Convolutional dictionary learning based auto-encoders for natural exponential-family distributions

Bahareh Tolooshams[*1], Andrew H. Song[*2], Simona Temereanca[3], and Demba Ba[1]

[1]Harvard University [2]Massachusetts Institute of Technology [3]Brown University

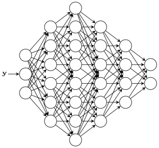
[*]Equal contributions

CRISP Group: https://crisp.seas.harvard.edu

ICML 2020

Motivation

Deep Learning



- Fast and scalable ✓
- Not interpretable ✗
- Memory and computationally expensive ✗

Signal Processing (SP)

Generative models

e.g., sparse coding model

$p(\mathbf{y} \mid \mathbf{x}) = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}, \quad \mathbf{x}$ is sparse

- Slow and not scalable ✗
- Interpretable ✓
- Memory efficient ✓

⬇

- Benefit from scalability of deep learning for traditional SP tasks.
- Guide to design interpretable and memory efficient networks.

Convolutional Dictionary Learning (CDL)

Generative model for each data j

$$\mathbf{y}^j = \sum_{c=1}^{C} \mathbf{h}_c * \mathbf{x}_c^j + \boldsymbol{\epsilon}^j = \mathbf{H}\mathbf{x}^j + \boldsymbol{\epsilon}^j, \quad \boldsymbol{\epsilon}^j \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

where $\mathbf{x}_c^j$ is sparse.

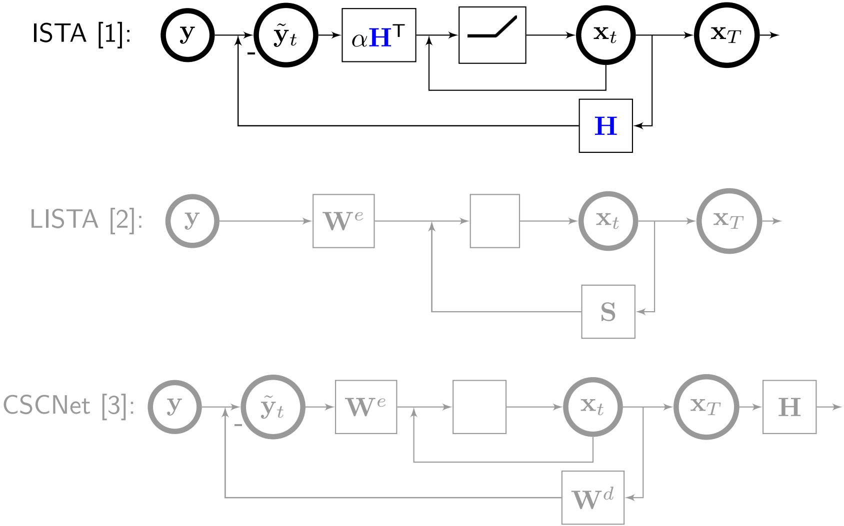
Goal: Learn $\mathbf{H}$ that maps sparse representation $\mathbf{x}^j$ to data $\mathbf{y}^j$.

$$\min_{\{\mathbf{h}_c\}_{c=1}^{C}, \{\mathbf{x}^j\}_{j=1}^{J}} \frac{1}{2} \sum_{j=1}^{J} \|\mathbf{y}^j - \mathbf{H}\mathbf{x}^j\|_2^2 + \lambda \|\mathbf{x}^j\|_1$$

- min w.r.t. $\mathbf{x}^j$ $\rightarrow$ *Convolutional Sparse Coding (CSC)*.
- min w.r.t. $\mathbf{H}$ and $\mathbf{x}^j$ $\rightarrow$ *Convolutional Dictionary Learning (CDL)*.

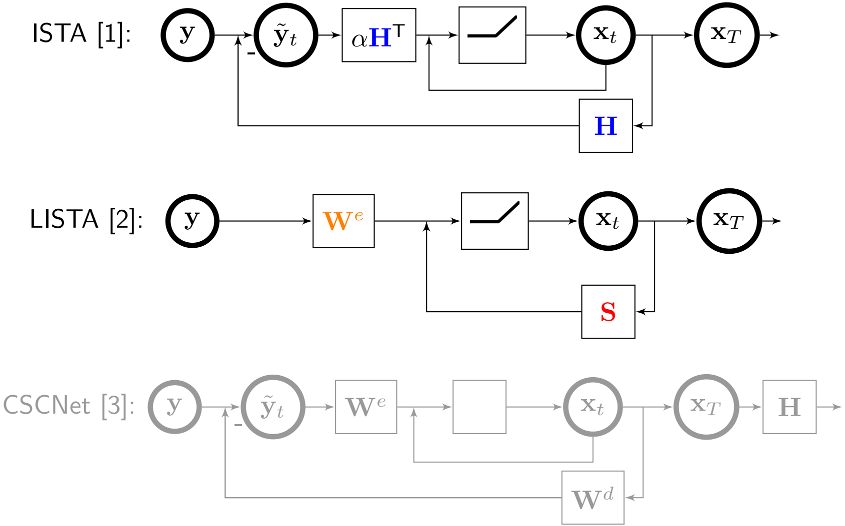
Unfolding Networks

Solve CSC and CDL by iterative proximal gradient algorithm.



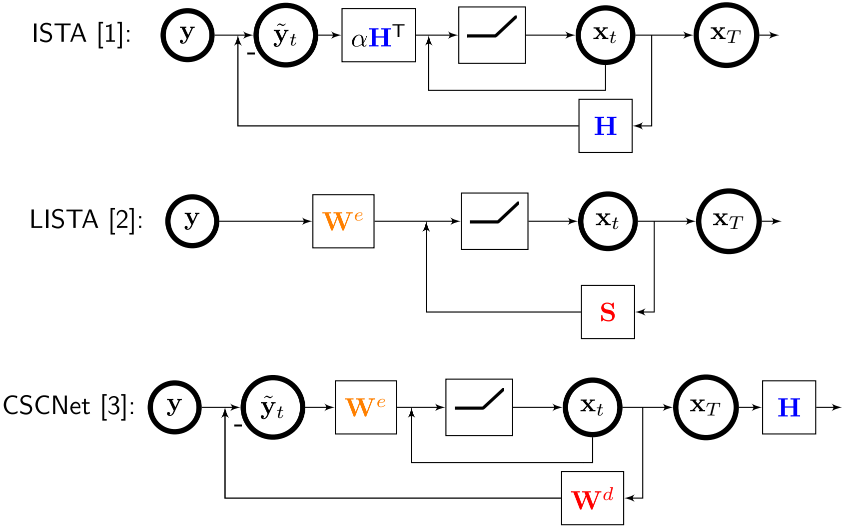
Unfolding Networks

Solve CSC and CDL by iterative proximal gradient algorithm.



Unfolding Networks

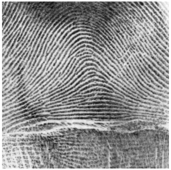
Solve CSC and CDL by iterative proximal gradient algorithm.



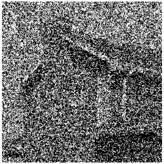
What if the observations are no longer Gaussian?

Count-valued data

Fingerprint



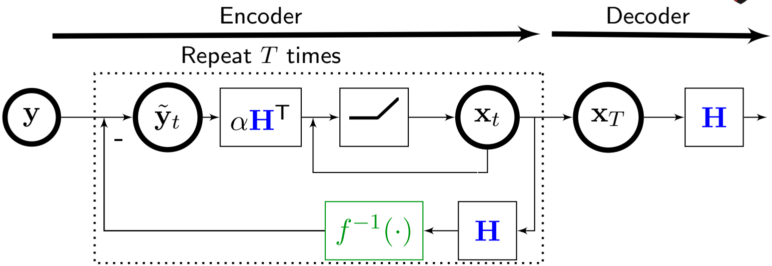
Photon-based imaging



Classical CDL approach: Alternating minimization with a Poisson generative model [4, 5].

- Unsupervised ✓
- Follows a generative model $\Rightarrow$ interpretable ✓
- Not scalable (can take minutes $\sim$ hours to denoise single image) ✗

Our Contributions



- Auto-encoder inspired by CDL, termed **D**eep **C**onvolutional **E**xponential **A**uto-encoder (**DCEA**), for non real-valued data
- Demonstration of the flexibility of DCEA for both
  - *unsupervised* task, e.g., CDL
  - *supervised* task, e.g., Poisson denoising problem
- Gradient dynamics of shallow exponential auto-encoder (SEA)
  - Prove that SEA recovers parameters of the generative model.

Deep Convolutional Exponential Auto-encoder

Problem description

Natural exponential family with convolutional generative model:

$$\log p(\mathbf{y}|\boldsymbol{\mu}) = f(\boldsymbol{\mu})^{\mathsf{T}}\mathbf{y} + g(\mathbf{y}) - B(\boldsymbol{\mu}), \text{ where } f(\boldsymbol{\mu}) = \mathbf{H}\mathbf{x}, \ \mathbf{x} \text{ is sparse.}$$

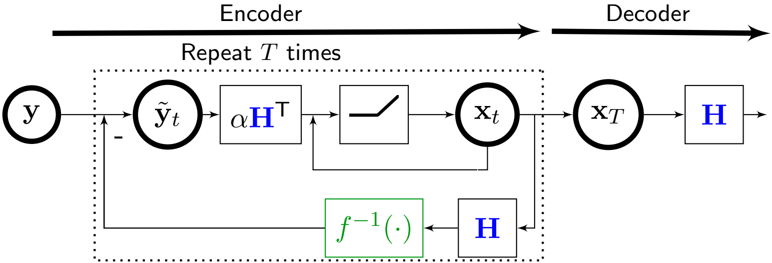
	$\mathbf{y}$	B(z)	Inverse link: $f^{-1}(\cdot)$
Gaussian	$\mathbb{R}$	$\mathbf{z}^{\mathsf{T}}\mathbf{z}$	$I(\cdot)$
Binomial	$[0..M]$	$-\mathbf{1}^{\mathsf{T}}\log(\mathbf{1}-\mathbf{z})$	sigmoid$(\cdot)$
Poisson	$[0..\infty)$	$\mathbf{1}^{\mathsf{T}}\mathbf{z}$	$\exp(\cdot)$

Exponential Convolutional Dictionary Learning (ECDL):

$$\min_{\mathbf{H},\mathbf{x}} \overbrace{-\log p(\mathbf{y}|\boldsymbol{\mu})}^{\text{negative log-likelihood}} + \overbrace{\lambda\|\mathbf{x}\|_1}^{\text{code sparsity constraint}}$$

Deep Convolutional Exponential Auto-encoder

Network architecture

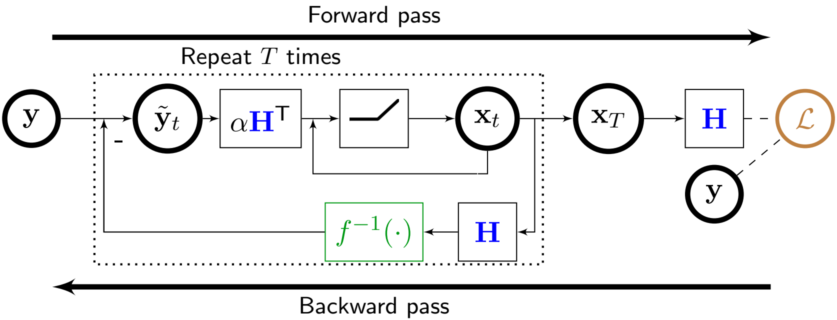


Components for different distributions

	$\mathbf{y}$	$f^{-1}(\cdot)$	Encoder Unfolding ($\mathbf{x}_t$)	Decoder ($f(\hat{\boldsymbol{\mu}})$)
Gaussian	$\mathbb{R}$	$I(\cdot)$	$\mathcal{S}_b\left(\mathbf{x}_{t-1} + \alpha\mathbf{H}^\mathsf{T}\widetilde{\mathbf{y}}_t\right)$	$\mathbf{H}\mathbf{x}_T$
Binomial	$[0..M]$	sigmoid$(\cdot)$	$\mathcal{S}_b\left(\mathbf{x}_{t-1} + \alpha\mathbf{H}^\mathsf{T}(\frac{1}{M}\widetilde{\mathbf{y}}_t)\right)$	$\mathbf{H}\mathbf{x}_T$
Poisson	$[0..\infty)$	$\exp(\cdot)$	$\mathcal{S}_b\left(\mathbf{x}_{t-1} + \alpha\mathbf{H}^\mathsf{T}(\text{Elu}(\widetilde{\mathbf{y}}_t))\right)$	$\mathbf{H}\mathbf{x}_T$

Deep Convolutional Exponential Auto-encoder

Training & inference



Training

- **Forward pass**: Estimate code $\mathbf{x}_T$ & compute loss function.
- **Backward pass** (back-propagation): Estimate dictionary $\mathbf{H}$.
- Equivalent to *alternating minimization* in CDL.

Inference: Once trained, the inference (forward pass) is fast.

Unsupervised → Supervised

Repurpose DCEA for supervised tasks with two modifications

1. **Loss function**: Any supervised loss function, e.g., reconstruction MSE loss or perceptual loss.
2. **Architecture**: Relax the constraints → Untie the weights of encoder and decoder, learn the bias b.

	Encoder	Decoder
Original	$\mathbf{x}_t = \mathcal{S}_b\big(\mathbf{x}_{t-1} + \alpha \mathbf{H}^\mathsf{T}(\mathbf{y} - f^{-1}\big(\mathbf{H}\mathbf{x}_{t-1}\big))\big)$	$\mathbf{H}\mathbf{x}_T$
Relaxed	$\mathbf{x}_t = \mathcal{S}_b\big(\mathbf{x}_{t-1} + \alpha (\mathbf{W}^e)^\mathsf{T}(\mathbf{y} - f^{-1}\big(\mathbf{W}^d\mathbf{x}_{t-1}\big))\big)$	$\mathbf{H}\mathbf{x}_T$

- Further relaxations possible, i.e., deep & non-linear decoder.

Experiments

Poisson image denoising

Baseline frameworks

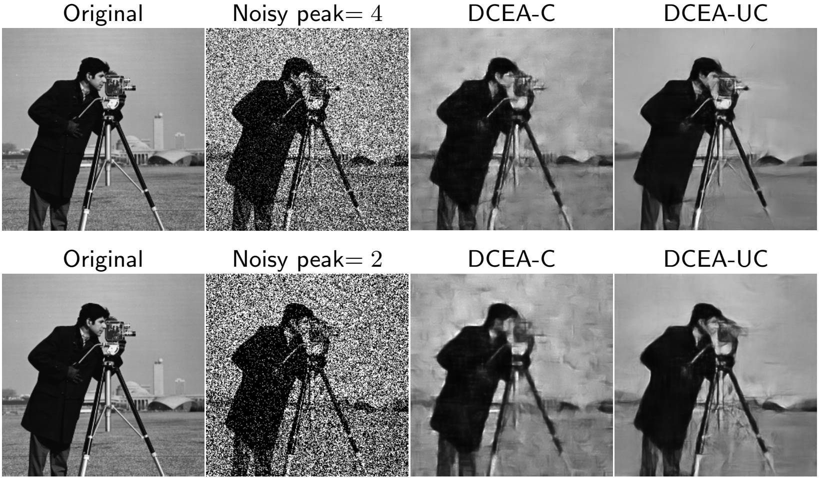
	Supervised?	Description
SPDA [5]	✗	ECDL + patch-based
CA [6]	✓	denoising NN
DCEA-C (ours)	✓	constrained DCEA (tied weights)
DCEA-UC (ours)	✓	unconstrained DCEA (untied weights)

PSNR performance on test dataset

		SPDA	CA	DCEA-C	DCEA-UC
Peak 1	Set12	20.39	**21.51**	20.72	21.37
	BSD68	.	21.78	21.27	**21.84**
Peak 2	Set12	21.70	**22.97**	22.02	22.79
	BSD68	.	22.90	22.31	**22.92**
Peak 4	Set12	22.56	**24.40**	23.51	24.37
	BSD68	.	23.98	23.54	**24.10**
# of Params		160K	655K	20K	61K

Experiments

Poisson image denoising



Experiments

Poisson image denoising

		SPDA	CA	DCEA-C	DCEA-UC
Peak 1	Set12	20.39	**21.51**	20.72	21.37
	BSD68	.	21.78	21.27	**21.84**
Peak 2	Set12	21.70	**22.97**	22.02	22.79
	BSD68	.	22.90	22.31	**22.92**
Peak 4	Set12	22.56	**24.40**	23.51	24.37
	BSD68	.	23.98	23.54	**24.10**
# of Params		160K	655K	20K	61K

- **Classical ECDL**: SPDA vs. **DCEA-C**

  ⇒ better denoising + *much more efficient*

  ⇒ classical inference task leveraging **scalability of NN**

- **Denoising NN**: CA vs. **DCEA-UC**

  ⇒ competitive denoising + *much less parameters*

  ⇒ NN architecture leveraging **generative model paradigm**

Experiments

CDL for simulated binomial data

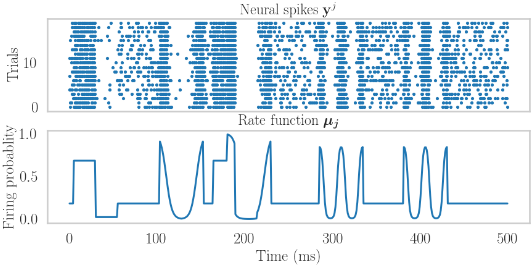


Figure: Example of simulated neural spikes and the rate (truth)

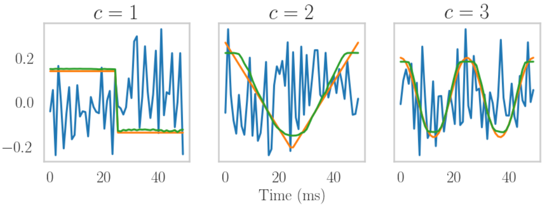
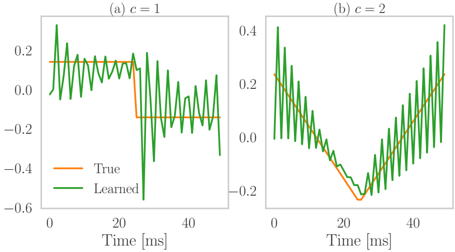


Figure: Random initialized (Blue), true (Orange), and learned templates (Green)

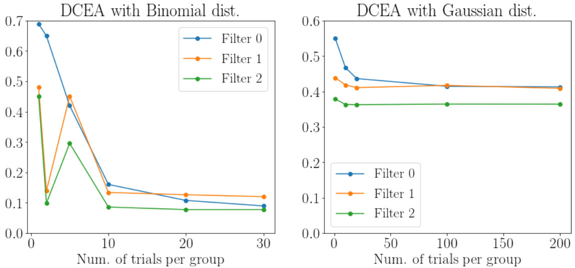
Experiments

CDL for simulated binomial data

- If we untie the weights, i.e., relax generative model constraints



- If we treat binomial data as Gaussian obs., i.e., model mismatch



Conclusion

In conclusion, **D**eep **C**onvolutional **E**xponential **A**uto-encoder (DCEA)

- is a class of NN based on a generative model for CDL, using data from natural exponential family.
- shows competitive performance in Poisson denoising tasks against SOTA frameworks, *with an order of magnitude fewer* trainable parameters (**supervised task**).
- is able to learn accurate convolutional patterns in ECDL task with simulated binomial and real neural spiking observations (**unsupervised task**).

Reference

I. Daubechies, M. Defrise, and C. De Mol.
An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.
Communications on Pure and Applied Mathematics, 57(11):1413–1457, 2004.

Karol Gregor and Yann Lecun.
Learning fast approximations of sparse coding.
In *International Conference on Machine Learning*, pages 399–406, 2010.

D. Simon and M. Elad.
Rethinking the CSC model for natural images.
In *Proc. Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2019.

Joseph Salmon, Zachary Harmany, Charles-Alban Deledalle, and Rebecca Willett.
Poisson noise reduction with non-local pca.
Journal of Mathematical Imaging and Vision, 48(2):279–294, Feb 2014.

Raja Giryes and Michael Elad.
Sparsity-based poisson denoising with dictionary learning.
IEEE Transactions on Image Processing, 23(12):5057–5069, 2014.

Tal Remez, Or Litany, Raja Giryes, and Alexander M. Bronstein.
Class-aware fully-convolutional gaussian and poisson denoising.
CoRR, abs/1808.06562, 2018.