LightFF: Lightweight Inference for Forward-Forward Algorithm

Amin Aminifar*, Baichuan Huang†, Azra Abtahi†, Amir Aminifar†

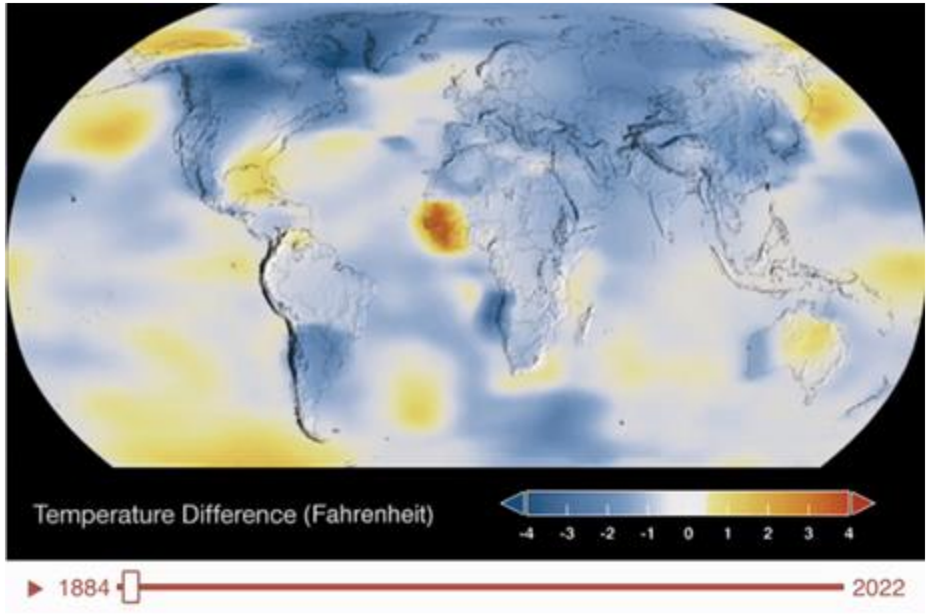
* Institute of Computer Engineering, Heidelberg University, Germany

† Department of Electrical and Information Technology, Lund University, Sweden

This research has been partially supported by the Swedish Wallenberg AI, Autonomous Systems and Software Program (WASP), the Swedish Research Council (VR), Swedish Foundation for Strategic Research (SSF), the ELLIIT Strategic Research Environment, and the European Union (EU).

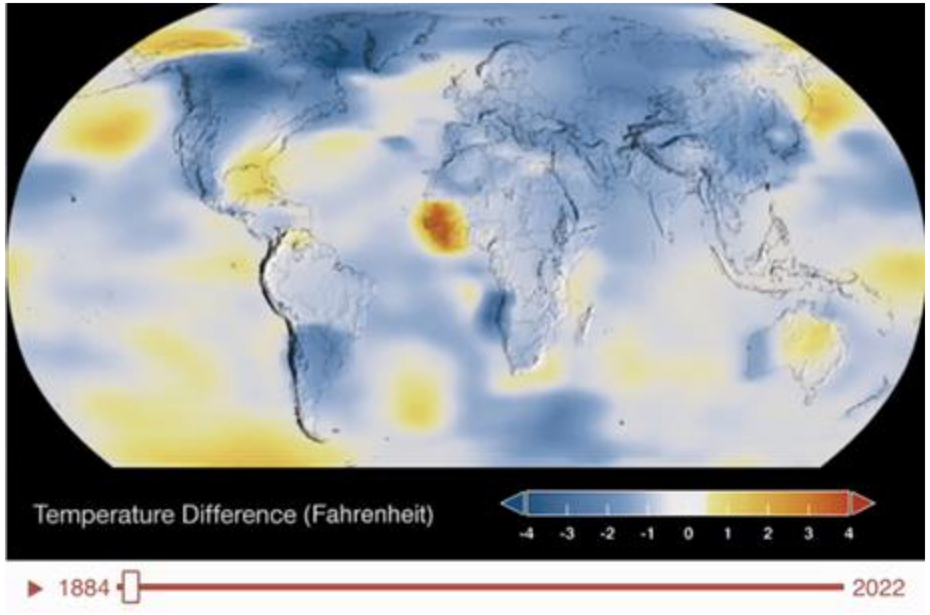
Introduction and Background

Global Warming



Source: https://giphy.com/; NASA; https://en.wikipedia.org/wiki/Climate_change_in_Europe

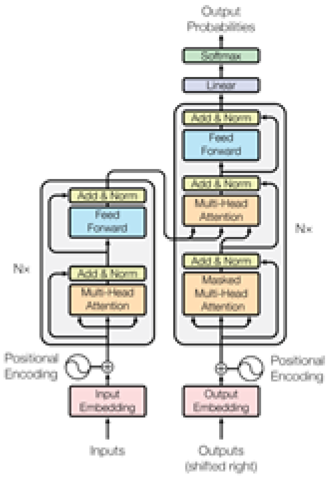
Global Warming



Europe: an average rise of 2.3°C compared to pre-industrial levels 1°C higher than the global average.

Source: https://giphy.com/; NASA; https://en.wikipedia.org/wiki/Climate_change_in_Europe

Environmental Impact of Deep Learning

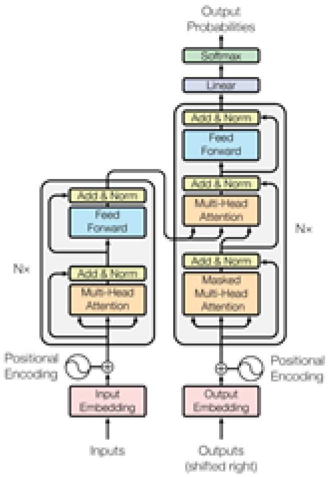


Training Transformer (Strubell E. 2020)

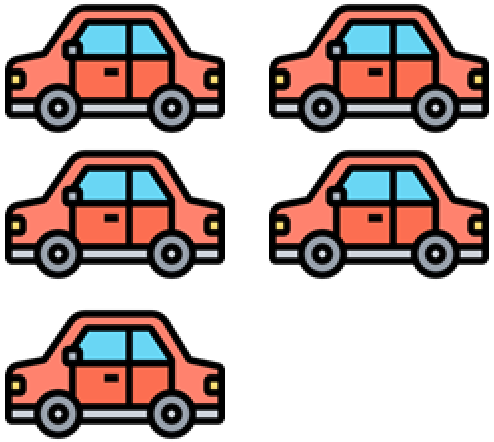
626,155 lbs

Strubell E, et al. Energy and policy considerations for modern deep learning research. AAAI, 2020.
Vaswani A. Attention is all you need. NeurIPS, 2017.
https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/

Environmental Impact of Deep Learning



Training Transformer (Strubell E. 2020)

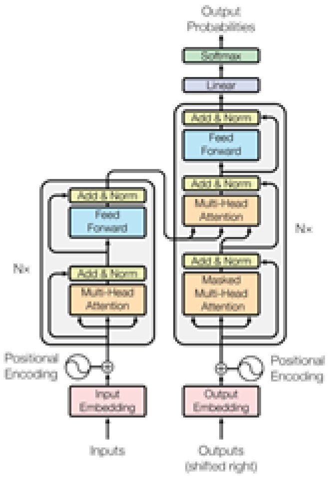


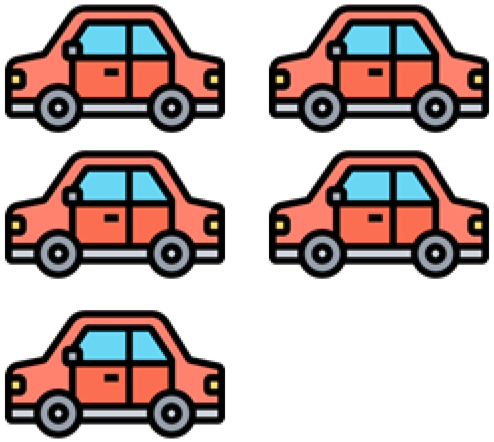
Total Lifetime of a Car

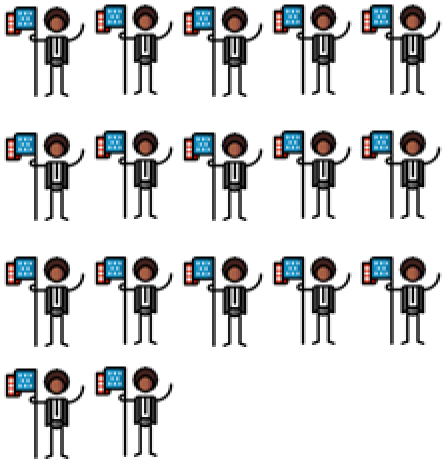
626,155 lbs = 5 × 126,000 lbs

Strubell E, et al. Energy and policy considerations for modern deep learning research. AAAI, 2020.
Vaswani A. Attention is all you need. NeurIPS, 2017.
https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/

Environmental Impact of Deep Learning





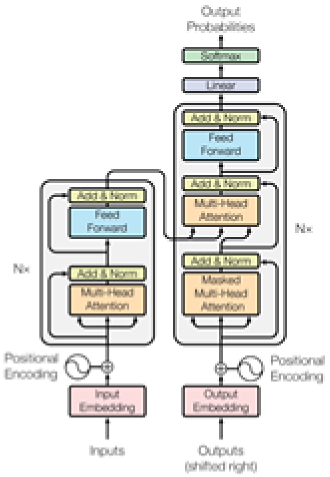


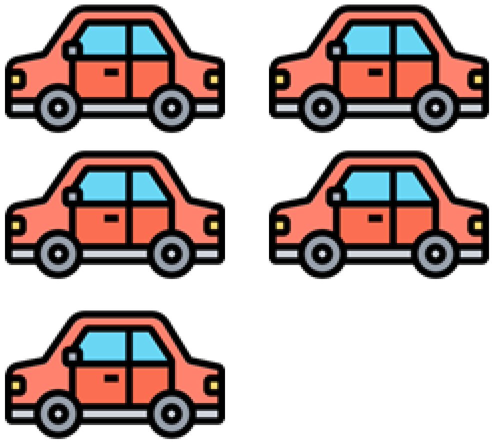
Training Transformer (Strubell E. 2020) | Total Lifetime of a Car | Average American in a Year

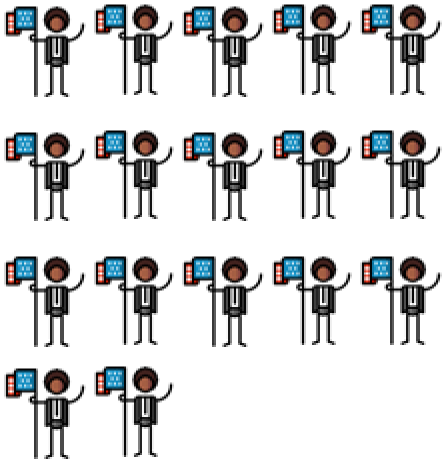
626,155 lbs = 5 × 126,000 lbs = 17 × 36,156 lbs

Strubell E, et al. Energy and policy considerations for modern deep learning research. AAAI, 2020.
Vaswani A. Attention is all you need. NeurIPS, 2017.
https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/

Environmental Impact of Deep Learning







Training Transformer (Strubell E. 2020) | Total Lifetime of a Car | Average American in a Year

CO_2 626,155 lbs = 5 × 126,000 lbs = 17 × 36,156 lbs

The computational resources needed to produce a best-in-class AI model has on average doubled every 3.4 months.

Strubell E, et al. Energy and policy considerations for modern deep learning research. AAAI, 2020.
Vaswani A. Attention is all you need. NeurIPS, 2017.
https://www.forbes.com/sites/robtoews/2020/06/17/deep-learnings-climate-change-problem/

Energy Consumption of Training

GPT-3

GPT-4

D. Patterson, et al. Carbon emissions and large neural network training, 2021.
https://tinyml.substack.com/p/the-carbon-impact-of-large-language
Data sources: U.S. Energy Information Administration, Electric Power Research Institute (EPRI)

Energy Consumption of Training

	GPT-3		GPT-4
CO_2	1,216,950 lbs	×13	15,238,333 lbs

D. Patterson, et al. Carbon emissions and large neural network training, 2021.
https://tinyml.substack.com/p/the-carbon-impact-of-large-language
Data sources: U.S. Energy Information Administration, Electric Power Research Institute (EPRI)

Energy Consumption of Training

	GPT-3		GPT-4
CO_2	1,216,950 lbs	×13	15,238,333 lbs
Energy	1,287 Megawatt-Hour	× 48	62,318 Megawatt-Hour

D. Patterson, et al. Carbon emissions and large neural network training, 2021.
https://tinyml.substack.com/p/the-carbon-impact-of-large-language
Data sources: U.S. Energy Information Administration, Electric Power Research Institute (EPRI)

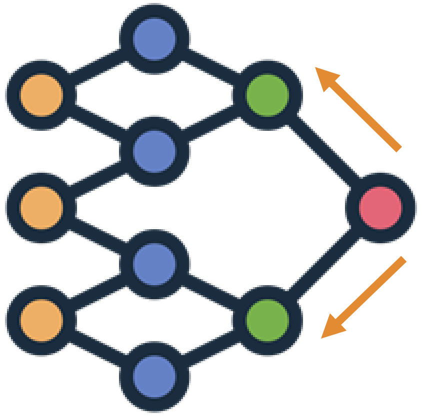
Biologically Plausible Alternatives

Human Brain
(~20 Watts)

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Biologically Plausible Alternatives

Human Brain
(~**20** Watts)

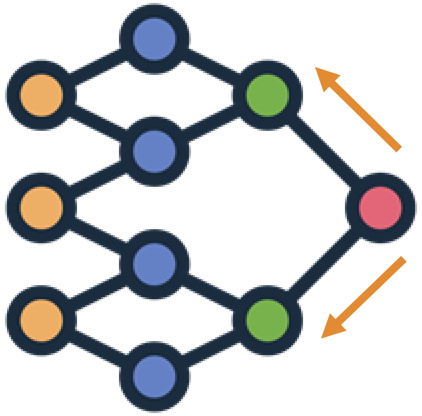


Back-Propagation
(Bio-**Implausible**)

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Biologically Plausible Alternatives

Human Brain
(~**20** Watts)



Back-Propagation
(Bio-**Implausible**)

Forward-Forward Algorithm
(Bio-**Plausible**)

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Energy Consumption of Inference

LUND UNIVERSITY

~60%

80%-90%

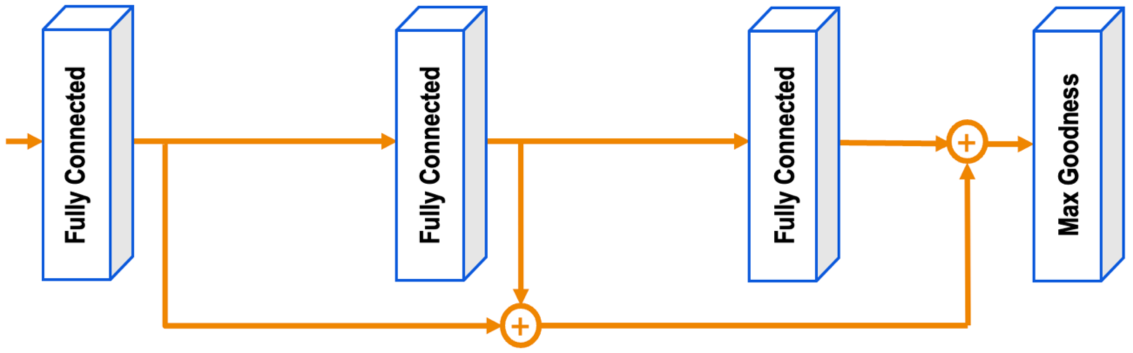
~90%

The percentage of inference in the total ML energy consumed.

D. Patterson, et al. Carbon emissions and large neural network training, 2021.
Patterson, David, et al. "The carbon footprint of machine learning training will plateau, then shrink." Computer, 2022.

Lightweight Inference for Forward-Forward Algorithm

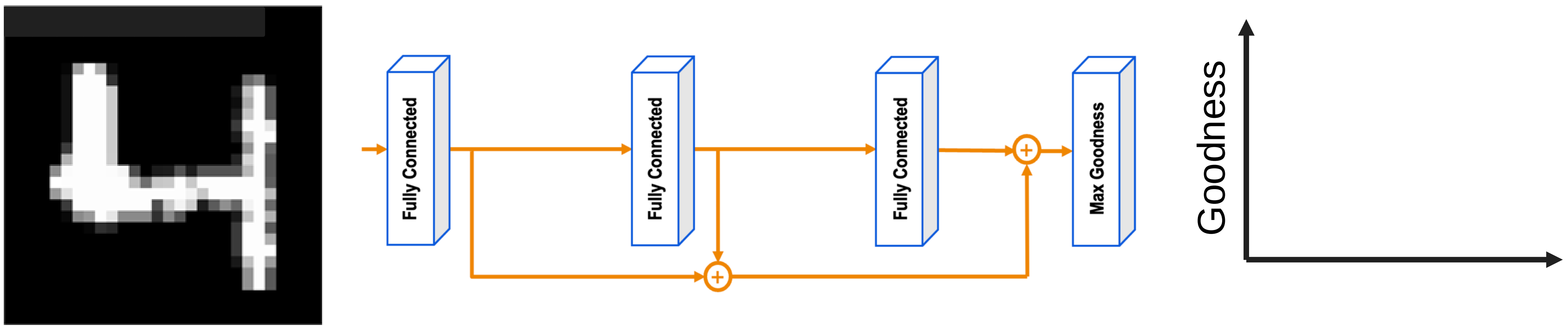
Glance at Hinton's FF Inference



The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

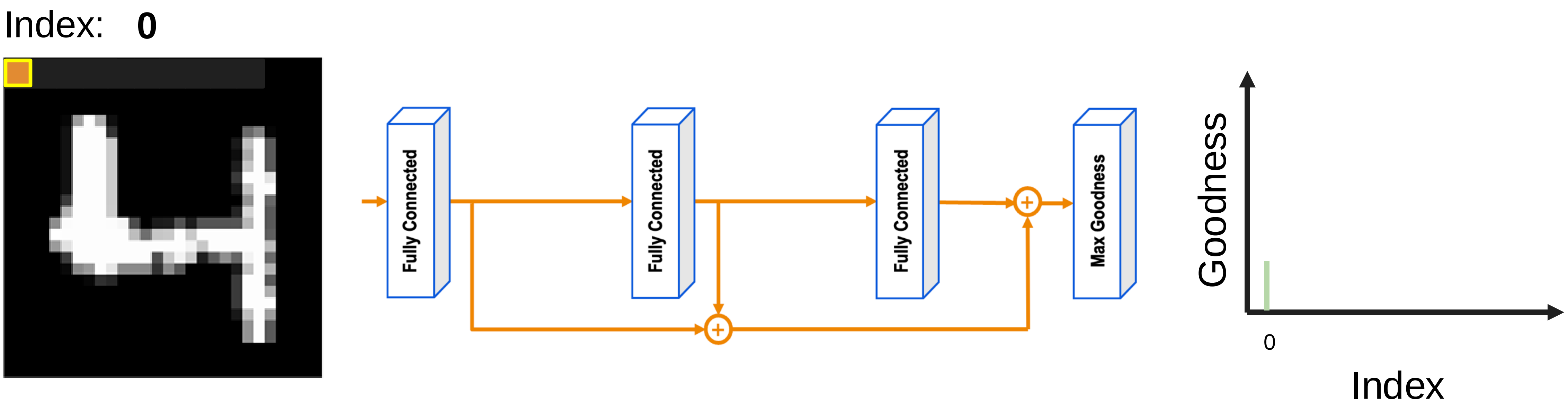
Glance at Hinton's FF Inference



The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

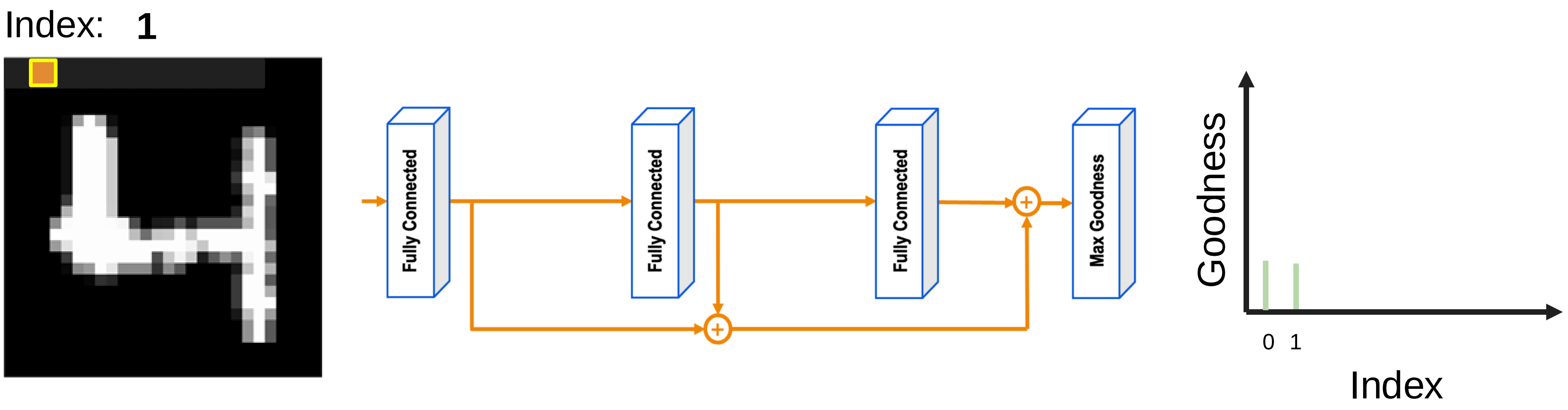
Glance at Hinton's FF Inference



The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

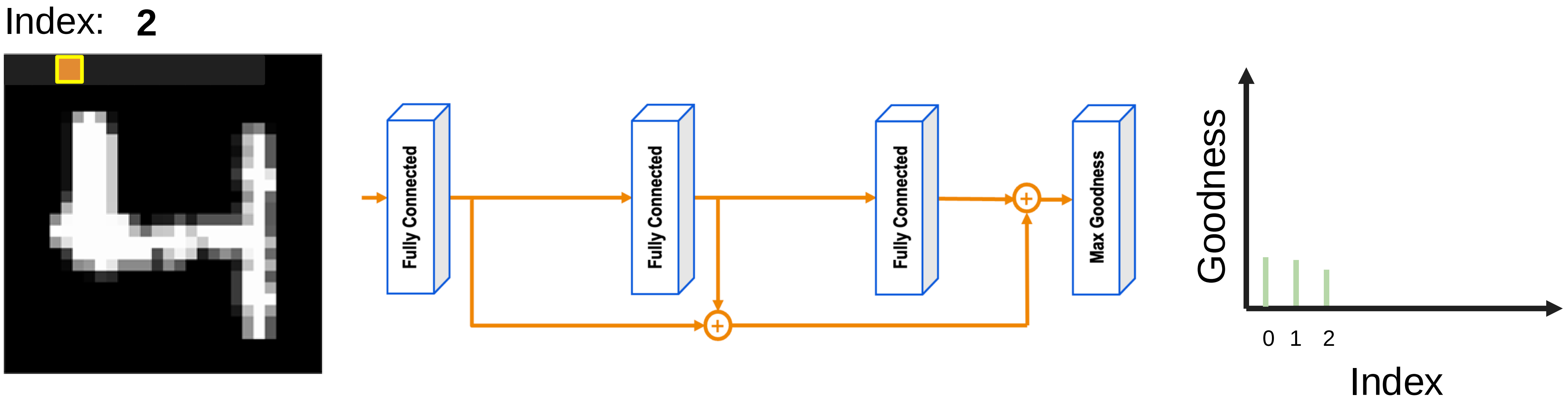
Glance at Hinton's FF Inference



The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

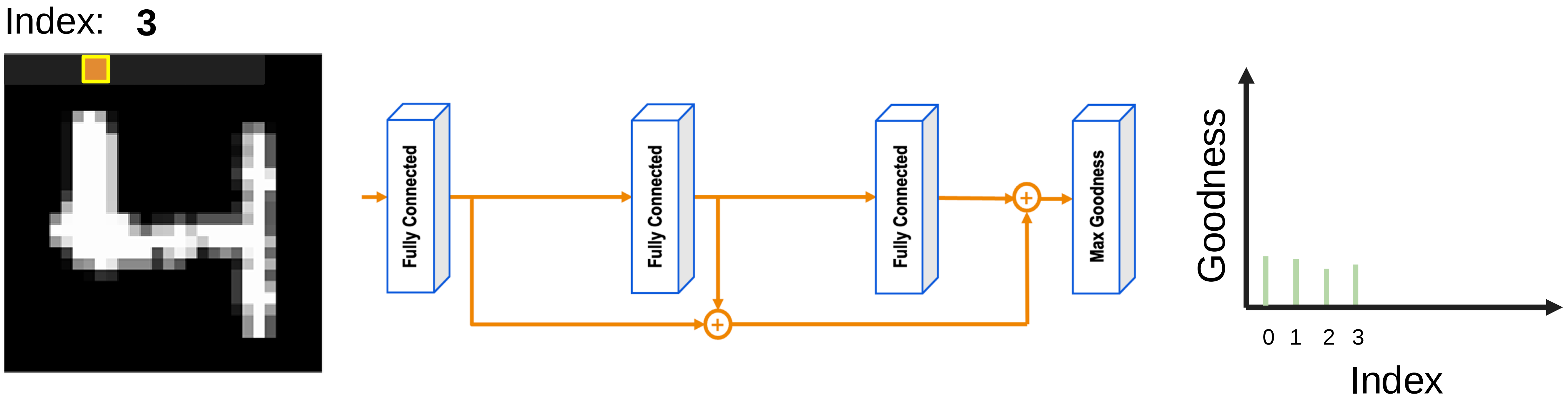
Glance at Hinton's FF Inference



The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

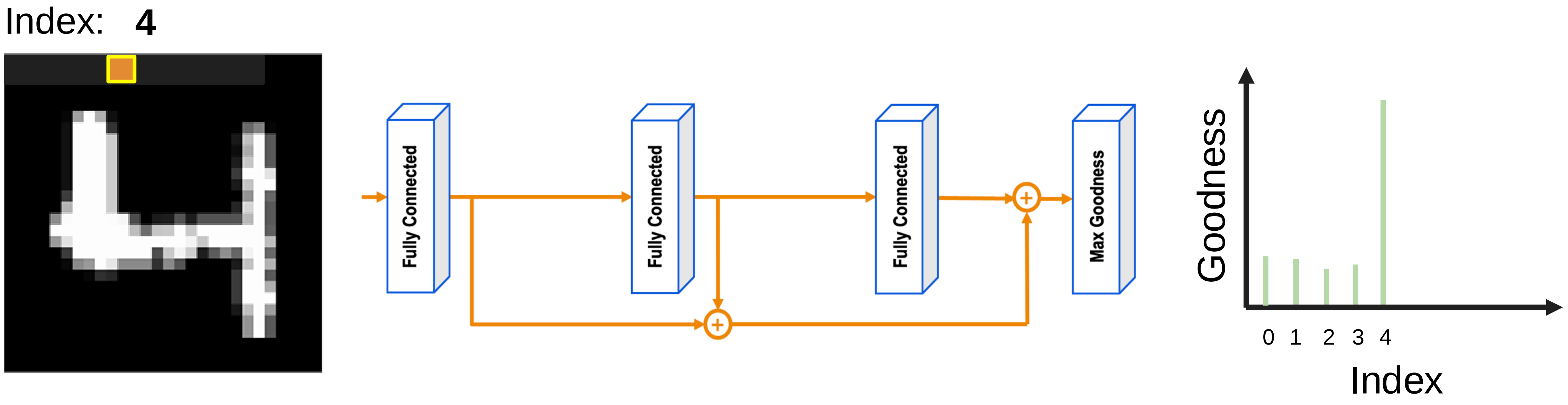
Glance at Hinton's FF Inference



The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

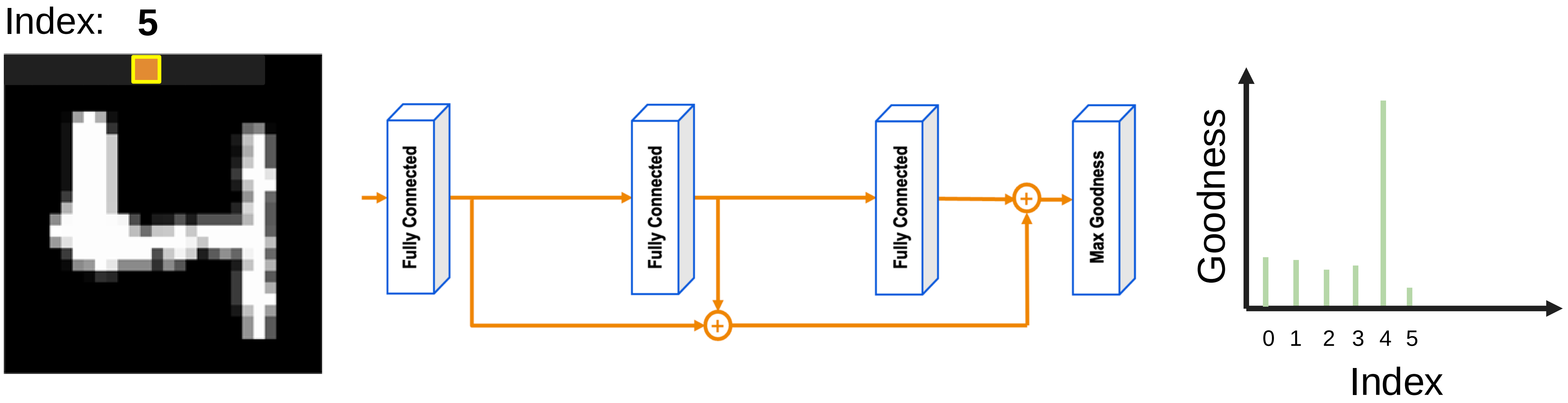
Glance at Hinton's FF Inference



The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

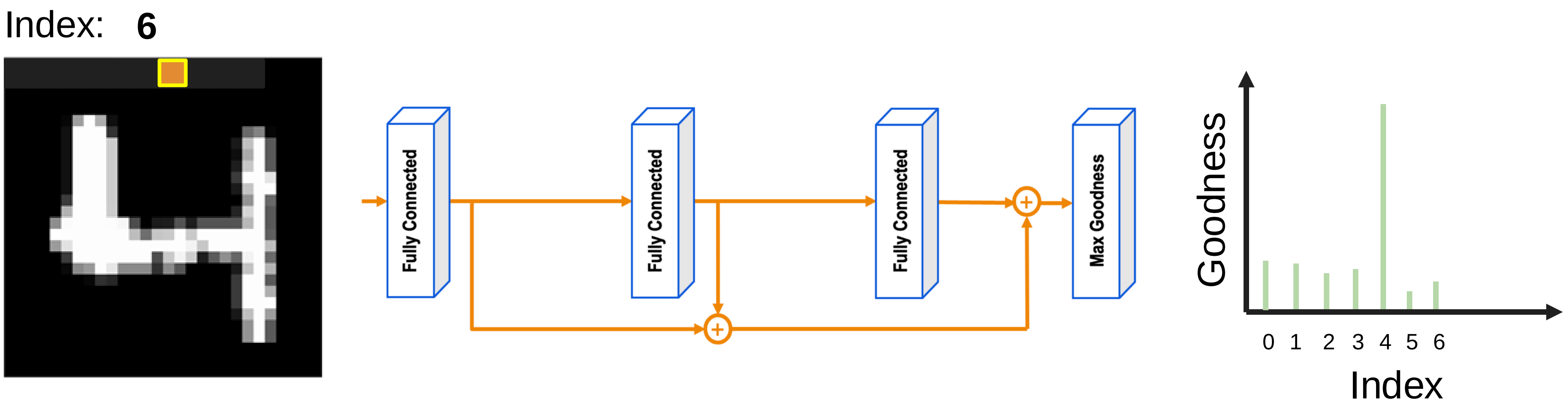
Glance at Hinton's FF Inference



The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

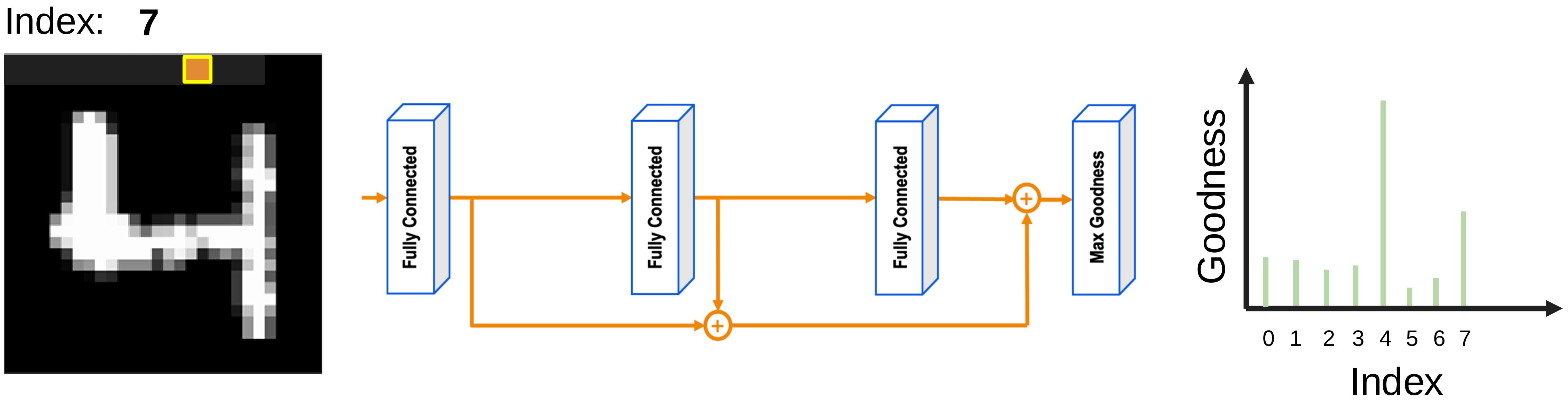
Glance at Hinton’s FF Inference



The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Glance at Hinton's FF Inference

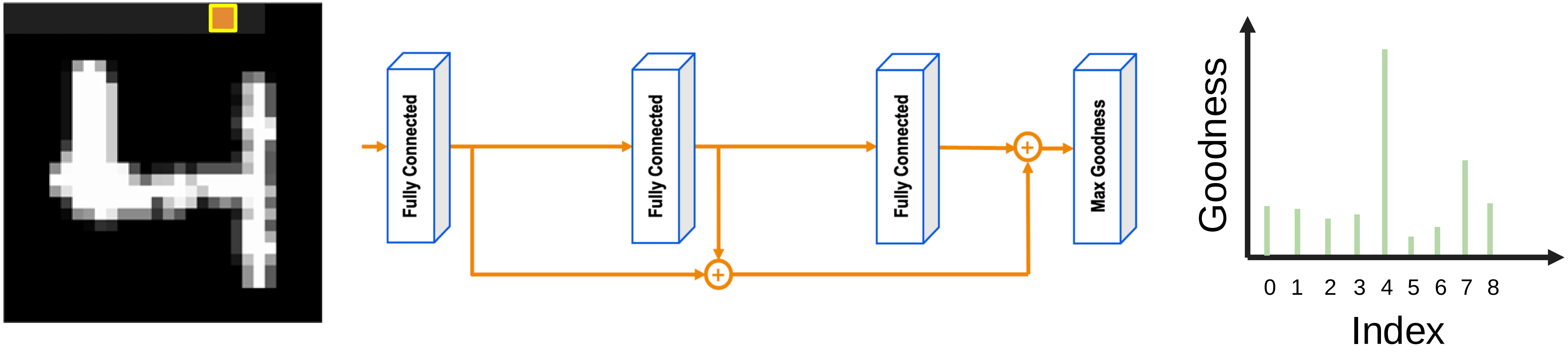


The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Glance at Hinton's FF Inference

Index: **8**

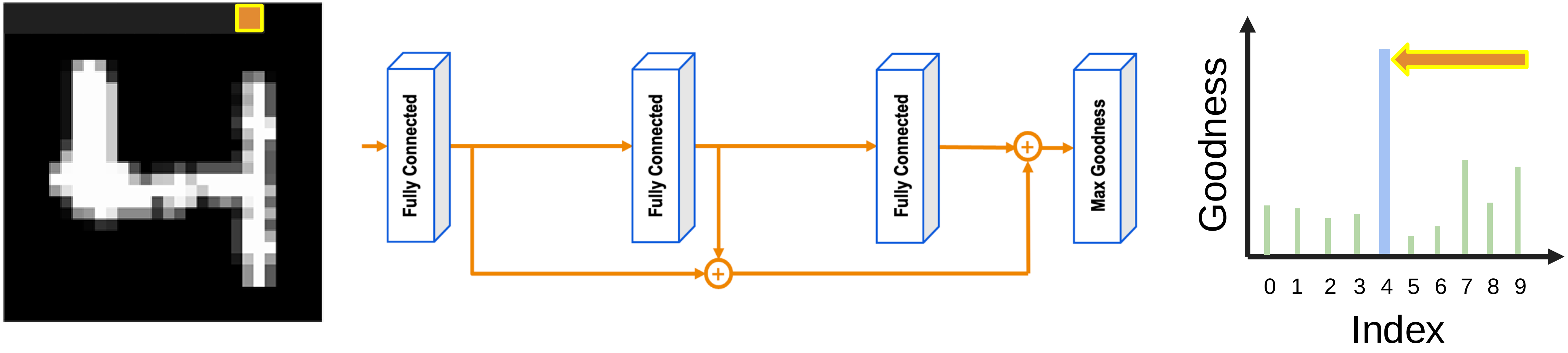


The Goodness is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Glance at Hinton's FF Inference

Index: **9**

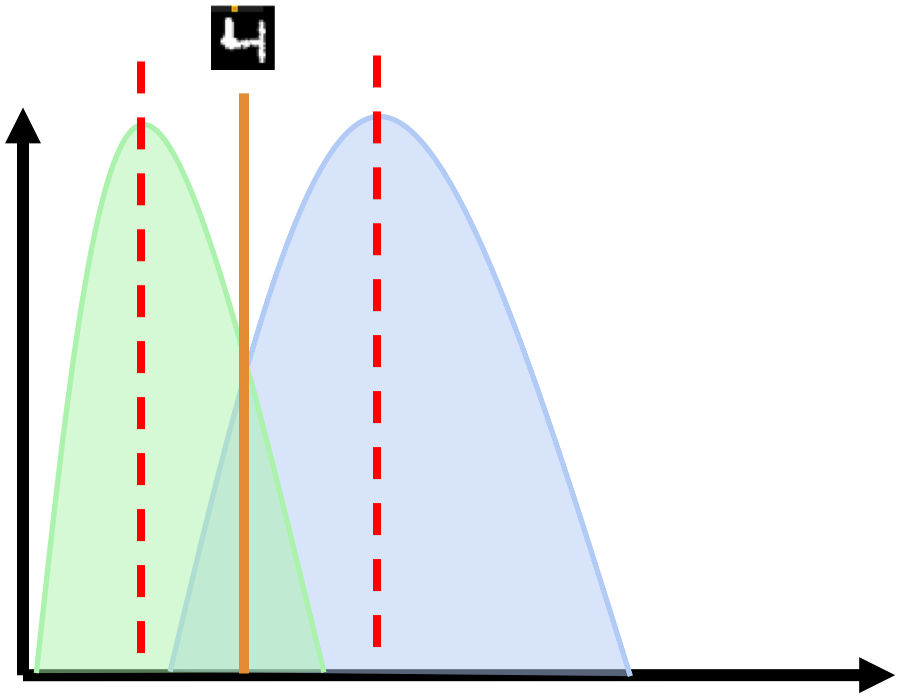


The <u>Goodness</u> is the sum of the square of the activity of each hidden neuron.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Insights From FF

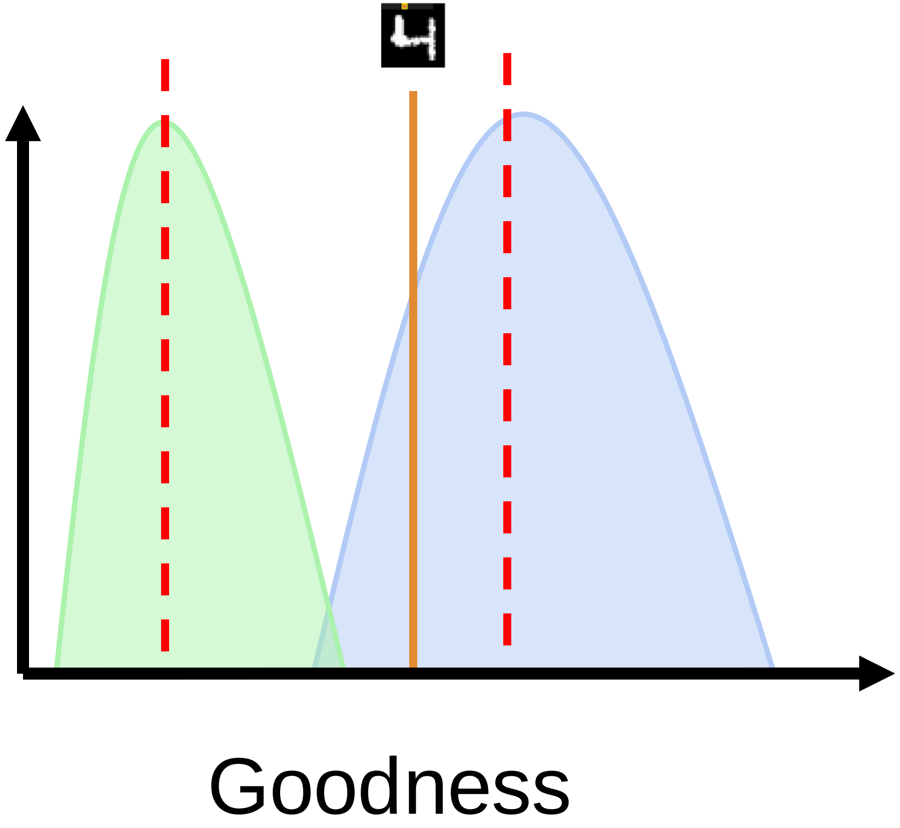
Current Layer



Goodness

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

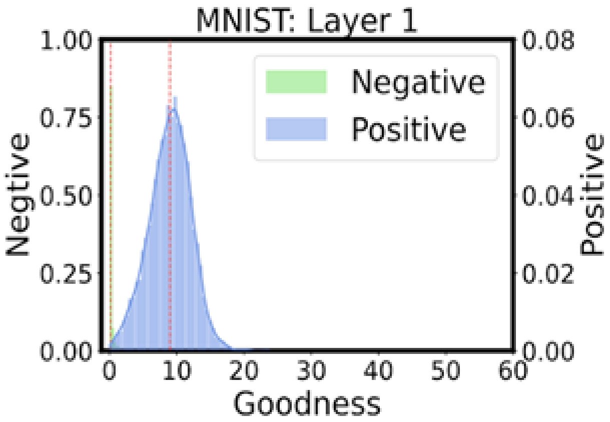
Insights From FF



Next Layer

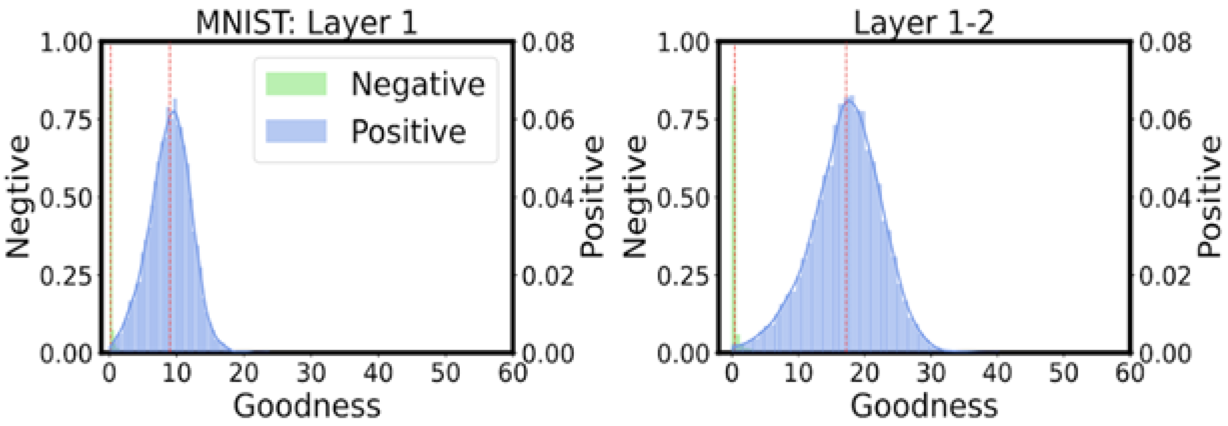
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Observation on FF



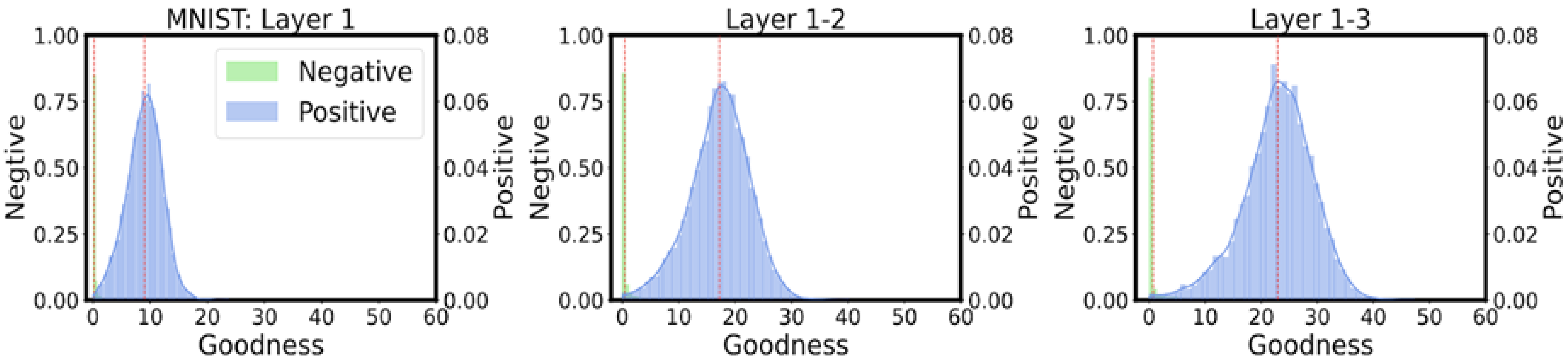
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Observation on FF



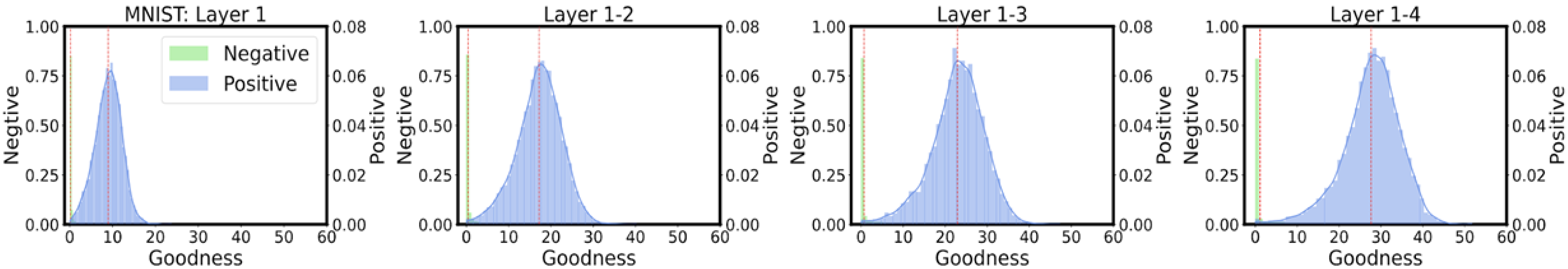
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Observation on FF



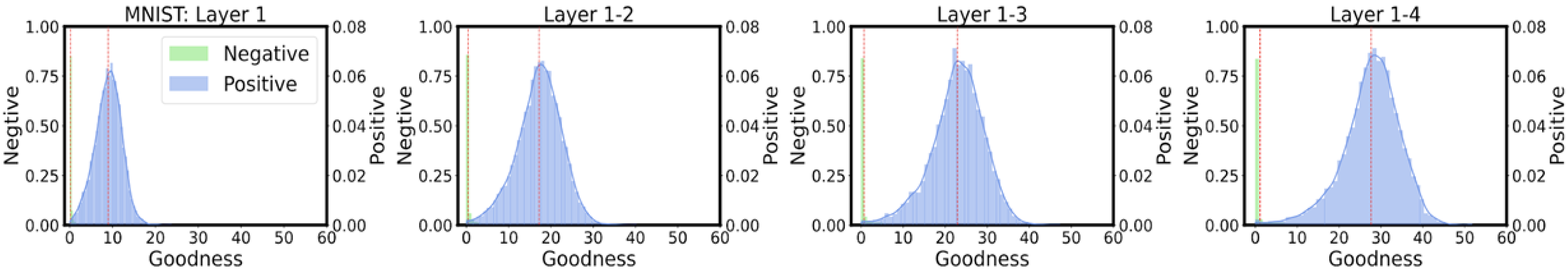
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Observation on FF



G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

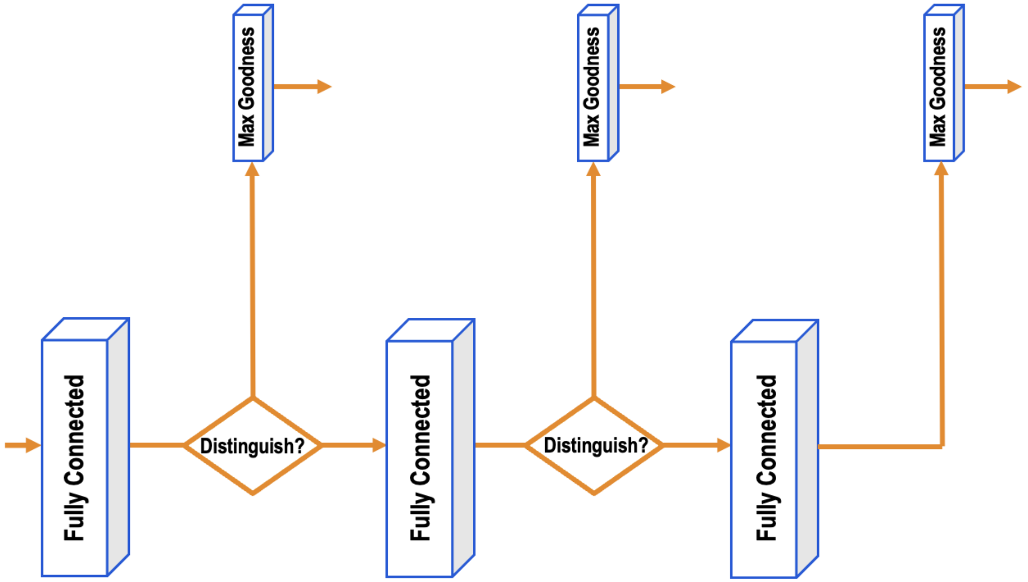
Observation on FF



The distance between the mean Goodness of the negative and positive samples/distributions increases as we consider more layers.

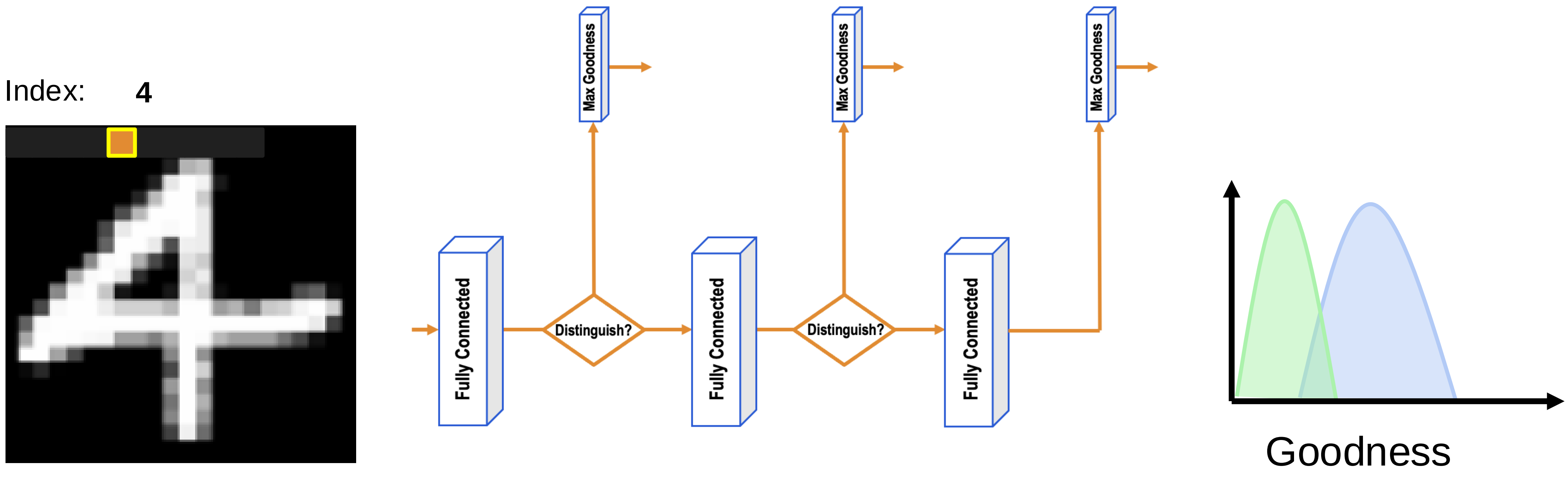
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF



G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

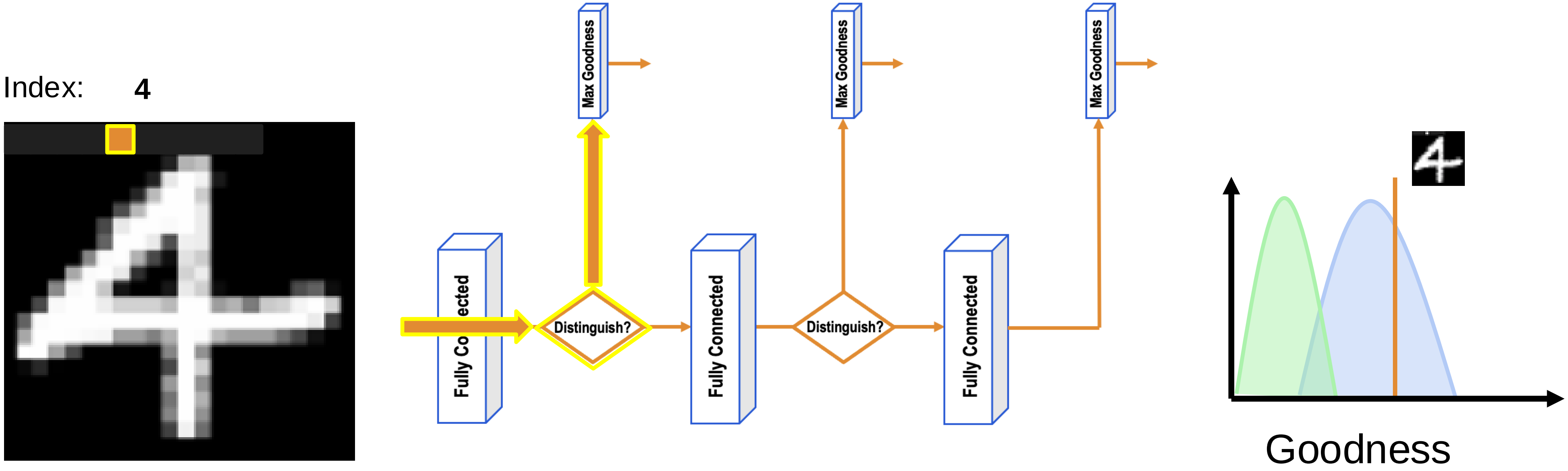
LightFF



G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF

Index: **4**

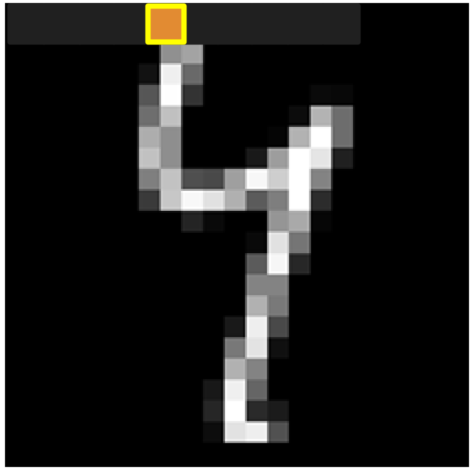


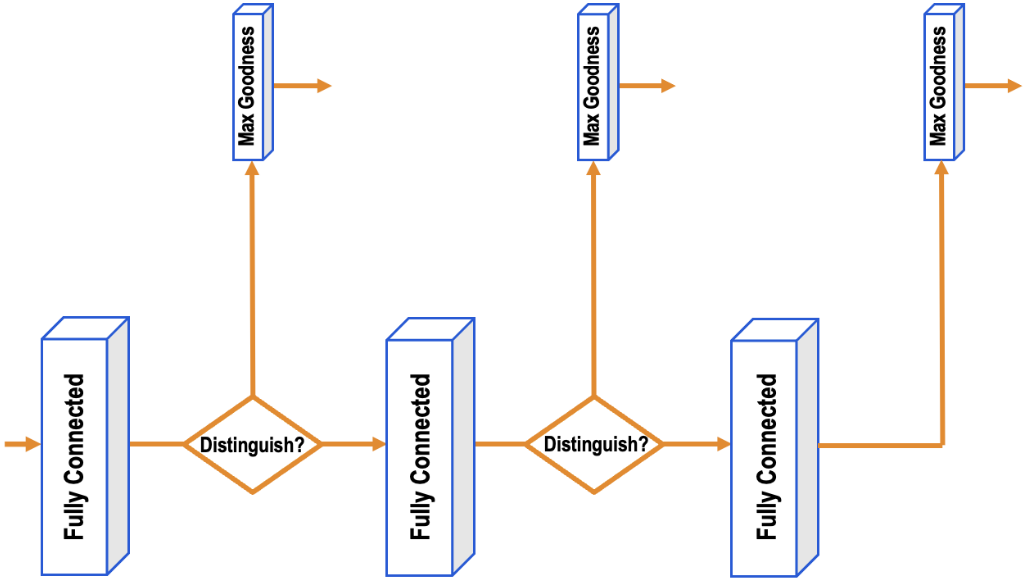
Distinguishable

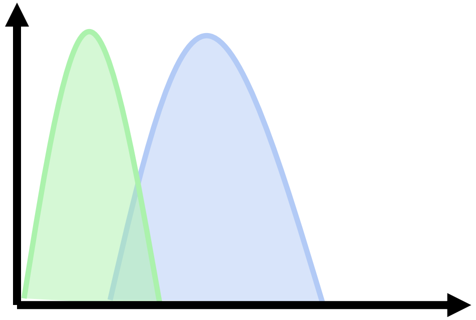
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF

Index: **4**



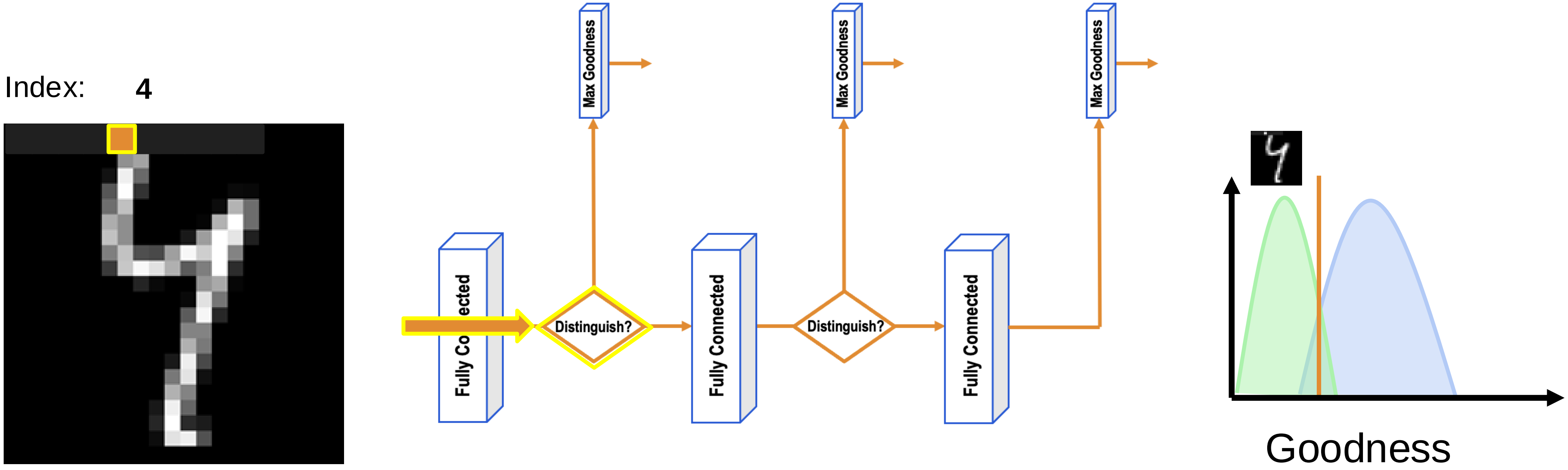




G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF

Index: **4**

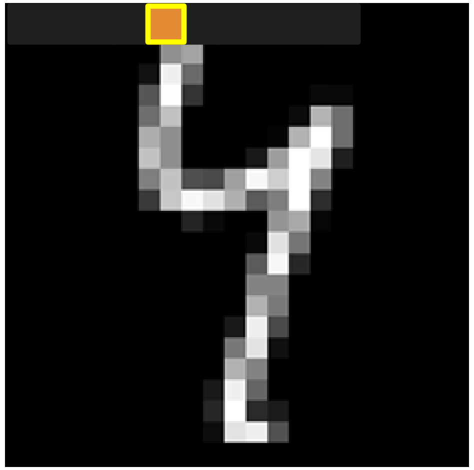


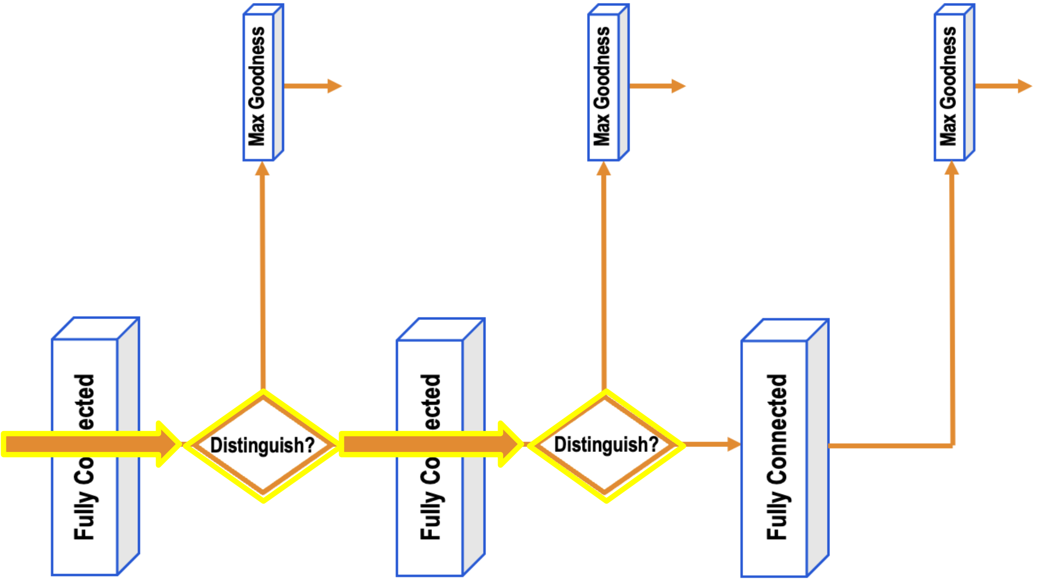
Indistinguishable

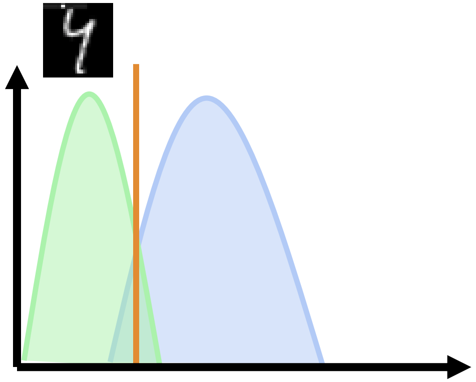
G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF

Index: **4**





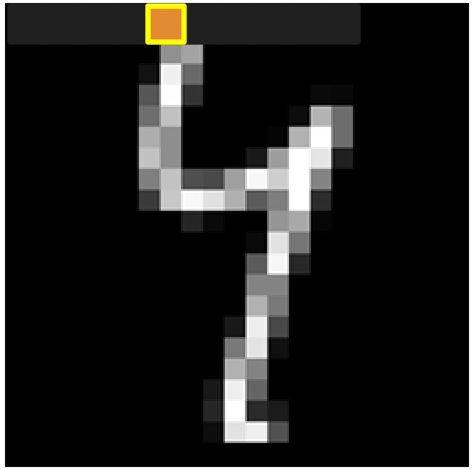


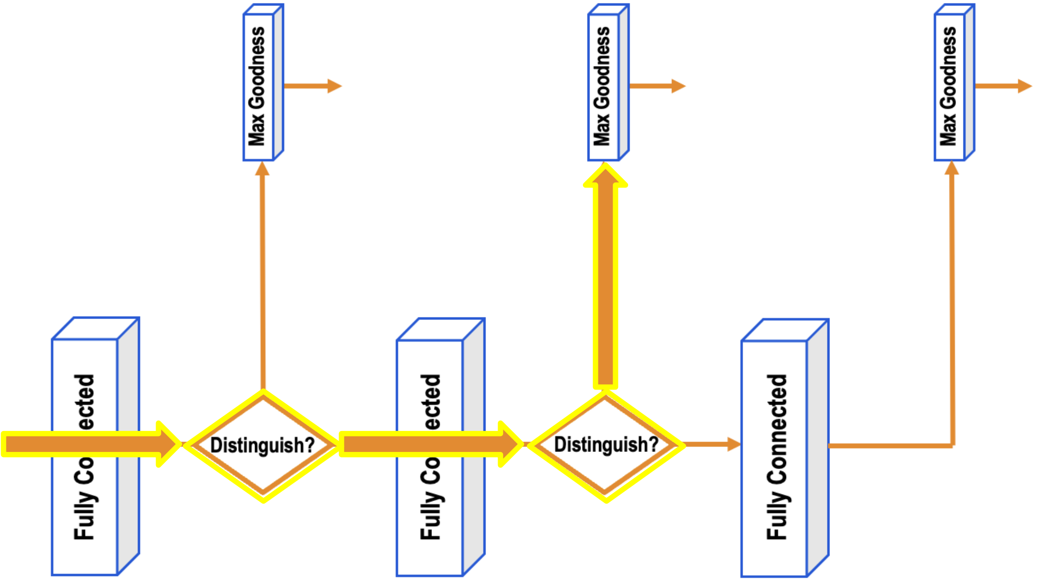
Goodness

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

LightFF

Index: **4**





Distinguishable

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Evaluation and Results

Dataset and Application

MNIST
Grayscale
Image

CIFAR-10
RGB
Images

A. H. Shoeb. Application of machine learning to epileptic seizure onset detection and treatment. PhD thesis, MIT, 2009.
R. Mark, et al. An annotated ecg database for evaluating arrhythmia detectors. IEEE Transactions on Biomedical Engineering, 1982.

Dataset and Application

MNIST
Grayscale
Image

CIFAR-10
RGB
Images

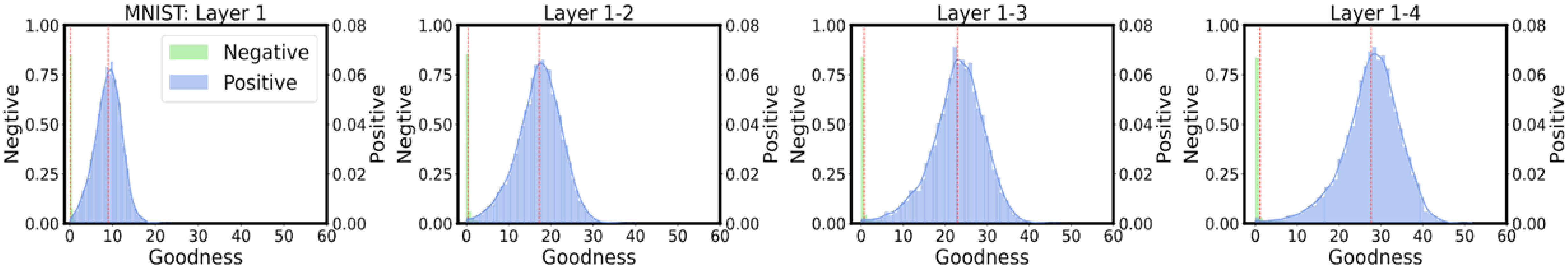
CHB-MIT
Electroencephalogram
(EEG)

MIT-BIH
Electrocardiogram
(ECG)

Real-world wearable applications:
Complexity overhead/energy consumption is a major constraint.

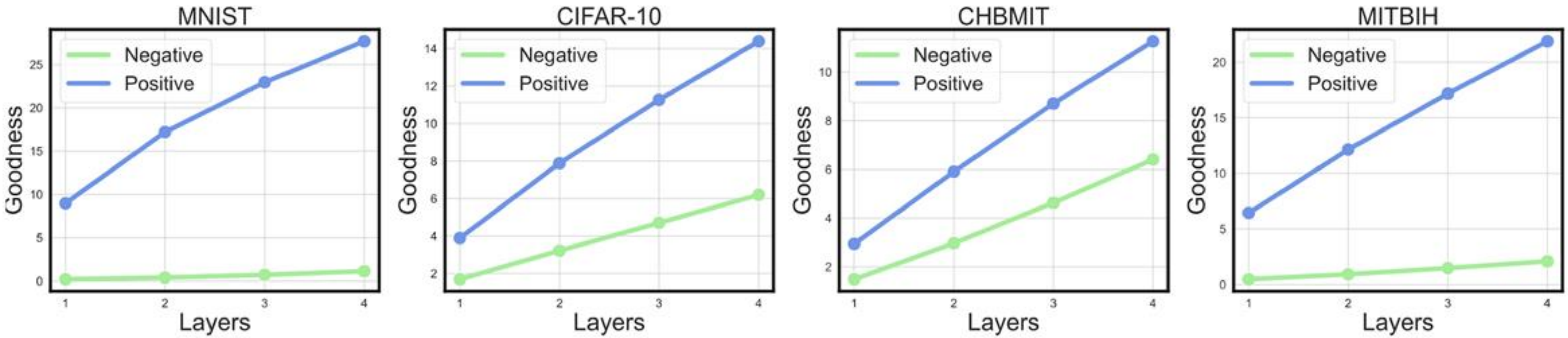
A. H. Shoeb. Application of machine learning to epileptic seizure onset detection and treatment. PhD thesis, MIT, 2009.
R. Mark, et al. An annotated ecg database for evaluating arrhythmia detectors. IEEE Transactions on Biomedical Engineering, 1982.

Generalized Insights



G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Generalized Insights



The distance between the mean Goodness of the negative and positive samples increases as more layers are taken into account.

G. Hinton. The forward-forward algorithm: Some preliminary investigations, 2022.

Classification Performance

Dataset	Error	
	FF	**LightFF**
MNIST	**1.51%**	**1.51%**
CIFAR-10	50.65%	**46.05%**
CHB-MIT	39.62%	**34.83%**
MIT-BIH	10.74%	**10.07%**

LightFF achieves a comparable classification error, compared to the Forward-Forward algorithm.

Lightweight Inference for BP

Dataset	Error	
	BP	LightBP
MNIST	**1.33%**	5.21%
CIFAR-10	**43.62%**	54.22%
CHB-MIT	**25.63%**	40.69%
MIT-BIH	**8.25%**	11.55%

Lightweight Inference for BP

LUND UNIVERSITY

Dataset	Error		
	BP		LightBP
MNIST	**1.33%**	<	5.21%
CIFAR-10	**43.62%**	<	54.22%
CHB-MIT	**25.63%**	<	40.69%
MIT-BIH	**8.25%**	<	11.55%

Goodness is a good metric for lightweight inference in LightFF.

Inference Complexity

Dataset	MACs	
	FF	**LightFF**
MNIST	12.94M	**2.95M**
CIFAR-10	17.30M	**10.31M**
CHB-MIT	13.39M	**8.39M**
MIT-BIH	11.90M	**1.99M**

Inference Complexity

Dataset	MACs			
	FF		LightFF	
MNIST	12.94M	>	**2.95M**	4.4x
CIFAR-10	17.30M	>	**10.31M**	1.7x
CHB-MIT	13.39M	>	**8.39M**	1.6x
MIT-BIH	11.90M	>	**1.99M**	6.0x

LightFF improves computational efficiency, compared to the Forward-Forward algorithm.

Execution Time

Dataset	Execution Time	
	FF	**LightFF**
MNIST	22.81ms	**3.39ms**
CIFAR-10	29.57ms	**16.38ms**
CHB-MIT	4.73ms	**2.85ms**
MIT-BIH	11.26ms	**1.83ms**

Execution Time

Dataset	Execution Time			
	FF		**LightFF**	
MNIST	22.81ms	>	**3.39ms**	6.7x
CIFAR-10	29.57ms	>	**16.38ms**	1.8x
CHB-MIT	4.73ms	>	**2.85ms**	1.7x
MIT-BIH	11.26ms	>	**1.83ms**	6.2x

LightFF reduces the execution time, compared to the Forward-Forward algorithm.

Conclusion

Challenge

Deep Learning
to be **Efficient**

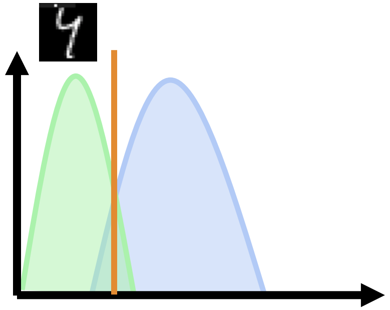
Conclusion

Challenge

Deep Learning
to be **Efficient**

Approach

Lightweight Inference
For Forward-Forward



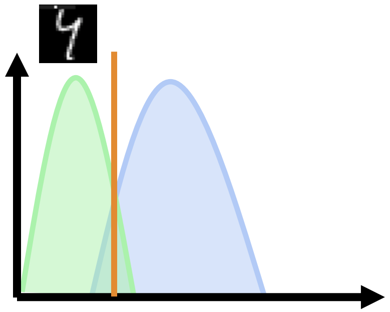
Conclusion

Challenge

Deep Learning
to be **Efficient**

Approach

Lightweight Inference
For Forward-Forward



Performance

Improves Efficiency
Comparable Error

Dataset	MNIST		
	FF		**LightFF**
Error	**1.51%**	=	**1.51%**
Execution Time	22.81ms	>	**3.39ms**
MACs	12.94M	>	**2.95M**

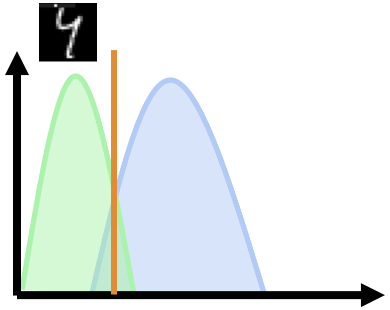
Conclusion

Challenge

Deep Learning
to be **Efficient**

Approach

Lightweight Inference
For Forward-Forward



Performance

Improves Efficiency
Comparable Error

Dataset	MNIST		
	FF		**LightFF**
Error	**1.51%**	=	**1.51%**
Execution Time	22.81ms	>	**3.39ms**
MACs	12.94M	>	**2.95M**

Paper

Wednesday Outreach Activities

Thank you!

Code