



CZECH TECHNICAL UNIVERSITY IN PRAGUE  
Faculty of Nuclear Sciences and Physical Engineering



**English Title**

**Czech Title**

Diploma thesis

Author: **Miroslav Kovář**

Supervisor: **M.Sc. M.A. Sebastián Basterrech, Ph.D.**

Academic year: 2018/2019



- Zadání práce -

- Zadání práce (zadní strana) -

*Acknowledgment:*

Some acknowledgment here.

*Author's declaration:*

I declare that this research project is entirely my own work and I have listed all the used sources in the bibliography.

Prague, August 10, 2018

Miroslav Kovář



*Název práce:*

**Czech Title**

*Autor:* Miroslav Kovář

*Obor:* Aplikace přírodních věd

*Zaměření:* Matematická informatika

*Druh práce:* Diplomová práce

*Vedoucí práce:* M.Sc. M.A. Sebastián Basterrech, Ph.D., Artificial Intelligence Center, FEE, CTU Prague

*Abstrakt:*

*Klíčová slova:*

*Title:*

**English Title**

*Author:* Miroslav Kovář

*Abstract:*

*Key words:*





# Contents

<b>1</b>	<b>Classical EEG signal analysis methods</b>	<b>13</b>
1.1	EEG signal . . . . .	13
1.2	Non-linear time series analysis . . . . .	15
1.2.1	Attractor . . . . .	15
1.2.2	Embedding . . . . .	18
1.2.3	Measures of non-linear interdependency . . . . .	19
1.2.4	Recurrence plot . . . . .	19
1.3	Applications . . . . .	20
<b>2</b>	<b>CNNs and CapsNets</b>	<b>21</b>
2.1	CNNs . . . . .	21
2.1.1	History . . . . .	21
2.1.2	Description . . . . .	22
2.1.3	Applications . . . . .	23
2.2	CapsNets . . . . .	23
<b>3</b>	<b>Experiments</b>	<b>25</b>
3.1	Dataset . . . . .	25
3.2	Results . . . . .	25



# Introduction



# Chapter 1

## Classical EEG signal analysis methods

### 1.1 EEG signal

Nature, processes in the brain, way of measuring, limitations, complications

Electroencephalography (EEG) is a noninvasive method of measuring fluctuations of electric potentials near the skull caused by synchronized firing of neurons in the upper cortical layers. Electroencephalogram is a record of these fluctuations measured over a period of time. [17]

Although EEG has significantly lower spatial resolution in comparison with other diagnostic techniques such as functional magnetic resonance sampling (fMRI) and magnetoencephalography (MEG) [22] and enables measuring only neural activity near the cortical surface, as a depression diagnostic tool, it has numerous benefits. Importantly, its significantly lower costs [26] [7], high portability, and ease of operation imply increased availability to the patients [20]. Moreover, it is perfectly noninvasive, which means less complications such as claustrophobia or anxiety [15].

Although the science of EEG signal analysis as a diagnostic tool brings compelling clinical promise as a result of the aforementioned benefits, it also presents multiple technical and conceptual challenges.

**Definition 1** ([19]). A series  $\{X_t\}_{t \in \mathbb{Z}}$  is called **stationary**, if  $\{X_t\}_{t \in \mathbb{Z}}$  for any set of times  $t_1, t_2, \dots, t_n$  and any  $k \in \mathbb{N}$ ,  $P[X_{t_1}, X_{t_2}, \dots, X_{t_n}] = P[X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}]$ , i.e. the joint probability distribution of  $\{X_t\}_{t \in \mathbb{Z}}$  is not a function of time. It is called **non-stationary**, if it is not stationary.

**Definition 2** ([1]). A series  $\{X_t\}_{t \in \mathbb{Z}}$  is called (noisy chaotic) **non-linear**, if it satisfies the relation

$$X_t = f(X_{t-1}) + \epsilon_t \quad (1.1)$$

for a general  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

EEG signals are prone to be infected with *noise* due to imperfect isolation from surrounding environment. They are known to be *transient*, *non-Gaussian*, *non-stationary* and *nonlinear* [10] [23]. Since some patterns do not activate relative to a stimulus, a successful classifier must be able to detect a pattern *regardless of its starting time*, or find one. And finally, EEG records are relatively high dimensional - 16 electrodes sampling at 256 Hz result 4096 data points par second.

Moreover, due to the phenomenon of neural oscillations, patterns may appear in multiple frequency bands, from slow cortical potentials of  $\delta$ -waves at 0.5-4 Hz, to high  $\gamma$  frequency band at 70-150 Hz.

Patterns of oscillatory activity in various frequency band have been linked to various mental states [3] [2] and diseases such as epilepsy [21], tremor [14], Parkinson's disease and depression [13]. Many of the diseases, including depression, share common oscillatory patterns known as thalamocortical dysrhythmia, characterized by decrease in normal resting-state  $\alpha$  (8-12 Hz) activity slowing down to  $\theta$  (4-8 Hz) frequencies, accompanied by increase in  $\beta$  and  $\gamma$  (25-50 Hz) activity. [25]



Figure 1.1: A comparison of stationary and non-stationary time series. (Courtesy: Protonk)

## 1.2 Non-linear time series analysis

Modern neuroscientific work suggests that the most plausible research target for explaining the brain dynamics are the assemblies of coupled and synchronously active neurons, and since majority of those assemblies are describable by non-linear differential equations, it is common to apply principles derived from nonlinear dynamics to characterize these neuronal systems.[10] In this section, we will describe some of those principles.

Attractors, Poincare plots, recurrence plots, Lyapunov coefficients, fractal dimension, Hurst exponent, etc.

**Definition 3** ([9]). *Let  $\mathbf{x} \in \mathbb{R}^m$  be an  $m \in \mathbb{N}$  dimensional state space vector dependent on time. A deterministic **dynamical system** is described by a set of  $m$  first-order differential equations*

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)), \quad t \in \mathbb{R} \quad (1.2)$$

In other words, a dynamical system is a model that determines the evolution of a system only given by the initial state, and current state is a function of the previous state. Hence, a total description of a dynamical system is given by the initial state and a set of equations. A non-linear dynamical system is a system where the differential equations describing its dynamics are non-linear. Unlike in a linear system, changes in the initial state of a non-dynamical system are allowed to have a non-linear relationship to the state space trajectory of the system. [10]

### 1.2.1 Attractor

**Definition 4** ([9]). *A dynamical system is called dissipative, when it is the case that*

$$E[\text{div}\mathbf{f}] < 1 \quad (1.3)$$

In other words, average space space volume (containing the initial state) is contracted as the system evolves.

Dissipative dynamical system have the property that a set of initial states (of positive measure) will converge to a subspace of the overall state space. This subspace is a geometrical object called an **attractor**. Example of four basic attractors can be seen on Figure 1.2.

Since most physiogenerated signals are chaotic, their analysis is concerned primarily with *chaotic* (strange) *attractors*. These attractors are complex, and exhibit what is known as *fractal geometry* (an example of a self-similar attractor is shown on Figure 1.3). . For our purposes, this means they can be characterized as having quantifiable self-similarity.<sup>1</sup> However, the following definition will be useful:

**Definition 5** ([5]). *Let  $F$  be any non-empty bounded subset of  $\mathbb{R}^n$ , and let  $N_\epsilon(F)$  be the smallest number of sets of diameter at most  $\epsilon$  which can cover  $F$ . Then, the **box-counting dimension** (also known as Minkowski–Bouligand dimension) is defined as*

$$d_F = \lim_{\epsilon \rightarrow 0} \frac{\log N_\epsilon(F)}{\log \frac{1}{\epsilon}}, \quad (1.4)$$

*if it exists.*

Intuitively, the number of mesh cubes of side  $\epsilon$  intersecting  $F$  gives an indication about how irregular the set is when inspected at scale  $\epsilon$ , and the box-counting dimension reflects “how rapidly” the irregularities develop as  $\epsilon \rightarrow 0$ . [5]

<sup>1</sup>Cantor set being a canonical example of self-similarity.

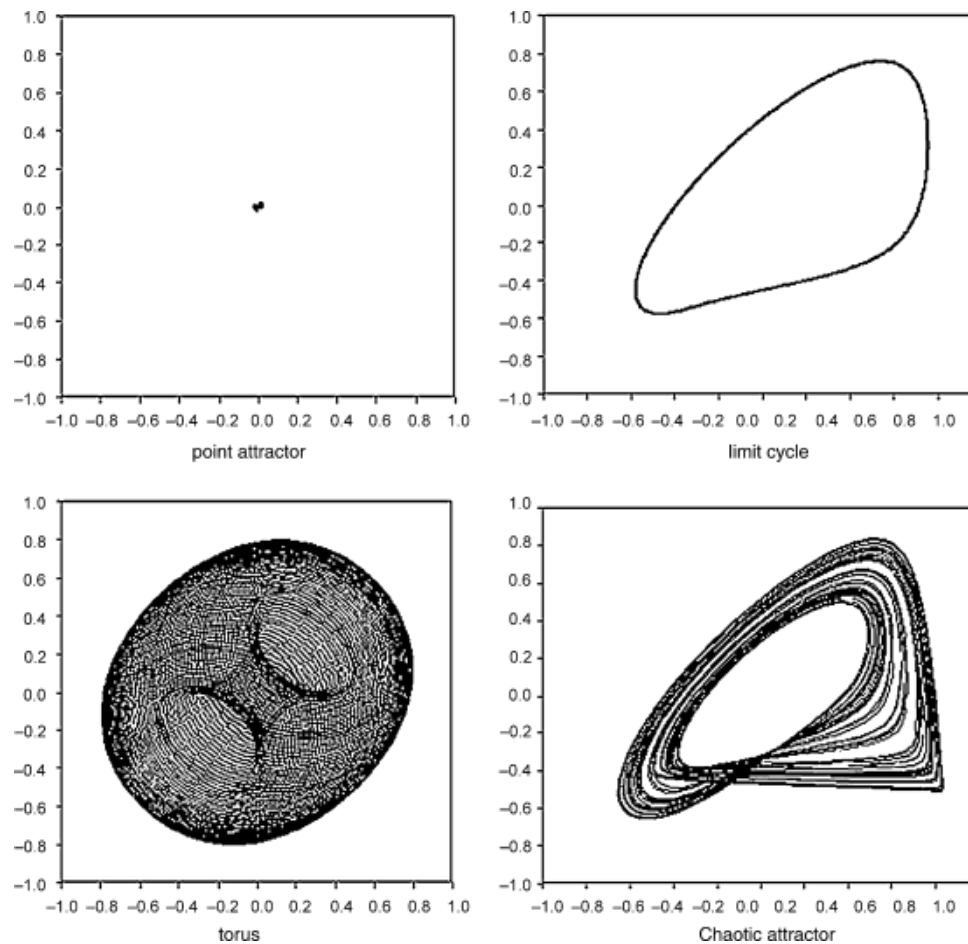


Figure 1.2: Visualization of four common attractor types (units are arbitrary). Left to right, top to bottom: *Point attractor* is the only type of attractor of linear deterministic dissipative systems. *Limit cycle* corresponds to a periodic system. *Torus attractor* corresponds to a quasi-periodic system, resulting (in this example) from a superposition of two periodic oscillations. *Chaotic (strange) attractor*, characterized by strong dependence on the original conditions. ([23])



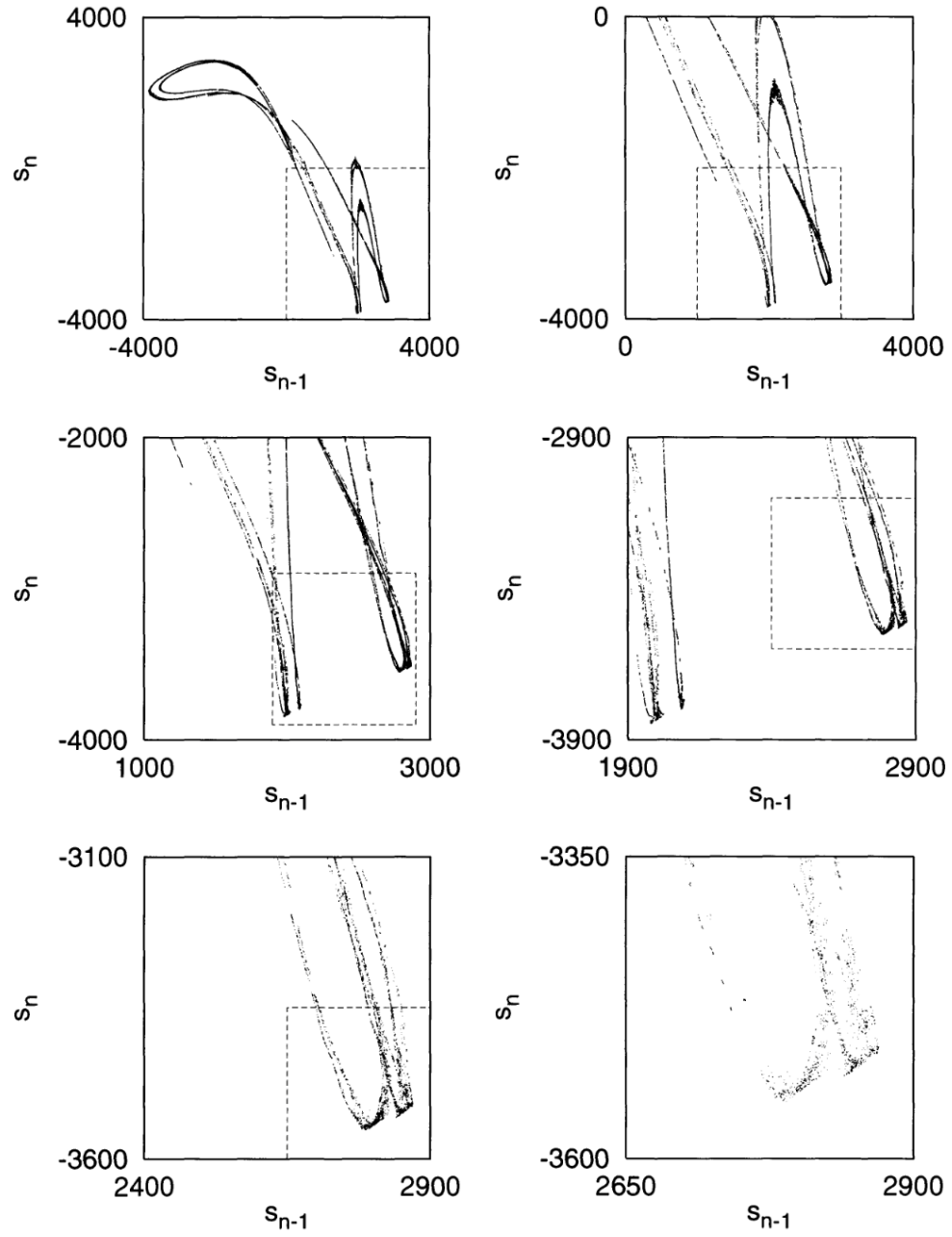


Figure 1.3: Noise-reduced visualization of successive enlargements of highly self-similar attractor. ([9])

### 1.2.2 Embedding

One possible approach to non-linear time series analysis consists of the following steps:

1. reconstruction of the dynamics of given system from recorded data,
2. characterization of the reconstructed attractor,
3. checking validity of the results with surrogate data testing. [23]

In the following text, we will describe multiple techniques of characterizing the properties of attractors (and hence ways of describing the dynamics of the dynamical system) using these steps.

In order to reconstruct the dynamics of given system from recorded data, it is necessary to deal with the fact that our data, however rich, rarely represent complete information about the studied system. In the case of EEG signals, the complete state of the system at any moment is determined by many variables, and the sensors are only able to collect traces of their cumulative effects (and noise). The process known as *state (phase) space reconstruction via the method of time delays* is a way of reconstructing the attractor of given dynamical system from single (or multiple) sequences of scalar measurements.

The following result from 1936 was important for theoretical understanding of time delay embeddings:

**Theorem 1** (Whitney embedding theorem, [27]). *Every  $n$ -dimensional differentiable manifold can be embedded in  $\mathbb{R}^{2n}$ .*

N. H. Packard has proposed in [18] a process by which a time series of measurements sampled at regular intervals can be embedded into a  $n$ -dimensional space in such a way that the series of reconstructed vectors is representative of the dynamics of the system. A year later, in [24], F. Takens has proved theoretically, using Whitney's result, that the reconstructed attractor has the same dynamical properties (entropy, dimension, Lyapunov spectrum) as attractor of the original system.

Takens delay embedding theorem is an important result of non-linear time series analysis and can be stated as follows:

**Theorem 2** ([24]). *Let  $M$  be a compact<sup>2</sup> manifold specifying the state space of a deterministic dynamical system of dimension  $n \in \mathbb{N}$ ,  $s : M \rightarrow \mathbb{R}$ ,  $s \in C^2$  a smooth measurement function,  $f : M \rightarrow M$ ,  $f \in C^2$  a smooth diffeomorphic state evolution function with a strange attractor of box-counting dimension  $d_A$ , and let  $k > 2d_A$ . Then a map  $\phi : M \rightarrow \mathbb{R}^k$ , defined by*

$$\phi(x) = (s(\mathbf{x}), s(f(\mathbf{x})), \dots, s(f^{k-1}(\mathbf{x}))) \quad (1.5)$$

*is an embedding.*

In what follows, we will describe how these results can be applied in practice. Let us have a time series of scalar measurements of a quantity depending on the current state of the system:

$$s_n = s(\mathbf{x}(n\Delta t)) + \eta_n, \quad (1.6)$$

where  $\mathbf{x}$  is a state space vector,  $s(\cdot)$  is a measurement function and  $\eta_n$  is a measurement noise.

Then, the time delay reconstruction is then simply formed by the following vectors:

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \dots, s_{n-\tau}, s_n), \quad (1.7)$$

---

<sup>2</sup>This theorem can be proved for  $M$  non-compact provided less restrictions are imposed on  $s$ .

for  $n > (m - 1)\tau$ . [9]

Of course, this method requires proper choice of parameters  $m$  and  $\tau$ . From Theorem 2 follows that vectors  $\{s_n\}$  are diffeomorphic to the attractor  $A$  of the system  $M$  if  $m$  is sufficiently large, specifically  $m > 2d_A$  (note that  $d_A$  may be much smaller than  $\dim M$ ). On the other hand,  $m$  adds redundancy, makes successive algorithms less efficient. Moreover, the total time delay  $m\tau$  is limited by the recording time.

There are multiple methods of choosing the parameters. We will start with a description of an algorithm for determining the minimal sufficient embedding dimension  $m$ . It uses concept known as *false nearest neighbors*.

(TODO: False nearest neighbors? The problem is I don't know if I need to go deep into this topic because we may simply end up using meta-recurrence plots. We may try computing non linear analysis parameters and train a NN classifier on them and compare (which seems more reasonable than CapsNets). I like the idea.)

(TODO: Mutual information algorithm for  $\tau$ )

(TODO: Mention spatial embedding. Is it worth using for EEG analysis? Pro: it achieves considerable reduction. Con: we may lose valuable information.)

### 1.2.3 Measures of non-linear interdependency

Phase synchronization, synchronization measures - we can use them to quantify interactions between brain regions. May be useful.

### 1.2.4 Recurrence plot

When presented with a task of finding regularities in seemingly chaotic data, one possible approach is analysing at least approximate repetitions of simple patterns, which can be further used for reconstruction of more complicated rules. Recurrence plot is a method of visualizing obtained state-space trajectory segments in relation to each other to achieve this goal. Furthermore, it can be used to test necessary conditions for validity of dynamical parameters derivable from a non-linear time series such as the information dimension, entropy, Lyapunov exponents, dimension spectrum, etc. The information contained in recurrence plots is not easily obtainable by other known methods. [4]

**Definition 6** ([4]). *Let  $N$  be the length of given time series,  $s_i$  for  $i \in \{1, 2, \dots, N\}$  be a  $i$ -th delay vector of any integer embedding dimension,  $\|\cdot\|$  a norm,  $\Theta(\cdot)$  a Heaviside step function, and  $\epsilon \in \mathbb{R}_0^+$  a tolerance parameter. Then, **recurrence plot** is the matrix*

$$M_{ij} = \Theta(\epsilon - \|s_i - s_j\|). \quad (1.8)$$

In other words,  $M_{ij}$  is a symmetric<sup>3</sup> binary  $N \times N$  matrix, where  $M_{ij} = 1$  when  $i$ -th and  $j$ -th points of the reconstructed trajectory enter each other's  $\epsilon$  neighborhood.

The essential drawback of recurrence plot is their size - it is quadratic in the length of the time series. A simple way of reducing its dimension is to partition the time series into disjointed segments, and let  $M_{ij}$  represent the distance between those two segments. This is known as **meta-recurrence plot**. [9] (TODO: Find a justification for using them.)

(TODO: Cross-recurrence plots may be useful? Only between two series. Joint recurrence plots may be used to detect phase synchronization.)

---

<sup>3</sup>Although this is true for our definition, it may not be true for an alternative definition using a more general topology instead of a norm.

## **1.3 Applications**

Results in applying classical methods and deep learning to EEG signal analysis

The number of studies applying methods of non-linear time series analysis for clinical depression diagnosis is limited. Abnormalities in sleep EEG have been found, as well as increased predictability of waking EEG. [16]

## Chapter 2

# CNNs and CapsNets

### 2.1 CNNs

#### 2.1.1 History

The classical approach to image pattern recognition consists of the following stages:

**preprocessing:** suppressing unwanted distortions and noise, enhancement beneficial for further processing,

**object segmentation:** separating disparate objects from the background,

**feature extraction:** gathering relevant information about the properties of the objects, removing irrelevant variations,

**classification:** categorizing segmented objects based on obtained features into classes.

The preprocessing step may require additional assumptions about the data or further processing, which are potentially too restrictive or too broad. Getting around this limitation requires dealing with complications such as high dimensionality of the input (number of pixels) and desirability of invariance towards a number of allowable distortions and geometrical transformations.

Artificial neural networks in combination with gradient-based learning are one possible solution to the problem. By gradually optimizing a set of weights based on a training data set using a differentiable error function, they provide a framework for learning a suitable set of assumptions automatically from the data.

One of the oldest neural network architectures, fully connected multi-layer perceptron (FC-MLP), can be used for image pattern recognition. However, it has the following drawbacks:

**parameter explosion:** the number of parameters of such network is exponential in the number of layers, increasing the capacity of the network and therefore need for more data,

**no invariance:** no invariance even with respect to common geometrical transformation such as translation, rotation and scaling,

**ignoring input topology:** natural images exhibit strong local structure and high correlation between intensities of neighboring pixels, but FC-MLPs are unstructured - inputs can be presented in any order.

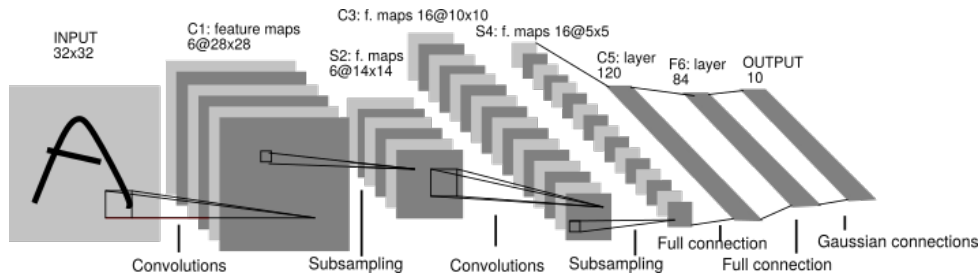


Figure 2.1: LeNet-5 architecture [12].

Although the main idea dates back 1980 with K. Fukushima's neocognitron [6], the back-propagation algorithm was not known at the time. The first convolutional architecture successfully applied on an image pattern recognition problem by attempting to solve the aforementioned problems, dubbed LeNet-5, was proposed in 1998 by Y. LeCun, L. Bottou, Y. Bengio and P. Haffner [11].

### 2.1.2 Description

Bearing resemblance to visual processing in biological organisms<sup>1</sup>, LeNet-5 proposed the following design principles to enforce *shift, scale and distortion invariance*: [12]

**local receptive fields:** each neuron in a layer receives input from a small neighborhood in the previous layer,

**shared weights:** each layer is composed of neurons organized in planes within which each neuron have the same weight vector (feature map),

**spatial subsampling:** adding a subsampling layers, which reduce the resolution of the previous layer by averaging or taking the maximal value of neighboring pixels in the previous layer.

*Local receptive fields* enable the network to synthesize filters that produce strong response to elementary salient features in the early layers (such as lines, edges and corners in a visual input, and equivalence in other modalities), and then learn to combine them in the subsequent layers to produce higher-order feature detectors.

*Shared weights* principle exploits the fact that translation-invariant features have the property that one feature detector can be used accross the entire image. Since neural units in a layer with differing receptive fields possess the same feature map, the same feature detecting operation (convolution with feature map kernel followed by additive bias and a application of a non-linear function) is performed on differing parts of the image. A single convolutional layer is composed of multiple feature detecting planes.

*Spatial subsampling* is purposed to ensure scale and distortion invariance<sup>2</sup> by reducing the precision at which a feature is encoded in a feature map by reducing its resolution - when scale and distortion invariance is assumed, the exact location of a feature becomes less important and is allowed to exhibit slight positional variance. Although this simple solution performs well in many practical situations, it

<sup>1</sup>As early as in 1968, D. H. Hubel and T.N. Wiesel discovered that some cells (called simple cells) in cat's primary visual cortex (V1) with small receptive fields (shared by neighboring neurons) are sensitive to straight lines and edges of light of particular orientation, and other cells (called complex cells) with larger receptive fields further in the visual cortex also respond to straight lines and edges, but with invariance to translation [8].

<sup>2</sup>Whether it achieves this goal has been famously doubted by Geoffrey Hinton: "The pooling operation used in convolutional neural networks is a big mistake and the fact that it works so well is a disaster." []

has multiple drawbacks, as mentioned in section 2.2 (relative representation of features is not encoded in the feature map representation)

### **2.1.3 Applications**

Maybe mention an example of how LeNet-5 was improved on subsequently (AlexNet, ResNet, etc.)

## **2.2 CapsNets**

Does it make sense trying them? I found a only a few successful implementations. Maybe it would be better to try those after we have some results already, because it seems risky - we might end up with nothing.





## **Chapter 3**

# **Experiments**

### **3.1 Dataset**

Size of our dataset, conditions during trials, labels, etc.

### **3.2 Results**



# Conclusion



# Bibliography

- [1] Peter J Bickel and Peter Bühlmann. What is a linear process? *Proceedings of the National Academy of Sciences*, 93(22):12128–12131, 1996.
- [2] György Buzsáki and Andreas Draguhn. Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929, 2004.
- [3] Ryan T Canolty, Erik Edwards, Sarang S Dalal, Maryam Soltani, Srikantan S Nagarajan, Heidi E Kirsch, Mitchel S Berger, Nicholas M Barbaro, and Robert T Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *science*, 313(5793):1626–1628, 2006.
- [4] J.-P Eckmann, S. Oliffson Kamphorst, and D Ruelle. Recurrence Plots of Dynamical Systems. *Europhysics Letters (EPL)*, 4(9):973–977, 1987.
- [5] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- [6] Kunihiro Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.
- [7] Matti Hämäläinen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413, 1993.
- [8] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 1968.
- [9] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.
- [10] Alexander Ya Kaplan, Andrew A Fingelkurts, Alexander A Fingelkurts, Sergei V Borisov, and Boris S Darkhovsky. Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges. *Signal processing*, 85(11):2190–2212, 2005.
- [11] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998.
- [12] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*, pages 319–345. Springer, 1999.

- [13] Rodolfo R Llinás, Urs Ribary, Daniel Jeanmonod, Eugene Kronberg, and Partha P Mitra. Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proceedings of the National Academy of Sciences*, 96(26):15222–15227, 1999.
- [14] JH McAuley and CD Marsden. Physiological and pathological tremors and rhythmic central motor control. *Brain*, 123(8):1545–1567, 2000.
- [15] Kieran J Murphy and James A Brunberg. Adult claustrophobia, anxiety and sedation in mri. *Magnetic resonance imaging*, 15(1):51–54, 1997.
- [16] Jean Louis Nandrino, Laurent Pezard, Jacques Martinerie, Farid El Massioui, Bernard Renault, Roland Jouvent, Jean François Allilaire, and Daniel Widlöcher. Decrease of complexity in EEG as a symptom of depression. *NeuroReport*, 5(4):528–530, 1994.
- [17] Paul L Nunez, Ramesh Srinivasan, et al. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.
- [18] Norman H Packard, James P Crutchfield, J Doyne Farmer, and Robert S Shaw. Geometry from a time series. *Physical review letters*, 45(9):712, 1980.
- [19] Maurice Bertram Priestley. Non-linear and non-stationary time series analysis. 1988.
- [20] Teal L Schultz. Technical tips: Mri compatible eeg electrodes: advantages, disadvantages, and financial feasibility in a clinical setting. *The Neurodiagnostic Journal*, 52(1):69–81, 2012.
- [21] Vladimir Shusterman and William C Troy. From baseline to epileptiform activity: a path to synchronized rhythmicity in large-scale neural networks. *Physical Review E*, 77(6):061911, 2008.
- [22] Ramesh Srinivasan. Methods to improve the spatial resolution of eeg. *International Journal of Bioelectromagnetism*, 1(1):102–111, 1999.
- [23] C. J. Stam. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clinical Neurophysiology*, 116(10):2266–2301, 2005.
- [24] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.
- [25] Sven Vanneste, Jae-Jin Song, and Dirk De Ridder. Thalamocortical dysrhythmia detected by machine learning. *Nature communications*, 9(1):1103, 2018.
- [26] Paul M Vespa, Val Nenov, and Marc R Nuwer. Continuous eeg monitoring in the intensive care unit: early findings and clinical efficacy. *Journal of Clinical Neurophysiology*, 16(1):1–13, 1999.
- [27] Hassler Whitney. Differentiable manifolds. *Annals of Mathematics*, pages 645–680, 1936.