

Bridging Deep Learning and Many-Body Quantum Physics via Tensor Networks

Yoav Levine

The Hebrew University of Jerusalem

APS Workshop, October 2018

Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

Y. Levine, D. Yakira, N. Cohen and A. Shashua (*ICLR 2018*)

Benefits of Depth for Long-Term Memory of Recurrent Networks

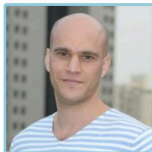
Y. Levine, O. Sharir and A. Shashua (*ICLR 2018 workshop*)

Bridging Deep Learning and Many-Body Quantum Physics via Tensor Networks

Y. Levine, O. Sharir, N. Cohen and A. Shashua (*arXiv Mar'2018*)



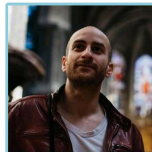
Prof. Amnon Shashua



Nadav Cohen



Or Sharir



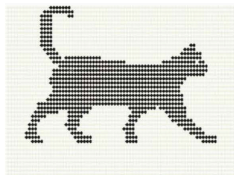
David Yakira

Machine Learning and Many-Body Physics

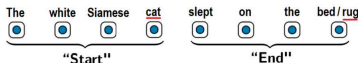
Distinct disciplines, share a need for “Many-Object” Expressiveness:

ML

image
classification:

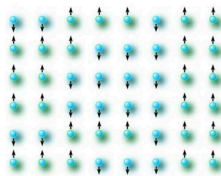


text
prediction:



Physics

2D spin
system:



wave-
function
 $|\psi^{(2D)}\rangle$

1D spin
system:



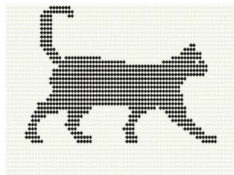
$|\psi^{(1D)}\rangle$

Machine Learning and Many-Body Physics

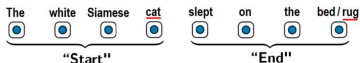
Distinct disciplines, share a need for “Many-Object” Expressiveness:

ML

image
classification:

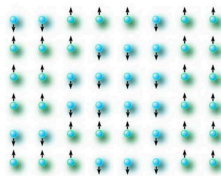


text
prediction:



Physics

2D spin
system:



1D spin
system:



wave-
function
 $|\psi^{(2D)}\rangle$

$|\psi^{(1D)}\rangle$

pixels $\longleftrightarrow$ particles

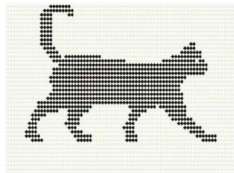
dependency $\longleftrightarrow$ entanglement

Machine Learning and Many-Body Physics

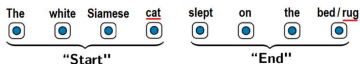
Distinct disciplines, share a need for “Many-Object” Expressiveness:

ML

image
classification:

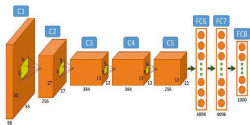


text
prediction:

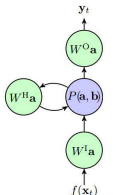


Tools:

ConvNet

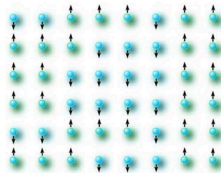


RNN



Physics

2D spin
system:



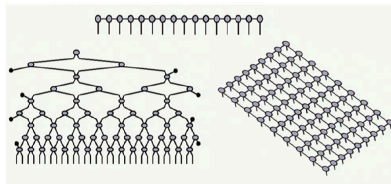
wave-
function
 $|\psi^{(2D)}\rangle$

1D spin
system:



$|\psi^{(1D)}\rangle$

Tensor Networks

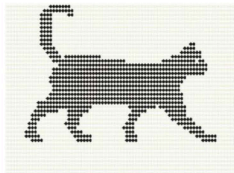


Machine Learning and Many-Body Physics

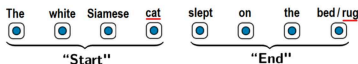
Distinct disciplines, share a need for “Many-Object” Expressiveness:

ML

image
classification:



text
prediction:



Tools:

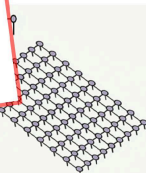
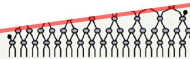
ConvNet

RNN

Tensor

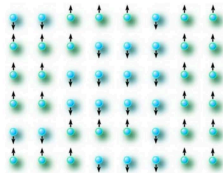


$$f(x_1, x_2, \dots, x_N)$$



Physics

2D spin
system:



1D spin
system:



wave-
function
 $|\psi^{(2D)}\rangle$

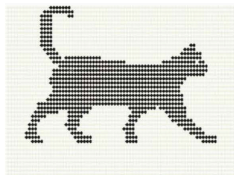
$|\psi^{(1D)}\rangle$

Machine Learning and Many-Body Physics

Distinct disciplines, share a need for “Many-Object” Expressiveness:

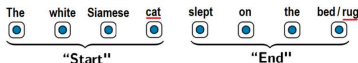
ML

image
classification:



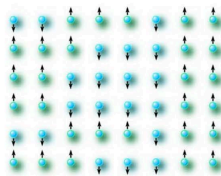
text

prediction:



Physics

2D spin
system:



1D spin
system:



wave-
function
 $|\psi^{(2D)}\rangle$

$|\psi^{(1D)}\rangle$

Strength

Unprecedented empirical
achievements

Gap

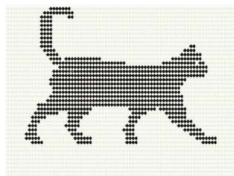
Formal understanding of
success is lacking

Machine Learning and Many-Body Physics

Distinct disciplines, share a need for “Many-Object” Expressiveness:

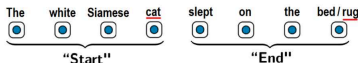
ML

image
classification:



text

prediction:



Strength

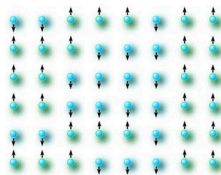
Unprecedented empirical
achievements

Gap

Formal understanding of
success is lacking

Physics

2D spin
system:



1D spin
system:



wave-
function
 $|\psi^{(2D)}\rangle$

$|\psi^{(1D)}\rangle$

Strength

Decades of theoretical
development

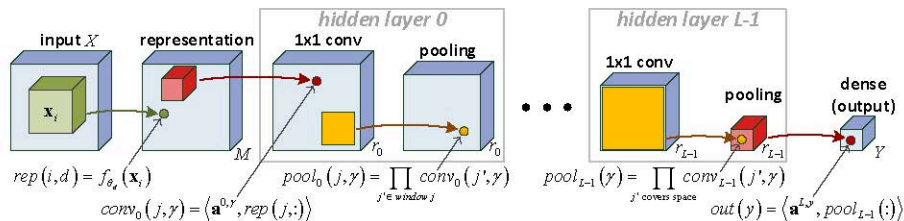
Gap

No efficient computation
schemes in 2D and 3D

Outline

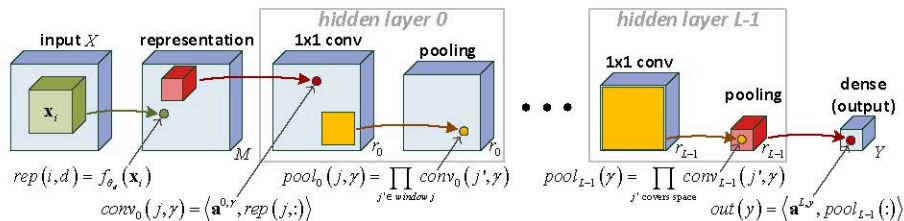
- 1 Quantum Wave Functions $\leftrightarrow$ Convolutional/Recurrent Networks
- 2 ML Result: Network Design via Quantum Entanglement
(Levine+Yakira+Cohen+Shashua@ICLR2018)
- 3 ML Result: Long-Term Memory in Deep Recurrent Networks
(Levine+Sharir+Shashua@ICLR2018workshop)
- 4 Physics Result: Quantum WF Representations via Deep Networks
(Levine+Sharir+Cohen+Shashua@arXiv'18)
- 5 Conclusion

Baseline Architecture - Convolutional Arithmetic Circuit



- $\mathbf{x}_1 \dots \mathbf{x}_N$ – input patches (e.g. pixels in an input image)
- $y \in [Y]$ – outputs (e.g. class scores)
- *Linear activation* ($\sigma(z) = z$), *product pooling*

Baseline Architecture - Convolutional Arithmetic Circuit



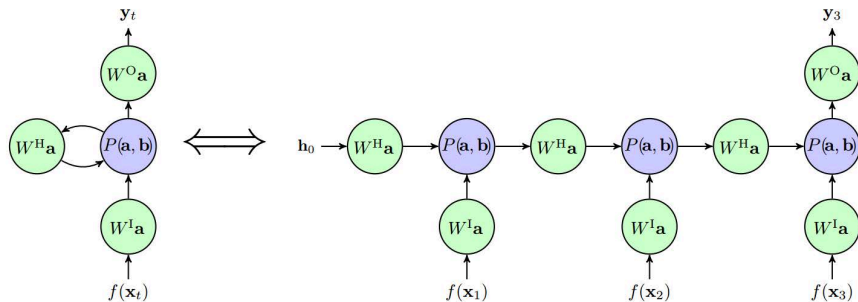
Function realized by network:

$$\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

$\mathcal{A}^{\text{conv}}$ - **conv weights tensor**

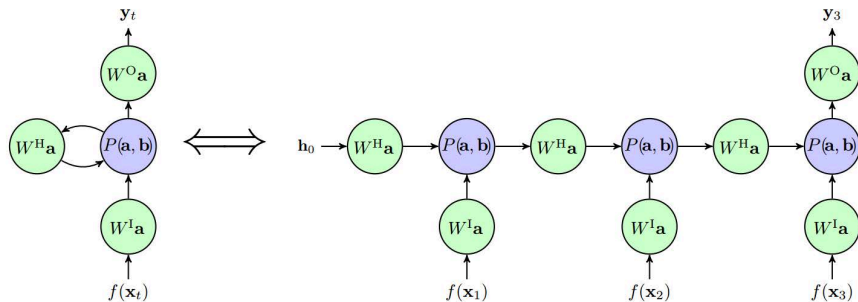
$\mathcal{A}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ - **inputs tensor** ($\mathcal{A}^{(\text{rank } 1)} = \mathbf{f}(\mathbf{x}_1) \otimes \dots \otimes \mathbf{f}(\mathbf{x}_N)$)

Baseline Architecture - Recurrent Arithmetic Circuit



- $\{\mathbf{x}_t\}_{t=1}^N$ – sequential inputs (e.g. words in a sentence)
- $y \in [Y]$ – outputs (e.g. score for next word)
- *Multiplicative Integration*: $P(\mathbf{a}, \mathbf{b}) := \mathbf{a} \odot \mathbf{b}$
 $\longrightarrow \mathbf{h}_t = W^H \mathbf{h}_{t-1} \odot W^I \mathbf{f}(\mathbf{x}_t)$

Baseline Architecture - Recurrent Arithmetic Circuit



Function realized by network:

$$\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{rec}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

$\mathcal{A}^{\text{rec}}$ - **recurrent weights tensor**

$\mathcal{A}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ - **inputs tensor** ($\mathcal{A}^{(\text{rank } 1)} = \mathbf{f}(\mathbf{x}_1) \otimes \dots \otimes \mathbf{f}(\mathbf{x}_N)$)

Equivalence to a ConvAC/RAC

- Many-body WF:

$$|\psi^{\text{conv/rec}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

- Constructed product state:

$$|\psi^{\text{ps}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N) |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

Equivalence to a ConvAC/RAC

- Many-body WF:

$$|\psi^{\text{conv/rec}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

- Constructed product state:

$$|\psi^{\text{ps}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N) |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

Inner product is equal to:

$$\langle \psi^{\text{ps}} | \psi^{\text{conv/rec}} \rangle = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

**Exactly reproducing the function
realized by a ConvAC/RAC!**

Equivalence to a ConvAC/RAC

- Many-body WF:

$$|\psi^{\text{conv/rec}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

- Constructed product state:

$$|\psi^{\text{ps}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N) |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

Inner product is equal to:

$$\langle \psi^{\text{ps}} | \psi^{\text{conv/rec}} \rangle = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

**Exactly reproducing the function
realized by a ConvAC/RAC!**

WF coeff tensor $\leftarrow \mathcal{A}^{\text{conv/rec}} \rightarrow$ conv/rec weights tensor

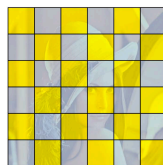
pixels $\longleftrightarrow$ particles

Dependencies Modeled by Networks – Previous Measure

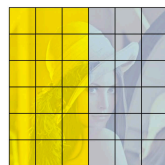
Network dependencies previously analyzed by **separation rank**:

$sep(y; A, B)$ – distance from separability of $y(x_1, \dots, x_N)$

A-  B- 



Interleaved



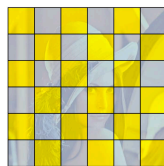
Left-Right

Dependencies Modeled by Networks – Previous Measure

Network dependencies previously analyzed by **separation rank**:

$sep(\mathbf{y}; A, B)$ – distance from separability of $\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N)$

A-  B- 



Interleaved



Left-Right

Claim

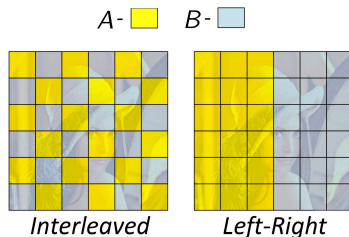
$sep(\mathbf{y}; A, B) = \text{Schmidt entanglement measure of } |\psi^{\text{conv/rec}}\rangle \text{ w.r.t. } (A, B)$

$$|\psi^{\text{conv/rec}}\rangle := \sum_{d_1 \dots d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv/rec}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

Dependencies Modeled by Networks – Previous Measure

Network dependencies previously analyzed by **separation rank**:

$sep(\mathbf{y}; A, B)$ – distance from separability of $\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N)$



Claim

$sep(\mathbf{y}; A, B) = \text{Schmidt entanglement measure of } |\psi^{conv/rec}\rangle \text{ w.r.t. } (A, B)$

interpretation:

separable state $\longleftrightarrow$ separable function

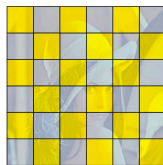
highly entangled state $\longleftrightarrow$ function w/ high sep rank

Dependencies Modeled by Networks – Previous Measure

Network dependencies previously analyzed by **separation rank**:

$sep(\mathbf{y}; A, B)$ – distance from separability of $\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N)$

A-  B- 



Interleaved



Left-Right

Claim

$sep(\mathbf{y}; A, B) = \text{Schmidt entanglement measure of } |\psi^{conv/rec}\rangle \text{ w.r.t. } (A, B)$

interpretation:

separable state $\longleftrightarrow$ separable function

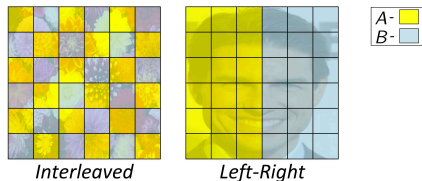
highly entangled state $\longleftrightarrow$ function w/ high sep rank

Entanglement quantifies dependencies supported by the network

Measures of Entanglement for Deep Learning Archs

Design network such that relevant dependencies are modeled:

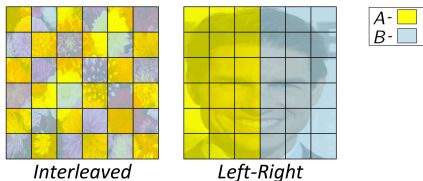
- **Convolutional networks:**



Measures of Entanglement for Deep Learning Archs

Design network such that relevant dependencies are modeled:

- **Convolutional networks:**

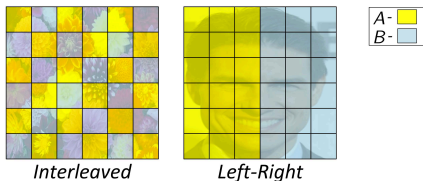


**ConvNet should support high entanglement measures
for given input dependencies**

Measures of Entanglement for Deep Learning Archs

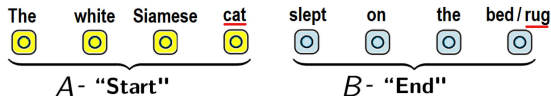
Design network such that relevant dependencies are modeled:

- Convolutional networks:**



**ConvNet should support high entanglement measures
for given input dependencies**

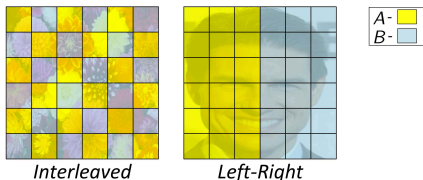
- Recurrent networks:**



Measures of Entanglement for Deep Learning Archs

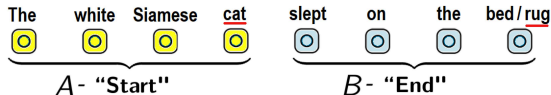
Design network such that relevant dependencies are modeled:

- Convolutional networks:**



**ConvNet should support high entanglement measures
for given input dependencies**

- Recurrent networks:**



**RNN should support high Start-End entanglement
to have 'long-term memory' capacity**

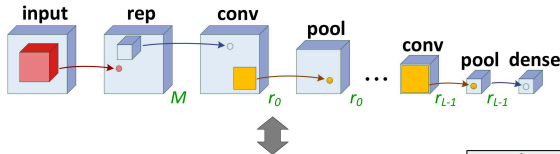
Convolutional Arithmetic Circuits $\longleftrightarrow$ Tensor Networks

Computation of ConvAC can be cast as a **Tree Tensor Network**:

Convolutional Arithmetic Circuits $\leftrightarrow$ Tensor Networks

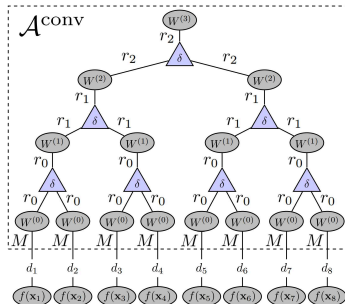
Computation of ConvAC can be cast as a **Tree Tensor Network**:

- Degree-1 nodes correspond to input patches



edge weights
correspond to ConvAC
layer widths

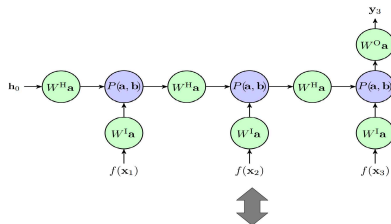
$$\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{conv}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$



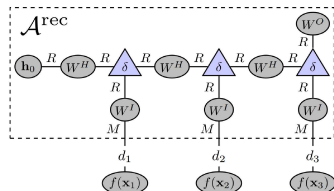
Recurrent Arithmetic Circuits $\longleftrightarrow$ Tensor Networks

Computation of RAC can be cast as an **MPS Tensor Network**:

- Edge weights hold layer width (# of hidden channels)
- Degree-1 nodes correspond to input patches



$$\mathbf{y}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{d_1, \dots, d_N=1}^M \mathcal{A}_{d_1 \dots d_N}^{\text{rec}} \mathcal{A}_{d_1 \dots d_N}^{(\text{rank } 1)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$



Outline

- 1 Quantum Wave Functions $\leftrightarrow$ Convolutional/Recurrent Networks
- 2 ML Result: Network Design via Quantum Entanglement
(Levine+Yakira+Cohen+Shashua@ICLR2018)
- 3 ML Result: Long-Term Memory in Deep Recurrent Networks
(Levine+Sharir+Shashua@ICLR2018workshop)
- 4 Physics Result: Quantum WF Representations via Deep Networks
(Levine+Sharir+Cohen+Shashua@arXiv'18)
- 5 Conclusion

Dependencies $\longleftrightarrow$ Min-Cut

Theorem (adapted from Cui *et al.* (2016))

Schmidt entanglement ConvAC/RAC models between input sets (A, B) is equal to min cut in respective TN separating nodes of (A, B)

Dependencies $\longleftrightarrow$ Min-Cut

Theorem (adapted from Cui *et al.* (2016))

Schmidt entanglement ConvAC/RAC models between input sets (A, B) is equal to min cut in respective TN separating nodes of (A, B)

Corollary

Controlling entanglement (dependencies) modeled by ConvAC/RAC is equivalent to controlling min cuts in respective TN

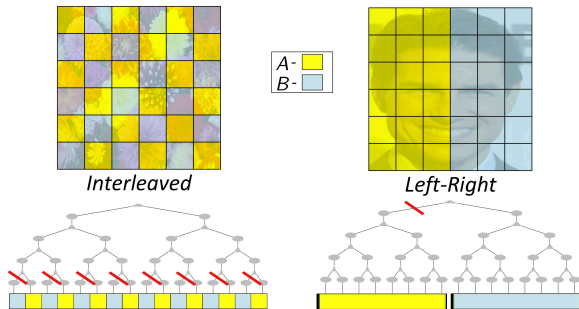
Dependencies $\longleftrightarrow$ Min-Cut

Theorem (adapted from Cui *et al.* (2016))

Schmidt entanglement ConvAC/RAC models between input sets (A, B) is equal to min cut in respective TN separating nodes of (A, B)

Corollary

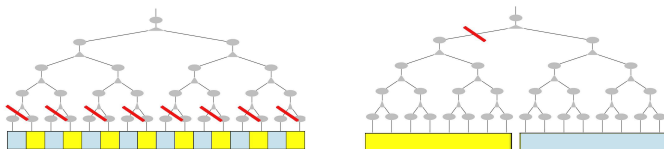
Controlling entanglement (dependencies) modeled by ConvAC/RAC is equivalent to controlling min cuts in respective TN



Controlling Dependencies – Layer Widths

Claim

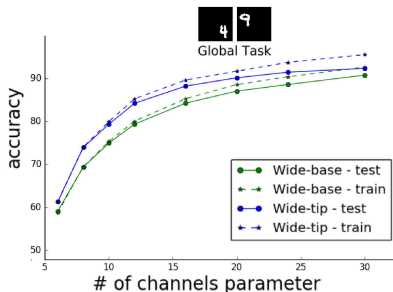
Deep (early) layer widths are important for long (short)-range dependencies



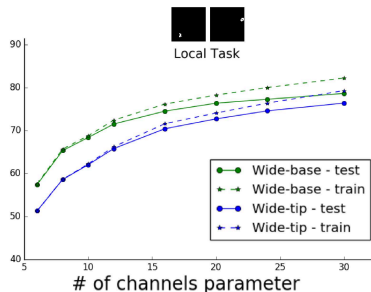
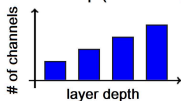
Controlling Dependencies – Layer Widths

Claim

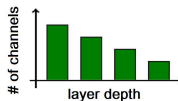
Deep (early) layer widths are important for long (short)-range dependencies



Wide-tip (schematic)



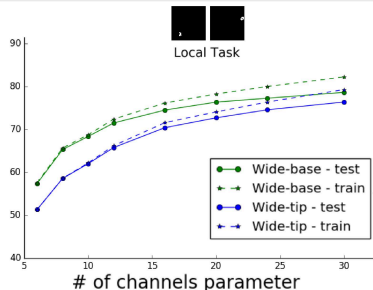
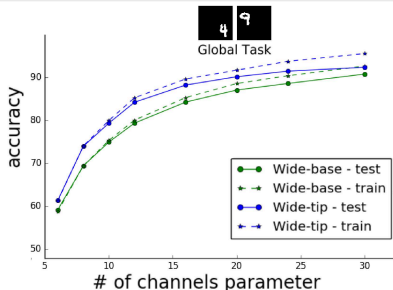
Wide-base (schematic)



Controlling Dependencies – Layer Widths

Claim

Deep (early) layer widths are important for long (short)-range dependencies

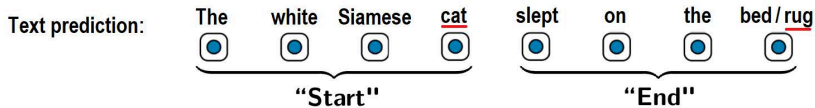


Tailor architecture to fit dependencies of a given task

Outline

- 1 Quantum Wave Functions $\leftrightarrow$ Convolutional/Recurrent Networks
- 2 ML Result: Network Design via Quantum Entanglement
(Levine+Yakira+Cohen+Shashua@ICLR2018)
- 3 ML Result: Long-Term Memory in Deep Recurrent Networks
(Levine+Sharir+Shashua@ICLR2018workshop)
- 4 Physics Result: Quantum WF Representations via Deep Networks
(Levine+Sharir+Cohen+Shashua@arXiv'18)
- 5 Conclusion

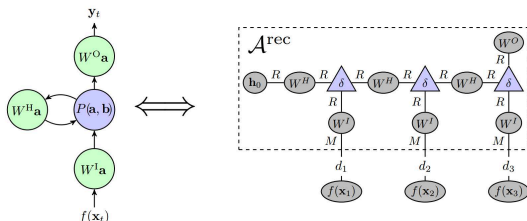
Start-End Entanglement in Recurrent Networks



Start-End entanglement quantifies long-term memory of the network

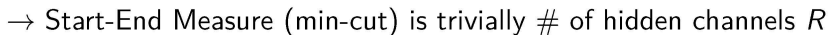
The white Siamese cat slept on the bed / rug

For shallow RACs:



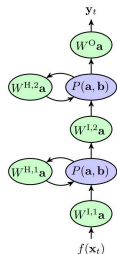
Start-End entanglement quantifies long-term memory of the network

For shallow RACs:



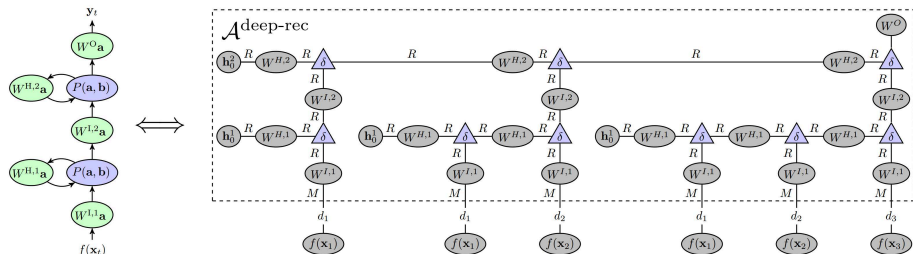
The white Siamese cat slept on the bed / rug

Start-End entanglement quantifies long-term memory of the network
Deep recurrent networks significantly out perform shallow ones, yet are seemingly ‘innocent’:



The white Siamese cat slept on the bed / rug

Start-End entanglement quantifies long-term memory of the network
Deep recurrent networks significantly out perform shallow ones, yet are seemingly ‘innocent’:



But due to information re-use their TN is recursive and exploding

Exponential Memory Capacity for Deep Networks

Theorem

- *Upper bound on Start-End entanglement of a shallow RAC:*

$$R \text{ (\# of params)}$$

- *Lower bound on Start-End entanglement of depth $L = 2$ RAC:*

$$\text{super-poly}(R)$$

Super-polynomial separation in memory capacity between shallow and deep recurrent networks!

Exponential Memory Capacity for Deep Networks

Theorem

- *Upper bound on Start-End entanglement of a shallow RAC:*

$$R \text{ (\# of params)}$$

- *Lower bound on Start-End entanglement of depth $L = 2$ RAC:*

$$\text{super-poly}(R)$$

Super-polynomial separation in memory capacity between shallow and deep recurrent networks!

Conjecture

Lower bound on Start-End entanglement of a depth L RAC :

$$\text{super-poly}(L)$$

Memory capacity increases super-polynomially w/depth!

Outline

- 1 Quantum Wave Functions $\leftrightarrow$ Convolutional/Recurrent Networks
- 2 ML Result: Network Design via Quantum Entanglement
(Levine+Yakira+Cohen+Shashua@ICLR2018)
- 3 ML Result: Long-Term Memory in Deep Recurrent Networks
(Levine+Sharir+Shashua@ICLR2018workshop)
- 4 Physics Result: Quantum WF Representations via Deep Networks
(Levine+Sharir+Cohen+Shashua@arXiv'18)
- 5 Conclusion

Physicists' Need for Expressivity

The quantum many-body wave-function:

$$|\psi\rangle = \sum_{d_1 \dots d_N=1}^M \underbrace{\mathcal{A}_{d_1 \dots d_N}}_{\text{encapsulates entanglement}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

Physicists' Need for Expressivity

The quantum many-body wave-function:

$$|\psi\rangle = \sum_{d_1 \dots d_N=1}^M \underbrace{\mathcal{A}_{d_1 \dots d_N}}_{\text{encapsulates entanglement}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

- General WFs in d -dimensions:

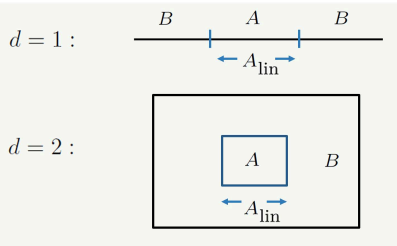
Volume law entanglement–

Entanglement $\propto A_{\text{lin}}^d$

- Some WFs of interest:

Area law entanglement–

Entanglement $\propto A_{\text{lin}}^{d-1}$



Physicists' Need for Expressivity

The quantum many-body wave-function:

$$|\psi\rangle = \sum_{d_1 \dots d_N=1}^M \underbrace{\mathcal{A}_{d_1 \dots d_N}}_{\text{encapsulates entanglement}} |\psi_{d_1}\rangle \otimes \dots \otimes |\psi_{d_N}\rangle$$

- General WFs in d -dimensions:

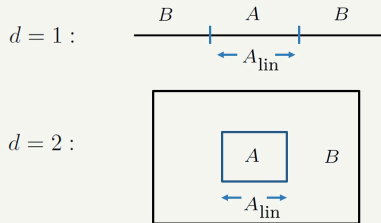
Volume law entanglement–

Entanglement $\propto A_{\text{lin}}^d$

- Some WFs of interest:

Area law entanglement–

Entanglement $\propto A_{\text{lin}}^{d-1}$

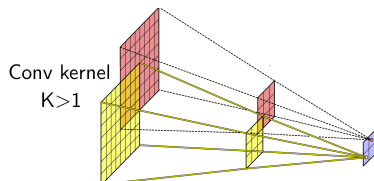
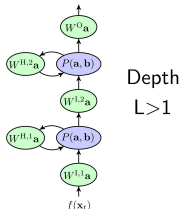


Physicists are in search for computational representations that efficiently capture high entanglement

- Leading tool: Tensor Networks
- More recently: Neural Networks (RBMs)

TN Constructions of Prominent Deep Learning Archs

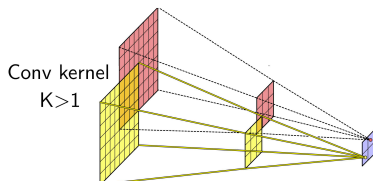
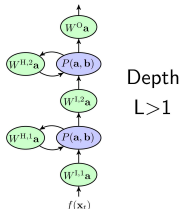
Deep recurrent networks and Overlapping convolutional networks



enhancements to above archs that involve
information re-use

TN Constructions of Prominent Deep Learning Archs

Deep recurrent networks and Overlapping convolutional networks



enhancements to above archs that involve
information re-use $\rightarrow$
 cannot be naively represented as TNs:

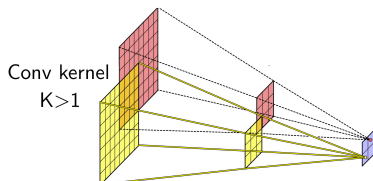
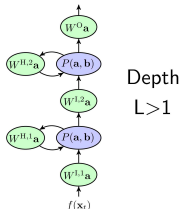
Claim

No Tensor Network can intake a vector $\mathbf{v}$ and duplicate it to $\mathbf{v} \otimes \mathbf{v}$



TN Constructions of Prominent Deep Learning Archs

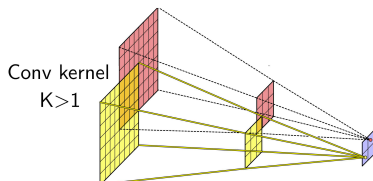
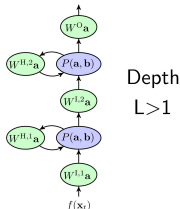
Deep recurrent networks and Overlapping convolutional networks



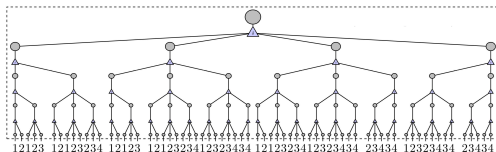
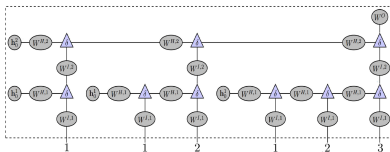
information re-use in TNs
via external duplication

TN Constructions of Prominent Deep Learning Archs

Deep recurrent networks and Overlapping convolutional networks



information re-use in TNs
via external duplication

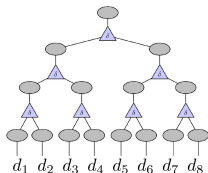


Information Re-Use Vs. Loops

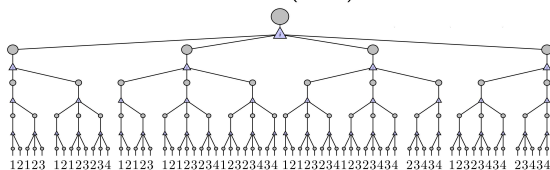
Competing mechanisms for expressivity enhancement of naive Tree:

Machine learning

Non-overlapping
ConvAC ($K=1$)



Overlapping
ConvAC ($K>1$)

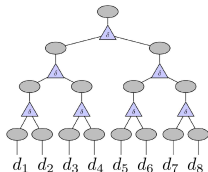


Information Re-Use Vs. Loops

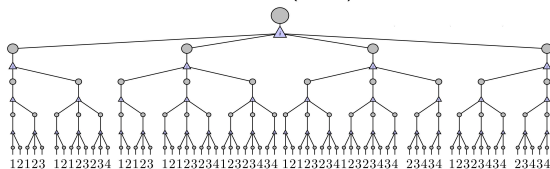
Competing mechanisms for expressivity enhancement of naive Tree:

Machine learning

Non-overlapping
ConvAC ($K=1$)

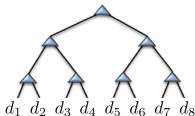


Overlapping
ConvAC ($K>1$)

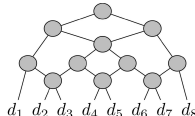


Physics

Tree
Tensor Network



MERA
Tensor Network



Results - Deep Learning Archs Support High Entanglement

Theorem

Deep recurrent networks support log corrections to area law in 1D:

$$\text{Entanglement Entropy} \propto \log\{(A_{lin})\}$$

Match MERA, with efficient WF amplitude calculation

Results - Deep Learning Archs Support High Entanglement

Theorem

Deep recurrent networks support *log corrections to area law in 1D*:

$$\text{Entanglement Entropy} \propto \log\{(A_{lin})\}$$

Match MERA, with efficient WF amplitude calculation

Theorem

Overlapping convolutional networks support *volume law entanglement in 1D and 2D, with*:

$$\# \text{ of params} \propto A_{lin} (\propto \sqrt{N} \text{ in } 2D)$$

Access to volume law entanglement, polynomially more efficient in 2D than proposed NN based representations

Outline

- 1 Quantum Wave Functions $\leftrightarrow$ Convolutional/Recurrent Networks
- 2 ML Result: Network Design via Quantum Entanglement
(Levine+Yakira+Cohen+Shashua@ICLR2018)
- 3 ML Result: Long-Term Memory in Deep Recurrent Networks
(Levine+Sharir+Shashua@ICLR2018workshop)
- 4 Physics Result: Quantum WF Representations via Deep Networks
(Levine+Sharir+Cohen+Shashua@arXiv'18)
- 5 Conclusion

Conclusion – Common Need: Many-Object Expressiveness

- **Drawn Quantum Physics and Deep Learning Analogies**

- Function realized by ConvAC/shallow-RAC $\leftrightarrow$ many-body WF
- Entanglement measures dependencies in deep networks
- ConvAC/RAC as a Tensor Network
- Dependencies bound by min-cut in TN graph

Conclusion – Common Need: Many-Object Expressiveness

- **Drawn Quantum Physics and Deep Learning Analogies**

- Function realized by ConvAC/shallow-RAC $\leftrightarrow$ many-body WF
- Entanglement measures dependencies in deep networks
- ConvAC/RAC as a Tensor Network
- Dependencies bound by min-cut in TN graph

- **Deep Learning Results**

- Tailor architecture to fit task dependencies
- Super-polynomial contribution of depth for long-term memory capacity

Conclusion – Common Need: Many-Object Expressiveness

- **Drawn Quantum Physics and Deep Learning Analogies**

- Function realized by ConvAC/shallow-RAC $\leftrightarrow$ many-body WF
- Entanglement measures dependencies in deep networks
- ConvAC/RAC as a Tensor Network
- Dependencies bound by min-cut in TN graph

- **Deep Learning Results**

- Tailor architecture to fit task dependencies
- Super-polynomial contribution of depth for long-term memory capacity

- **Many-Body Physics Results**

- Powerful DL archs can model highly entangled states

THANK YOU

Questions?