

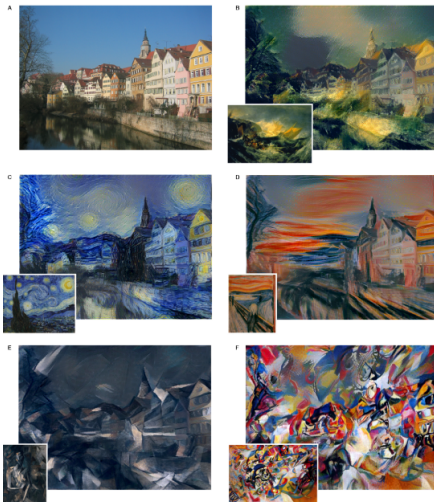
Why does classification help generation and not the other way around?

Roland Memisevic

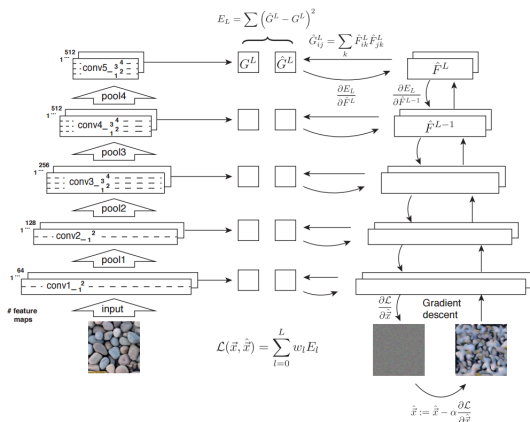
University of Montreal & TwentyBN

MIT Deep Learning Workshop 2016

Gatys, Ecker, Bethge 2015

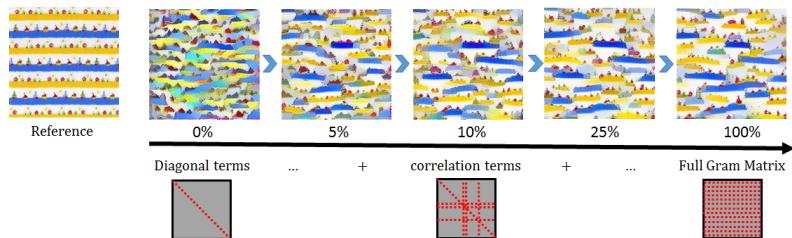


Gatys, Ecker, Bethge 2015

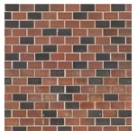
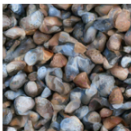


- ▶ Compute the spatial average of all pairs of products of features for source image x (“Gramians”)
- ▶ Optimize another image \hat{x} so as to make the Gramians match those of x

What do the Gramians represent? (joint work with Guillaume Berger)



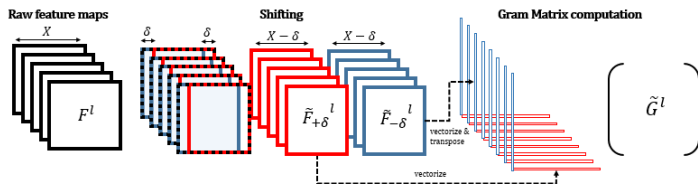
- ▶ They are static
- ▶ They represent local coincidences across features
- ▶ On the lowest layer, this can represent the relative “arrival phases” of individual Fourier components.



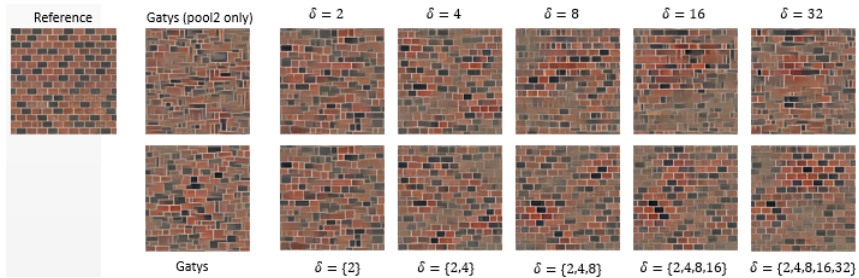
Depuis le 1er juin 2017
d'eye tracking a été pu
Dynamox au GABU de
Bretonneau, service a
testé auprès d'une dizaine
membres du personnel
paramédical ont tous
explique le Dr Rodot-

Depuis le 1er juin 2017
d'eye tracking a été pu
Dynamox au GABU de
Bretonneau, service a
testé auprès d'une dizaine
membres du personnel
paramédical ont tous
explique le Dr Rodot-

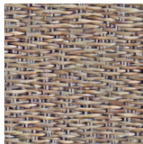
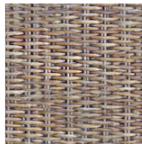
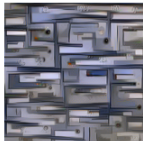
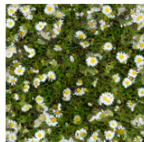
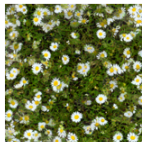
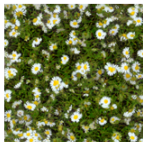
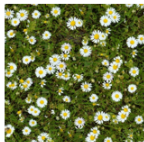
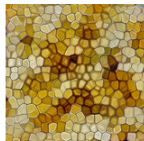
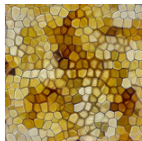
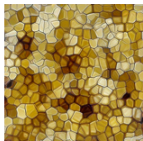
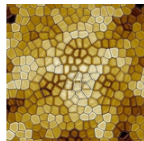
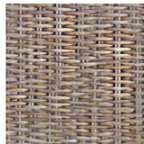
Adding long-range structure

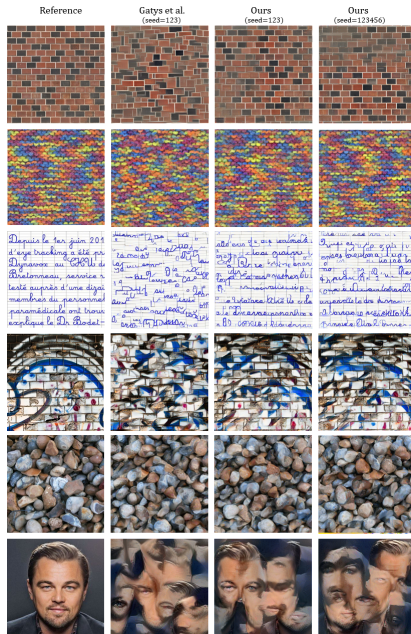


- ▶ Static: yes
- ▶ Local: no
- ▶ To account for long-range correlations, measure coincidences between features spatially transformed features.

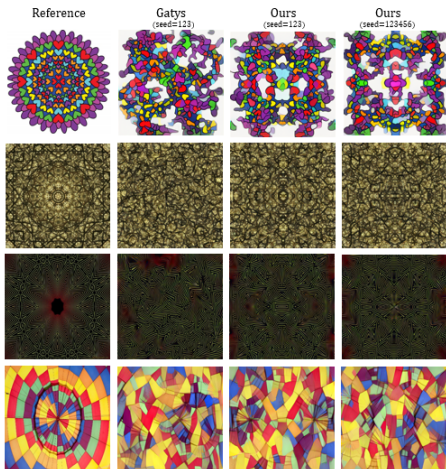


Reference

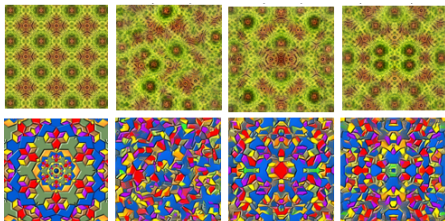
Gatys et al.
(seed=123)Ours
(seed=123)Ours
(seed=123456)



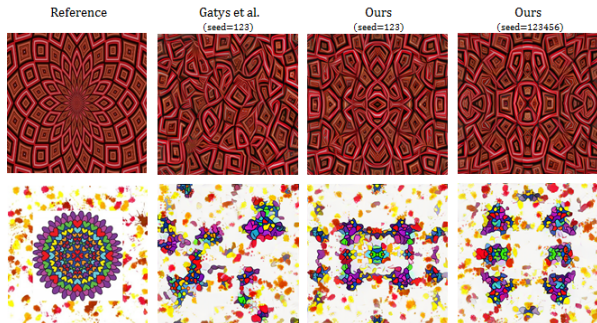
Symmetric textures



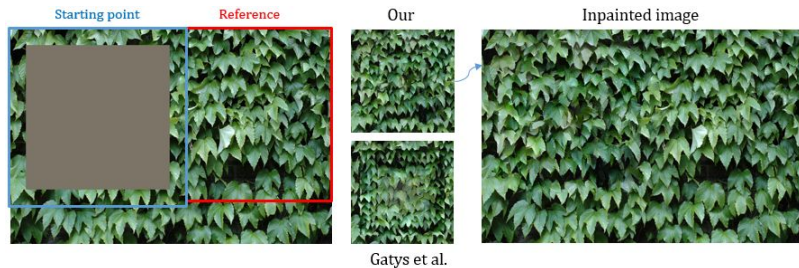
Symmetric textures



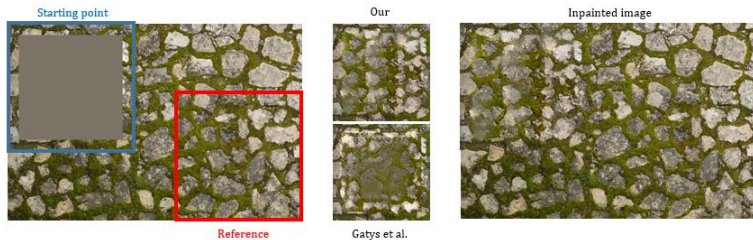
Symmetric textures



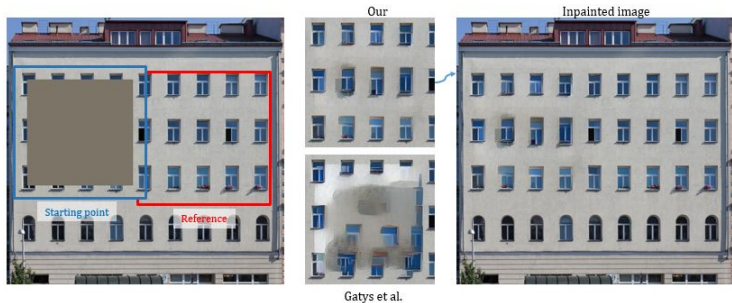
In-painting



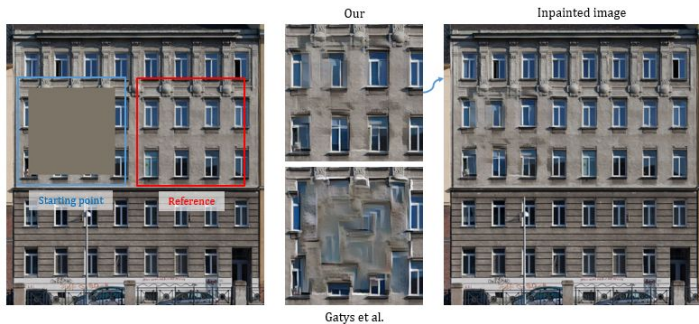
In-painting



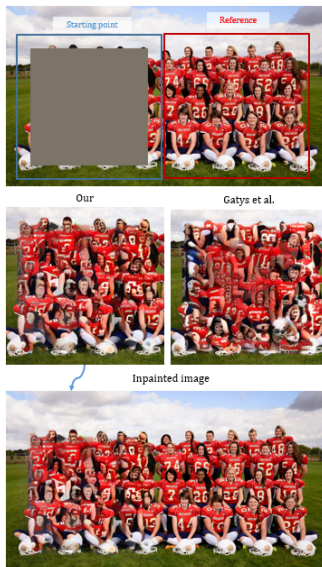
In-painting



In-painting



In-painting



Season transfer

Style



Content (& init)



Ours



Gatys



Season transfer

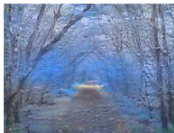
Style



Content (& init)



Ours



Gatys



Season transfer

Style



Content (& init)



Ours



Gatys



Daytime transfer

Style



Content (& init)

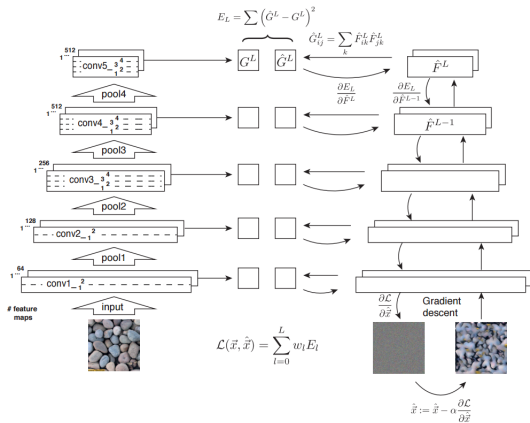


Ours

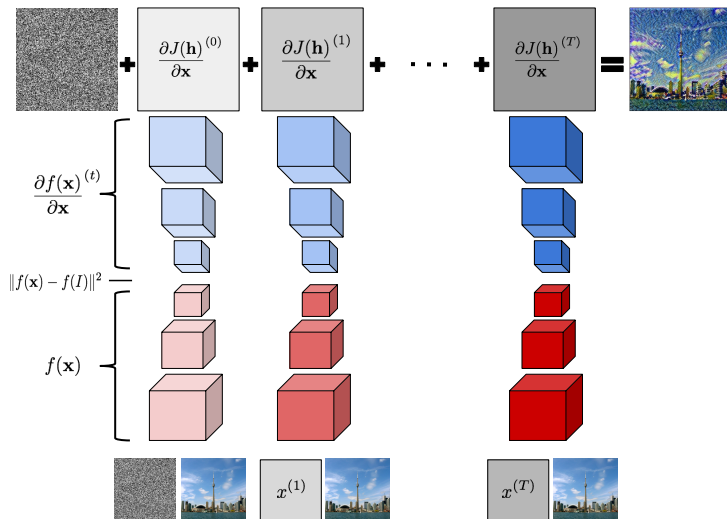


Gatys

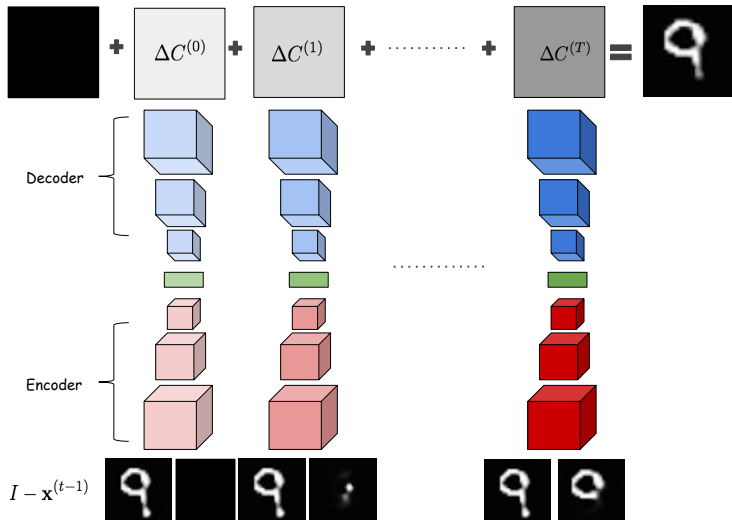




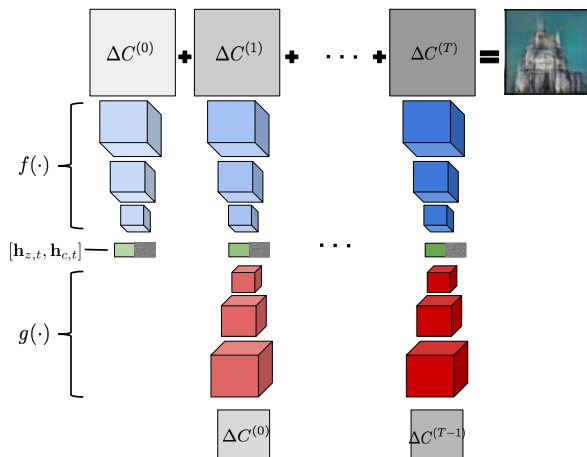
Gatys et al. as a recurrent net (joint work with Daniel Im)



The DRAW network (Gregor et al 2015)



Recurrent generative adversarial networks



- ▶ Train adversarially (Goodfellow et al. 2014)
- ▶ cf. Denton et al. 2015

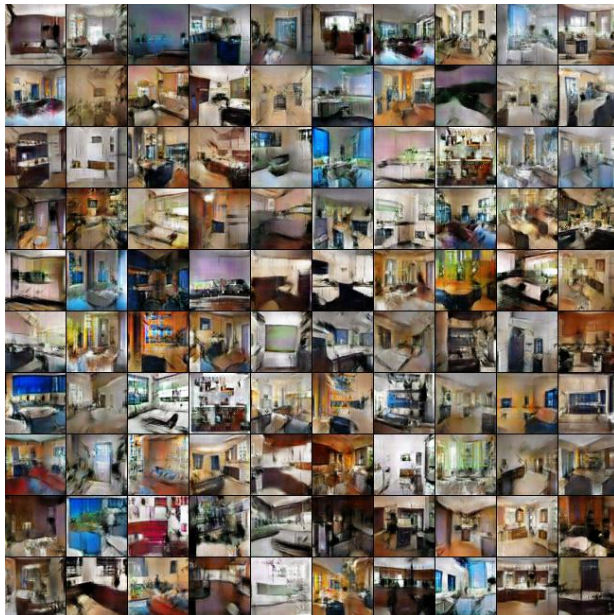
Cifar10 samples generated by GRAN



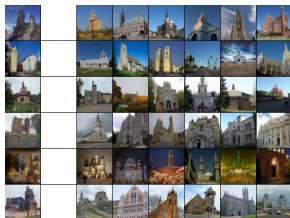
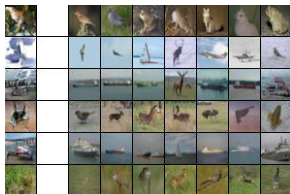
LSUN samples generated by GRAN



LSUN samples generated by GRAN



Nearest neighbors



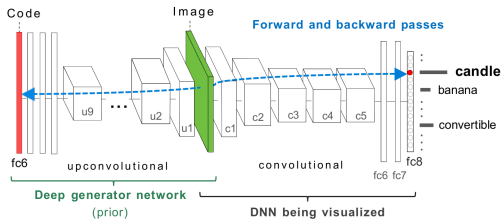
Intermediate samples



Quantitative comparison

Data set	Battler	r_{test}	r_{sample}	Winner
MNIST	GRAN1 vs. GRAN3	0.79	1.75	GRAN3
	GRAN1 vs. GRAN5	0.95	1.19	GRAN5
CIFAR10	GRAN1 vs. GRAN3	1.28	1.001	GRAN3
	GRAN1 vs. GRAN5	1.29	1.011	GRAN5
	GRAN3 vs. GRAN5	1.00	2.289	GRAN5
LSUN	GRAN1 vs. GRAN3	0.95	13.68	GRAN3
	GRAN1 vs. GRAN5	0.99	13.97	GRAN5
	GRAN3 vs. GRAN5	0.99	2.38	GRAN5

Nguyen et al. 2016



What is wrong with unsupervised learning?

- ▶ The irony: We are now using image classification to improve rendering, *not* the other way around.

What is wrong with unsupervised learning?

- ▶ The irony: We are now using image classification to improve rendering, *not* the other way around.
- ▶ Why does unsupervised learning not work? A possible answer:
There is more to natural images than we need to know.

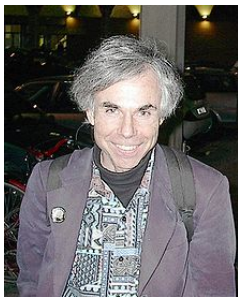
What is wrong with unsupervised learning?

- ▶ The irony: We are now using image classification to improve rendering, *not* the other way around.
- ▶ Why does unsupervised learning not work? A possible answer:
There is more to natural images than we need to know.
- ▶ For supervised learning, how do we deal with the data dilemma?

What is wrong with unsupervised learning?

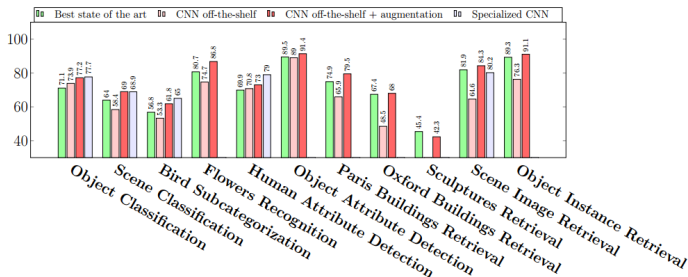
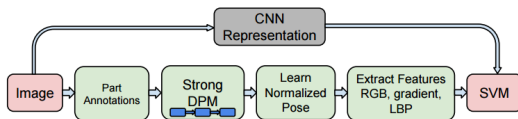
- ▶ The irony: We are now using image classification to improve rendering, *not* the other way around.
- ▶ Why does unsupervised learning not work? A possible answer:
There is more to natural images than we need to know.
- ▶ For supervised learning, how do we deal with the data dilemma?
- ▶ We want more data per parameter. → **Share neural circuitry**

Douglas Hofstadter: Intelligence is analogy-making *on all levels*



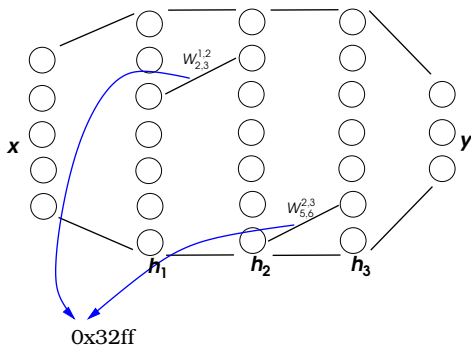
- ▶ “The brain is a computer”
- ▶ “The company will weather this storm”
- ▶ “What is the Washington of France”
- ▶ “This is a hammer”
- ▶ “Move x to the other side”
- ▶ “This is an elevator”
- ▶ “This is an edge”

Transfer learning: sharing a CNN across tasks

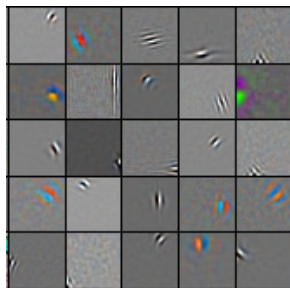
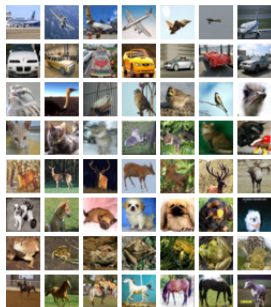


(Razavian, Azizpour, Sullivan, Carlsson; 2014), (Donahue et al, 2013), (Thrun 1996), ...

Sharing weights



Visual features *want* to be shared



features from a fully-connected net

- ▶ Fully connected, square: ≈ 9 mio parameters
- ▶ Local receptive fields: ≈ 1 mio parameters
- ▶ With weight sharing: ≈ 30 k parameters

Sparsity is good

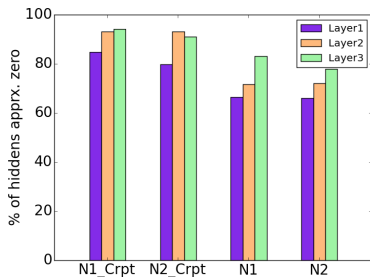


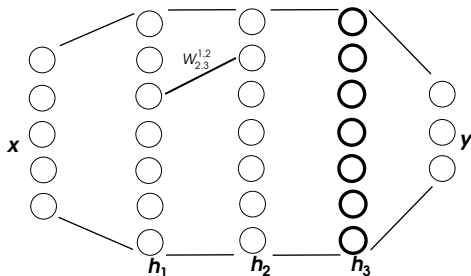
figure by Kishore Konda

Sparsity levels in two networks trained on CIFAR-10. N1=(1000-2000-3000), N2=(2000-2000-2000 units). (N1_Crpt, N2_Crpt trained with dropout).

- Conv-nets really maximize *the ratio*

$$\frac{\#neurons}{\#parameters}$$

Sharing lower layers (deep learning)



- ▶ Higher-layer neurons share lower-layer weights.
- ▶ The function multi-layer nets can represent may be restricted and unnatural (see Surya's and Tomaso's talks), but this is may be our only chance.

Sharing networks in RL policies

- ▶ Rusu et al. 2016
- ▶ Parisotto et al. 2016
- ▶ Romoff et al. 2016

Just how generic is human cognition?

$$3 + x = 7$$

- ▶ We solve equations like these using “dexterous manipulation” involving our motor cortex.
- ▶ Why is this useful?

Just how generic is human cognition?

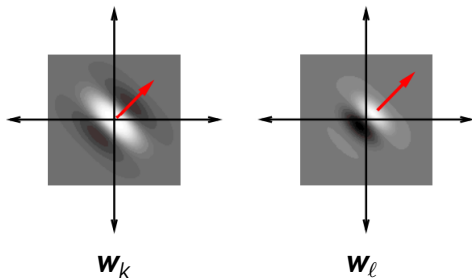
$$3 + x = 7$$

- ▶ We solve equations like these using “dexterous manipulation” involving our motor cortex.
- ▶ Why is this useful?

Cognition is analogy making, because this can make supervised learning work.

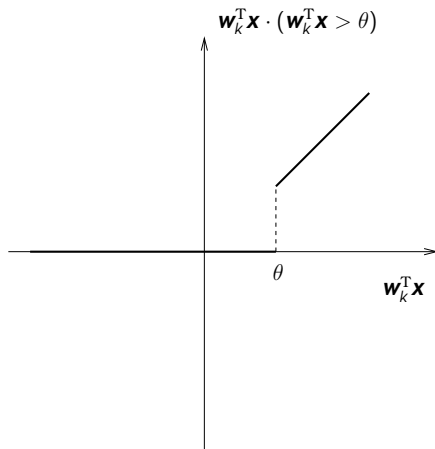
Thank you
Questions?

Another opportunity for sharing

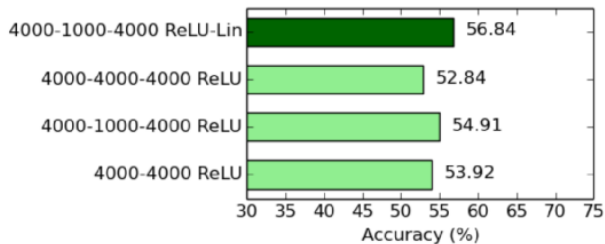


figures by Javier Movellan

Zero-bias activation function

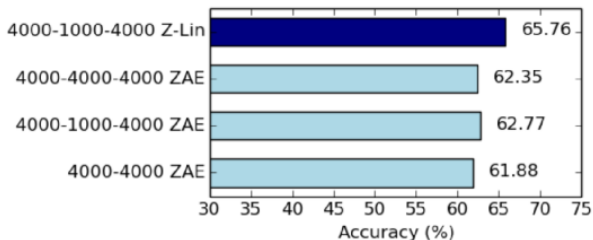


Deep fully-connected CIFAR-10



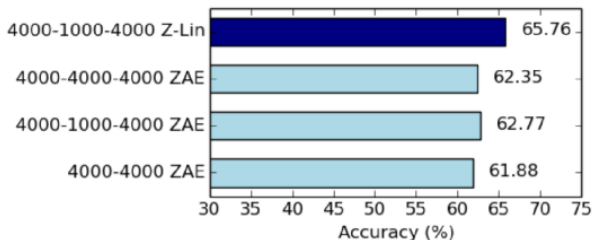
(Zhouhan Lin)

Deep fully-connected CIFAR-10



(Zhouhan Lin)

Deep fully-connected CIFAR-10



(Zhouhan Lin)

8 layers and dropout: **69.62%**

Training with deformations (not perm-invariant):
78.62%

Multi-task learning research

- ▶ increase network capacity (non-linearities, biases)
- ▶ multi-node (and scaling up in general)
- ▶ multi-task recurrent nets
- ▶ find opportunities for sharing
- ▶ multi-modality (e.g. language and images; or NMT and image classification)
- ▶ model-zoo
- ▶ pool datasets