Learned Convolutional Sparse Coding

Hillel Sreter & Raja Giryes

ICASSP 2018





Outline

- Motivation and applications
- Brief sparse coding background
- (ASC) Approximate sparse coding
- (CSC) Convolution sparse coding
- (ACSC) Approximate Convolution sparse coding

ACSC - real-time image processing

Sparse coding is a strong image processing tool

Sparse coding algorithms tend to be quite slow

Train an approximate efficient sparse coding model

Keep the quality. Drop the runtime

ACSC - advantages

Sparse coding has been shown to work well for Image processing task. Yet, it tends to be slow \implies Requires multiple minutes on CPU.

Approximate convolutional sparse coding cuts the runtime to a fraction while maintaining the sparse coding performance.
 ➡ Requires less than a second on a CPU.

ACSC - Image denoising example



ACSC (0.7 sec) PSNR 29.88 dB KSVD (70 sec) PSNR 28.67 dB

x100 speedup compared to OMP sparse coding in KSVD

ACSC - Document S&P denoising

CHAFTER II.



UT returning to our argument, since all knowledge and moral purpose aims at some good, what is the good which we say the political science has in view? or, in other

words, what is the highest of all the goods attainable by action? Now, as fat as the name goes, there is a

pretty general agreement: for both the eduented and the uncducated call it happiness, and 'living well' and 'doing well' they conscive to be synonymous with 'being happy' but as to the nature of this happiness, they disagree, and the multitude do not in their necount of it agree with the wise. The mass of manified define, it as something palpable and apparent as pleasure or wealth or honoursome one and some another way, oftentines the same man gives a different definition of it at different times; for if he is ill, he thinks

CHAFTER II



UT returning to our argument, since all knowledge and moral purpose aims at some good, what is the good which we say the political science has in view? or, in other

words, what is the highest of all the goods attainable by action ?

Now, as far as the name goes, there is a pretty general agreement: for both the educated and the uneducated call it happiness, and 'living well' and 'doing well' they conceive to be synonymous with 'being happy;' but as to the nature of this happiness, they disagree, and the multitude do not in their account of it agree with the wise. The mass of mankind define it as something palpable and apparent, as pleasure or wealth or honour; some one and some another: nay, oftentimes the same man gives a different definition of it at different times; for if he is ill, he thinks

ACSC (0.43 sec) PSNR 30.2 dB

ACSC - Document inpainting

CHAPTER IC



JT returning to out argument, since all knowledge and moral purpose since at some good, what is the good which we say the political science has in view 2 or, in other

words, what is the highest of all the goods atminuble by action $\hat{\gamma}$

Now, as far as the name goes, there is a protty general agreement. for both the educated and the uncducated call it Lappiness, and 'living well' and 'doing well' they conorive to be synonymous with 'heing happy :' but as to the nature of this happiness, they disagree, and the maliitude do not in their account of it agree with the wise. The mass of menkind define it as something paipable and apparent, as pleasure or weath or honeur; some one and some another: nay, oftentiones the same man gives a different definition of it at different times; for if he is ill, be thinks

CHAPTER II



UT returning to our argument, since all knowledge and moral purpose aims at some good, what is the good which we say the political science has in view? or, in other

words, what is the highest of all the goods attainable by action ?

Now, as far as the name goes, there is a pretty general agreement: for both the educated and the uneducated call it happiness, and 'living well' and 'doing well' they concoive to be synonymous with 'being happy;' but as to the nature of this happiness, they disagree, and the multitude do not in their account of it agree with the wise. The mass of mankind define it as something palpable and apparent, as pleasure or wealth or honour; some one and some another: nay, offentimes the same man gives a different definition of it at different times; for if he is ill, he thinks

ACSC (0.43 sec) PSNR 30.13 dB

Sparse Code problem setup

Given An input X and an over complete matrix D Find Z^* such that:

 $X = DZ^* \ ||Z^*||_0 < ||Z||_0 \ \forall Z \in \{Z|X = DZ\}$

Sparse Code problem setup

Given An input X and an over complete matrix D Find Z^* such that:

 $X = DZ^* ~~ ||Z^*||_0 < ||Z||_0 ~~ orall Z \in \{Z|X = DZ\}$

Solving the above is intractable due to the ℓ_0 norm. Can be approximated by ℓ_1 relaxation Leads to the LASSO minimization problem

$$Z^*_{lasso} = \mathbf{argmin}_Z ||X - DZ||_2^2 + \lambda ||Z||_1$$

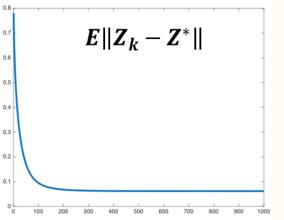
Sparse Code setup - solving LASSO

Use the proximal gradient technique -

$$egin{aligned} f(x) &= g(x) + h(x) \ x^{(k)} &= \mathrm{prox}_{t_k,h}(x^{(k-1)} - t_k
abla g(x^{(k-1)})) \ \| \mathrm{ISTA} \ \| \mathbf{Z}_{k+1} &= S_{\lambda/L}(Z_k + D^T(X - DZ_k)) \end{aligned}$$

This is an iterative method

Many iterations may be required for convergence!



Sparse Code setup - solving LASSO

Long convergence time of ISTA can be an issue for time sensitive applications

→ An acceleration strategy that approximates the solution is required

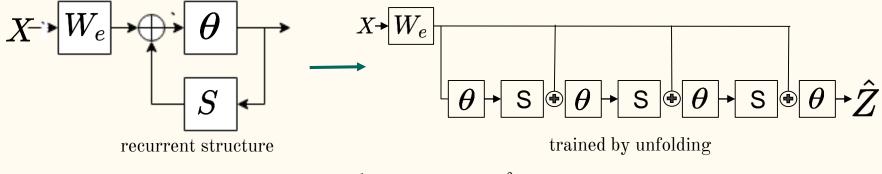
Learned ISTA (LISTA): Replaces linear operations and threshold in ISTA with learned ones

$$egin{aligned} Z_{k+1} &= S_{\lambda/L}((I-D^TD)Z_k+D^TX) \ &igsquare{1} \ &igstyute{1} \ &igst$$

Using Learned-ISTA as a SC approximation

Given a fixed number of iterations learn the matrices S and W and the thresholds $oldsymbol{ heta}$

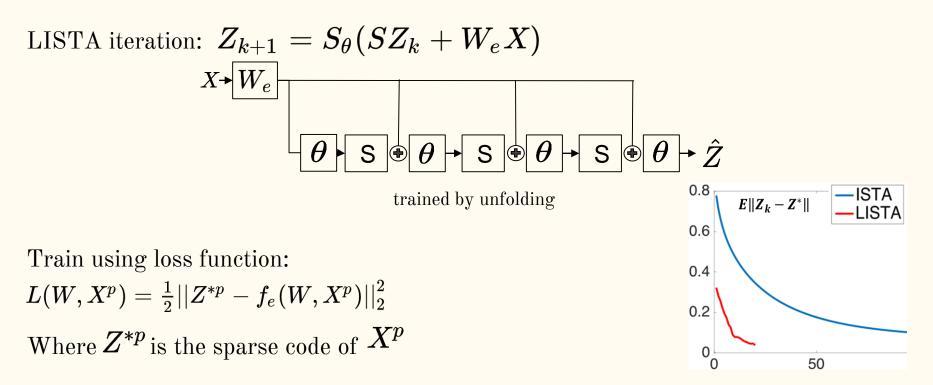
$$Z_{k+1} = S_{ heta}(SZ_k + W_eX)$$



 $L(W,X^p) = rac{1}{2} ||Z^{*p} - f_e(W,X^p)||_2^2$

Learning Fast Approximations of Sparse Coding - Karol Gregor and Yann LeCun 12

LISTA training



Learning Fast Approximations of Sparse Coding - Karol Gregor and Yann LeCun ¹³

LISTA disadvantages

LISTA is a patch based method.

Therefore, we have

- Loss of spatial information.
- Inefficient for large patches
- Image properties not incorporated (i.e. translation invariance).

Solution: Use convolutional structure.

Approximate Convolutional Sparse Coding (ACSC)

Learn a **convolutional** sparse coding of the whole image instead of patches → A global algorithm:

- Image is processed as whole.
- Efficient.
- Convolution based \rightarrow inherently incorporates image properties.

This overcomes the disadvantages of LISTA being a patch based method:

- Loss of spatial information.
- Not Inefficient for large patches
- Image properties not incorporated (i.e. translation invariance)

Standard Convolutional Sparse Coding (CSC)

In convolutional sparse coding we replace the regular dictionary with a convolutional one:

Given input X, solve:

$$\mathbf{argmin}_{d,z} = ||X - \sum_{k=1}^{K} d_k * Z_k||_2^2 + \beta \sum_{k=1}^{K} ||Z_k||_1 \quad s. t. ||d_k||_2 \le 1$$

In other word --

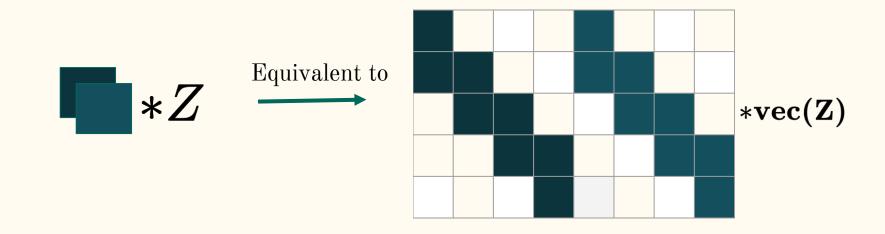
Find a set of kernels d that reconstruct X with sparse feature maps $\mathbf Z$

$$X = \sum_{k} \mathbf{z}_{k} \times \mathbf{z}_{k}$$

Convolutional Sparse Coding as Regular Sparse Coding

Convolutional Sparse Coding (CSC) \longleftrightarrow Regular Sparse Coding (SC)

A convolutional dictionary is a concatenation of toeplitz matrices



From SC to CSC

Rewrite ISTA as a solution to CSC problem

Assuming **D** is a concatenation of toeplitz matrices we can replace matrix multiplication with convolution

D is a toeplitz matrix

$$egin{aligned} & Z_{k+1} = S_{\lambda/L}((I-D^TD)Z_k+D^TX) \ & \downarrow ext{Convolutional ISTA} & \hat{d} = d^T \ & Z_{k+1} = S_{\lambda/L}(Z_k - \hat{d} * (d * Z_k) + \hat{d} * X) \ & d \in \mathbb{R}^{s imes s imes M imes C} & \hat{d} \in \mathbb{R}^{s imes s imes C imes M} \end{aligned}$$

From CSC to Approximate CSC

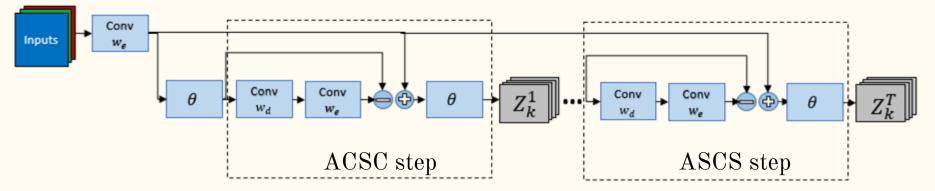
Approximate Convolutional sparse coding (ACSC): Replaces convolutional operations and threshold in Convolutional ISTA solution with learned ones

$$egin{aligned} &Z_{k+1} = S_{\lambda/L}(Z_k - \hat{d} * (d * Z_k) + \hat{d} * X) \ &\downarrow & \mathsf{ACSC} \ &\downarrow & \mathsf{ACSC} \ &Z_{k+1} = S_{ heta}(Z_k - w_e * (w_d * Z_k) + w_e * X) \ &w_d \in \mathbb{R}^{s imes s imes M imes C} &w_e \in \mathbb{R}^{s imes s imes C imes M} \end{aligned}$$

Approximate CSC architecture

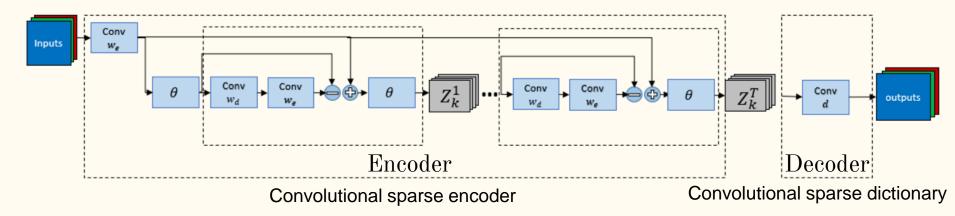
$$Z_{k+1} = S_{ heta}(Z_k - w_e * (w_d * Z_k) + w_e * X)$$

Proposed recurrent architecture for CSC with unfold of T steps



Approximate CSC with dictionary learning

Train end to end as a sparse autoencoder



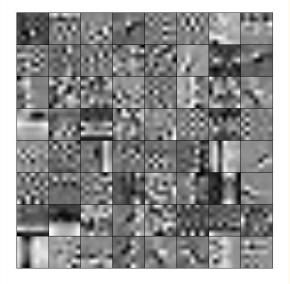
We learn in our framework both the sparse coding and the dictionary In LISTA, only the sparse coding is being learned

Training ACSC autoencoder

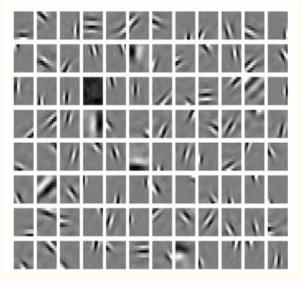
- Initialization is important $\longrightarrow w_d = d = \mathbf{flipud}(\mathbf{fliplr}(w_e))$
- We found kernel of size 7x7 to give best results.
- We found Unfolding 3 time step to be sufficient
- Use loss function $L(x, \hat{x}) = \alpha(1 \mathbf{ms_ssim}(x, \hat{x})) + (1 \alpha)||x \hat{x}||_1$ H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks,"

Approximate CSC and dictionary learning

Learned convolution dictionary



Classical sparse dictionary



Classical approach tends to learn the same atom with a translation. Convolution dictionary a single atom covers all translation.

Example result on denoising task



ACSC (0.56 sec) PSNR 32.11 dB KSVD (57 sec) PSNR 32.09 dB **x100 speedup** compared to OMP sparse coding in KSVD ²⁵

Example result on denoising task



ACSC (0.6 sec) PSNR 30.14 dB KSVD (64 sec) PSNR 30.05 dB **x100 speedup** compared to OMP sparse coding in KSVD

26

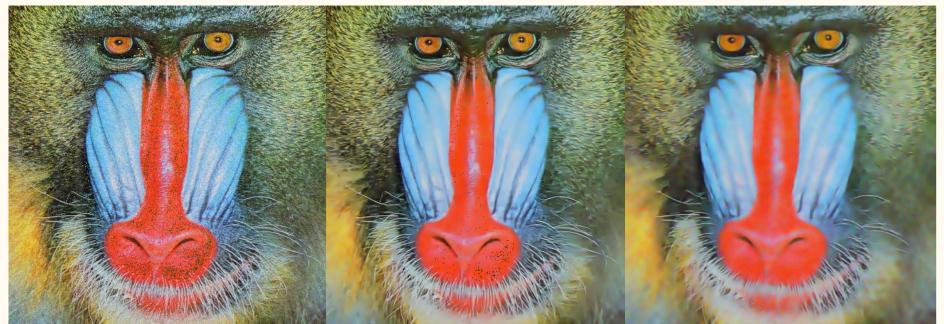
Example result on inpainting task





ACSC (0.57 sec) PSNR 32.3 dB

Example result on denoising task



ACSC (0.83 sec) PSNR 30.32 dB KSVD (78 sec) PSNR 30.21 dB **x100 speedup** compared to OMP sparse coding in KSVD ²⁸

Example result on inpainting task



ACSC (0.48 sec) PSNR 30.32 dB Heide et al.(320 sec) PSNR 30.21 dB

x650 speedup compared to Heide et al. (Fast and Flexible Convolutional sparse coding)

Example result on document denoising task

LEWKOWICZ	File No. Engl. 6
First Name Ludwig Sex m	Opening
Address Feldaring.DPC .	Date
Birthdate 8.9.23 Birthplace Warsohau	In transit from:
Nationality ; Present	Accompanied by
Occupation: PresentFormer	
Country of destination England.	Closing Date

INDEX CARD A.J.D.C. EMIGRATION SERVICE	
LEWKOWICZ	File No. Engl.6
First NameLudwigSex	Opening
Address Feldafing.DPC.	Date
Birthdate 8.9.23 Birthplace Warschau	In transit from:
Nationality: Present	Accompanied by
Occupation:	
Present	Closing
Country of destination	Date

ACSC (0.44 sec) PSNR 29.8dB

Example result on document denoising task

the Small Claims Division, Durinet Gener B(1, 1, 0), Dr. and T Trial Data Control Science Sci	CRS 24A 260	SMATH CLAR	IS COMPLAINT	and the second s
Angle of the Angle of the Addition of the	He Small Claims Division, Dis	JON COUNT	Henkins	County
Anomy Excess 8: A sector of the sector	The second second second second	trice choire bi to Di		A Contraction of the second
And the second set of the second set of the second set of the second secon	Theman & Prentle	w ar way	Amorney Thomas E. Thom	r
Indicary 116 - KY 4631 Indicary 116 - KY 463 Indicary 116	123 Four Center	94	(If Any) F. O. Sox 1245.	and the second second
Index of the second secon	Madisonville, K	<u>r 42431</u>		SACATO STORAGE
Brandington, LULLA, TURNUM, LAN, VA ADGOL MARKET, C. M. AND ALL AND	efendant Tromas Padgett			
20207. Becontary of State. Evolution: NY 40001 PARAMENT OF ACT DEPENDANT - READ CAMPTULEY Two are being used in the Statel Chaine Diverse of Diverse Came Problem Vision Biology and States States on the Assessing States and States States on the States of States States on the Assessing States and States States on the States of			and the second	<u></u>
Movement in the Sheal Chaine Diverse of Diverse Cases the Mainfull (19) how a larves. The performance of the Sheal Chaine Diverse of Diverse Cases and the Sheal Chaine and the Sheal Chaine and Sheal Chaine and Sheal Chaine Chain Chaine Chaine Chaine Chaine Chaine Chaine Chaine Chaine Chaine	FRVE: Secretary of Sta	ate, Frankfort, XY 40	0601	·
By the state of the state	1994 C 40 C 40 C 40 C	NOTICE TO ÉACH DEFEN	DANT - READ CAREFULLY	行行机器的表现
Benefit of the end of the en	ny decision in this case. You al ispute. However, if you (all to appea	r in court on the date shown	on the attached summons a court judgeme	ai aha ba takén sesinti
Complexit Planetific here the defendence of the above 25, 1997, employed 21 statistifie to oppresent here a conservation of the above 25, 1997, employed 21 statistifie or oppresent here a conservation of admission of the above 26 statistic of				
19. Planeff claims that the defendent on the photo: August 23, 1892, Environd. Mathield E. an employed for an environment of the	If you have questions or new	d assistance; consult the Sm	all Claims information pamphiet or can a	tt erförnet.
Planetif shine the defendent. On Dr. phott: August 23, 1892. Environd. Matrieff 2: no constants International Matrieff 2: No constants International Matrieff 2: No c				1.
him an Concentration and Section 2014 according of the protocol according to the sections. The sections are concentrational according to the section of the				
PhageWindows fabries our spect	and advinced expension ment of employment, pay \$1195.30 as the reinforment of employment.	ses regarding said to . Transa Padgett bin e ancient ciding to Pia spenses adamend by P starset with free the	failed and refused and contin intiff as attorney less for th Jainiff The aforesaid fee a most by Plaintiff and other at	ris to refuse to it services and id expenses are tomeys for
		n part.	and the start of the start was stored	21
Cart Paor Kaiha Adjust Gas Konnto	Macel	-1	States and a state of the state	
Cant	Plaintiff's Signature or	Attorney	Date of Filing	
Cant			A STATE A STATE AND A STATE AND A STATE	and the second
(See Rown)			the Diane Number	and the second se
	Clark		Phone Number	A 3/ 50 10 52
	Sty and Carton Broken		Phone Number	
	St. and Calif. Broken		Phone Number	
	Sty and Carton Landon			
	Sty and Carton Landon	(§		
	Sty and Carton Landon	(g		
	Sty and Carton Landon	(§		4
사람은 승규는 가장 방법을 가지 않는 것을 잘 들었다. 나는 것이 나는 것이 나는 것이 없는 것이 없는 것이 없는 것이 없다. 것이 없는 것이 없는 것이 없는 것이 없는 것이 없는 것이 없는 것이 없다. 것이 없는 것이 없다. 것이 없는 것이 없 않는 것이 없는 것이 없 않는 것이 없는 것이 없 않이 않는 것이 없는 것이 않는 것이 없는 것이 없는 것이 없는 것이 없는 것이 않는 것이 않이	Sy and Carthy Login in			x
	Sy and Carthy Login in			4

n the Small Claims Di	ision, District Cept	TPL COUNT	Bookins	County	
	Distance of Dis	10 0.00	Assigned Trial I	Date:	
Thomas &	Brantley 81-	- h		MA E. Tharmer	
	Center St.		AF 4	. Sox 1265	
Madisony	1110. KY 42431		Mad	(servilio, KY 42431	
Defendant Thomas I			Attorney		
	m Castlo Lane	the second se	(If Any)		
Manutan	on Mille Hicht	man 48336			
SERVE: Secrotan	y of State. Pra	micfort. Sy 40601	_		
	MATTER	O EACH DEFENDANT	- READ CAREF	ULLY	
				hown above. The judge has	

However, if you fail to appear in sourt on the data shown on the atached summens a court judgement can be taken against you for the money or property descaled in the following completen. This good is not summitment of your pay check and/or sale of you herma or of the t-belonging fudnet protected by her it you shift with folgement.

If you have outstituts or need assistance, consult the Small Claims Information pamphiet or call an attorney.

Complaint .

- 1. Plaintf chain that the defendant, Ph. OT. about August 25, 1992, employed Elatoritief on an advantage of the environment of the environment
- 11ke sorvices. 2. Planuit claim the following sum from the defendent for damages brought by the above somplaint \$ 1194.30 plan inforces they some at the rule of 18% per amount from Area 11, 1963

12.12

ACSC (0.38 sec) PSNR 29.4dB

Conclusion

- LISTA is good for accelerating sparse coding for patches
- Yet, it is slow when working on images as a whole
- Our proposed learned CSC is good for processing the whole image
- Provide comparable performance to other sparse coding based techniques on images with up to x100 speedup

Thank You!

