invertible neural networks, can effectively tackle image reconstruction challenges in CT and MRI. By incorporating radial distributions, the models significantly enhance reconstruction quality and provide valuable uncertainty estimations, improving the reliability of medical imaging outcomes.

Normalizing field flows: Solving forward and inverse stochastic differential equations using physics-informed flow models:

Introduction:

This paper introduces the *Normalizing Field Flow* (NFF) model, which uses deep learning to learn random fields from sparse data and tackle stochastic partial differential equations (SDEs). The NFF model establishes a bijective transformation between a Gaussian random field, structured through a Karhunen-Loève expansion, and the target stochastic field. By maximizing the log-likelihood of scattered measurements, the model is trained to handle forward and inverse problems in a unified framework. Unlike traditional methods, NFF does not require fixed sensor locations or prior knowledge of the input randomness, helping alleviate the curse of dimensionality. The approach also includes a physics-informed component, incorporating SDE loss to improve accuracy in data-driven forward, inverse, and mixed problem-solving. Numerical experiments demonstrate NFF's effectiveness for both Gaussian and non-Gaussian fields, highlighting its potential as a versatile tool in uncertainty quantification and data-driven modeling of complex systems.

Conclusion:

This work introduced the *Normalizing Field Flow* (NFF) model for learning random fields from sparse measurements, using a bijective transformation to solve forward, inverse, and mixed stochastic PDEs. Unlike traditional methods, NFF handles variable sensor locations and reduces the curse of dimensionality. Future directions include time-dependent and multi-scale problems.

Predicting Thermal Performance of an Enhanced Geothermal System From Tracer Tests in a Data Assimilation Framework:

Introduction:

This study presents a framework to predict the thermal performance of enhanced geothermal systems (EGS) by interpreting tracer data to characterize complex fracture flow patterns. Due to the challenges in directly observing subsurface fractures and limited geological data, the framework uses data assimilation to optimize information extraction from tracer data and handle uncertainties. It includes three components: (a) principal component analysis (PCA) to reduce model dimensionality; (b) ensemble smoother with multiple data assimilation (ES-MDA) to estimate fracture aperture and flow fields, allowing joint data assimilation for uncertainty quantification; and (c) integration of inverted fracture fields into reservoir models for thermal prediction. Field-scale and mesoscale models validate the framework's ability to capture heterogeneous fracture flows, enhancing EGS optimization and risk management. The results demonstrate its effectiveness in extracting fracture flow details from tracer data for reliable thermal performance prediction.

Conclusion:

This study introduced a framework using tracer data to predict EGS thermal performance by inverting fracture aperture and flow patterns. Tested on synthetic and real cases, it accurately forecasted thermal performance and quantified uncertainty. Joint assimilation of conservative and sorptive tracer data is essential to avoid overestimating fluid-rock interactions.

Diagnosing and Fixing Manifold Overfitting in Deep Generative Models:

Introduction:

This paper examines the limitations of likelihood-based deep generative models, which aim to estimate high-dimensional densities but contradict the *manifold hypothesis*—the idea that data lies on a lower-dimensional manifold within high-dimensional space. The authors prove that maximum-likelihood training leads to "manifold overfitting," where the model captures the manifold's structure but not its true distribution. To address this, they propose a two-step approach: first reducing dimensionality, then performing maximum-likelihood density estimation. This method avoids manifold overfitting and achieves accurate distribution recovery in the nonparametric regime. Additionally, it enables density estimation on manifolds learned by implicit models like GANs, addressing key limitations. The approach unifies various recent methods, theoretically justifying a range of deep generative modeling techniques that align with the manifold hypothesis.

Conclusion:
This paper has highlighted the limitations of likelihood-based deep generative models in the presence of high-dimensional data that lie on lower-dimensional manifolds, as described by the manifold hypothesis. We demonstrated that maximum-likelihood training can lead to *manifold overfitting,* where models learn the structure of the manifold but fail to capture the distribution on it.

Conditional Injective Flows for Bayesian Imaging:

Introduction:

This paper addresses Bayesian modeling for computational imaging, focusing on uncertainty quantification (UQ) in ill-posed inverse problems. In such scenarios, we model the unknown object of interest x and observed measurements y as random vectors with a joint distribution, often considering cases like Gaussian or Poisson noise in forward operators. Standard machine learning approaches often provide single-point estimates, such as the minimum-mean-squared-error (MMSE) estimator, which may be misleading for complex, multimodal problems. For instance, in radio interferometric imaging of astronomical phenomena, multiple solutions may fit the data, making single estimates insufficient. Bayesian UQ offers a solution by capturing the distribution of potential solutions, allowing more reliable interpretations and aiding decision-making in fields like medical imaging, where understanding uncertainty is essential for accurate diagnoses and subsequent actions.

Conclusion:

C-Trumpets, a conditional flow model designed for efficient uncertainty quantification in imaging, offers reduced memory and compute costs while producing accurate posterior samples. Limitations include arbitrary latent space dimensions, slower likelihood estimation, and theoretical gaps in posterior distribution modeling. Future work may address these through architectural adjustments and exploring universality in conditional distribution modeling.

Invert to Learn to Invert

Abstract:

Iterative learning to infer approaches have become popular solvers for inverse problems. However, their memory requirements during training grow linearly with model depth, limiting in practice model expressiveness. In this work, we propose an iterative inverse model with constant memory that relies on invertible networks to avoid storing intermediate activations. As a result, the proposed approach allows us to train models with 400 layers on 3D volumes in an MRI image reconstruction task. In experiments on a public data set, we demonstrate that these deeper, and thus more expressive, networks perform state-of-the-art image reconstruction.

Conclusion:

Our approach reduces memory usage for iterative inverse models, enabling deep training on large-scale imaging tasks. The invertible layer stabilizes deep model training and supports structured prediction, improving image reconstruction quality. Future work will explore unsupervised or semi-supervised training to further enhance model versatility.

A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis:

Introduction:

In recent years, deep neural networks (DNNs) have gained prominence across various fields, particularly in medical image analysis, where convolutional neural networks (CNNs) process imaging data such as

MRI, X-rays, and CT scans for disease diagnosis. While DNNs have the potential to significantly enhance healthcare systems, they are also susceptible to adversarial attacks—small, imperceptible perturbations in the input data that can lead to incorrect model predictions. This paper reviews existing methods for adversarial attacks, detection, and defense strategies in the context of medical imaging. We highlight how these attacks can degrade model performance without being detectable by human observers, raising concerns about the safety and reliability of DNNs in critical applications. Despite the vulnerabilities, certain effective defense mechanisms can mitigate risks to some extent. The paper concludes with a discussion on the current state of the field and future challenges in ensuring the robustness of deep learning models in medical image analysis.

Conclusion:

Deep learning has transformed medical image analysis, yet adversarial attacks pose significant risks to the efficacy of these models, endangering patient safety. This paper reviews existing adversarial attacks and defenses specific to medical imaging. We highlight the need for the research community to address these vulnerabilities to facilitate the safe integration of deep learning in real-world medical applications. Future work will involve comparing attacks on a common database, examining various imaging modalities for robustness, and assessing deep learning models specifically designed for medical images to develop safer solutions in this critical field.

End-to-End Variational Networks for Accelerated MRI Reconstruction:

Introduction:

Magnetic Resonance Imaging (MRI) is often hampered by slow acquisition speeds, necessitating advancements in techniques such as Parallel Imaging (PI) and Compressed Sensing (CS) to enable faster scans. This paper introduces a novel end-to-end learning approach for reconstructing images from undersampled multi-coil MRI data, building on existing variational methods. Our method circumvents the unrealistic assumption of a completely known forward process, enhancing reconstruction fidelity. We extend the variational network model to learn the forward process alongside reconstruction, optimizing intermediate representations and neural network architectures for improved outcomes. Extensive experiments on the fastMRI dataset demonstrate that our approach achieves state-of-the-art results for both brain and knee MRI reconstructions, paving the way for more efficient MRI scanning techniques.

Conclusion:

In this paper, we introduced End-to-End Variational Networks for multi-coil MRI reconstruction. While MRI reconstruction can be posed as an inverse problem, multi-coil MRI reconstruction is particularly challenging because the forward process (which is determined by the sensitivity maps) is not completely known. We alleviate this problem by estimating the sensitivity maps within the network, and learning fully end-to-end. Further, we explored the architecture space to identify the best neural network layers and intermediate representation for this problem, which allowed our model to obtain new state-of-the art results on both brain and knee MRIs.

Reversible Vision Transformers:

Introduction:

We introduce Reversible Vision Transformers, a memory-efficient architecture for visual recognition that significantly reduces the GPU memory footprint without compromising model complexity or accuracy. By decoupling memory usage from model depth, our architecture allows for the scalable training of transformers. We adapt popular models like the Vision Transformer (ViT) and Multiscale Vision Transformer (MViT) into reversible forms, achieving memory savings of up to 15.5× while maintaining performance across tasks such as image classification, object detection, and video classification. Our results show that the computational overhead of recomputing activations is outweighed by the efficiency gains, particularly in deeper models where throughput can increase by up to 3.9×. We also identify that reversible transformers exhibit stronger inherent regularization, leading to refined training recipes that enhance their performance. This work positions Reversible Vision Transformers as a robust backbone for resource-constrained environments in deep learning applications.

Conclusion:

We introduce Reversible Vision Transformers, memory-efficient adaptations of ViT and MViT models that excel in various tasks, including image classification, object detection, and video classification. Our Rev-ViT and Rev-MViT models achieve accuracy comparable to non-reversible variants while significantly reducing memory usage—up to 15.5× lighter for ViT and 4.5× lighter for MViT—without sacrificing training throughput. Future work will focus on leveraging these reversible models to develop deeper, memory-efficient visual recognition architectures.

Accelerated MRI With Un-Trained Neural Networks:

Introduction:

I. INTRODUCTION

Convolutional neural networks (CNNs) have emerged as powerful tools for various image reconstruction tasks, often surpassing traditional methods in applications such as image denoising, compressive sensing, and image compression. Typically, CNNs rely on extensive training datasets to learn and generalize effectively. However, the Deep Image Prior (DIP) introduced a novel perspective by demonstrating that CNN architecture can serve as a strong prior for image reconstruction, even in the absence of training data. Untrained networks have shown promise in denoising, compressive sensing, phase retrieval, and video reconstruction, yet their effectiveness in practical medical imaging contexts remains underexplored.

In this study, we investigate the feasibility of using untrained neural networks to accelerate Magnetic Resonance Imaging (MRI). MRI is a critical, non-invasive medical imaging modality, but its slow acquisition times present significant challenges. Compressed sensing techniques aim to address these limitations by enabling image reconstruction from a reduced number of measurements, thereby accelerating scan times. While traditional methods such as l_1-norm minimization and Total Variation (TV) norm minimization are prevalent and untrained, they have been largely outperformed by deep learning approaches, as evidenced by competitions like the fastMRI challenge. This work seeks to evaluate the performance of untrained networks in MRI reconstruction and compare them with their trained counterparts.

Conclusion:

In summary, this research demonstrates that untrained neural networks can effectively accelerate MRI reconstruction, rivaling traditional methods and some trained models. Their inherent capabilities as strong priors offer significant potential in medical imaging, warranting further exploration to optimize their performance in clinical applications.

Understanding and Mitigating Exploding Inverses in Invertible Neural Networks:

Introduction:

This work investigates the stability of Invertible Neural Networks (INNs), highlighting issues such as exploding inverses that can lead to numerical non-invertibility. We demonstrate that these failures affect various applications, including generative modeling and memory-efficient gradient computation. Through analyzing common INN architectures, we establish bi-Lipschitz properties and propose solutions to mitigate these issues. For tasks requiring local invertibility, we introduce a flexible regularizer, while for global invertibility tasks—such as normalizing flows on out-of-distribution data—we emphasize the need for stable INN building blocks. Our findings provide critical insights into ensuring reliable performance in applications relying on invertibility.

Cocnlusion:

Invertible Neural Networks (INNs) are vital in deep learning, yet their analytical invertibility can fail in numerical computations. For local stability, our finite difference penalty serves as a robust stabilizer. For global stability, especially in normalizing flows, architectures like Residual Flows are crucial. Understanding both forward and inverse stability is essential for effective INN application.

Invertible DenseNets with Concatenated LipSwish

Abstract:

We present Invertible Dense Networks (i-DenseNets), an efficient extension of Residual Flows that enforces invertibility through Lipschitz continuity analysis. By implementing a learnable weighted concatenation, we enhance model performance and highlight the significance of concatenated representations. Additionally, we introduce the Concatenated LipSwish activation function, which maintains the Lipschitz condition and improves performance. i-DenseNets surpass Residual Flows and

other flow-based models in density estimation, evaluated in bits per dimension, while maintaining an equal parameter budget. Our model also excels in hybrid training as both a generative and discriminative model, showcasing its versatility and effectiveness.

Conclusion:

In this paper, we introduced i-DenseNets as a parameter-efficient alternative to Residual Flows, enforcing invertibility through Lipschitz continuity in dense layers. Our method features a learnable concatenation of representations, enhancing model performance. Trained on CIFAR10 and ImageNet32, i-DenseNets outperformed Residual Flows and other flow-based models in density estimation and hybrid modeling, demonstrating their strong potential as effective flow-based architectures.

Densely connected normalizing flows:

Introduction:

In this paper, we introduce DenseFlow, an advanced normalizing flow architecture that enhances model expressivity by incrementally padding intermediate representations with noise. By employing cross-unit coupling, where noise is preconditioned based on prior invertible units, and integrating a densely connected block with Nyström self-attention, our approach significantly improves likelihood evaluation and sampling efficiency. DenseFlow leverages both cross-unit and intra-module couplings to maximize capacity while maintaining a manageable computational budget. Our experimental results demonstrate that DenseFlow achieves state-of-the-art performance in density estimation on ImageNet32 and ImageNet64, showcasing its potential as a powerful tool in deep generative modeling.

Conclusion:

Normalizing flows allow principled recovery of the likelihood by evaluating factorized latent activations. However, their efficiency is hampered by the bijectivity constraint since it determines the model width. We propose to address this issue by incremental augmentation of intermediate latent representations. The introduced noise is preconditioned with respect to preceding representations throughout cross-unit affine coupling. We also propose an improved design of intra-module coupling transformations within glow-like invertible modules.

Invertible Monotone Operators for Normalizing Flows:

Introduction:

In this paper, we introduce Monotone Flows, a novel approach to normalizing flows that addresses the limitations of ResNet-based models regarding Lipschitz constants. By leveraging monotone operators, we derive a monotone formulation that maintains architectural flexibility while significantly loosening Lipschitz constraints. Our approach involves parameterizing the Cayley operator of maximally monotone operators to ensure effective density transformation. We also present a new activation function, Concatenated Pila (CPila), designed to enhance gradient flow and mitigate saturation issues. Extensive experiments demonstrate that Monotone Flows achieve superior performance on various density estimation benchmarks, including MNIST, CIFAR-10, ImageNet32, and ImageNet64, outperforming existing models while also showcasing robust expressive power through theoretical analysis and ablation studies.

Conclusion:

We presented Monotone Flows, a novel parametrization of normalizing flows based on monotone operators (the monotone formulation) combined with a new activation function called Concatenated Pila (CPila). Our theoretical analysis elucidated why the monotone formulation is more expressive than the residual or inverse residual formulations. Experimental results demonstrated the effectiveness of the proposed method on various density estimation benchmarks.

Learning Data Representations with Joint Diffusion Models:

Introduction:

In this work, we propose a joint diffusion model that integrates classification and generation tasks into a single framework, addressing the uneven performance and training instability often observed in existing models. By leveraging the powerful internal representations generated by diffusion-based models, we introduce a classifier that shares parameters with the diffusion model, enabling stable end-to-end training. This approach not only improves performance across both tasks but also allows for conditional sampling, enhancing the quality of generated samples. Our empirical analysis validates the effectiveness of the shared representations, leading to state-of-the-art results on various benchmarks for both classification and generation. Additionally, we demonstrate how our model can facilitate visual counterfactual explanations, providing insights into model decision-making by highlighting necessary changes to input data for desired outputs. Overall, our joint diffusion model offers a significant advancement in hybrid modeling, combining the strengths of generative and discriminative tasks seamlessly.

Conclusion:

We introduced a joint model that merges a diffusion model with a classifier using shared parameters. Our experiments demonstrated that diffusion models learn meaningful representations suitable for classification. This approach enhances both classification and generation tasks, enabling high-quality outputs and conditional generations with classifier guidance. The model achieves state-of-the-art performance among joint models and facilitates visual counterfactual explanations seamlessly.

Monotone, Bi-Lipschitz, and Polyak-Łojasiewicz Networks:

Introduction:

This paper introduces the BiLipNet, a novel bi-Lipschitz invertible neural network that allows for controlled adjustments of both its Lipschitzness (sensitivity to input changes) and inverse Lipschitzness (distinguishability of inputs from outputs). The second key contribution is the PLNet, a scalar-output network formed by combining BiLipNet with a quadratic potential. We demonstrate that PLNet adheres to the Polyak-Łojasiewicz condition, making it suitable for learning non-convex surrogate losses with a unique, efficiently computable global minimum. Central to these innovations is an invertible residual layer characterized by strong monotonicity and Lipschitz properties, which, when integrated with orthogonal layers, constructs the BiLipNet. We certify these properties using incremental quadratic constraints, yielding tighter bounds than spectral normalization allows. Additionally, we present an efficient method for calculating the inverse of a BiLipNet, framing it as a series of three-operator splitting problems amenable to rapid algorithms.

Conclusion:

In conclusion, the BiLipNet and PLNet present significant advancements in the design of invertible neural networks with certified properties. By enabling fine control over Lipschitzness and ensuring robust learning through the Polyak-Łojasiewicz condition, these models enhance the stability and performance of neural network training. The introduction of a novel invertible residual layer, coupled with the use of incremental quadratic constraints for property certification, offers tighter bounds and improved reliability

ENTROPY-INFORMED WEIGHTING CHANNEL NORMALIZING FLOW:

Introduction:

In this work, we introduce the Entropy-Informed Weighting Channel Normalizing Flow (EIW-Flow), a novel approach to enhancing the expressive power of Normalizing Flows (NFs) while addressing their substantial memory requirements. Traditional multi-scale architectures reduce latent dimensions through static splitting, which can limit the model's ability to capture complex relationships within the data. Our proposed Shuffle operation generates feature-dependent channel-wise weights, adaptively shuffling latent variables before splitting them. This dynamic method guides the latent variables toward entropy increase, improving overall model performance. Experimental results demonstrate that EIW-Flow achieves state-of-the-art density estimation and sample quality across various benchmarks, including CIFAR-10, CelebA, and ImageNet, with minimal additional computational costs. Overall, EIW-Flow represents a significant advancement in the field of deep generative models, combining efficiency with enhanced expressiveness.

Conclusion:

This paper presents the Entropy-Informed Weighting Channel Normalizing Flow (EIW-Flow), integrating a reversible and regularized Shuffle operation into a multi-scale architecture. The Shuffle operation consists of three components: solver, guider, and shuffler, enhancing entropy through principles like the Central Limit Theorem. EIW-Flow demonstrates superior density estimation and image generation compared to state-of-the-art models, with minimal additional computational overhead. Ablation studies further explore hyperparameter impacts and the Shuffle operation's effectiveness.

2.Invertible Convolutional Flow

Introduction:

This work introduces novel normalizing flows leveraging circular and symmetric convolutions to enhance generative models' efficiency and flexibility. Traditional normalizing flows rely on computationally demanding Jacobian determinant and inverse calculations, typically simplified by constraining the architecture to produce diagonal or triangular Jacobians. Here, circular and symmetric convolutions provide an alternative that enables these operations in $O(N \log N)$ time, while element-wise multiplication is used for added flexibility. Additionally, an analytic approach for designing nonlinear, invertible transformations introduces implicit regularization, improving layer properties in the model.

Compared to previous flow-based models, which often employ low-rank or block diagonal Jacobians, the proposed adaptive convolution filters allow greater expressiveness. This is achieved by dynamically adjusting the convolutional kernel based on input features, enhancing the model's ability to capture intricate data structures. These approaches are particularly suited for high-dimensional generative tasks, such as image and audio synthesis, providing a practical framework for constructing effective normalizing flow models. The proposed transformations not only maintain tractability but also facilitate invertibility, making them viable for complex generative applications without compromising computational efficiency.

Conclusion:

This work demonstrates that circular and symmetric convolutions enable efficient, invertible normalizing flows with adaptive coupling layers and analytically derived nonlinearities. These components improve model flexibility, regularize intermediate activations, and yield state-of-the-art performance for fast, high-quality sampling in generative tasks.

Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models

Abstract—Deep generative models encompass various techniques that utilize deep neural networks to learn training sample distributions. This compendium explores energy-based models, variational autoencoders, generative adversarial networks, autoregressive models, and normalizing flows, along with hybrid approaches. It analyzes the trade-offs among these methods, such as run-time, diversity, and architectural constraints, while reviewing current state-of-the-art advancements and implementations in the field, highlighting their interrelations and foundational principles.

Conclusion:

While GANs excel in sample quality, emerging models are narrowing the gap, offering reduced mode collapse and simpler training. Hybrid models provide a balance but add complexity. Innovations in data augmentation, linear attention, and implicit networks enhance scalability and efficiency, paving the way for unified generative models across various data types and resolutions.

Why Normalizing Flows Fail to Detect Out-of-Distribution Data

Abstract:
This study explores the limitations of normalizing flows in detecting out-of-distribution (OOD) data, highlighting their tendency to assign high likelihoods to unrelated data, such as handwritten digits when trained on clothing images. The research identifies that flows learn local pixel correlations and generic transformations rather than dataset-specific features. By modifying the architecture of coupling layers, the authors improve flows' ability to learn the semantic structure of target data, enhancing OOD detection while revealing that characteristics promoting high-fidelity image generation may hinder OOD performance.

Conclusion:
This work highlights the impact of inductive biases in deep learning, particularly regarding normalizing flows in out-of-distribution (OOD) detection. While these models can generate semantically similar samples to training data, they often learn generic graphical features and local pixel correlations instead of capturing semantic structures specific to the dataset. The authors note that flows may overfit low-level image features, raising questions about their sampling behavior and memorization of training data. They suggest that further investigation into this aspect of normalizing flows is a promising area for future research.

Neural-network-based regularization methods for inverse problems in imaging

Abstract:
This review provides an overview of neural-network-based regularization methods for solving inverse problems in imaging, targeting readers with a background in applied mathematics and some familiarity with neural networks. It highlights the significant impact of deep learning techniques on the field, particularly through convolutional neural networks (CNNs). The article explains key concepts, such as learned generators and learned priors, including diffusion models, while focusing on the function space analysis of neural network approaches. Inverse problems aim to infer an unknown signal, typically an image, from noisy measurements, described by the equation $y = A u + \zeta$, where A is a forward operator and ζ represents noise. Traditional solutions often involve variational regularization, ensuring the problem is well-posed, and adopting a Bayesian perspective to model the variables involved. The review emphasizes that while comprehensive, it cannot cover every aspect due to the rapid development of the field and encourages readers to consult additional literature for further insights.

Conclusion:

This article offers an overview of state-of-the-art neural-network-based regularization methods for inverse problems in imaging, emphasizing conceptual understanding over an exhaustive list of existing works. It highlights the transformative impact of data-driven techniques over the past decade, noting a shift towards statistical modeling that enhances generalizability and uncertainty quantification. The review indicates that many recent approaches prioritize interpretable model structures rather than merely optimizing specific problem outcomes. Looking ahead, the authors predict that future developments will focus on integrating data with insights from physical and statistical modeling, fostering structured methods that improve understanding of uncertainty in real-world applications.

Understanding and Mitigating Exploding Inverses in Invertible Neural Networks
Abstract:

This work investigates the stability of invertible neural networks (INNs) and identifies critical failures such as exploding inverses that can lead to numerical non-invertibility. These issues affect various applications, including normalizing flow models, memory-efficient gradient computation, and posterior flexibility in variational autoencoders. The authors demonstrate that common INN architectures can produce significant reconstruction errors, particularly when handling in-distribution and out-of-distribution data, undermining their theoretical invertibility. The study reveals that these failures may not be evident during training, resulting in silent non-invertibility that compromises exact likelihood computation and efficient sampling. To address these challenges, the authors propose a flexible regularizer for tasks requiring local invertibility and emphasize the need for stable INN building blocks for scenarios necessitating global invertibility. By deriving bi-Lipschitz properties of INN components, the research provides insights that could lead to more robust and reliable INN architectures in future applications.

Conclusion:

Invertible Neural Networks (INNs) are an increasingly popular component of the modern deep learning toolkit. However, if analytical invertibility does not carry through to the numerical computation, their underlying assumptions break. When applying INNs, it is important to consider how the inverse is used. If local stability is sufficient, like for memory-efficient gradients, our finite difference penalty is sufficient as an architecture agnostic stabilizer. For global stability requirements e.g. when using INNs as normalizing flows, the focus should be on architectures that enable stable mappings like Residual Flows (Chen et al., 2019). Altogether we have shown that studying stability properties of both forward and inverse is a key step towards a complete understanding of INNs.

Renormalization group flow as optimal transport

Introduction:

The renormalization group (RG) is fundamental in quantum field theory, allowing insight into how a system's effective description evolves with measurement precision. This paper discusses exact renormalization group (ERG) equations, particularly Polchinski's, which can be reformulated as a gradient flow of relative entropy using a functional generalization of the Wasserstein-2 metric from optimal transport. The authors show that this framework enhances the understanding of RG flows, reveals a new RG monotone, and offers a variational formula applicable in numerical methods, extending beyond Polchinski's equation and elucidating features of various ERG schemes.
Conclusion:

In this paper we have provided a new approach to the exact renormalization group using the tools of optimal transport theory. In so doing, we defined new, nonperturbative RG monotones, developed a novel variational formula for RG flows, and suggested new numerical algorithms.

GENERATIVE TIME-SERIES MODELING WITH FOURIER FLOWS
Abstract:
This paper addresses the challenge of generating synthetic time-series data, particularly in privacy-

sensitive fields like medicine, where access to real data is restricted. The authors propose a novel explicit likelihood model called Fourier flow, which utilizes a discrete Fourier transform to convert variable-length time-series into fixed-length spectral representations. This approach enables efficient computation of Jacobian determinants and inverse mapping, outperforming traditional GAN-based methods while ensuring data privacy.

Conclusion:

This paper presents Fourier flows, an explicit likelihood model utilizing normalizing flows to represent stochastic time-series data in the frequency domain. By leveraging the properties of the discrete Fourier transform (DFT), Fourier flows efficiently compress temporal information into low-dimensional spectral representations, enabling more effective distribution learning. They facilitate complex transformations without additional Jacobian computation costs and incorporate recurrent neural networks (RNNs) for capturing the sequential nature of spectral data. Unlike implicit likelihood models, Fourier flows offer exact likelihood computation, potentially enhancing privacy by reducing the risk of memorizing training data, although further research on their generalization properties is needed.

Learning Efficient and Robust Ordinary Differential Equations via Invertible Neural Networks
Abstract:

This paper introduces a novel approach for learning ordinary differential equations (ODEs) by treating their dynamics as a vector field related to a base vector field through a diffeomorphism, represented by an invertible neural network (INN). This method offloads some modeling complexity onto the INN, allowing for a base ODE that is easier to integrate. The authors demonstrate significant speed improvements and enhanced robustness when integrating trajectories from learned ODEs, achieving speed-ups of up to two orders of magnitude in training and evaluating benchmark ODE systems and continuous-depth neural network models.

Conclusion:

We have proposed a novel approach to learning ODEs with unknown dynamics, which uses invertible neural networks to learn a diffeomorphism relating a desired target ODE to a base ODE that is often easier to integrate. We have investigated using a base ODE that is linear or parameterized by a neural network. By leveraging the closed form solution of linear ODEs, our method provides significant speed-ups and allows for asymptotic property constraints on the learned ODEs. Alternatively, by using a base ODE parameterized by a neural network, our approach can learn “difficult” ODEs, with simpler networks modeling their dynamics. We have validated our method by learning ODEs on synthetic and real-world data, on robotic learning problems and within continuous-depth neural network models. Future work could explore more on how to balance offloading the burden of learning to the invertible neural network and the base ODE.

CONSTRUCTING ORTHOGONAL CONVOLUTIONS IN AN EXPLICIT MANNER
Abstract:
This paper introduces explicitly constructed orthogonal (ECO) convolutions, which achieve an orthogonal input-output Jacobian matrix, ensuring robustness to input perturbations and preserving gradient norms during back-propagation. By constructing the convolution kernel to enforce singular values of the

Jacobian to be 1, ECO convolutions reduce computational costs to those of standard dilated convolutions while maintaining efficiency over existing methods like skew orthogonal convolution (SOC). Experiments on CIFAR-10 and CIFAR-100 show that ECO convolutions offer faster evaluations with competitive accuracy in standard and certified robust settings, highlighting their potential for stable training in deep neural networks.
Conclusion:

This study presents explicitly constructed orthogonal convolutions (ECO) that achieve an orthogonal Jacobian matrix, ensuring 1-Lipschitz stability and gradient preservation. ECO convolutions match the computational cost of standard dilated convolutions while outperforming the state-of-the-art SOC method in efficiency, demonstrating competitive accuracy on CIFAR-10 and CIFAR-100 datasets.

Relative gradient optimization of the Jacobian term in unsupervised deep learning
Abstract:

Unfortunately, however, typical strategies employed in neural networks training do not scale well for objective functions like the aforementioned ones; in fact, through the change of variable formula, the logarithm of the absolute value of the determinant of the Jacobian appears in the objective. Its exact computation, let alone its optimization, quickly gets prohibitively computationally demanding as the data dimensionality grows. A large part of the research on deep density estimation, generally referred to under the term autoregressive normalizing flows, has therefore been dedicated to considering a restricted class of transformations such that the computation of the Jacobian term is trivial [14, 44, 15, 34, 25, 12], thus imposing a tradeoff between computation and expressive power. While such models can approximate arbitrary probability distributions, the extracted features are strongly restricted based on the imposed triangular structure, which prevents the system from learning a properly disentangled representation. Other strategies involve the optimization of an approximation of the exact objective [5], and continuous-time analogs of normalizing flows for which the likelihood (or some approximation thereof) can be computed using relatively cheap operations [13, 19].
Conclusion:

This work introduces a method using relative gradients for exact optimization of objective functions involving the log-determinant of the Jacobian in neural networks. Unlike traditional normalizing flows, it supports fully connected layers without structural limitations, allowing for greater expressiveness in learning inverse transformations, particularly in identifiable nonlinear ICA models.

4. Towards Understanding the Invertibility of Convolutional Neural Networks

Abstract:

This paper investigates the approximate invertibility of Convolutional Neural Networks (CNNs) by developing a theoretical model for sparse signal recovery, relating it to CNNs with random weights. The authors establish a connection to model-based compressive sensing, demonstrating that CNNs can effectively reconstruct inputs from hidden activations. They introduce three key concepts: a model of sparse linear combinations of learned filters consistent with natural images, the model-RIP property of

effective matrices capturing convolutions, and the interpretation of CNN layers as Iterative Hard Thresholding (IHT) iterations for model-based compressive sensing. The authors provide empirical evidence that large-scale CNNs align with their theoretical analysis, suggesting that CNNs, despite being trained for classification, exhibit significant reconstructive capabilities. Additionally, they highlight gaps between the assumptions of their model and practical CNN training scenarios, indicating the need for further theoretical exploration to bridge the gap between empirical success and mathematical understanding of deep learning architectures.

Conclusion:

The authors propose three concepts linking compressive sensing models, learned filter properties, and the approximate invertibility of CNNs. Their experiments reveal that trained CNN filters align with the proposed mathematical properties, while hidden units exhibit more complexity than predicted. They suggest that adopting a compressive model for hidden units could better capture this complexity. Additionally, significant information in switch units remains unexplored, indicating areas for future research.

GAN Inversion: A Survey

Abstract:

GAN inversion is a technique that converts real images into the latent space of a pretrained Generative Adversarial Network (GAN), enabling the faithful reconstruction of these images by the GAN's generator. This method is vital for bridging the gap between real and synthetic images, facilitating real image editing applications using models like StyleGAN and BigGAN. By inverting an image into the latent space, users can manipulate specific attributes while retaining others, thus applying controllable transformations to real images. The paper surveys various GAN inversion algorithms and their applications in image restoration and manipulation, highlighting trends and challenges for future research. Additionally, it provides a curated list of GAN inversion methods, datasets, and related resources, underscoring GAN inversion's role in advancing both image editing capabilities and the understanding of GAN latent spaces.

Conclusion:

This paper provides a comprehensive survey of GAN inversion methods, highlighting their algorithms and applications. It explores how deep generative models like GANs learn to capture the variations in training data through weak supervision in image generation, facilitating interpretable latent representations for diverse image editing tasks. The authors summarize the essential properties of GAN latent spaces and introduce four categories of GAN inversion methods along with their key characteristics. The survey also covers various applications of GAN inversion, including image manipulation, generation, restoration, and emerging uses beyond image processing. Additionally, it discusses the challenges and future directions in the field of GAN inversion.

Compressed Sensing using Generative Models

Abstract:

Compressed sensing aims to reconstruct a vector from an underdetermined set of noisy linear measurements by leveraging prior knowledge about the vector's structure. Traditionally, this structure is represented by sparsity in a well-defined basis. This paper introduces a new approach that achieves guarantees similar to standard compressed sensing without relying on sparsity. Instead, it assumes that the vectors are near the range of a generative model $G: \mathbb{R}^k \rightarrow \mathbb{R}^n$. The main theorem states that if G is L-Lipschitz, approximately $O(k \log L)$ random Gaussian measurements are sufficient for an ℓ_2/ℓ_2 recovery guarantee. This method demonstrates significant efficiency, requiring 5-10 times fewer measurements than Lasso for equivalent accuracy. The paper showcases the results using generative models from existing variational autoencoders and generative adversarial networks. The research highlights a shift in the understanding of recovery guarantees in compressed sensing, suggesting that leveraging generative models can provide effective solutions even in the absence of sparsity assumptions, making it applicable to a wider range of problems including those in medical imaging and signal recovery.

Conclusion:

This work demonstrates how to apply compressed sensing using neural network-based generative models, which can represent data distributions more efficiently than traditional sparsity models. Their differentiability facilitates rapid signal reconstruction, enabling significant reductions in the number of required measurements. Theorems and experiments indicate that signal reconstruction approaches optimality with relatively few measurements, and larger generative models could further enhance performance as measurement numbers increase. Ongoing advancements in generative models will likely improve reconstruction quality. The framework's adaptability to various generative models allows for continuous enhancements in reconstruction efficiency, potentially serving as a benchmark for evaluating model quality.

Large-capacity Image Steganography Based on Invertible Neural Networks

Abstract:

This paper presents the Invertible Steganography Network (ISN) for image steganography, addressing the challenge of increasing payload capacity without making the container image detectable. The ISN operates as a single invertible network, utilizing forward and backward propagation to simultaneously manage the embedding and extraction of hidden images. By sharing parameters within the architecture, the ISN efficiently generates high-quality container images while significantly enhancing the payload capacity by increasing the number of channels in the hidden image branch. The proposed method not only ensures imperceptibility—making it difficult for steganalysis to identify the hidden data—but also achieves state-of-the-art performance in visual and quantitative assessments. This approach promises improvements in practical applications of image steganography, balancing the need for high payload capacity with the critical requirement for the concealment of hidden data.

Conclusion:

In this paper, we have proposed an Invertible Steganography Network (ISN) for image steganography, where the forward and backward propagation operations of the same network are leveraged to embed and extract hidden images, respectively. Our method significantly improves the steganography payload capacity, and can be easily adapted to hide multiple images with high imperceptibility. Comprehensive experiments demonstrate that with significant improvement of the steganography payload capacity, our ISN method achieves state-of-the-art both visually and quantitatively.

A Study on Overfitting in Deep Reinforcement Learning

Abstract:
This paper investigates the overfitting behavior of deep Reinforcement Learning (RL) agents, emphasizing the importance of understanding their generalization in critical applications like healthcare and finance. Despite advancements in deep RL, powered by large neural networks and novel algorithms, the risk of overfitting increases as model complexity grows. The study reveals that standard RL techniques, which often introduce stochasticity, do not effectively prevent or detect overfitting, leading to significant variations in test performance, even among agents achieving optimal training rewards. The authors highlight the need for more rigorous evaluation protocols in RL, as current practices fall short of addressing the vulnerabilities and reproducibility issues in testing. By utilizing a configurable maze environment, the research systematically analyzes the generalization capabilities of RL agents. The findings underscore the necessity of establishing a deeper understanding of representation learning, long-term planning, exploration, and adaptation before deploying deep RL models in real-world settings. This work serves as a call for improved methodologies to assess and ensure the robustness of RL agents against overfitting, ultimately enhancing their reliability in practical applications.

Conclusion:
Large neural networks can effectively memorize extensive training sets, but characterizing their inductive bias and understanding generalization in over-parameterized agents remain challenging. Careful evaluation protocols are essential to detect overfitting, and isolating training and test data is recommended, even in noisy and non-deterministic environments.

Robust Invertible Image Steganography

Abstract:
This paper introduces a novel flow-based framework for image steganography called Robust Invertible Image Steganography (RIIS). Traditional steganography methods often struggle with hiding capacity and robustness, particularly under distortions like Gaussian noise and JPEG compression. RIIS leverages a conditional normalizing flow to model the distribution of high-frequency components based on the container image, enhancing the robustness of the hidden data reconstruction. A Container Enhancement Module (CEM) further aids in this process. The framework incorporates Distortion-Guided Modulation (DGM) to adjust network parameters for various distortion levels, making it a versatile, one-size-fits-all solution. Extensive experiments demonstrate that RIIS significantly improves robustness while maintaining imperceptibility and capacity, addressing limitations in existing learning-based methods. This is the first learning-based approach to enhance robustness in image steganography, broadening its applicability in real-world scenarios. The proposed method is effective against a range of distortions, making it suitable for various applications, including face-swap detection and grayscale colorization.

Conclusion:

This paper introduces a robust invertible image steganography (RIIS) framework designed to effectively hide a large number of secret images within a limited-size container image, even under challenges such as noise and JPEG compression. The framework employs a Container Adaptive Normalization and Processing (CANP) module and a Container Enhancement Module (CEM), along with a strategic training approach, to mitigate distortion effects during the steganography process. Experimental results demonstrate that the model achieves high performance and robustness, enhancing the applicability of steganography in real-world scenarios. Additionally, the efficiency of the RIIS design is validated in other low-level inverse problems, such as decolorization. Future work will involve supporting RIIS on MindSpore, a new deep learning computing framework.

Reversible Architectures for Arbitrarily Deep Residual Neural Networks

Abstract:

This paper explores deep residual networks (ResNets) through the lens of ordinary differential equations (ODEs), offering a theoretical framework for their stability and reversibility. By interpreting ResNets as dynamical systems, the authors develop three new reversible neural network architectures that can be trained to arbitrary depths with minimal memory usage, as they do not require storing activations for most hidden layers. The stability of these architectures allows for effective training even with modest computational resources. Empirical results on CIFAR-10, CIFAR-100, and STL-10 datasets demonstrate that these architectures achieve state-of-the-art performance, often outperforming or matching strong baselines while showing superior generalization with limited training data. This research addresses gaps in the theoretical understanding of ResNets and provides practical architectures that enhance deep learning applications in various domains.

Conclusion:

The authors introduce three stable and reversible architectures that integrate stable ordinary differential equations (ODEs) with deep residual neural networks. This approach leverages the reversibility property for a memory-efficient implementation that does not require storing activations for most hidden layers, enabling deeper networks to be trained with limited computational resources. Evaluations on three public datasets show that these architectures achieve superior or comparable performance to state-of-the-art methods, particularly excelling in accuracy with limited training data. The robustness with small datasets is attributed to the inherent stability of the Hamiltonian neural network architecture used in their models.

Generalization Bounds and Representation Learning for Estimation of Potential Outcomes and Causal Effects

Abstract:

This work addresses the challenge of evaluating decisions based on non-experimental observational data across fields like healthcare, economics, and education. It focuses on estimating individual-level potential outcomes and causal effects, such as a patient's response to different medications, using historical

records of interventions, contexts, and outcomes. The authors provide generalization bounds for estimating outcomes based on distributional distances between re-weighted samples of treatment groups and establish conditions for tight bounds, connecting these to unsupervised domain adaptation. They develop algorithms that minimize these bounds by regularizing the treatment group distance, facilitating information sharing between groups. Experimental evaluations on real and synthetic datasets demonstrate the effectiveness of their proposed representation architecture and regularization methods in improving causal effect estimation.

Conclusion:

Uncertainty Quantification in Deep MRI Reconstruction

Abstract:
This study addresses the critical need for reliable MRI image reconstruction by quantifying uncertainty in deep learning (DL) models, particularly in the context of undersampling and the complexity of these models. The researchers utilize variational autoencoders (VAEs) to create a probabilistic reconstruction scheme that converts low-quality, aliased MRI scans into diagnostic-quality images. The VAE encodes acquisition uncertainty in a latent space and provides a posterior distribution from which pixel variance maps can be generated using Monte-Carlo sampling. To assess prediction risk, Stein's Unbiased Risk Estimator (SURE) is employed as a proxy for mean-squared error (MSE). Experiments focused on knee MRI reconstruction reveal that adversarial losses increase uncertainty, while recurrent unrolled networks help reduce prediction uncertainty and associated risk. The findings highlight the importance of uncertainty quantification in enhancing the accuracy and interpretability of MRI reconstructions, ultimately aiding radiologists in making more informed decisions during diagnoses.

Conclusion:

This paper presents methods for analyzing uncertainty in deep-learning-based compressive MRI image recovery, aimed at enhancing medical imaging workflows for improved diagnosis and acquisition. The authors develop a probabilistic variational autoencoder (VAE) model to produce realistic and data-consistent images with low error. A Monte Carlo approach is employed to quantify pixel variance and generate uncertainty maps alongside image reconstruction. To evaluate reconstruction risk, Stein's Unbiased Risk Estimator is applied to density-compensated zero-filled images, which are assumed to have zero-mean Gaussian residuals. The study validates these methods using a knee MRI dataset, finding that increased adversarial loss correlates with higher uncertainty, highlighting a trade-off between high-frequency retrieval and uncertainty. In contrast, employing multiple recurrent blocks reduces uncertainty, suggesting a way to enhance robustness in reconstructions.

Future research directions include exploring other sources of uncertainty related to data and knowledge, conducting extensive evaluations with various MRI datasets and acquisition strategies, and analyzing uncertainty in pathological cases. Additionally, enhancing pixel-wise SURE to improve variance estimation could lead to better performance. Ultimately, the integration of spatial risk maps and Monte

Carlo results with global SURE scores could significantly enhance the medical imaging workflow, particularly in diagnosis and automated image quality assessment.

Modeling Sparse Deviations for Compressed Sensing using Generative Models

Abstract:
This paper introduces Sparse-Gen, a framework designed for compressed sensing that combines the strengths of sparsity-based methods and generative models. In compressed sensing, the goal is to reconstruct an unknown signal using a limited number of linear measurements. Traditional approaches rely on assumptions like sparsity or predefined generative models, which can limit recovery to a small support set of signals. Sparse-Gen addresses this limitation by allowing sparse deviations from the support set, enabling reconstruction across the entire signal space.

Theoretically, Sparse-Gen defines a new class of signals for compressed sensing, which generalizes traditional sparse vector recovery while avoiding the restrictive support constraints imposed by generative models. Empirically, the framework demonstrates consistent improvements in reconstruction accuracy over existing methods, particularly in scenarios where a generative model trained on a data-rich source domain aids in sensing for a data-scarce target domain. This approach significantly enhances the efficiency of data acquisition in various fields, such as MRI and geophysical applications, where measurement costs are high.

Conclusion:
This paper discusses the integration of deep generative models as priors for compressed sensing, leading to a novel framework called Sparse-Gen. This framework enhances traditional algorithms for data acquisition by allowing sparse deviations from the range of a generative model, thereby generalizing both sparse vector recovery and recovery using generative models. The advantages of this approach are demonstrated both theoretically and empirically.

Future work aims to develop algorithms that can more effectively model the structure within these sparse deviations, leveraging existing research in structured sparse vector recovery. Additionally, a deeper understanding of the non-convexity challenges associated with generative model-based recovery could yield stronger theoretical guarantees and improved optimization methods. The authors also express interest in extending the Sparse-Gen framework to other data modalities, such as graphs, for applications in network tomography and reconstruction, where real-world networks often exhibit sparsity that can be effectively modeled using deep generative models.

6. Invertible Residual Networks

Abstract:
This paper presents a method to make standard ResNet architectures invertible, enabling them to be

utilized for multiple tasks, including classification, density estimation, and generation. Traditionally, achieving invertibility in neural networks involves complex architectures or dimensional partitioning. However, the authors propose a simpler approach that only requires adding a normalization step during training, which is already a common practice in standard frameworks.

The resulting invertible ResNets define a generative model that can be trained using maximum likelihood on unlabeled data. To facilitate the computation of likelihoods, the authors introduce a tractable approximation for the Jacobian log-determinant of a residual block. Empirical evaluations demonstrate that invertible ResNets achieve competitive performance compared to state-of-the-art image classifiers and flow-based generative models, marking a significant advancement in integrating discriminative and generative capabilities within a single architecture.

Overall, this work bridges the gap between discriminative and generative models by leveraging the continuous dynamics of invertible ResNets, thus enhancing their applicability in various machine learning domains.

Conclusion:

The paper introduces i-ResNets, a new architecture that allows for flexible layer designs while still providing tractable density estimates. Unlike traditional partitioning-based models, i-ResNets utilize unrestricted Jacobians to enable expansion and contraction through residual blocks, eliminating the need for affine blocks and scaling layers to maintain non-volume preservation.

However, the authors acknowledge several challenges for future work. First, the log-determinant estimator currently used is biased, but recent advancements in unbiased estimators (Han et al., 2018) could enhance the performance of their generative model. Second, implementing and designing networks with a Lipschitz constraint poses difficulties, as it requires constraining individual linear layers rather than directly controlling the Lipschitz constant for an entire block. The authors reference Anil et al. (2018) as a promising approach to tackle this issue.

Normalizing Flows for Probabilistic Modeling and Inference

Abstract:

This review paper provides a unified perspective on normalizing flows, which are used to define expressive probability distributions by applying a series of bijective transformations to a simple base distribution. The authors argue that the field of normalizing flows has matured, necessitating a comprehensive understanding of their design principles, expressive power, and computational trade-offs.

The paper begins by discussing the importance of well-specified probabilistic models in statistics and machine learning, emphasizing the role of normalizing flows in achieving richer probabilistic descriptions. The authors explain how normalizing flows can transform a simple density into complex, multi-modal distributions, making them suitable for various statistical tasks, including modeling, inference, and simulation.

The review highlights the need for a cohesive framework to understand the advancements in normalizing flows and their relationship to prior work. While acknowledging existing surveys, the authors aim to provide a more comprehensive tutorial that addresses open problems, such as the extension of flows to discrete variables and Riemannian manifolds. The review systematically covers the formal and conceptual structure of normalizing flows, their construction methods for both finite and infinitesimal variants, and concludes with broader implications for their application in diverse fields.

Conclusion:

The review summarizes the foundational principles of normalizing flows and their applications in probabilistic modeling and inference. It highlights key concepts such as expressive power and the chain rule of probability, providing guidance for practitioners. The authors emphasize the importance of computational constraints in implementing flows, noting the trade-offs between autoregressive and non-autoregressive designs. Ultimately, the article aims to aid practitioners in making informed choices when utilizing normalizing flows for various tasks.

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Abstract:

This work introduces a stochastic differential equation (SDE) framework for generative modeling, enabling the transformation of complex data distributions into a known prior by gradually adding noise, and a reverse SDE for sampling data by removing noise. The reverse-time SDE relies solely on the gradient field, or score, of the perturbed data distribution, which can be accurately estimated using neural networks. This framework integrates previous methods in score-based generative modeling and diffusion probabilistic modeling, facilitating new sampling techniques and enhanced modeling capabilities. A novel predictor-corrector method is introduced to refine the discretized reverse SDE's evolution. Additionally, an equivalent neural ordinary differential equation (ODE) is derived to allow for exact likelihood computation and improved sampling efficiency. Experimental results demonstrate the framework's effectiveness in class-conditional generation, image inpainting, and colorization, achieving state-of-the-art performance in unconditional image generation on CIFAR-10 with high-quality image outputs.

Conclusion:

The proposed SDE-based framework enhances score-based generative modeling by providing new sampling techniques and capabilities. It achieves state-of-the-art performance in various tasks, highlighting its potential for effective image generation and other applications.

CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion

Abstract:

The paper presents a novel multi-modality image fusion network called Correlation-Driven feature Decomposition Fusion (CDDFuse), designed to integrate features from different imaging modalities, such as infrared and visible images, while preserving their unique characteristics. CDDFuse utilizes Restormer blocks for shallow feature extraction and employs a dual-branch architecture comprising Lite Transformer (LT) blocks for low-frequency global features and Invertible Neural Networks (INN) for high-frequency local features. A correlation-driven loss function is introduced to ensure that low-frequency features are correlated, while high-frequency features remain uncorrelated. The output from global and local fusion layers results in a fused image that enhances clarity for downstream tasks like semantic segmentation and object detection. Extensive experiments demonstrate CDDFuse's effectiveness across various fusion tasks, showcasing improved performance compared to existing methods in both infrared-visible image fusion and medical image fusion applications. The code for CDDFuse is publicly available for further exploration.

Conclusion:

In this paper, we propose a dual-branch TransformerCNN architecture for multi-modal image fusion. With the help of Restormer, Lite transformer and invertible neural network blocks, modality-specific and -shared features are better extracted, and the decomposition for them is more intuitive and effective by the proposed correlation-driven decomposition loss. Experiments demonstrate the fusion effect of our CDDFuse, and the accuracy of downstream multi-modal pattern recognition tasks can be also improved.

Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift

Abstract:

This paper addresses the challenge of quantifying predictive uncertainty in deep learning models, particularly when there is a shift in input distributions from the training data. Accurate uncertainty estimates are essential for making reliable predictions in critical applications such as medical diagnostics and autonomous driving. The authors conduct a large-scale empirical comparison of various state-of-the-art methods for uncertainty quantification in classification tasks, focusing on how these methods perform under dataset shifts. The study reveals that traditional post-hoc calibration methods often fall short, while certain techniques that marginalize over models demonstrate robust performance across a range of tasks. The findings highlight the need for effective uncertainty quantification strategies to enhance the reliability of machine learning models in real-world applications, especially when faced with non-stationary data distributions. This research contributes to the understanding of how different methods handle uncertainty in changing environments, ultimately guiding improvements in predictive modeling practices.

Conclusion:

The study highlights the importance of accurately quantifying predictive uncertainty in deep learning models, especially under dataset shifts. It reveals that traditional post-hoc calibration methods often underperform, while model marginalization techniques yield strong results across various tasks. This research emphasizes the necessity for effective uncertainty quantification strategies to ensure the reliability of machine learning applications in real-world settings, particularly in high-stakes scenarios requiring informed decision-making.

Normalizing Flows: An Introduction and Review of Current Methods

Abstract:
The primary goal of statistics and machine learning is to model probability distributions from samples, a process known as generative modeling, which is essential due to the abundance of unlabeled data. This approach finds applications in density estimation, outlier detection, prior construction, and dataset summarization. Various generative modeling methods exist, including direct analytic approaches, variational methods, expectation maximization, and graphical models. Recent advances include generative neural networks like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which have shown impressive results in learning distributions of natural images but face limitations such as difficulties in evaluating probability densities, training instability, and mode collapse.

Normalizing Flows (NF) offer a solution with tractable distributions that allow for efficient and exact sampling and density evaluation. NFs have applications across diverse fields, including image, video, and audio generation, as well as reinforcement learning and computer graphics. This article aims to review the literature on Normalizing Flows, offering foundational knowledge, a literature overview, and highlighting open questions and future research directions. The structure includes sections on the introduction to NFs, their training, constructions, performance evaluation, and identification of research opportunities in this growing field.

Conclusion:
In conclusion, Normalizing Flows represent a powerful and flexible approach to generative modeling, overcoming many limitations faced by traditional methods such as GANs and VAEs. With their ability to provide exact density evaluations and efficient sampling, NFs have wide-ranging applications across various domains, including image and audio generation. This article not only reviews the current state of research on Normalizing Flows but also identifies key open questions and potential future directions, encouraging further exploration and innovation in this promising area of study.

Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty

Self-supervised learning has emerged as a powerful approach to improve representations in situations where labeled data is scarce. While traditionally viewed as a means to reduce reliance on annotations, recent findings highlight its significant contributions to model robustness. Self-supervision enhances resilience against adversarial attacks, label corruption, and common input corruptions like fog and blur, demonstrating its potential beyond merely catching up to fully supervised methods. Notably, self-supervised techniques excel in out-of-distribution detection for challenging near-distribution outliers, outperforming fully supervised counterparts. These insights suggest that self-supervised learning should not be considered a mere supplement to supervised learning; rather, when combined, they provide robust regularization that can improve uncertainty estimation and overall model robustness without necessitating larger models or additional data. As self-supervised methods continue to evolve, it is essential to evaluate their effectiveness not only in terms of clean accuracy but also by their ability to enhance robustness and uncertainty quantification. The results encourage future research in self-supervised learning to prioritize these aspects, thereby broadening the scope of evaluation criteria for these methods. The authors have made their code and an expanded validation dataset for ImageNet available for further research.

Conclusion:

This paper demonstrates that self-supervised learning can enhance the robustness and uncertainty of deep learning models beyond what is achievable with solely supervised methods. Significant improvements were noted in resilience to adversarial examples, label corruption, and common input corruptions when current supervised techniques were supplemented with an auxiliary rotation loss. Additionally, self-supervised methods excelled in out-of-distribution detection for challenging near-distribution anomalies, outperforming fully supervised methods in experiments conducted on CIFAR and ImageNet. The most substantial gains were observed in ImageNet, where the larger input size facilitated the application of more complex self-supervised objectives. The findings indicate that future efforts in developing robust models and improved data representations could greatly benefit from incorporating self-supervised approaches.

Revisiting Self-Supervised Visual Representation Learning

Abstract:

Unsupervised visual representation learning poses significant challenges in computer vision, although self-supervised techniques have emerged as effective solutions. This paper revisits various self-supervised models and conducts a large-scale study to uncover critical insights into their performance. It challenges conventional practices in self-supervised visual representation learning, revealing that standard CNN design principles do not consistently apply. The study highlights the strong dependence of visual representation quality on the choice of CNN architecture used in self-supervised tasks. For example, ResNet50 v1 performs well with relative patch location tasks but underperforms with rotation tasks. By exploring these aspects, the authors significantly enhance the performance of existing techniques, surpassing previously published state-of-the-art results. This work underscores the importance of carefully selecting CNN architectures and pretext tasks to optimize the effectiveness of self-supervised learning in generating high-quality visual representations.

Conclusion:

This work explores self-supervised visual representation learning, revealing several key insights. First, architectural lessons from fully supervised settings do not always apply to self-supervised contexts. Second, in residual architectures, the final prelogits layer consistently yields better performance than previously favored architectures like AlexNet. Third, the widening factor of CNNs significantly impacts the effectiveness of self-supervised techniques. Fourth, training linear logistic regression with SGD can require an extended convergence time. The study demonstrates that existing self-supervision techniques can be consistently improved, reducing the performance gap with fully labeled supervision. Notably, the ranking of architectures varies across methods, and vice versa, indicating that pretext tasks should be evaluated alongside the chosen architectures for optimal results.

Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style

Abstract:

Self-supervised representation learning has achieved significant success across various domains, particularly through data augmentation techniques that employ hand-crafted transformations aimed at preserving the semantic content of the data while modifying other attributes. This paper aims to explore the theoretical underpinnings of this approach by modeling the augmentation process as a latent variable model (LVM). The model posits a partition of the latent representation into two components: a content component, assumed invariant to augmentations, and a style component, which can vary. Unlike previous studies on disentanglement, this work allows for both statistical and causal dependencies in the latent space. The authors investigate the identifiability of the latent representation, demonstrating sufficient conditions for identifying the invariant content component from pairs of views of observations in both generative and discriminative contexts. Their findings, supported by numerical simulations with dependent latent variables, align with their theoretical framework. Additionally, they introduce Causal3DIdent, a new dataset featuring high-dimensional, complex images with intricate causal dependencies, to further examine the impact of data augmentations in practice. Overall, this research provides valuable insights into the effectiveness of self-supervised learning with data augmentation from a theoretical perspective.

Conclusion:

In conclusion, this study enhances the understanding of self-supervised representation learning by rigorously examining the role of data augmentation through the lens of latent variable modeling. By distinguishing between content and style components within latent representations, the authors provide a theoretical foundation for identifying invariant features critical to learning robust representations. Their insights, supported by simulations and the introduction of the Causal3DIdent dataset, underscore the importance of carefully designed augmentations in preserving semantic information while adapting to nuisance variations. This work not only clarifies the mechanisms underlying the empirical success of data augmentation but also lays the groundwork for future research in self-supervised learning, emphasizing the need for a deeper exploration of augmentation strategies in high-dimensional domains.

Webpage 2

1.

CInC Flow: Characterizable Invertible 3×3 Convolution
Abstract:

Normalizing flows offer an efficient alternative to GANs for generative modeling by optimizing directly on maximum likelihood and allowing exact computation of latent vectors due to their invertible transformations. Traditional convolutions, like 3×3 CNNs, are typically non-invertible, making them challenging to use in flow-based models. Emergent convolutions attempted to address this by constructing invertible CNN layers using pairs of masked CNNs, but this approach was inefficient.

This paper explores conditions for standard 3×3 CNNs to be invertible, focusing on enabling efficient and expressive normalizing flows without needing complex workarounds like emergent convolutions. The

authors derive simple, necessary, and sufficient conditions for invertibility in padded CNNs, allowing invertibility to be maintained throughout training. Their approach requires only one CNN layer per invertible convolution, improving efficiency compared to previous methods. They also introduce "Quad-coupling," a new coupling method for invertible convolutions.

The study benchmarks this method, demonstrating similar performance to emergent convolutions while reducing computational demands. This research advances normalizing flow models by making invertible, expressive CNN layers feasible, potentially enhancing the model's capability to capture complex data distributions in generative modeling.

Conclusion:

This paper introduces an efficient method for invertible \(n \times n \) convolutions in normalizing flows, using a novel "Quad-coupling" approach. Unlike previous methods, this approach requires only one CNN layer per invertible convolution, improving both inference and sampling times. The method generalizes to any \(n \times n \) kernel and achieves better log-likelihood performance on standard image modeling benchmarks compared to emergent convolutions.

StyleFlow: Disentangle Latent Representations via Normalizing Flow for Unsupervised Text Style Transfer

Abstract:

The paper presents *StyleFlow*, a novel model for unsupervised text style transfer that uses a disentanglement-based approach to separate content and style representations. Unlike traditional encoder-decoder models, StyleFlow leverages Normalizing Flow, a reversible neural network, allowing exact reconstruction of content without loss during both encoding and decoding. The model introduces attention-aware coupling layers to handle words that carry both content and style information, ensuring better content preservation. Additionally, a Normalizing Flow-based data augmentation method improves robustness by generating pseudo-sentences near the original style. Experimental results show that StyleFlow outperforms previous models, achieving state-of-the-art results on multiple metrics across datasets. The model is notable for its effective content preservation, robustness, and improved performance on style transfer benchmarks.

Conclusion:

This paper introduces *StyleFlow*, a novel model for text style transfer that enhances content preservation through a disentanglement approach. Using the Normalizing Flow framework, StyleFlow implements a reversible encoder that encodes forward and decodes in reverse without content loss. This reversibility allows accurate cycle reconstruction, ensuring transferred sentences retain their original content. Additionally, data augmentation improves robustness. Experimental results show that StyleFlow outperforms previous models in content preservation and achieves state-of-the-art results on multiple metrics. Future work will focus on handling multi-style transfer and transferring multiple style attributes.

Generating Synthetic Data with The Nearest Neighbors Algorithm

Abstract:

This paper introduces the *Local Resampler (LR)*, a novel semiparametric approach that utilizes the k-nearest neighbors (kNN) algorithm for generating synthetic data. LR creates synthetic samples by forming subsamples from the original data through kNN and drawing values from locally estimated distributions. Unlike model-based synthetic data methods, LR requires minimal hyperparameter tuning and can accurately replicate nonlinear and non-convex distributions. It outperforms or matches popular synthetic data generation techniques, offering a flexible alternative that combines kNN's simplicity with robust synthetic data generation. This approach is especially beneficial for applications in causal inference, missing value imputation, and data confidentiality.

Conclusion:

This paper proposes an approach using kNN to create synthetic samples, called Local Resampler (LR). LR uses kNN to define subsamples and estimates distributions for each subsample sequentially. The synthetic values are drawn from the locally estimated distributions. The proposed method allows using both parametric and nonparametric distributions

2. Invertible Attention

Abstract:

This paper introduces *Invertible Attention*, a novel module designed to enable attention mechanisms in invertible networks, which previously lacked this capability due to the non-bijective nature of conventional attention layers. To achieve invertibility, the proposed module constrains the Lipschitz constant of the attention mechanism, making each layer a bijective transformation suitable for invertible networks. The authors mathematically validate the invertibility by imposing specific constraints on the response map and feature mapping, and by introducing a Lipschitz-constrained convolution in the residual branch.

Experiments on image reconstruction tasks using popular datasets (CIFAR-10, SVHN, and CelebA) demonstrate that this invertible attention module allows nearly perfect reconstruction of inputs, a key requirement for invertible networks. The module can be integrated into existing invertible architectures, making it versatile for generative and dense prediction tasks. Comparative results indicate that invertible attention performs similarly to non-invertible attention in capturing long-range dependencies and achieving high performance on dense prediction tasks. This work bridges the gap between invertible networks and attention mechanisms, expanding the applications of invertible architectures in tasks that benefit from attention's ability to capture global dependencies.

Conclusion:

In conclusion, this work successfully introduces a novel *Invertible Attention* mechanism that extends the capability of invertible networks to leverage attention's long-range dependency capture. By imposing constraints on the Lipschitz constant and modifying the attention module to be bijective, the proposed method overcomes the inherent non-invertibility of conventional attention mechanisms. Experimental results validate the effectiveness of this approach, showing that it maintains high performance on tasks requiring dense predictions, comparable to traditional non-invertible attention methods. Additionally, on image reconstruction tasks, the invertible attention module enables near-perfect reconstruction, further demonstrating its suitability for invertible architectures. This contribution not only enhances the receptive field of invertible networks but also broadens their applicability to generative tasks and complex dense prediction applications. Future work may explore further integration of invertible attention across various network architectures and its potential in fields beyond computer vision.

Generative Flows with Invertible Attentions

Abstract:

This paper presents *AttnFlow*, an innovative approach that introduces invertible attention mechanisms within flow-based generative models, a new solution to model long-range dependencies efficiently. AttnFlow leverages two invertible attention types—map-based (iMap) and transformer-based (iTrans)—using a masked scheme to maintain invertibility with tractable Jacobian determinants. The iMap attention models individual positional importance, while the iTrans attention captures interactions among distant features, enabling generative flows to model complex correlations at any position in the flow architecture.

Traditional flow models have limited capacity to capture global dependencies, often requiring many layers to approximate these non-linear relationships. AttnFlow's invertible attention layers, however, integrate seamlessly into the models, enabling more efficient and accurate data dependency modeling. The approach improves upon prior methods like Glow and Flow++, which capture only short-range dependencies within individual feature splits.

AttnFlow achieves high efficiency and accuracy, as shown in multiple image synthesis tasks, performing favorably against state-of-the-art unconditional and conditional generative flows. This work represents a significant step toward extending the power of flow-based models to broader, more complex applications in deep generative modeling.

Conclusion:

This paper introduces invertible map-based and transformer-based attentions for both unconditional and conditional generative normalizing flows. The proposed attentions are capable of learning network dependencies efficiently to strengthen the representation power of flow-based generative models. The evaluation on image generation, super-resolution and image translation show clear improvement of our proposed attentions over the used unconditional and conditional flow-based backbones

LATENT LINEAR ODES WITH NEURAL KALMAN FILTERING FOR IRREGULAR TIME SERIES FORECASTING

Abtsract:

This paper presents *LinODEnet*, a novel Neural Ordinary Differential Equation (ODE) model designed for forecasting irregularly sampled time series, addressing the high computational burden and potential failures of traditional numerical integration methods. By embedding observations into a latent space governed by a linear ODE with constant coefficients, LinODEnet simplifies state propagation to

computing the matrix exponential, which has stable numerical implementations. The model incorporates a Kalman filter-inspired state update mechanism that effectively handles missing values and maintains self-consistency, ensuring the model state only updates when predictions differ from observations. Evaluated on medical and climate benchmark datasets, LinODEnet outperforms existing state-of-the-art models by up to 30%. The paper highlights LinODEnet's capability to provide fast and reliable forecasting while fulfilling essential properties such as stability and handling of irregular time series data, offering a significant advancement in the field of time series forecasting.

Conclusion:

This paper introduces a novel forecasting model for irregularly sampled time series with missing values, which utilizes constant linear ODE dynamics to map the observation space to a latent space. The model employs a state estimation update inspired by the Kalman filter and simplifies the solution of the linear ODE to calculating matrix exponentials, bypassing the need for numerical integration. It guarantees forward stability at initialization and has been evaluated against existing benchmarks, demonstrating significant improvements over current models. Additionally, the framework allows for the integration of future covariates in forecasting and facilitates analysis and modification of dynamics through linear algebra, paving the way for promising future research directions.

Boosting Urban Prediction via Addressing Spatial-Temporal Distribution Shift

Abstract:

This paper presents the Shift-Aware Urban Prediction (SAUP) framework, designed to address the challenges of urban prediction tasks that involve modeling complex spatial and temporal patterns of urban indicators, such as weather and vehicle charging demand. Traditional methods often focus solely on spatial and temporal correlations, neglecting the impact of distribution shifts, which can hinder prediction accuracy. The SAUP framework includes a Shift Elimination Module that utilizes Spatial-Temporal Attention Flows (STAF) to unify the distribution of raw shifted data, effectively removing spatiotemporal shifts. Following this, the Correlation Processing Module captures key correlations through topological and geographic learning using Graph Convolutional Networks (GCNs) and Convolutional Neural Networks (CNNs). Additionally, a model-agnostic Forecasting Module allows the integration of various forecasting architectures. Extensive experiments on two real-world datasets demonstrate that SAUP consistently outperforms six state-of-the-art spatiotemporal forecasting models, highlighting its effectiveness in enhancing urban prediction accuracy.

In this work, we introduce the Shift-Aware Urban Prediction (SAUP) framework to address the intrinsic shifts in spatial-temporal data. The framework begins with a Shift-Elimination Module that employs invertible attention and coupling layers to mitigate spatial and temporal shift effects. We then develop a Correlation Processing Module that captures correlations in urban data by embedding Point of Interest (POI) information through convolutional layers and utilizing Graph Convolutional Networks (GCNs) to uncover hidden spatial patterns from the adjacency matrix. These components are integrated for enhanced feature fusion. Following this, a model-agnostic Forecasting Module is employed for making predictions. Finally, the inverse transformation of the Shift-Elimination Module is applied to recover the original shift information. Extensive experiments on real-world datasets demonstrate that the SAUP framework significantly enhances the performance of existing backbone models.

IMPROVING LEARNED INVERTIBLE CODING WITH INVERTIBLE ATTENTION AND BACK-PROJECTION

Abstract:

This paper presents advancements in learned image compression (LIC) using Invertible Neural Networks (INNs) that improve performance over traditional methods like Versatile Video Coding (VVC). The authors propose replacing convolution with attention mechanisms in the coupling layers of INNs to enhance information extraction while preserving reversibility. They also introduce a Back-Projection (BP) mechanism to refine dimensionality reduction before quantization, minimizing information loss. The model incorporates an advanced channel-wise autoregressive entropy model to further improve performance. Experimental results demonstrate that the proposed modifications lead to significant enhancements, particularly in high-bitrate scenarios, surpassing existing state-of-the-art models. The main contributions include the integration of an Invertible Attention structure for better nonlinearity, the design of stable sampling techniques using BP, and notable performance improvements in the high bitrate range.

Conclusion:

In this paper, we improve the current optimal learned invertible coding model by adding a newly designed Invertible Attention to the INNs part for stronger nonlinearity. To make the sampling process more compatible with the reversible mapping, we combined the original Average&Copy with BackProjection and a carefully designed branching structure for a fixed number of channels. Experimental results show that our proposed modules are effective and the improved transformation model is comparable to the current optimal transformation models for LIC

PROFITI: PROBABILISTIC FORECASTING OF IRREGULAR TIME SERIES VIA CONDITIONAL FLOWS

Abstract:

This paper introduces ProFITi, a novel model for probabilistic forecasting of irregularly sampled multivariate time series (IMTS) with missing values, utilizing conditional normalizing flows. Traditional forecasting methods often rely on differential equation-based models and assume specific target distributions, which can limit their effectiveness. In contrast, ProFITi learns a joint probability distribution of future values based on past observations and future time-channel information. The model features two innovative components: a sorted invertible triangular attention layer (SITA) that adapts to conditioning inputs, and an invertible activation function named Shiesh, which operates across the entire real line. Extensive experiments across four datasets reveal that ProFITi significantly outperforms existing baseline models, achieving an average likelihood that is four times higher than the best previous model. The research highlights the potential of ProFITi for more accurate and reliable forecasting in various fields, including astronomy, finance, and healthcare.

Conclusion:

In this work, we propose a novel model ProFITi for probabilistic forecasting of irregularly sampled multivariate time series with missing values using conditioning normalizing flows. ProFITi is a permutation invariant normalizing flow model for conditional permutation invariant structured distributions. We propose two novel model components, sorted invertible triangular self attention and Shiesh activation function in order to learn any random target distribution. Our experiments on 3 IMTS datasets demonstrate that ProFITi provides better likelihoods than a state-of-the-art IMTS forecasting baselines.

Conditional Invertible Neural Networks for Medical Imaging

Absttract:
Recent advancements in deep learning have made it a popular choice for addressing inverse problems, particularly in medical imaging tasks such as low-dose computed tomography (CT) and accelerated magnetic resonance imaging (MRI). Traditional methods often yield only point estimates for image reconstructions, which can be inadequate for ill-posed problems where understanding uncertainty is crucial. This work explores generative flow-based models using invertible neural networks to provide a statistical interpretation of the reconstruction process, aiming to estimate the entire conditional distribution of the image given noisy measurements. While conventional methods like Markov chain Monte Carlo and approximate Bayesian computation have been used to estimate this distribution, they can be computationally intensive for large-scale problems. Flow-based generative models, a newer approach, utilize invertible transformations to learn continuous probability densities, enabling exact likelihood computation and facilitating maximum likelihood training. The study includes various architectures of invertible neural networks and ablation studies, revealing that using a radial distribution instead of a standard Gaussian base can enhance reconstruction quality. This research contributes to the understanding of uncertainties in image reconstruction, providing a framework for more reliable medical imaging applications.
Conclusion:

This study examined various architectures and best practices for implementing conditional flow-based methods in medical image reconstruction, specifically for CT and MRI. It focused on two prominent invertible network designs: the iUNet, inspired by the widely-used UNet architecture, and a multiscale architecture utilized in major normalizing flow frameworks like Glow and NICE. The researchers developed a conditioning Invertible Neural Network (cINN) framework, which merges the benefits of memory-efficient invertible networks with normalizing flows for uncertainty estimation, allowing for an integration of model-based and data-driven approaches. Initial investigations into using a radial Gaussian distribution instead of the traditional Gaussian base distribution indicated potential advantages. Looking ahead, the study suggests developing novel invertible network architectures based on existing models. While state-of-the-art deep learning methods for medical image reconstruction often rely on unrolled iterative methods, there is potential to explore the integration of invertible architectures into flow-based models, which could enhance memory-efficient backpropagation techniques and improve reconstruction outcomes.

A Survey on Medical Image Segmentation Based on Deep Learning Techniques
Abstract:
This survey explores the role of deep learning techniques in medical image segmentation, highlighting their rapid rise as a preferred method for analyzing medical images. It discusses the fundamental concepts of deep learning and its various applications, including image categorization, object recognition, segmentation, and registration. The paper introduces basic deep learning frameworks and examines the challenges in medical image segmentation, such as low accuracy in classification, low segmentation resolution, and insufficient image enhancement.

Convolutional neural networks (CNNs) are emphasized for their effectiveness in feature extraction and image recognition, marking a significant advancement in medical data processing. The paper notes that while deep learning has the potential to outperform traditional diagnostic methods, there remains a lack of comprehensive analyses of certain models, such as restricted Boltzmann machines (RBMs), generative adversarial networks (GANs), recurrent neural networks (RNNs), and deep neural networks (DNNs).

It calls for future research to address existing gaps and improve medical image segmentation techniques, offering insights that can guide researchers in overcoming current challenges. The study serves as a resource for understanding state-of-the-art deep learning applications in medical imaging and their implications for healthcare.

Conclusion:

Deep learning has made significant strides in medical image analysis, enhancing classification, segmentation, and registration processes. However, practical applications still face limitations due to challenges such as low accuracy in image classification, poor segmentation resolution, and inadequate image enhancement. While advancements in processing capabilities and structured data availability have yielded promising results, several barriers must be addressed before fully automating these processes.

This review discusses various frameworks and technologies available for developing deep learning systems tailored to medical image analysis. Convolutional neural networks (CNNs) are identified as critical in improving classification accuracy, segmentation resolution, and image enhancement. Given that deep learning is still evolving, there remains potential for developing models with enhanced accuracy and effectiveness in disease detection.

Future research should focus on expanding current medical image datasets, as their limited scale contributes to overfitting in deep learning models. Enhancing feature extraction techniques with high-resolution data and exploring advanced methods and applications will be essential for improving deep learning's contributions to medical challenges. The paper encourages continued research to advance the effectiveness of diagnostic tools and modeling in medical imaging.

POSTERIOR SAMPLING BASED ON GRADIENT FLOWS OF THE MMD WITH NEGATIVE DISTANCE KERNEL

Abstract:

This paper introduces conditional flows based on the maximum mean discrepancy (MMD) with a negative distance kernel for posterior sampling and conditional generative modeling. MMD, or energy distance, offers efficient computation through slicing and sorting techniques. The authors approximate the joint distribution of ground truth images and observations using discrete Wasserstein gradient flows and establish an error bound for the posterior distributions. They demonstrate that their particle flow method is a valid Wasserstein gradient flow for an appropriate functional. The effectiveness of the approach is showcased through numerical examples, including conditional image generation and inverse problems such as super-resolution, inpainting, and computed tomography in low-dose and limited-angle settings.

The study highlights the growing interest in generative models for inverse imaging problems, where reconstruction from noisy observations is challenging due to the ill-conditioned nature of forward operators. By adopting a Bayesian framework, the authors aim to sample from posterior distributions to account for uncertainties in reconstructions, allowing for the identification of uncertain image regions. The paper discusses the advantages of using MMD with a negative distance kernel, including translation and scale equivariance, efficient computation, and improved sample complexity, facilitating effective gradient flow simulations in high-dimensional probability distributions.

Conclusion:

In conclusion, the proposed conditional flows based on the maximum mean discrepancy with a negative distance kernel offer a robust framework for posterior sampling and conditional generative modeling in medical image analysis. By leveraging discrete Wasserstein gradient flows, the method effectively addresses the challenges of uncertainty in image reconstruction from noisy observations, enabling improved accuracy in applications such as super-resolution, inpainting, and computed tomography. The advantageous properties of the MMD, including efficient computation and scale equivariance, enhance the performance of the proposed approach, making it a valuable tool for tackling ill-posed inverse problems in imaging. The numerical examples demonstrate the method's capability to generate reliable reconstructions while providing insights into pixel-wise uncertainties. Overall, this research contributes to the advancement of generative modeling techniques in medical imaging, paving the way for future developments that could further refine diagnostic capabilities and enhance the quality of image analyses in clinical settings.

A Conditional Normalizing Flow for Accelerated Multi-Coil MR Imaging

Abstract:

Accelerated magnetic resonance (MR) imaging attempts to reduce acquisition time by collecting data below the Nyquist rate. As an ill-posed inverse problem, many plausible solutions exist, yet the majority of deep learning approaches generate only a single solution. We instead focus on sampling from the posterior distribution, which provides more comprehensive information for downstream inference tasks. To do this, we design a novel conditional normalizing flow (CNF) that infers the signal component in the measurement operator’s nullspace, which is later combined with measured data to form complete images. Using fastMRI brain and knee data, we demonstrate fast inference and accuracy that surpasses

recent posterior sampling techniques for MRI.
Conclusion:

In this work, we present the first conditional normalizing flow for posterior sample generation in multi-coil accelerated MRI. To do this, we designed a novel conditional normalizing flow (CNF) that infers the signal component in the measurement operator's nullspace, whose outputs are later combined with information from the measured space. In experiments with fastMRI brain and knee data, we demonstrate improvements over existing posterior samplers for MRI. Compared to score/Langevin-based approaches, our inference time is four orders-of-magnitude faster. We also illustrate how the posterior samples can be used to quantify uncertainty in MR imaging. This provides radiologists with additional tools to enhance the robustness of clinical diagnoses. We hope this work motivates additional exploration of posterior sampling for accelerated MRI.

A Comparative Study of Variational Autoencoders, Normalizing Flows, and Score-based Diffusion Models for Electrical Impedance Tomography

Abstract:

This study explores the application of deep generative models (DGMs) in Electrical Impedance Tomography (EIT), a technique used in various fields such as industrial inspection, geophysical prospecting, and medical imaging. The inherent nonlinearity and ill-posedness of EIT image reconstruction pose significant challenges for traditional regularization techniques, particularly in selecting appropriate regularization terms and the reliance on prior knowledge. The research investigates three specific DGMs—variational autoencoder networks, normalizing flow, and score-based diffusion models—aimed at learning implicit regularizers for EIT imaging.

The paper begins with an overview of EIT imaging and its inverse problem formulation, followed by the introduction of three algorithms based on the respective DGMs for solving EIT inverse problems. Numerical and visual experiments reveal two key findings: first, there is no single DGM that consistently outperforms the others across all scenarios; second, the conditional normalizing flow (CNF) model demonstrates superior generalization for low-level noise when reconstructing objects with anomalies from a training dataset, while the conditional score-based diffusion model (CSD*) excels in high-level noise conditions. The authors hope these preliminary results will stimulate further research in leveraging DGMs for EIT imaging challenges.

Conclusion:

In conclusion, this study evaluated the effectiveness of three deep generative models—conditional variational autoencoder (CVAE), conditional normalizing flow (CNF), and conditional score-based diffusion model (CSD*)—for Electrical Impedance Tomography (EIT) image reconstruction, generalization, and efficiency. The results indicate that CVAE performs well in high-level noise scenarios with fewer anomalies, while CSD* is more effective with a greater number of anomalies under similar noise conditions. CNF excels in low-level noise regardless of the number of anomalies. However, generalization results for CVAE were less satisfactory, while CNF and CSD* showed stronger performances in their respective noise conditions.

From an efficiency standpoint, CVAE has a smaller architecture and fewer parameters, making it the most efficient model, whereas CNF, with its larger parameter size, is more time-consuming. CSD* maintains reasonable efficiency by utilizing the score-based diffusion model as a post-processing step, thus avoiding the need for repetitive training with changing datasets. Importantly, no single model consistently outperformed the others across all noise scenarios. Future research directions include scaling the methodology to larger images, using hierarchical DGMs for enhanced robustness, integrating graph neural networks for improved flexibility, and incorporating physical models into DGMs for better convergence and interpretability. Additionally, the application of these methods to other imaging challenges, such as optical diffraction tomography and electromagnetic imaging, is also proposed.

4.

IUNETS: LEARNABLE INVERTIBLE UP- AND DOWNSAMPLING FOR LARGE-SCALE INVERSE PROBLEMS

Abstract:

U-Nets have been established as a standard neural network architecture for image-to-image problems such as segmentation and inverse problems in imaging. For high-dimensional applications, as they for example appear in 3D medical imaging, U-Nets however have prohibitive memory requirements. Here, we present a new fully-invertible U-Net-based architecture called the iUNet, which allows for the application of highly memory-efficient backpropagation procedures. As its main building block, we introduce learnable and invertible up- an downsampling operations. For this, we developed an open-source implementation in Pytorch for 1D, 2D and 3D data.

Conclusion:

This work presents a fully invertible U-Net (iUNet) featuring learnable invertible up- and downsampling through orthogonal convolutional operators. The iUNet demonstrated superior performance in 3D learned post-processing for CT reconstructions and volumetric segmentation, particularly excelling in post-processing tasks due to its increased depth and width. Future research aims to explore the iUNet's applicability to various tasks, including normalizing flows, which necessitate full invertibility, highlighting its potential for generative applications.

Deep learning in photoacoustic tomography: current approaches and future directions

Abstract:

Biomedical photoacoustic tomography (PAT) is an emerging imaging technique that utilizes optical absorption to produce high-resolution 3D images of soft tissues, moving towards clinical applications. However, challenges arise from the need for rapid image formation and limitations in data acquisition due to clinical workflows. Traditional image reconstruction methods struggle with incomplete or imperfect data, while deep learning (DL) has gained traction as a promising solution. This review examines current trends in learned image reconstruction, framing these approaches within classical methods from a Bayesian perspective. It includes a tutorial demonstration of three prototypical DL approaches, with code and datasets available for research. The authors suggest that DL could significantly impact in vivo applications where data is sparse and fast imaging is essential. Future research directions are also discussed, emphasizing the potential of DL in overcoming the complexities of PAT image reconstruction.

Conclusion:

The diversity of the work that has been done on learned image reconstruction in PAT in just the last few years, and the increasing rate at which it is being produced, suggests that the field will continue to develop for some time. In particular, we notice that already a move has begun from straightforward proof-of-concept applications of DL to more sophisticated approaches. Nevertheless, there are many issues that remain to be addressed. For instance, on the one hand there are model agnostic reconstruction pipelines using fully learned approaches that get a lot of attention due to low latency. On the other hand, as described already, there are learned reconstructions that use a physical model in combination with a network, which have been shown to be more stable and require less training data but are (considerably) slower in providing a reconstruction

WINNet: Wavelet-Inspired Invertible Network for Image Denoising

Abstract:

This paper presents a wavelet-inspired invertible network (WINNet) for image denoising, combining the strengths of model-based and learning-based approaches. WINNet consists of K-scale lifting inspired invertible neural networks (LINNs), a sparsity-driven denoising network, and a noise estimation network. LINNs leverage the lifting scheme from wavelets to learn a non-linear redundant transform with a perfect reconstruction property, aiding in noise removal. The denoising network applies sparse coding for effective denoising, while the noise estimation network assesses the noise level to adaptively adjust soft-thresholds in LINNs. By producing a redundant multi-scale representation, WINNet enables effective denoising and reconstructs the image using the inverse transform of LINNs. The simulation results demonstrate that WINNet offers high interpretability, strong generalization to unseen noise levels, and competitive performance in both non-blind and blind image denoising, as well as image deblurring.

Conclusion:

This paper presents the wavelet-inspired invertible network (WINNet), which combines K levels of lifting inspired invertible neural networks (LINNs) with sparsity-driven denoising networks. LINNs replicate wavelet transform properties to achieve perfect reconstruction, while the denoising network effectively reduces noise in detail coefficients and adapts to various noise levels. With an integrated noise estimation network, WINNet excels in blind image denoising and image deblurring tasks. Future research may involve using WINNet in other deep learning applications, enhancing interpretability, and improving efficiency, along with exploring its non-linear image approximation capabilities.

Wavelet-Based Texture Reformation Network for Image Super-Resolution
Abstract:
This paper introduces the Wavelet-based Texture Reformation Network (WTRN) for reference-based image super-resolution (RefSR). Unlike traditional methods that use raw features from a pretrained VGG encoder, WTRN leverages wavelet transformation to decompose texture features into low-frequency and high-frequency subbands. By performing feature matching on the low-frequency component and transferring wavelet-domain features based on correlation maps, WTRN enhances texture transfer. A wavelet-based texture adversarial loss further improves visual plausibility. Experimental results on four

benchmark datasets show that WTRN outperforms existing RefSR methods both quantitatively and qualitatively, providing a more effective approach to high-resolution image reconstruction.

Conclusion:
In this paper, we have proposed a Wavelet-based Texture Reformation Network (WTRN) for RefSR, which matches the low-frequency sub-band of features and progressively transfers wavelet-domain features as the spatial level increases. Besides, a wavelet-based texture adversarial loss, which fully focuses on the high-frequency sub-bands of the generated images, has been proposed to generate more vivid textures. The quantitative and qualitative evaluations on several large-scale public benchmarks have demonstrated the superior performance of our proposed method compared to previous methods.

WPPNets and WPPFlows: The Power of Wasserstein Patch Priors for Superresolution

Abstract:
This paper presents a novel approach for image super-resolution using Wasserstein patch priors (WPP) as data-driven regularizers within a variational framework. While WPP has shown promise in reconstructing high-resolution images from low-resolution observations, it traditionally requires solving a computationally expensive non-convex minimization problem for each input image. To address this, the authors propose two unsupervised neural network architectures based on WPP loss functions.

The first architecture, WPPNet, integrates convolutional neural networks (CNNs) to efficiently apply the learned model to any input image after training. The second approach incorporates conditional normalizing flows to facilitate uncertainty quantification in the reconstruction process. Numerical experiments demonstrate that WPPNet achieves excellent performance in super-resolution across various image classes, even when the forward operator is only approximately known.

The paper positions its method within the context of inverse problems, specifically super-resolution, and contrasts it with existing neural network-based techniques that often require large datasets of paired high- and low-resolution images. The proposed framework effectively leverages image patch distributions and allows for more efficient and robust image restoration compared to traditional methods.

Conclusion:

We introduced WPPNets, which are CNNs trained with a new loss function based on comparisons of empirical patch distributions via the quadratic Wasserstein distance and demonstrated its power by several numerical examples. In particular, we observed that WPPNets are very stable under inaccurate operators appearing in real-world applications. Due to the fact that WPPs require the knowledge of one high-resolution reference image, WPPs could also be interpreted as a method for one-shot learning, see [9, 68] and references therein. However, as no low-resolution correspondence to the reference image is given, we would call our WPP based methods an unsupervised learning method. We measured the uncertainty within the reconstructions by combining WPPs with conditional normalizing flows.

Manifold Learning by Mixture Models of VAEs for Inverse Problems

Abstract:

This paper proposes a method for representing high-dimensional data manifolds using a mixture model of variational autoencoders (VAEs). This approach allows for the representation of arbitrary topologies by using multiple encoder-decoder pairs, where each pair corresponds to a chart of the manifold. The authors introduce a loss function for maximum likelihood estimation of the model weights and select an architecture that provides analytical expressions for the charts and their inverses.

Once the manifold is learned, it can be applied to solve inverse problems by minimizing a data fidelity term constrained to the learned manifold. To tackle the resulting minimization problem, the authors propose a Riemannian gradient descent algorithm that operates on the learned manifold.

The paper highlights the limitations of traditional methods, such as principal component analysis (PCA), which assume low-dimensional subspaces for high-dimensional data. It contrasts these with the manifold hypothesis, which suggests that complex data sets reside within low-dimensional manifolds. The proposed method improves upon existing approaches that struggle with disconnected manifolds or require a priori knowledge of the manifold's topology.

Numerical experiments demonstrate the effectiveness of the proposed approach in low-dimensional toy examples, as well as in applications such as deblurring and electrical impedance tomography on specific image manifolds. Overall, this work contributes to the field of manifold learning by providing a flexible and efficient framework for representing complex data structures.

Conclusion:

In this paper we introduced mixture models of VAEs for learning manifolds of arbitrary topology. The corresponding decoders and encoders of the VAEs provide analytic access to the resulting charts and are learned by a loss function that approximates the negative log-likelihood function. For minimizing functions F defined on the learned manifold we proposed a Riemannian gradient descent scheme. In the case of inverse problems, F is chosen as a data-fidelity term. Finally, we demonstrated the advantages of using several generators on numerical examples