Adaptive artificial neural networks for neuroscience applications

PAULINE MOUCHÈS

pauline.mouches@inserm.fr

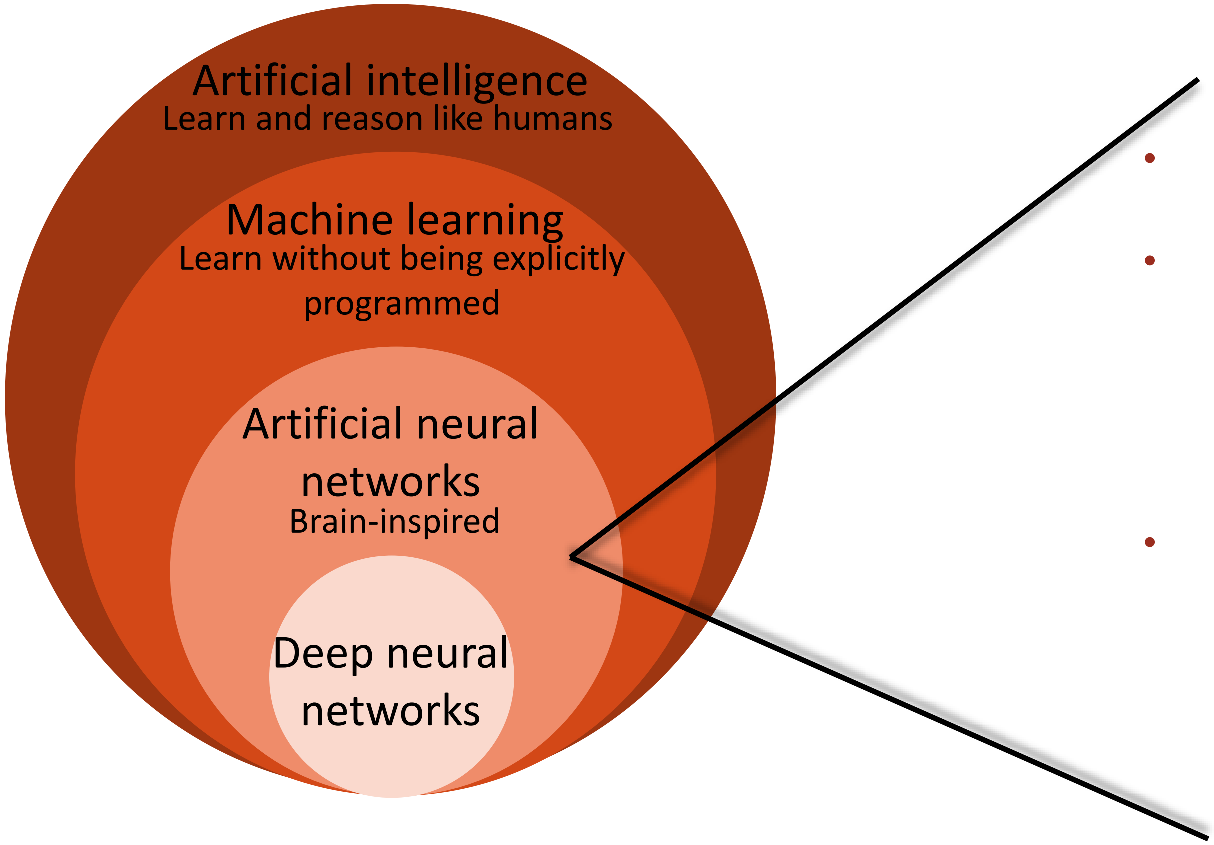
Summary

1. Introduction to artificial neural networks
   - Artificial neuron
   - Artificial neural networks
   - Deep convolutional neural networks
2. Comparing model's and neural responses
   - Compare internal representations
   - Study errors
3. Draw insights from artificial neural networks
   - Around the model
   - Into the model
4. Case study: Dementia in convolutional neural networks
5. Demo

Kietzmann, T. C., McClure, P., & Kriegeskorte, N. (2019). Deep neural networks in computational neuroscience. Oxford Research Encyclopedia of Neuroscience.

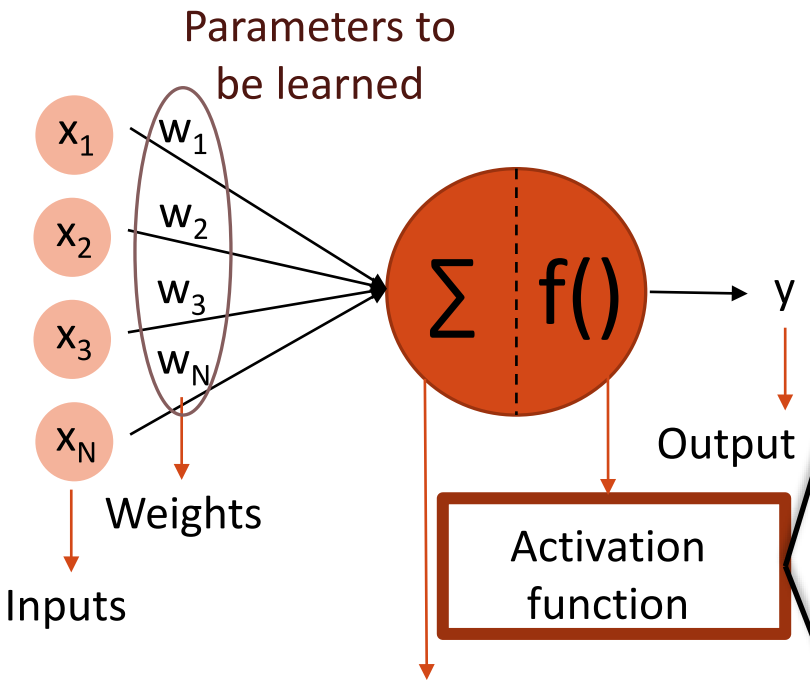
1- Introduction to artificial neural networks

Context



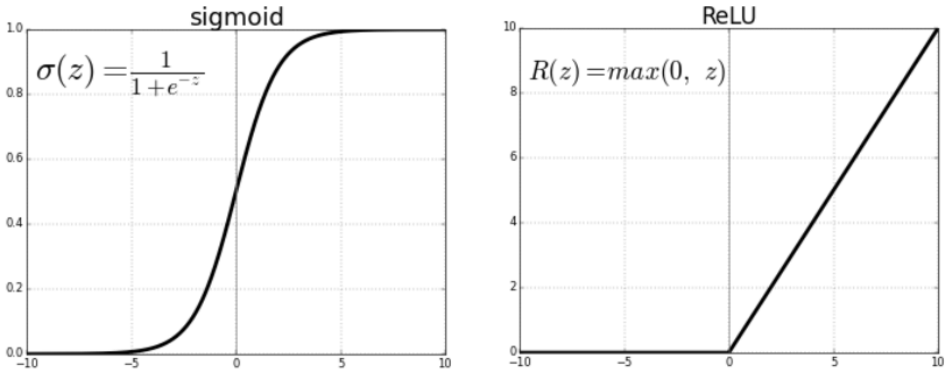
- Biologically inspired
- Trained to optimize a specific task BUT not derived from neural data
  - Underlying computations might employ different mechanisms
- Can solve various tasks
  - Visual object recognition
  - Speech recognition
  - Image generation

Artificial neuron



$$\left(\sum_{i=1}^{N} w_i \times x_i\right) + b$$

- Non-linear component of the model
- Describes when the neuron should fire
- Constraints the output to a certain range of values
- Examples of activation functions:

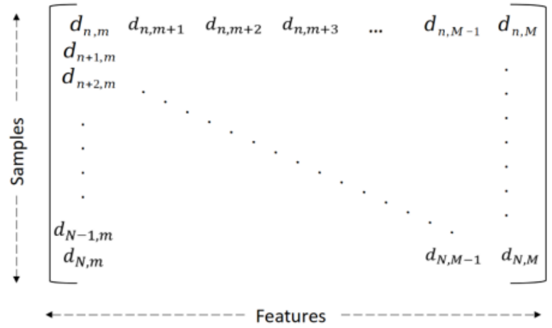


Activation Functions in Neural Networks | by SAGAR SHARMA | Towards Data Science

Artificial neural network

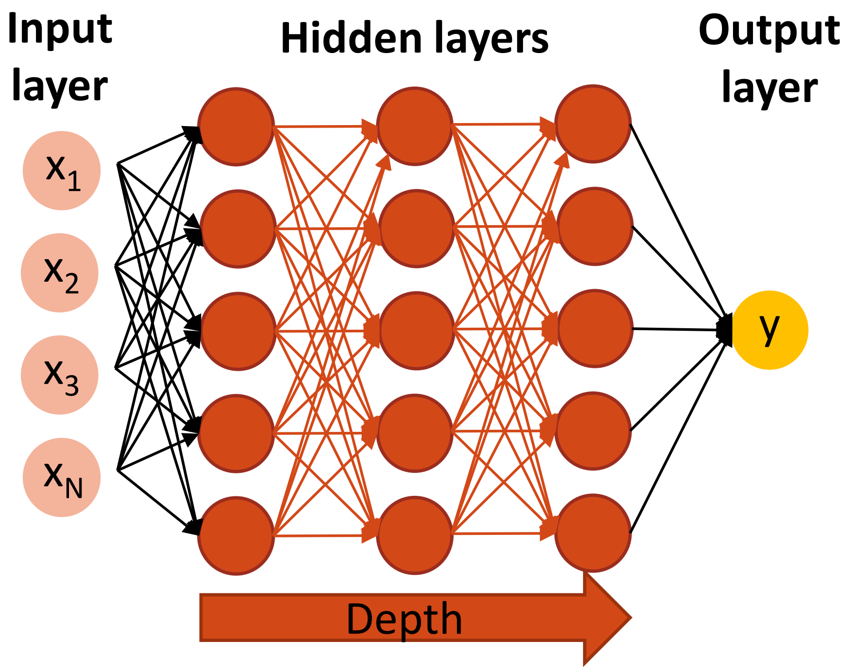
Working with tabular data:

Input:
- Clinical data
- Extracted measurements



Lo Vercio, L., et al. (2020). Supervised machine learning tools: a tutorial for clinicians. Journal of Neural Engineering, 17(6), 062001.

Spatial structure is ignored



Output:
Classification
- Categorical label

Regression
- Value

Supervised learning:
- Model trained on annotated data
- Loss function: minimize the error between ground truth and model prediction

Deep convolutional neural network

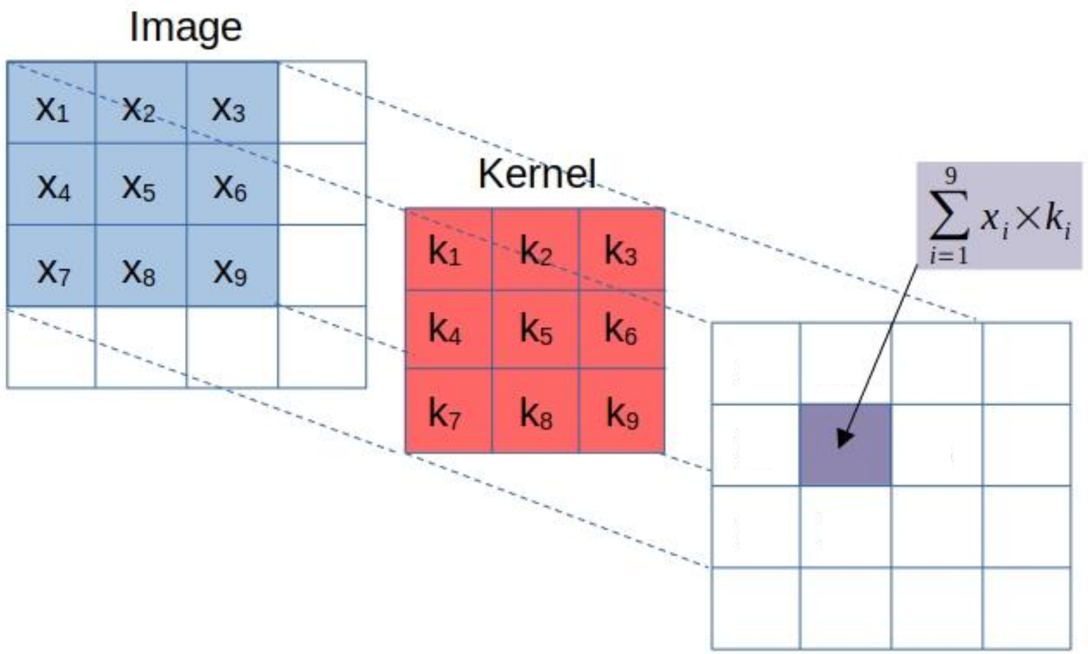
Working with spatially organized data:

- Deep convolutional neural networks (DCNNs) were originally inspired by the neuron and synaptic structure found in the mammalian visual cortex

- DCNNs have shown strong correlation to biological visual systems in terms of neural responses

Artificial convolutional neural network

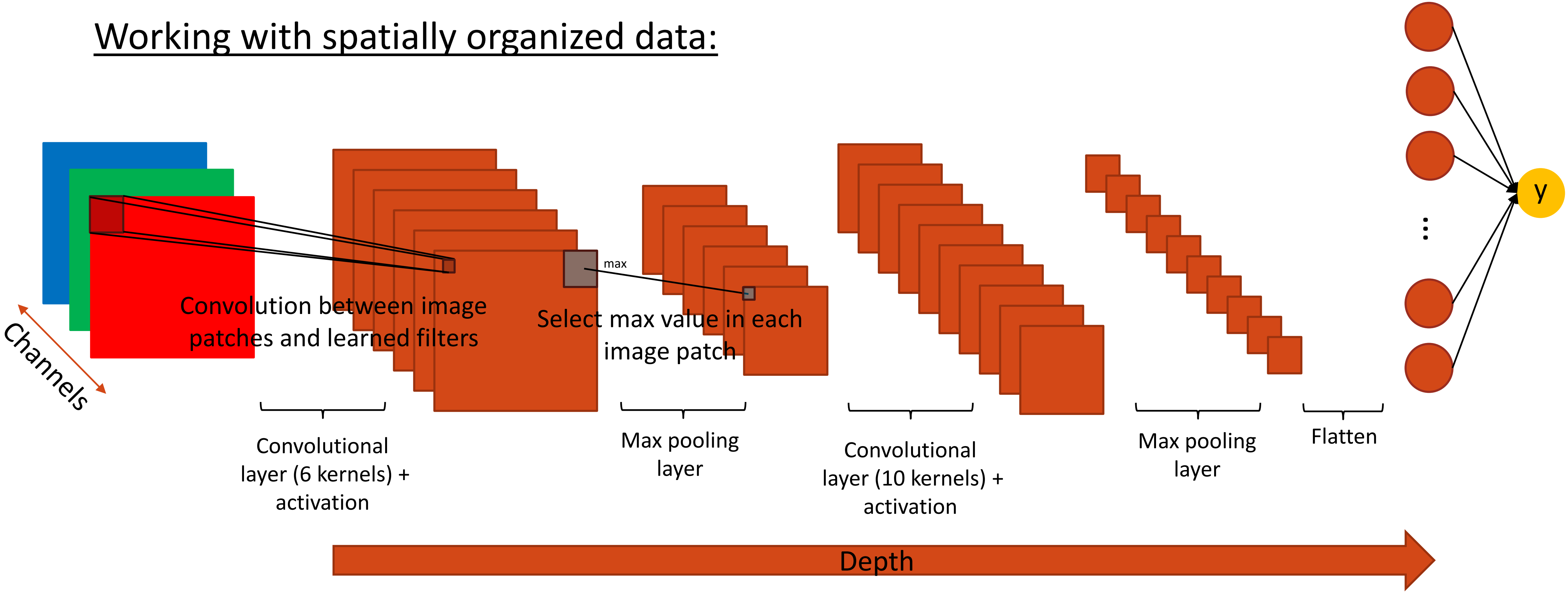
Working with spatially organized data:



Example of convolution

Artificial convolutional neural network

<u>Working with spatially organized data:</u>



Introduction to artificial neural networks

Take home message:

- Artificial neural networks can do prediction on tabular data
- Convolutional neural networks can do prediction on spatially organized data (1D, 2D, 3D)
- Model types can be combined to work on multimodal data

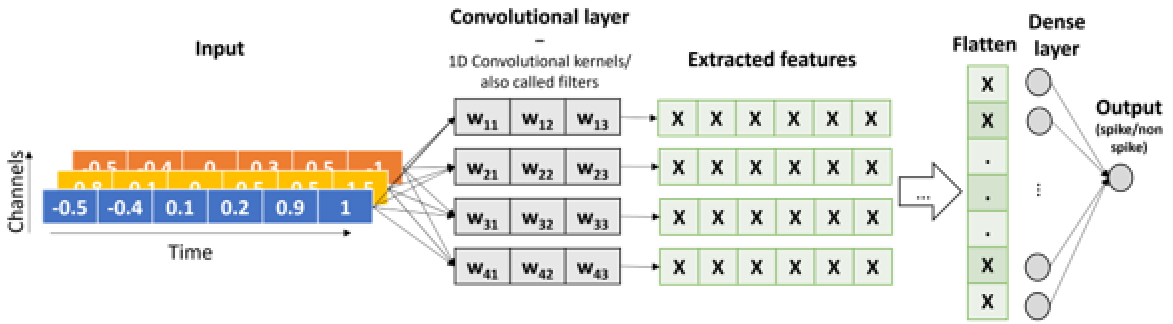
Questions?

Think about an application & the architecture of a model using the following input data:

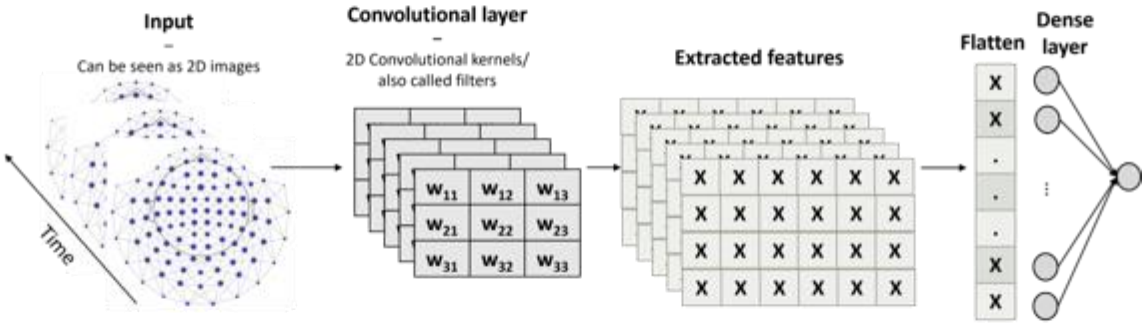
MEG/EEG signal
Image data and clinical measurements

EEG signal

1D convolutions over time

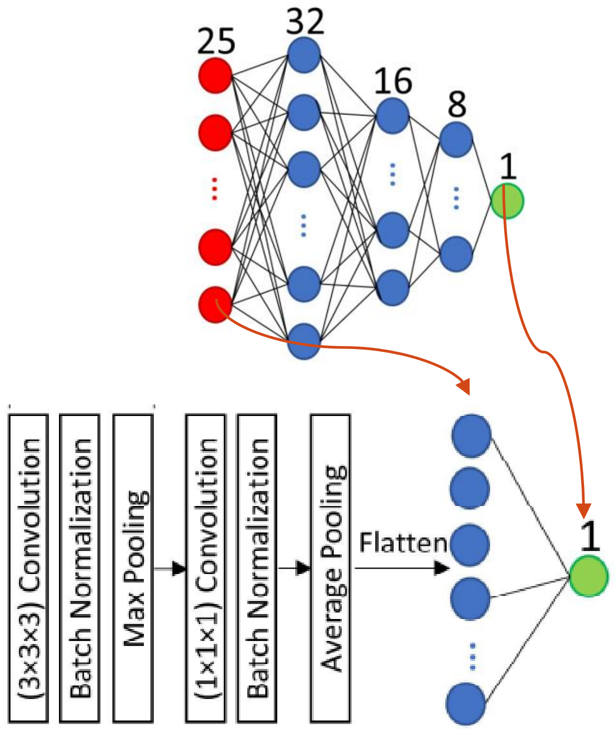


2D convolutions over MEG sensors



Brain CT scan and clinical measurements

Either 1 2D CNN where clinical data are added in the last layer or one ANN and one CNN for which we aggregate estimations



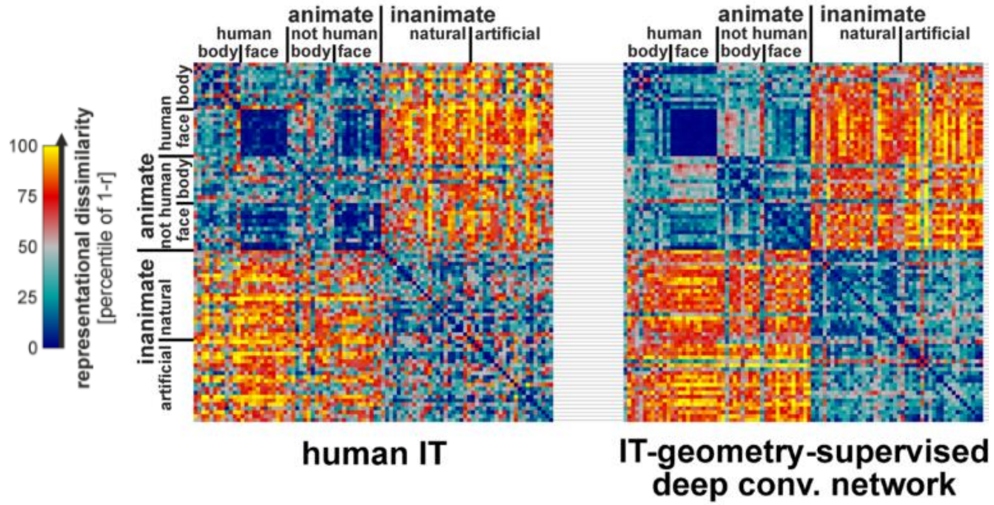
2- Comparing model's and neural responses

BUILDING BIOLOGICALLY PLAUSIBLE MODELS

Compare internal representations

Representational similarity analysis:

- Used to compare internal representations between deep neural networks (DNNs) and the brain
- Visualized using representation dissimilarity matrices (RDM)
  - Compare responses to different pairs of stimuli
- RDMs can then be compared
- RDMs can be computed
  - At each model layer for DNNs
  - From fMRI, MEG/EEG data, behavioral measures or perceptual judgement

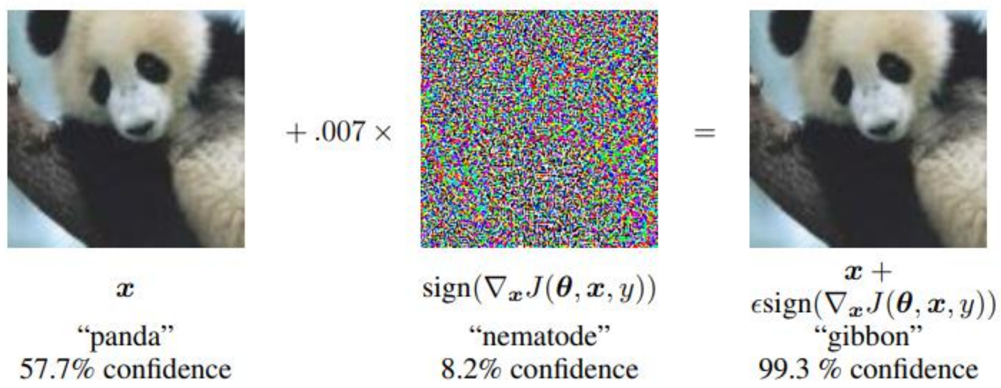


Adapted from Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS computational biology, 10(11), e1003915.

Study errors

Study the response sensitivity to change in the stimuli:

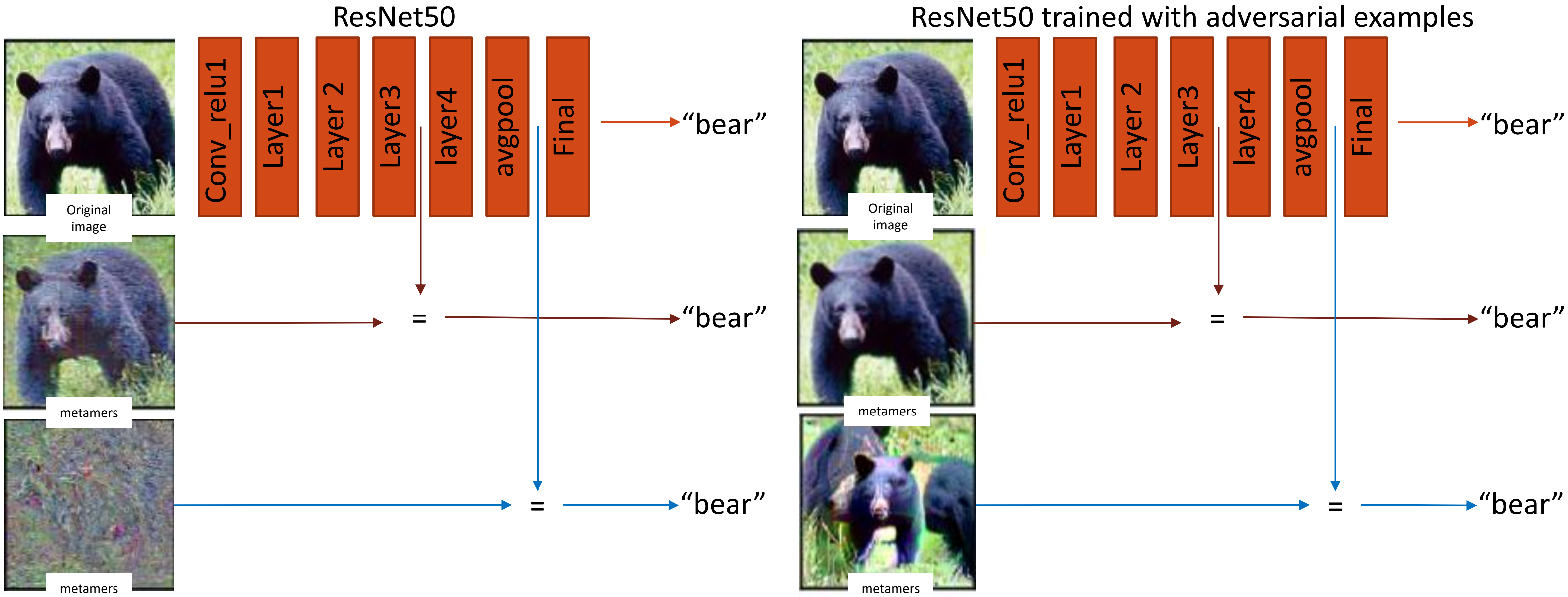
- Adversarial examples: A stimulus with minimal change leading to a different label prediction



Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

- Metamers examples: A physically different stimulus that is perceived to be the same

Study errors



Feather, J., Leclerc, G., Mądry, A., & McDermott, J. H. (2022). Model metamers illuminate divergences between biological and artificial neural networks. bioRxiv.

Comparing model’s and neural responses

Take home message:

- Although DCNN models seem to behave like human (in terms of prediction), their internal representation of a stimulus might be different from our brain internal representation
- Studying internal representations and errors are necessary to build biologically plausible models

3- Drawing insights from DNNs

UNDERSTANDING HOW THE MODEL MAKES ITS PREDICTIONS

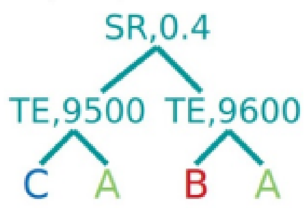
Around the model

DNNs are black box models BUT we can draw insights from:

- The input data
  - Study the model activation to different stimuli
- The model architecture
  - Compare model architectures
  - Alter model architectures
- Loss function
  - Try different loss functions
  - Classic loss function (classification error) requires thousands of labeled images to train → not biologically plausible

Into the model

- DNNs are black box models
  - We can't easily understand the internal computations (contrary to tree-based models for instance)

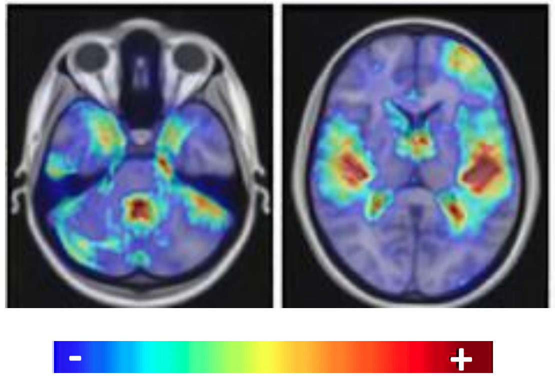


Lo Vercio, L., et al. (2020). Supervised machine learning tools: a tutorial for clinicians. Journal of Neural Engineering, 17(6), 062001.

- BUT we can look at
  - Input data importance (saliency maps):
  - Internal model representations:

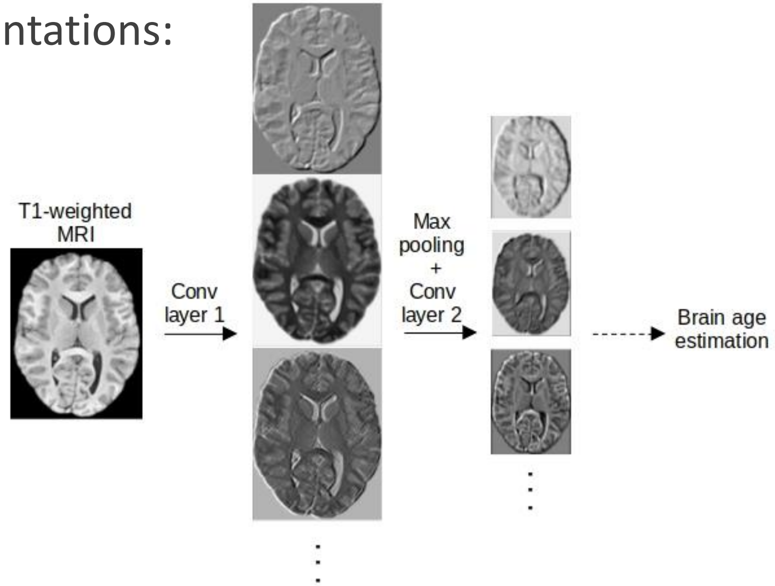
T1-weighted MRI

How would changing this voxel affect the prediction?



Importance

Mouches, P., et al. (2022). Multimodal biological brain age prediction using magnetic resonance imaging and angiography with the identification of predictive regions. Human brain mapping, 43(8), 2554-2566.



Drawing insights from DNNs

Take home message:

Even though many models are not directly comparable to the brain, they can be used (in medical applications) to:

- Understand better some mechanisms
- Discover new biomarkers (features that the model looks at but are not commonly used by humans)
- As decision support tools

Questions?

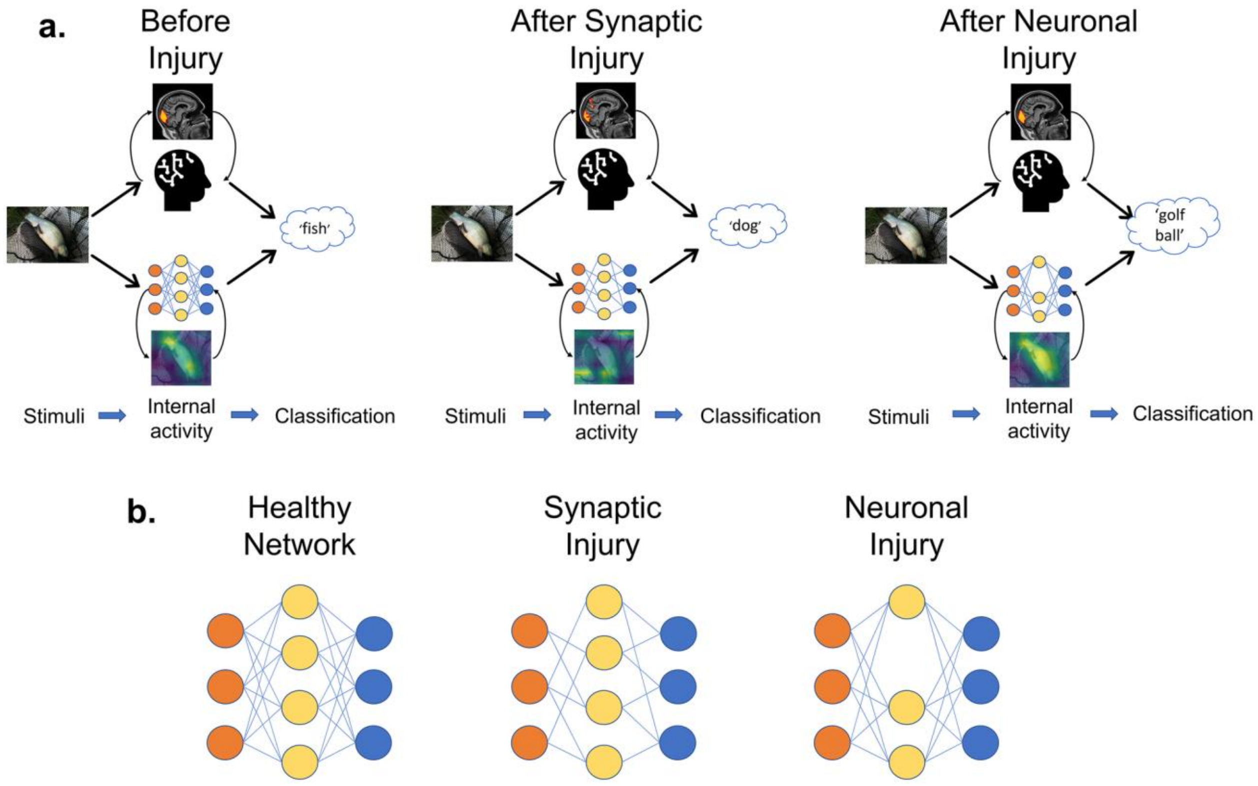
Case study: Dementia in Convolutional Neural Networks

Dementia in Convolutional Neural Networks: Using Deep Learning Models to Simulate Neurodegeneration of the Visual System

Moore, J.A., Tuladhar, A., Ismail, Z., Mouches, P., Wilms, M., Forkert, N.D.; *Neuroinformatics* (2022)

Aims:

- Use artificial neural networks to simulate Posterior cortical atrophy – PCA
  - Can accompany Alzheimer's
  - Affects the visual cortex of patients
- Build and train a biologically relevant DCNN, i.e., with strong correlation to biological neural activation data
- Create pipelines for different types of injury



Dementia in Convolutional Neural Networks: Using Deep Learning Models to Simulate Neurodegeneration of the Visual System

Material and Methods:

- VGG19 model architecture
  - Has one of the highest correlation values when compared to mammalian neuronal activation data
  - 16 convolutional layers + 3 dense layers
- Trained on imagenette database (subset of ImageNet)
  - 10 classes
  - Animate and inanimate objects

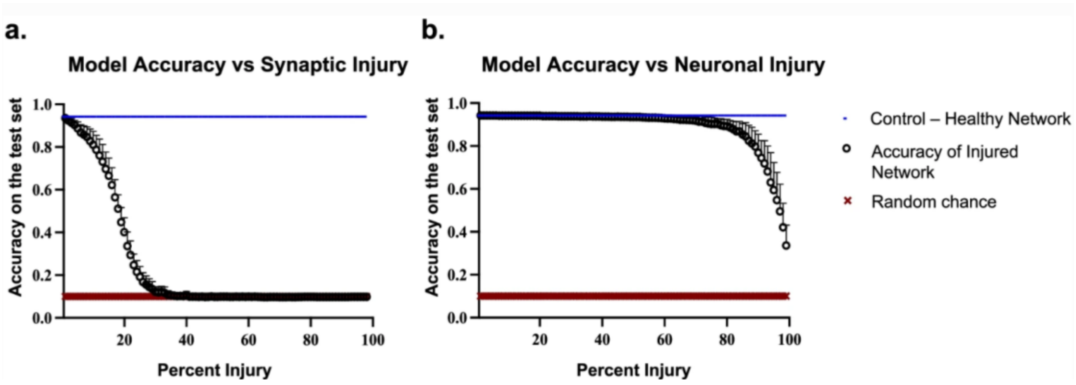
imagenette | TensorFlow Datasets

- Synaptic injury: set a weight to zero
- Neuronal injury: removing a node
  - Equivalent to filter in a DCNN
  - Implies the deletion of adjacent weights
- Comparison of the different levels of injury
  - Model accuracy
  - Saliency maps
  - Representational Dissimilarity Matrices

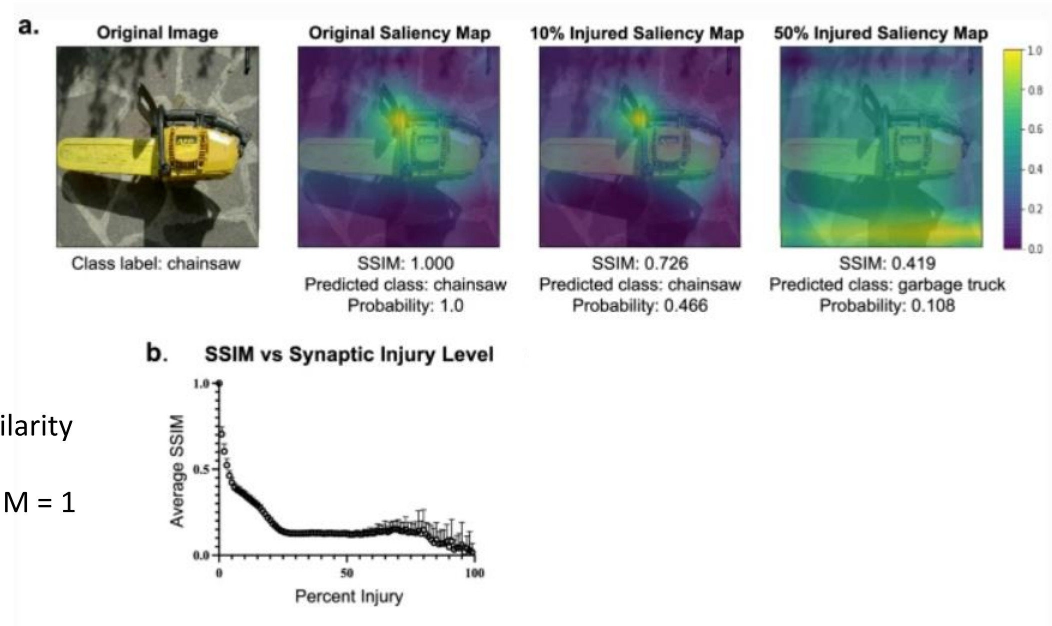
Dementia in Convolutional Neural Networks: Using Deep Learning Models to Simulate Neurodegeneration of the Visual System

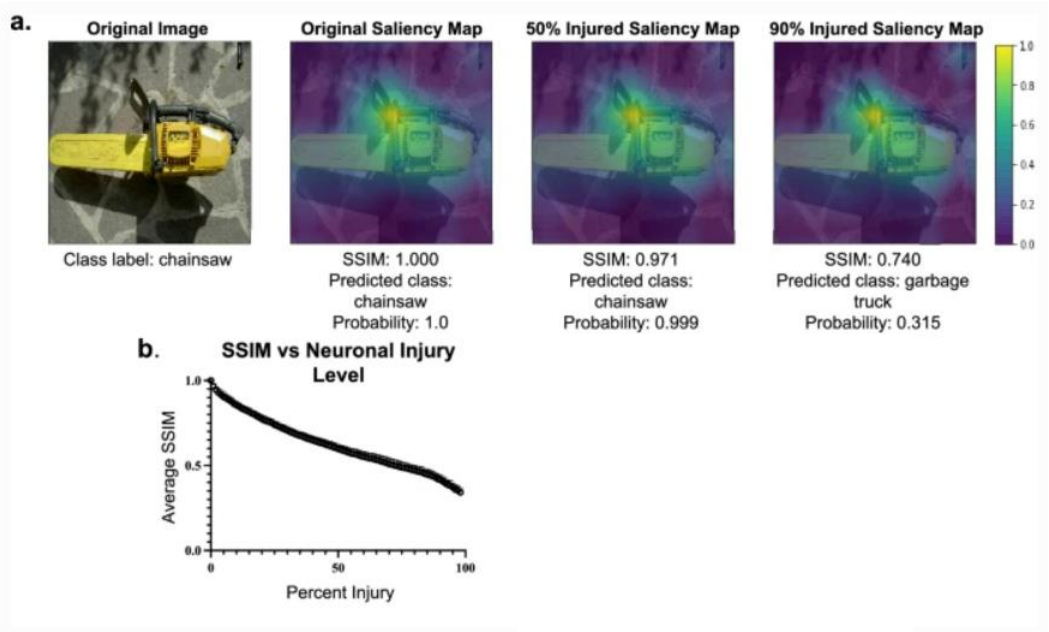
<u>Results:</u>

- Model accuracy



- Saliency maps



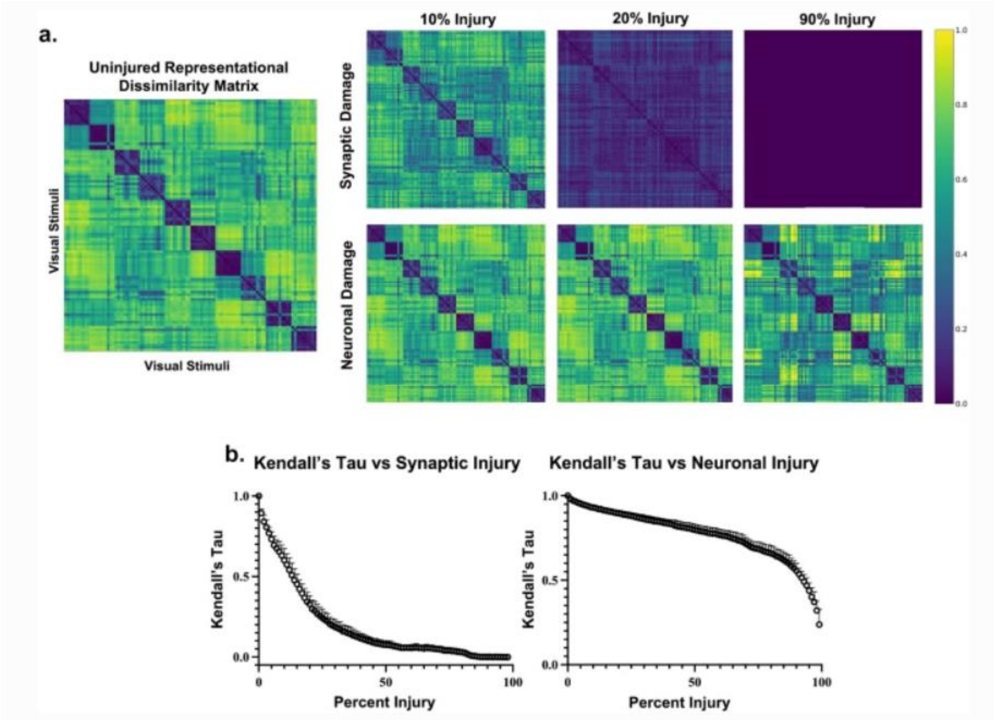


SSIM: Structural similarity index measure
Identical images: SSIM = 1

Dementia in Convolutional Neural Networks: Using Deep Learning Models to Simulate Neurodegeneration of the Visual System

Results:

- Representational dissimilarity matrices



A (left) The RDM generated by the healthy model showing pairwise comparisons of 10 images in each class in the test set of images. (right) RDMs generated by the network after imposing 10%, 20%, and 90% synaptic damage. **B** Kendall's tau as a function of synaptic injury. Synaptic injury is much more detrimental to the internal representations of the network. **C** Kendall's tau as a function of neuronal injury

Dementia in Convolutional Neural Networks: Using Deep Learning Models to Simulate Neurodegeneration of the Visual System

Discussion:

- Synaptic injury: When **removing weights in a randomly dispersed manner** in these static, feed-forward networks, the filters in the subsequent layers receive a **widespread lack of meaningful information**

- Neuronal injury: Removing a filter from a convolutional layer or a complete unit (neuron) from a dense layer in a deep learning network does not leave the subsequent layers with as much of a lack of information. Large convolutional neural networks, such as the **VGG19**, have proven to be quite **robust in model compression** studies, implying a certain level of **redundancy in the network**

- Clinical conclusions: **increasing levels of neurodegeneration** lead to gradual **loss of visual cognitive abilities**, similar to what can be observed clinically in patients with PCA. More precisely, at high levels of progressive neuronal injury and already at small levels of synaptic injuries, object recognition accuracy, correct attention focus, and categorized internal representation of objects all experience cognitive decline akin to patients suffering from PCA.

Questions?

Artificial neural networks in practice

Tools

Good practices in machine learning in general:

Lo Vercio, L., et al. (2020). Supervised machine learning tools: a tutorial for clinicians. Journal of Neural Engineering, 17(6), 062001

Python packages:

- scikit learn (traditional machine learning and basic neural networks)
- Tensorflow/Keras
- Pytorch

Demo:

Adapted from: Basic classification: Classify images of clothing | TensorFlow Core

Other resources:

1.17. Neural network models (supervised) — scikit-learn 1.1.3 documentation

Tutorials | TensorFlow Core