Moving brain tumor diagnosis to cost-efficient systems.

Brad Niepceron, Ahmed Nait-Sidi-Moh and Filippo Grassia. University of Picardie Jules Vernes

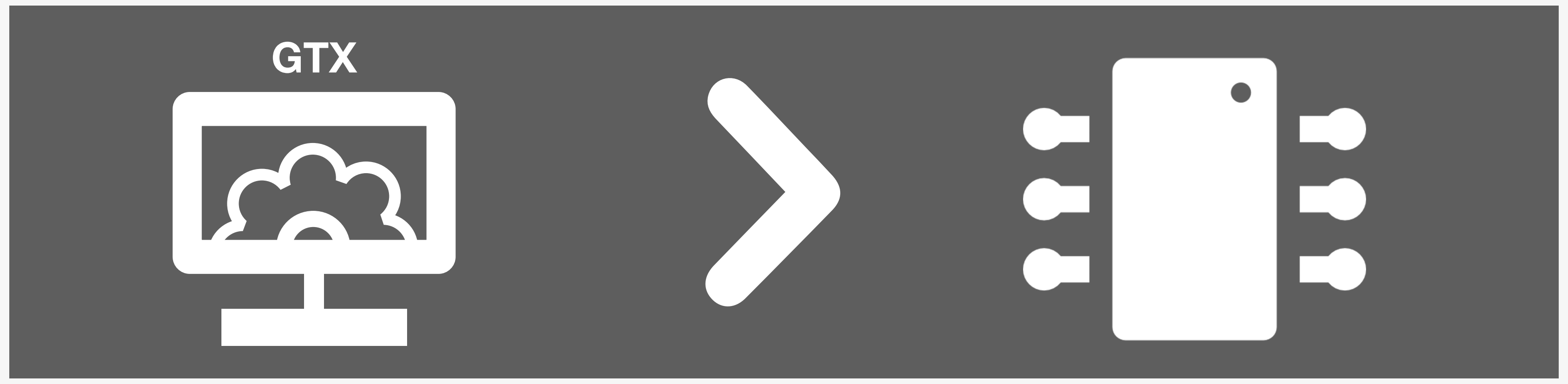
Speaker : *Brad Niepceron. brad.niepceron@etud.u-picardie.fr*

Laboratoire des Technologies Innovantes

INTRODUCTION

1

1 INTRODUCTION



Why moving brain tumor diagnosis computation to cost-efficient systems ?

1 INTRODUCTION

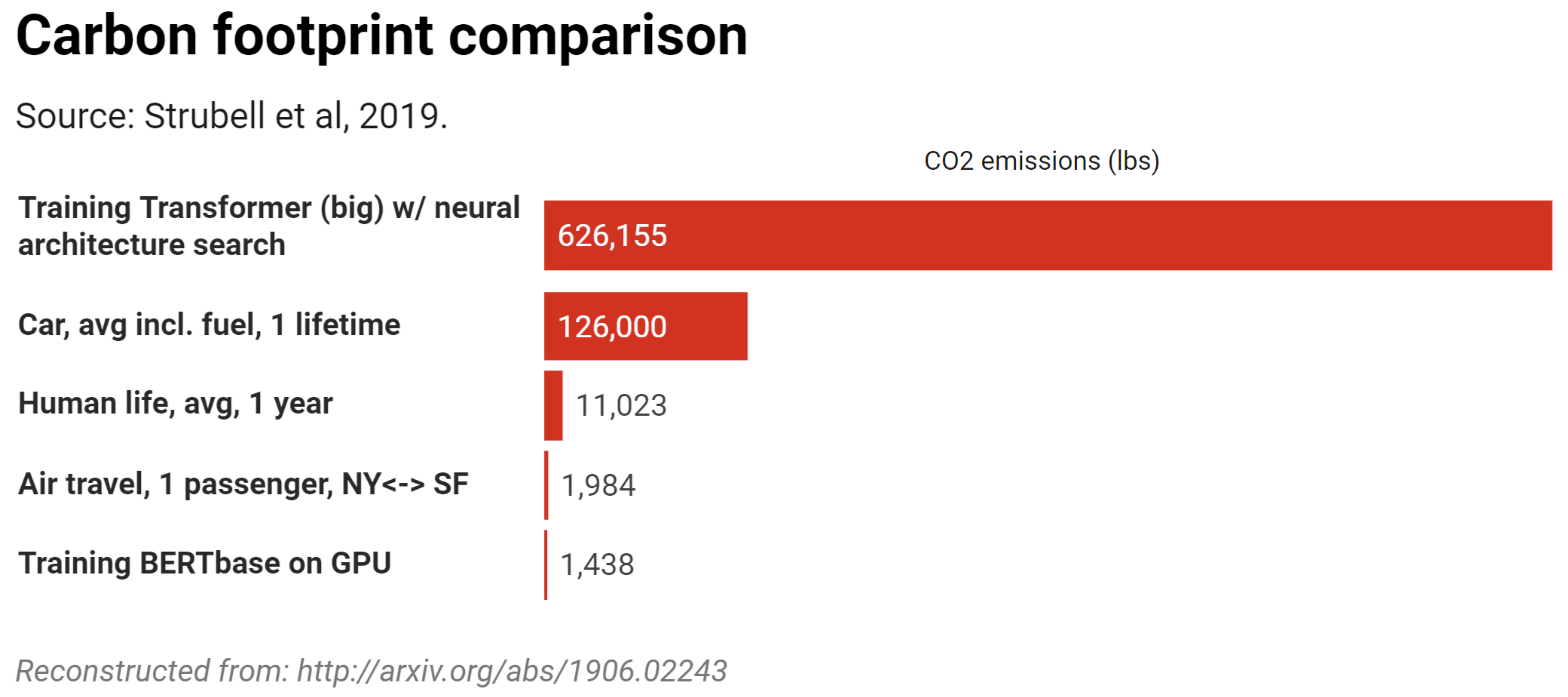


Fig 1. Carbon footprint benchmark including a DL model. Source : Strubell et al. Energy and Policy Considerations for Deep Learning in NLP. 2019.

Deep Learning issues

- Long training time
- GPU dependent
- High energy cost
- Blackbox effect
- No architecture rules

1 INTRODUCTION

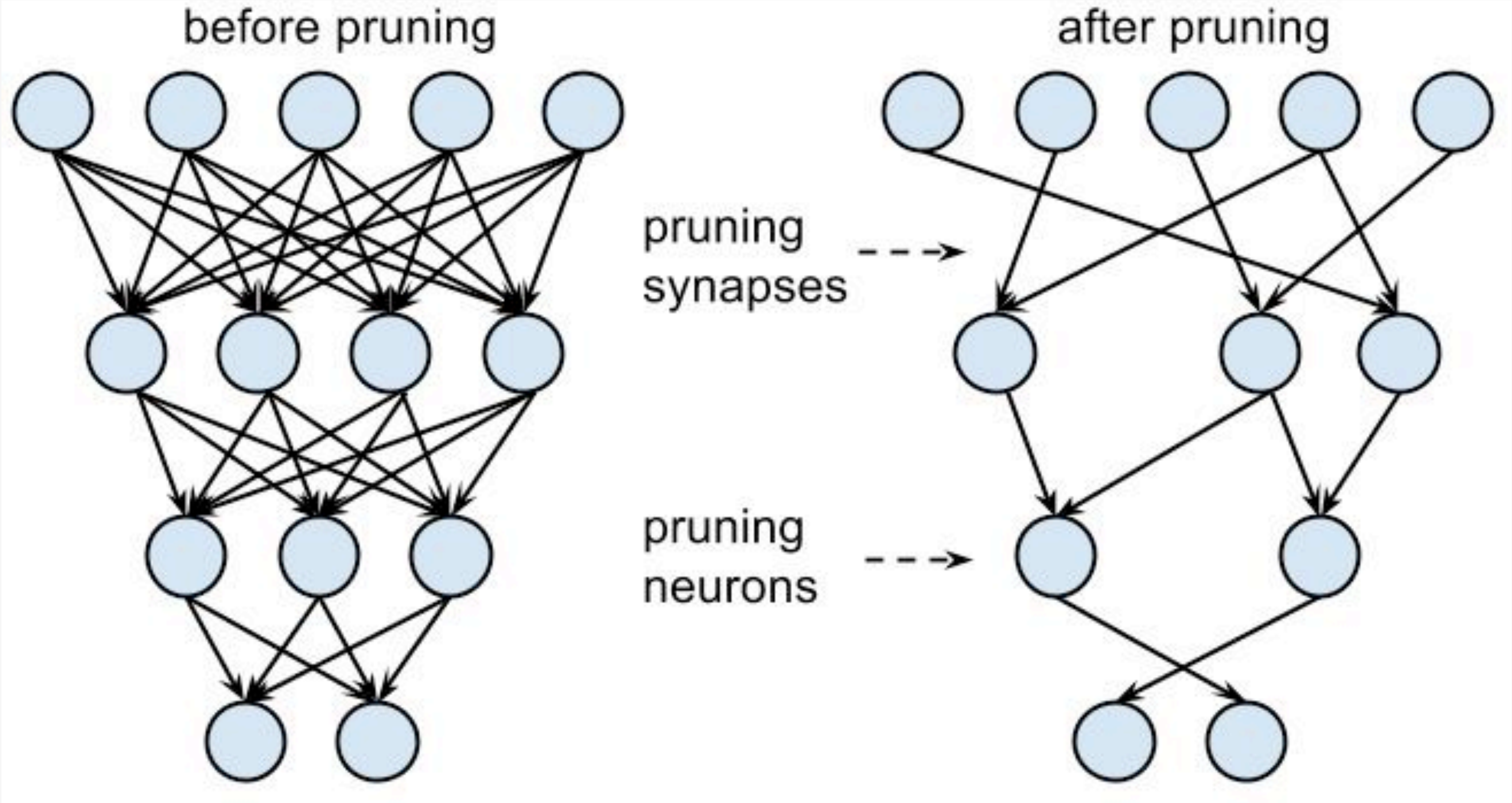


Fig2. Neural network pruning. Source : Song Han. Oreilly.com

Model Compression

- Reduce the number of parameters
- Speed-up computation
- Reduce memory footprint
- Reduce energy consumption

1 INTRODUCTION

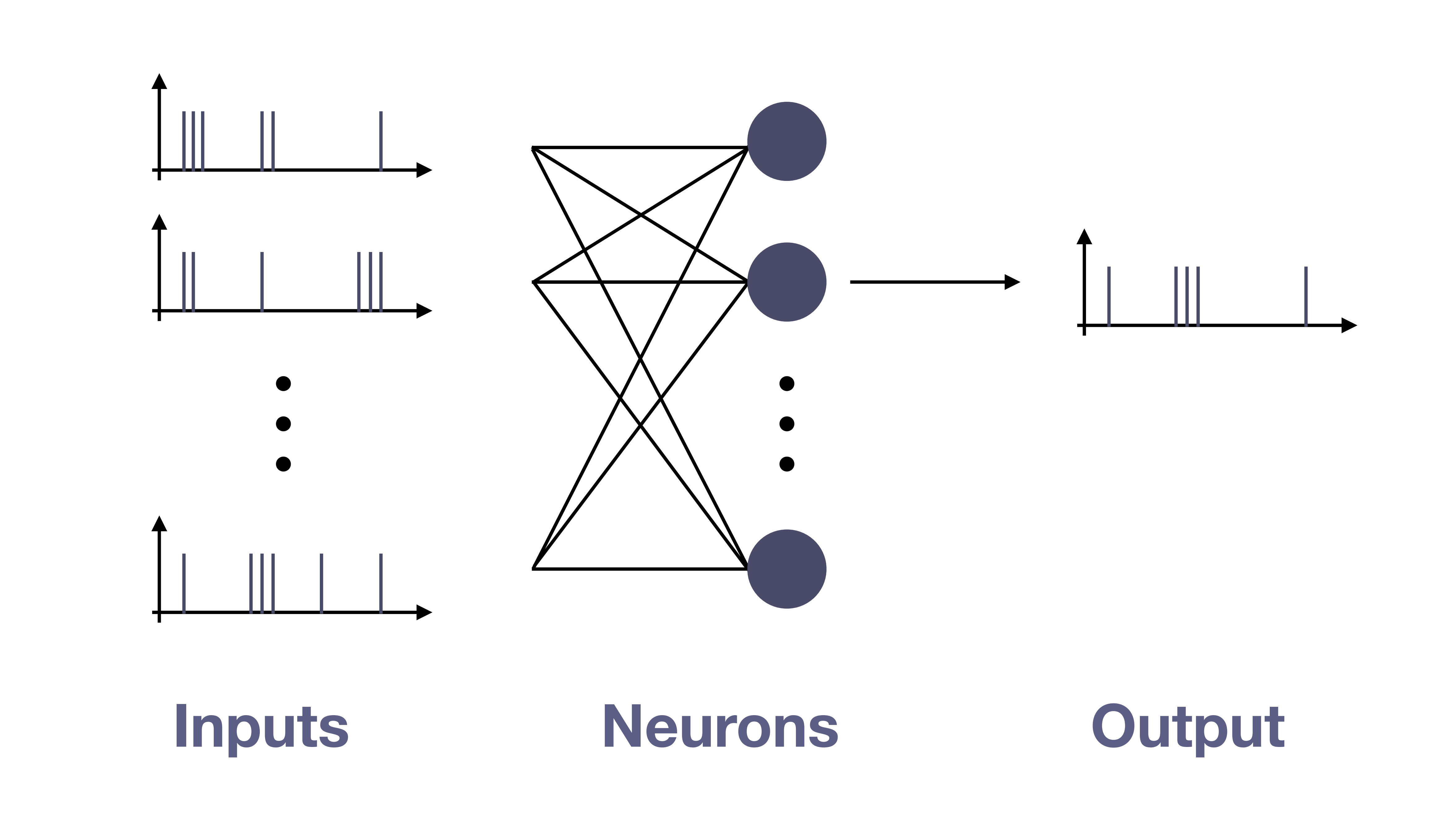


Fig 3. A Spiking Neural Network

- **Spike-based computation**
- Rise of Spiking Neural Nets. (SNN)
- Towards a non-abstract AI
- Cost-efficient models
- Easy large-scale deployment

METHODS

2 METHODS

CNN Compression

2

METHODS

1. CNN Compression

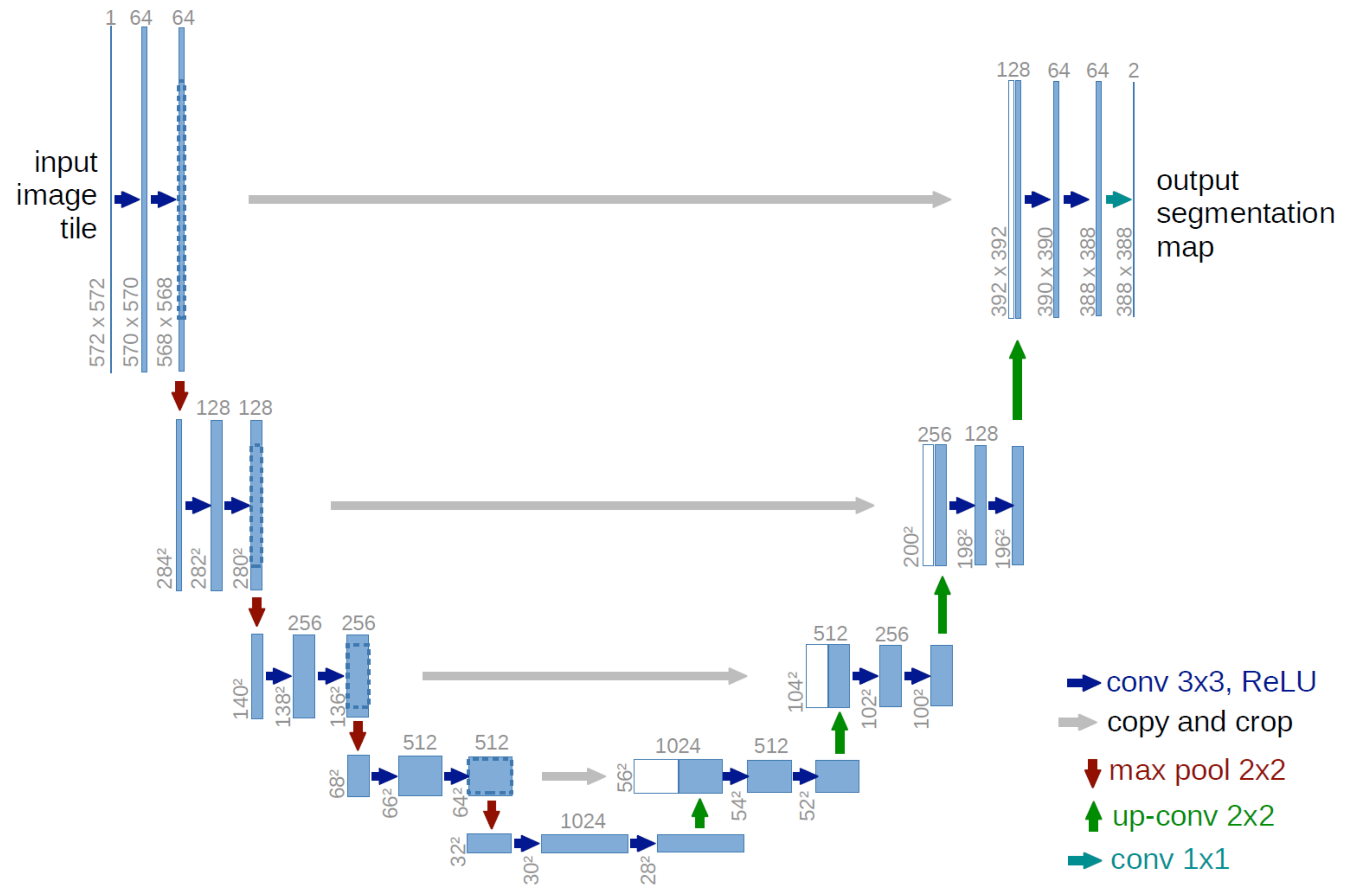


Fig 4. U-Net architecture for medical image segmentation. Source : Olaf Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

2

METHODS

1. CNN Compression

I Group Normalization

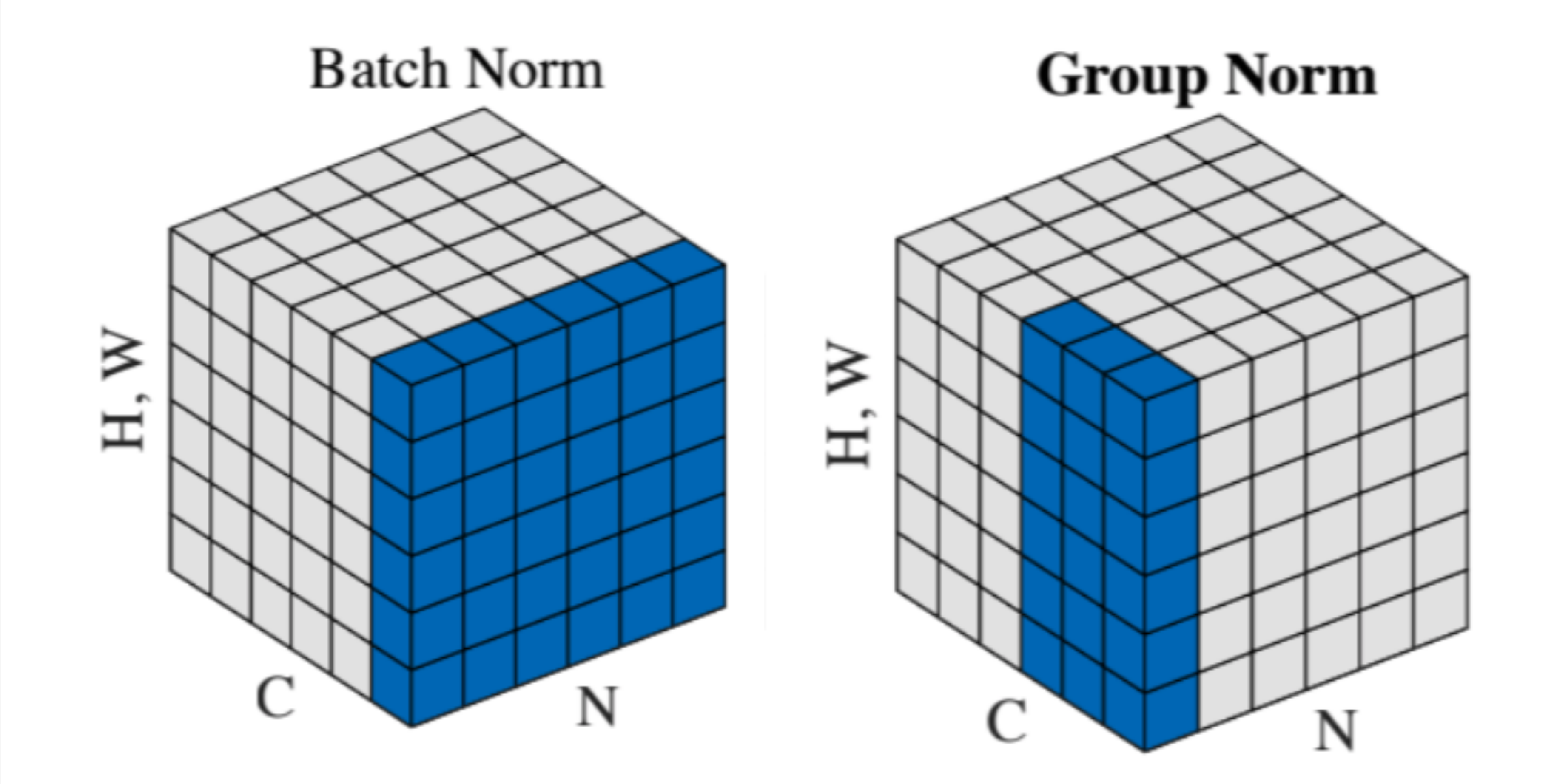


Fig 5. Normalization method comparison. Source : Yuxin Wu, Kaiming He. Group Normalization 2018.

- Deals with batch statistics estimation
- Prevents the use of large batches
- Ideal for resource-constrained systems

2

METHODS

1. CNN Compression

2 Depthwise Separable Convolution (DSC)

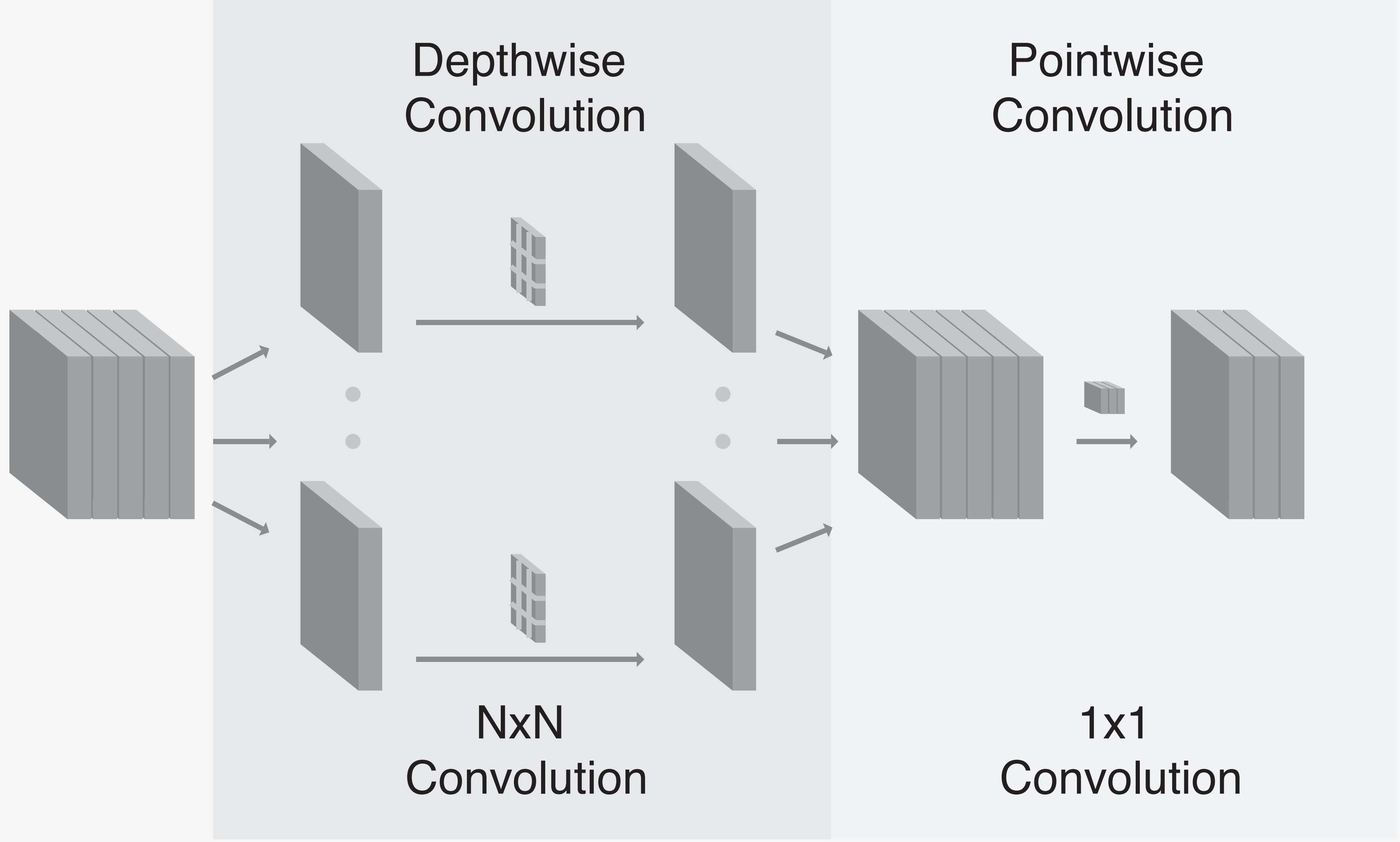


Fig 6. Depthwise Separable Convolution. Source : Niepceron et al. Moving Medical Image Analysis to GPU Embedded Systems: Application to Brain Tumor Segmentation. 2020.

- Factorization of the conv. process
- Conv. performed on individual channels
- Breaks down kernel dimension
- Reduces de number of parameters

3

Cost efficient conv. block.

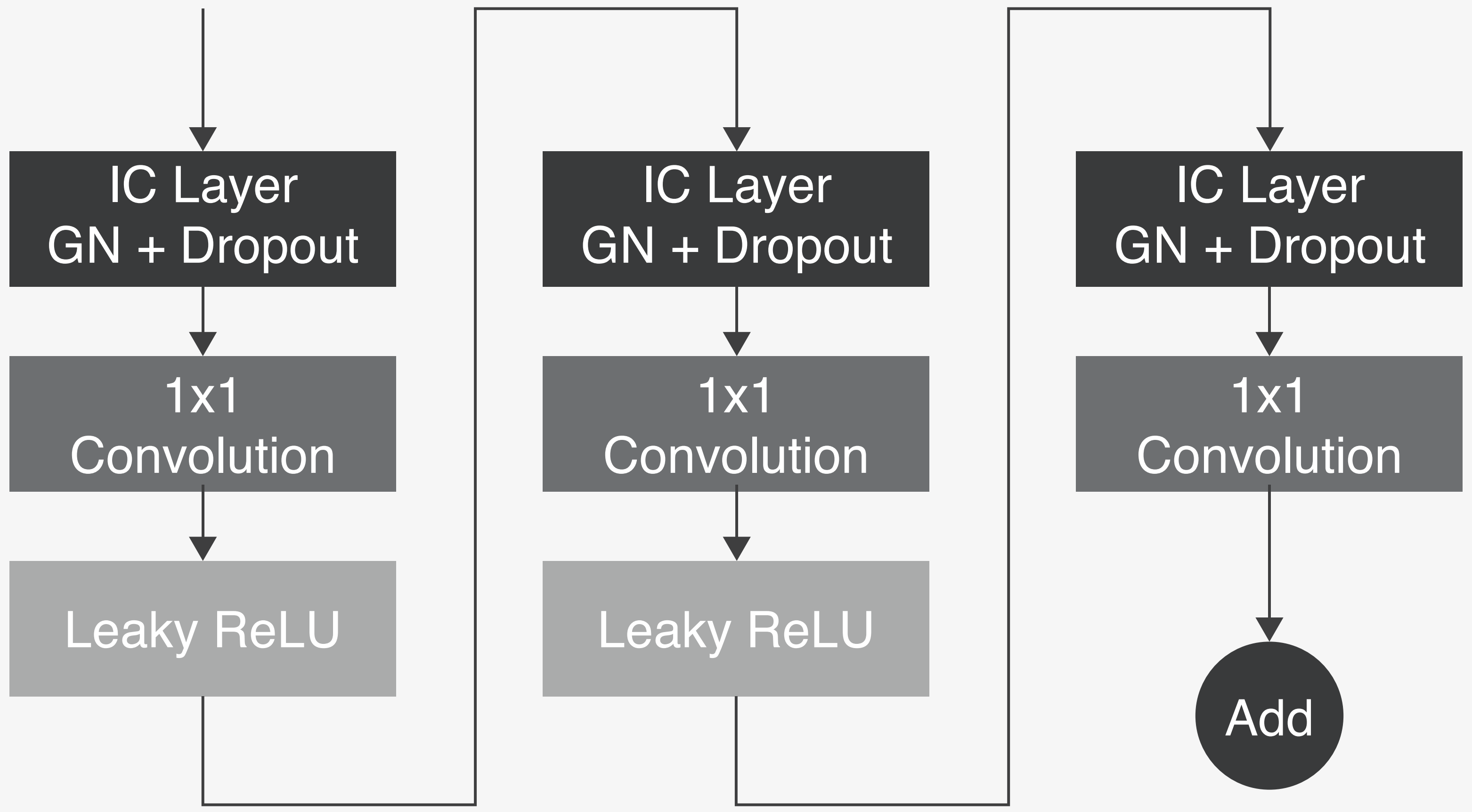


Fig 7. Depthwise Convolutional Block. Source : Niepceron et al. 2020.

- Independent Component Layer
- ResNet-like conv. block.
- Replacing each U-net conv.

METHODS

1. CNN Compression

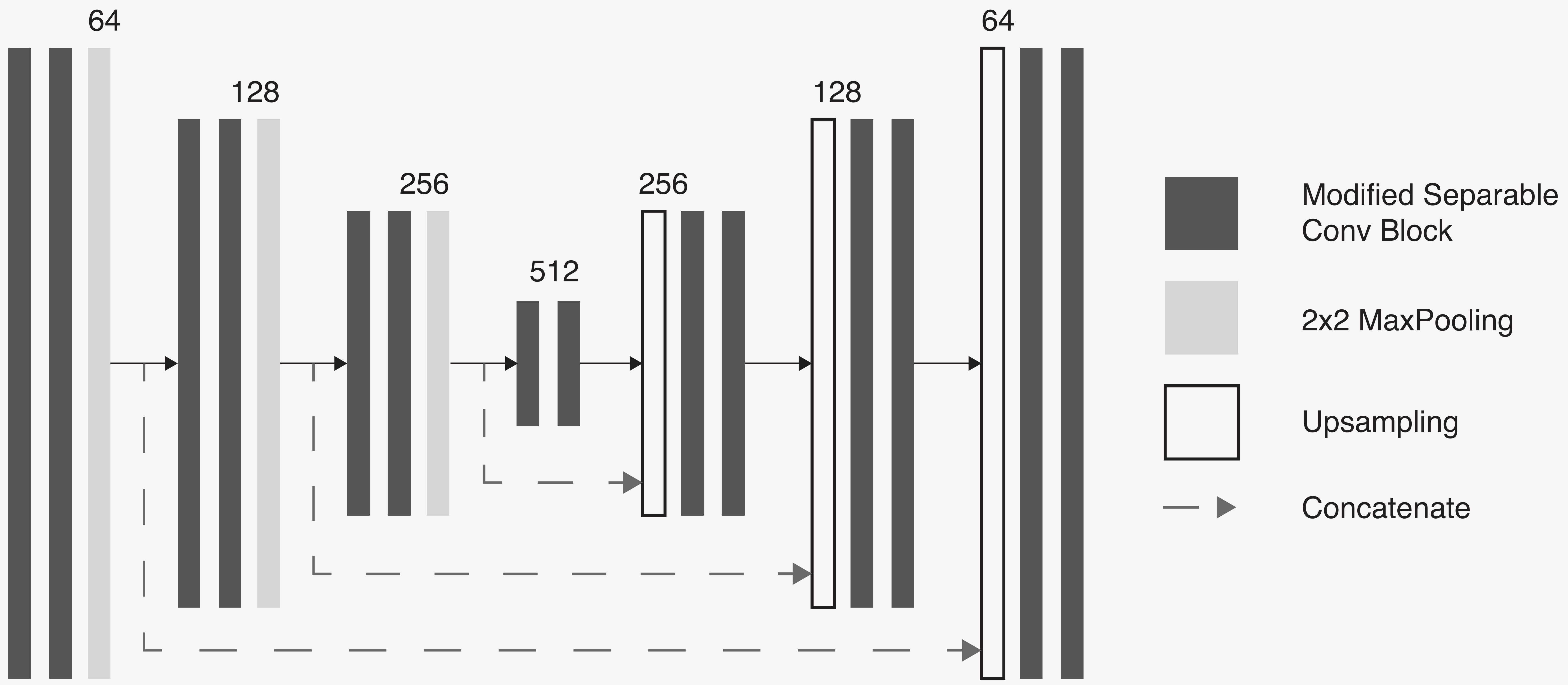


Fig 8. Compressed U-Net architecture. Source : Niepceron et al. 2020.

METHODS

1. CNN Compression

NVIDIA Jetson AGX Xavier

	MAXN	10W	15W	30W_ALL
CPU Cores	8	2	4	8
CPU Max Freq. (MHz)	1377	520	6700	900
GPU Max Freq. (MHz)	2265.6	1200	1200	1200
Memory Max Freq. (MHz)	2133.6	1066	1333	1600

Fig 9. NVIDIA Jetson Xavier properties. Source : Niepceron et al. 2020.

- Small weight and size
- Low power consumption
- Deep Learning Accelerator
- JetPack SDK
- Large memory

2 METHODS

Spike-based model for tumor recognition

2

METHODS

2. Spike-based model for tumor recognition

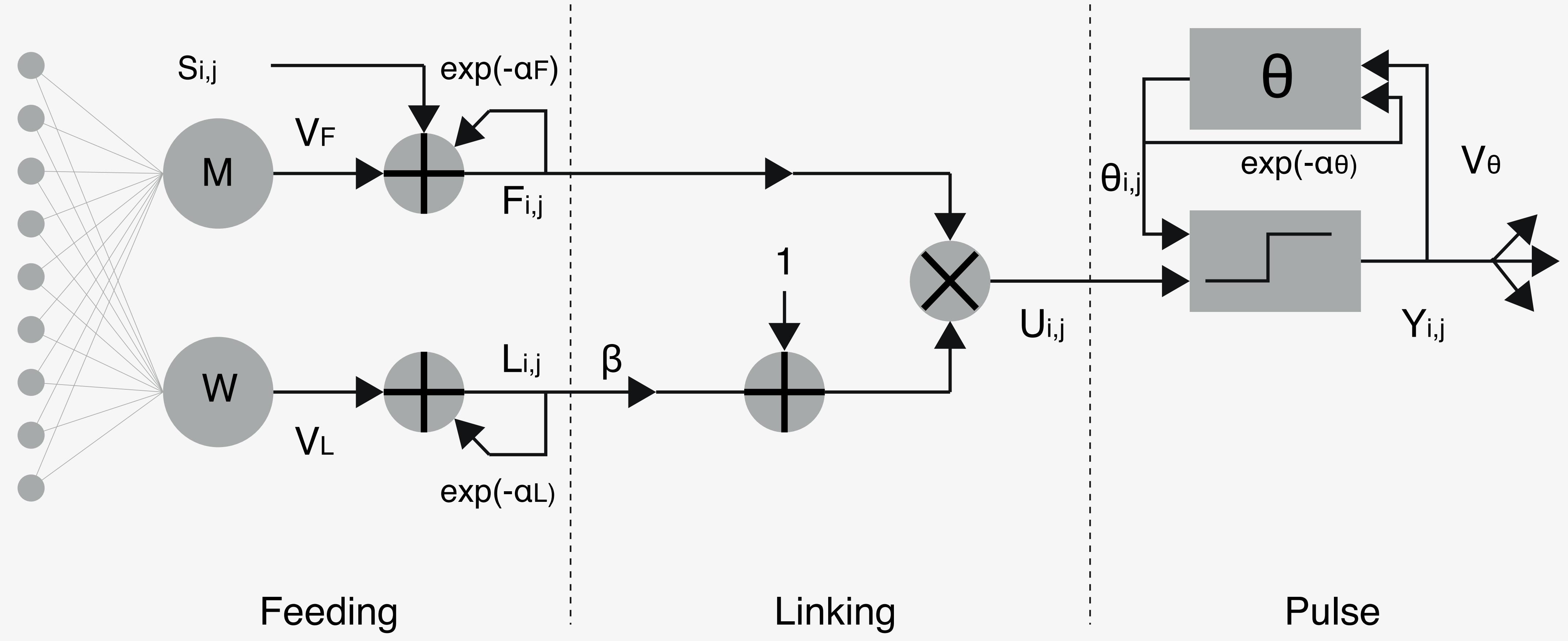


Fig 10. Standard PCNN. Source : Niepceron et al. 2021.

- Laterally-connected neurons
- 2D input image
- Iterative model
- Outputs a segmentation map
- **Feeding :** Computes voltage with input stimulus
- **Linking :** Updates neuron's internal activity
- **Pulse :** Fire if membrane potential exceeds threshold

2

METHODS

2. Spike-based model for tumor recognition

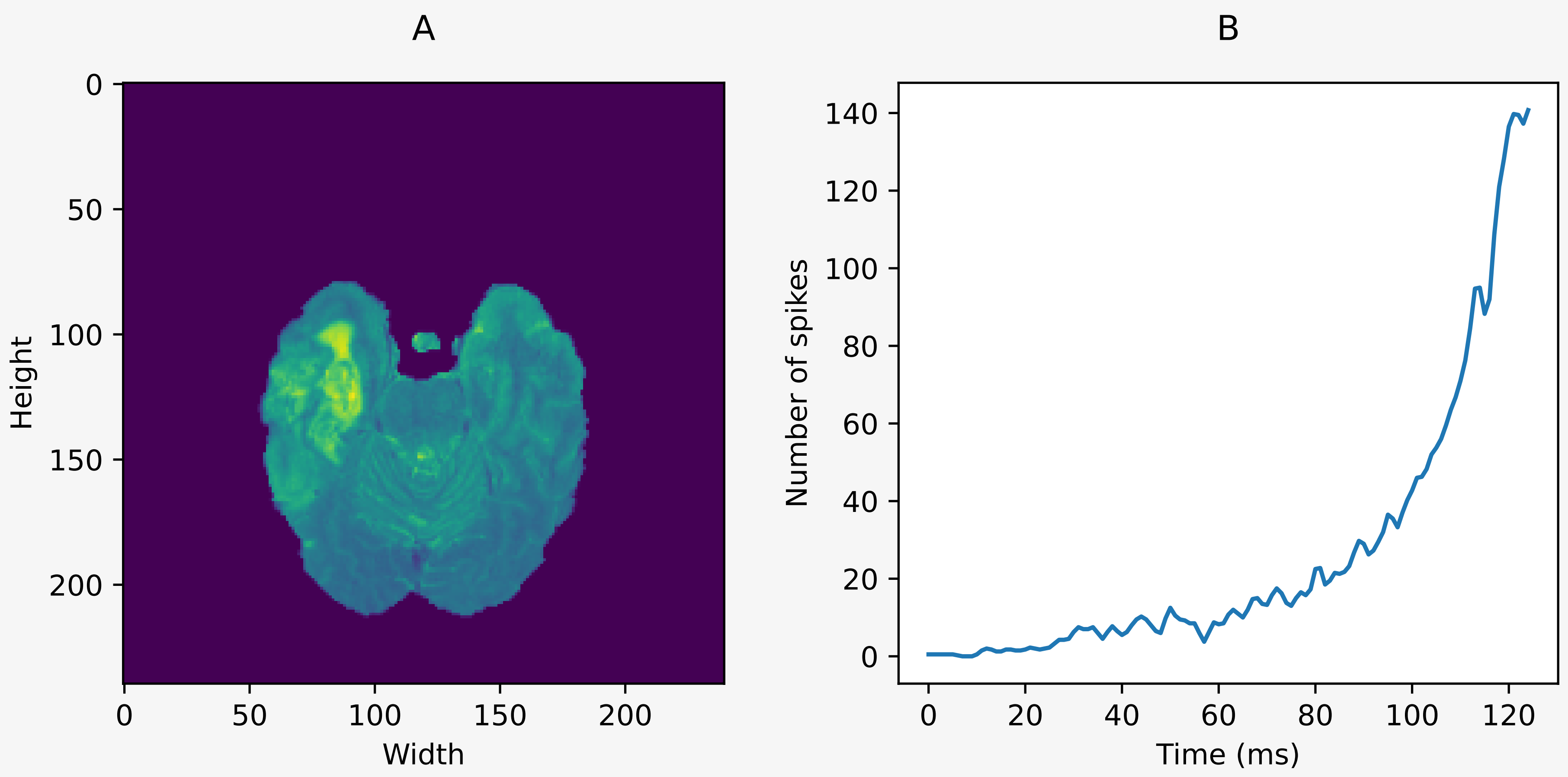


Fig 11. Image Signature (B) of an MRI Slice (A). Source : Niepceron et al. 2021.

- Image Signature as feature extraction
- Gets the firing activity as a time-series
- Comparison with regions of interest
- Already used for object detection without region selection methods

METHODS

2. PCNN for visual diagnosis

Fig 12. The Selective Search algorithm. Source : arthurdouillard.com.

2

METHODS

2. Spike-based model for tumor recognition

Step 1 : Compute Selective Search on M to create B_{ij} bounding boxes
Step 2 : Remove boxes with area greater than threshold θ
Step 3 : Extract image patches P_{ij} from the boxes
Step 4 : Convert each P_{ij} to signature S_{ij}
Step 5 : Compute the Euclidean distance between each S_{ij} and the signature M_{sign} obtained from M
Step 6 : Retain B_{min} the box that gives the smallest distance D_{min} as the complete tumor detection box

Fig 13. Tumor detection algorithm based on PCNN feature extraction. Source : Niepceron et al. 2021.

RESULTS

4

RESULTS

1. CNN Compression

DETAILS

Successful deployment
93 % of reduction
Decreased training time
20 mins per epoch

LIMITS

Feature Maps
DSC not GPU supported

	Dice Score		
Method	Complete	Core	Enhancing
Proposed	**0.81**	**0.77**	**0.53**
Zhao (Zhao et al., 2018)	0.82	0.72	0.62
Pereira (Pereira, Pinto, Alves, & Silva, 2016)	0.84	0.72	0.62
Dong (Dong et al., 2017)	0.86	0.86	0.65

Fig 14. Results of our compressed U-Net. Source : Niepceron et al. 2020.

RESULTS

2. PCNN for tumor segmentation

DETAILS (SEGMENTATION)	**Average time :** 17 sec

LIMITS	Single class Morphological processing

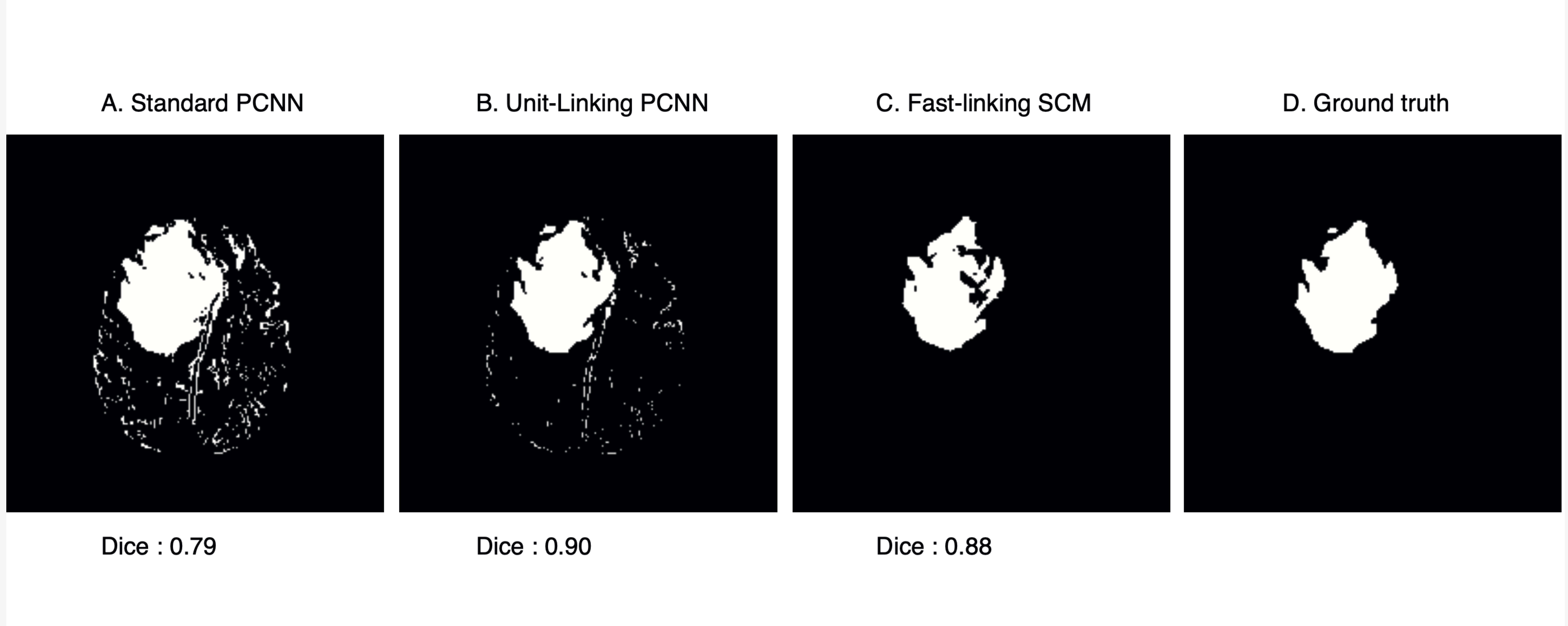


Fig 15. PCNN Segmentation results. Source : Niepceron et al. 2020.

4

RESULTS

2. PCNN for tumor detection

DETAILS (DETECTION)	Accurate bounding boxes Fast computation	LIMITS	Selective Search param. Small tumorous regions Whole tumor

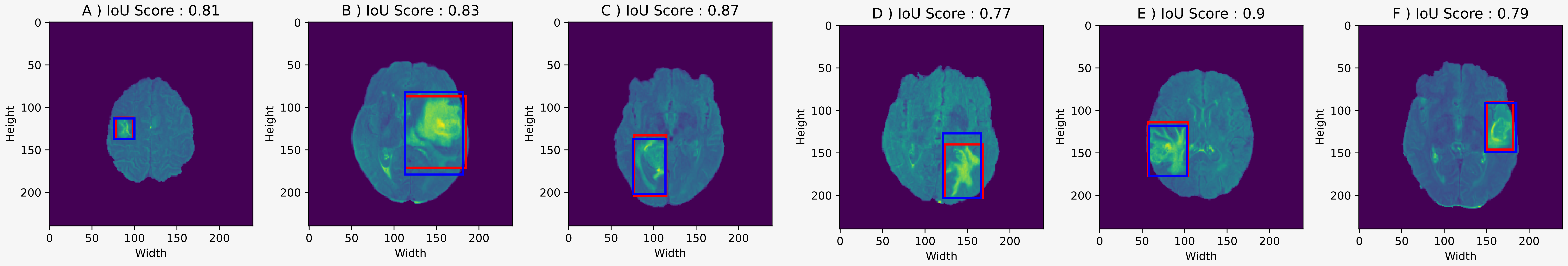


Fig 16. PCNN Detection results. Source : Niepceron et al. 2020.

CONCLUSION

IV

CONCLUSION

Model Compression

93% of compression obtained with minor penalties.
Leads to perspective for more profitable AI.

Parameter Pruning, Knowledge Distillation …

Spike-based computation

Explore new methods of light MRI segmentation.
Proven efficient for medical data.

Extend to multi-class

?

QUESTIONS