

UNIVERSIDADE FEDERAL DO PARANÁ

GUSTAVO HENRIQUE GOMES MATSUSHITA

AUTOMATIC IDENTIFICATION OF PHASIC DOPAMINE RELEASE

CURITIBA PR

2019

GUSTAVO HENRIQUE GOMES MATSUSHITA

AUTOMATIC IDENTIFICATION OF PHASIC DOPAMINE RELEASE

Dissertation submitted in partial fulfillment of the requirements for the degree of M.Sc in Computer Science. Graduate Program in Computer Science (PPGInf), sector of Exact Sciences, at Federal University of Paraná (UFPR).

Field: *Computer Science*.

Advisor: Luiz Eduardo Soares de Oliveira, Dr.

Co-advisor: Yandre Maldonado e Gomes da Costa, Dr.

CURITIBA PR

2019

Ficha Catalográfica

Esta folha deve ser substituída pela ficha catalográfica fornecida pela Biblioteca Central da UFPR, a pedido da secretaria do PPGInf/UFPR (vide arquivo `ficha.tex`).

Termo de aprovação

Esta folha deve ser substituída pela ata de defesa ou termo de aprovação devidamente assinado, que será fornecido pela secretaria do programa após a defesa ter sido concluída e aprovada (vide arquivo aprovacao.tex).

Acknowledgements

I would like to thank my advisor, Prof. Dr. Luiz Eduardo Soares de Oliveira, and co-advisor, Prof. Dr. Yandre Maldonado e Gomes da Costa, for their suggestions, corrections and patience. In addition to all the knowledge transmitted and trust deposited for the accomplishment of this work.

Thanks to Prof. Dr. Cláudio da Cunha and Adam Hideo Sugi for all assistance provided to the development of datasets.

I would also like to thank the members of the Vision, Robotics and Image Laboratory (VRI) for all collaboration and support. And the Department of Informatics of Federal University of Parana (UFPR), for the available infrastructure.

I must express my very profound gratitude to my family and friends for love and encouragement.

Finally, thanks to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), which financed in part this study - Finance Code 001.

Abstract

The study and analysis of dopamine (DA) release in the organism are of great importance due to the fact that this neurotransmitter directly influences processes such as cognition and drug abuse. It is also related to the physiopathology of some neurological diseases. The fast scan cyclic voltammetry technique allows efficient recording of the phasic release of DA. However, due to the high temporal resolution of the technique, the experiments generate large amounts of data, resulting in a slow and repetitive manual analysis. The present work is intended to develop and evaluate the performance of an automatic identification system starting from phasic dopamine release images, using different visual descriptors and convolutional neural network models, combining different approaches and classifiers searching for complementarity to improve the system overall performance. Two phasic dopamine release image datasets were created, and the best developed classifier obtained an accuracy of 98.31% using a combined approach of convolutional neural networks.

Keywords: phasic dopamine release, pattern recognition, texture, convolutional neural networks.

Resumo

O estudo e análise da liberação de dopamina (DA) no organismo são de grande importância devido ao fato deste neurotransmissor influenciar diretamente processos como os de aprendizado e dependência de drogas, além de ter relação com o desenvolvimento de diversas patologias neurológicas. A técnica de voltametria cíclica de varredura rápida permite o registro eficiente da liberação de DA fásica, entretanto os experimentos tendem a ter uma alta resolução temporal, gerando grandes quantidades de dados, resultando em uma análise manual demorada e repetitiva. O presente trabalho tem por finalidade apresentar e avaliar o desempenho de um sistema de identificação automática de liberação fásica de dopamina, utilizando diferentes descritores de características visuais e modelos de redes neurais convolucionais, combinando diferentes abordagens e classificadores no intuito de aproveitar informações complementares existentes e gerar melhores resultados. Duas bases de dados contendo imagens de liberação fásica de dopamina foram geradas, sendo que o melhor classificador desenvolvido obteve uma acurácia de 98.31% utilizando uma abordagem combinada de redes neurais convolucionais.

Palavras-chave: liberação de dopamina fásica, reconhecimento de padrões, textura, redes neurais convolucionais.

Contents

1	Introduction	14
1.1	Motivation	15
1.2	Challenges	15
1.3	Objectives	16
2	Theoretical Background	17
2.1	Dopamine	17
2.2	Fast-Scan Cyclic Voltammetry	18
2.3	Pattern Recognition	18
2.4	Representation	18
2.4.1	Local Binary Pattern (LBP)	19
2.4.2	Local Phase Quantization (LPQ)	20
2.5	Representation Learning	21
2.5.1	Inception	23
2.5.2	You Only Look Once (YOLO)	24
2.6	Classification	26
2.6.1	Classifiers Combination	26
2.7	Evaluation Measures	27
2.8	Receiver Operating Characteristic Curve	28
3	State-of-the-Art	29
4	Methodology	32
4.1	Datasets	32
4.1.1	Dataset I	32
4.1.2	Dataset II	34
4.2	Image Patches Approach	35
4.2.1	Texture Descriptors	35
4.2.2	Classification	35
4.2.2.1	Early and Late Fusion	36
4.2.3	Convolutional Neural Networks	37
4.2.4	Proposed Metrics	38
4.3	Combined Approach	39
4.4	Object Detection	40
5	Results	42
5.1	Dataset I	42
5.1.1	Manually Extracted Patches	42
5.1.1.1	Texture Descriptors	42

5.1.1.2	Early Fusion	43
5.1.1.3	Late Fusion	43
5.1.2	Automatically Extracted Patches	44
5.2	Dataset II	45
5.2.1	Texture Descriptors	45
5.2.2	Convolutional Neural Networks	47
5.2.2.1	YOLO	50
6	Final Considerations and Future Works	52
	Bibliography	54

List of Figures

1.1	Example of phasic dopamine release image.	15
2.1	Pattern recognition system steps.	19
2.2	Central pixel C with P neighboring ones.	20
2.3	Example of a neighborhood with LBP _{8,2}	20
2.4	Example of a convolutional neural network layers.	22
2.5	Example of transfer learning.	23
2.6	Inception module. Extracted from Szegedy et al. (2015).	24
2.7	Overview of Inception v3 network model.	24
2.8	The YOLO Detection System. Extracted from Redmon et al. (2016).	24
2.9	The model detection. Extracted from Redmon et al. (2016).	25
2.10	Example of ROC space.	28
4.1	DA release interval.	33
4.2	Peaks and common dopamine release region.	33
4.3	Example of automatically extracted patches.	34
4.4	Images generated using different background positions.	35
4.5	An overview of the image patches approach.	36
4.6	Overview of Late Fusion.	36
4.7	Representation of Early Fusion.	37
4.8	Overview of Early Fusion.	37
4.9	Examples of Data Augmentation.	37
4.10	Overview of Roecker et al. (2018) model.	38
4.11	An overview of the combined approach.	40
4.12	Examples of the original image and two zoning variation.	41
4.13	Simplified YOLO overview.	41
5.1	ROC curves of the best Dataset I tests.	45
5.2	ROC curve of the best test using entire images approach.	48
5.3	Example of True Positives and False Positives.	50
5.4	ROC curve of the best test using YOLO.	51
5.5	Example of True Positives and False Positives using YOLO.	51

List of Tables

2.1	YOLOv3 model (Darknet-53)	26
4.1	Architecture of Roecker et al. (2018) model	38
5.1	Results obtained with LBP/RLBP and SVM using manually extracted patches .	42
5.2	Results obtained with LPQ and SVM using manually extracted patches	43
5.3	Results obtained with Early Fusion using manually extracted patches	43
5.4	Results obtained with Late Fusion using manually extracted patches	44
5.5	Results with texture descriptors and SVM using automatically extracted patches	44
5.6	Results obtained with Early and Late Fusion using automatically extracted patches	44
5.7	Results from Dataset II obtained with texture descriptors	46
5.8	Tests using different zoned images	46
5.9	Results obtained using a combined training set	47
5.10	Results obtained using Late Fusion of background position A and C predictions	47
5.11	Results using texture descriptors and automatically extracted patches from Dataset II	47
5.12	Results obtained using Early and Late Fusion and automatically extracted patches from Dataset II	47
5.13	Results obtained using Inception v3	48
5.14	Results obtained using Roecker et al. (2018) CNN model	49
5.15	Results obtained with Data Augmentation using Roecker et al. (2018) CNN model	49
5.16	Results obtained using automatically extracted patches and CNN	49
5.17	Results obtained using Tiny YOLOv3 model	51

List of Acronyms

A	Amps
AUC	Area Under the Curve
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNN	Convolutional Neural Networks
DA	Dopamine
DFT	Discrete Fourier Transform
DINF	Department of Informatics
DWT	Discrete Wavelet Transform
FN	False Negatives
FP	False Positives
FPR	False Positive Rate
FSCV	Fast-Scan Cyclic Voltammetry
GLCM	Gray-Level Co-occurrence Matrices
GLRLM	Gray-Level Run-Length Matrix
Hz	Hertz
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LBP	Local Binary Pattern
L-DOPA	L-3-4-dihydroxyphenylalanine
LPQ	Local Phase Quantization
LReLU	Leaky Rectifier Unit
MLP	Multilayer Perceptron
NN	Nearest Neighbor
PNN	Probabilistic Neural Network
PPGINF	Graduate Program in Computer Science
RBF	Radial Basis Function
RGB	Red Green Blue
RLBP	Robust Local Binary Pattern
ROC	Receiver Operating Characteristic
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TN	True Negatives
TP	True Positives
TPR	True Positive Rate
UFPR	Federal University of Parana
UNC	University of North Carolina
V	Volts
VMAT	Vesicular Dopamine Transporter
VRI	Vision, Robotics and Image Laboratory

YOLO You Only Look Once

Chapter 1

Introduction

The neurotransmitter dopamine (DA) has a key role in the brain and body, participating in the modulation of functions of movement, behavior, emotion, learning, and memory (Da Cunha et al., 2012). Abnormal variations in DA levels may be related to some pathologies, such as Parkinson's disease and schizophrenia (Da Cunha et al., 2015). Its release at the synaptic terminals can happen in two distinct ways: phasic, where there is a large amount of short duration release, and in a tonic form, where the dopamine is released slowly and in low quantity (Grace, 1995).

Effects of behavioral influence, such as decision making, are related to dopamine phasic release. This kind of release promotes a motivation increase to initiate actions that result in immediate rewards. Thus, it is also related to stimuli that are naturally rewarding, such as the abuse of drugs use (Grace, 1995; Wanat et al., 2009).

The use and dependence of drugs like cocaine, amphetamine, nicotine, and alcohol are serious public health problems. These drugs increase the concentration of extracellular dopamine in the circuitry known as "reward pathway" by increasing dopamine release and slowing dopamine reuptake (Wanat et al., 2009). In addition, Phillips et al. (2003) presented relationships between this phasic release and drug-seeking behavior, which further evidence the relationship of dopamine to drugs dependence.

An electrochemical method known as fast-scan cyclic voltammetry (FSCV) consists of applying electrical potentials to microelectrodes, thus inducing oxidation and/or reduction of molecules like dopamine. This technique allows the data collection of the phasic DA release and store it in a numerical matrix, which can be processed and transformed in images.

Figure 1.1 shows an image generated from the data collected with FSCV, in which there is a dopamine release highlighted. This type of image is the visual representation as the raw data, but using a palette of false colors. Although there are variations in the data matrices and how to read them depending on the equipment and proprietary software used in the experiments, researchers usually represent this data in image forms using a standard false color palette. It facilitates not only the visualization of the information but also the application of powerful descriptors for the feature extraction and it standardizes all samples.

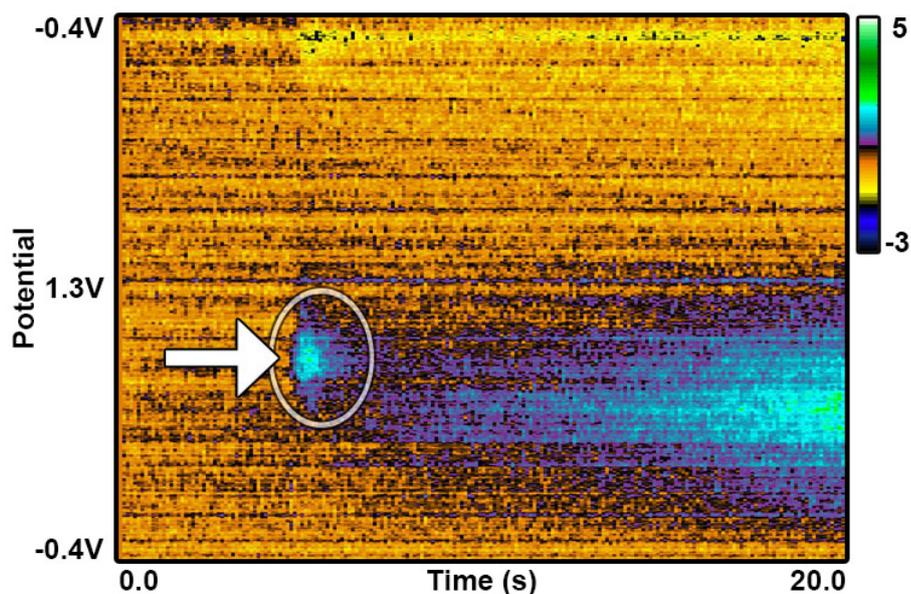


Figure 1.1: Example of phasic dopamine release image.

1.1 Motivation

Due to the dopamine behavioral impact, the study of the phasic release of this neurotransmitter becomes extremely important, as well as the need for methods and tools that allow and facilitate its analysis.

Although the fast-scan cyclic voltammetry stores the data of DA releases, many experiments have a high temporal resolution and consequently large amounts of data generated. Thereby, a manual analysis of these results requires a great deal of time, as well as being a laborious and repetitive task.

Some systems have already been developed to identify substances release such as transient adenosine in FSCV data (Borman et al., 2017). Nevertheless, as far as we know there is no similar system to automatically identify phasic dopamine release images using pattern recognition stages as preprocessing, feature extraction and classification.

1.2 Challenges

Each information collected in the experiments is stored in a matrix represented by the potential on the y-axis and the cycle (time) on the x-axis. But the identification of the phasic dopamine is difficult because it has a random nature of release over time and a certain voltage variation.

The nature of these events impacts the identification, being easily visualized when it is in considerable quantity or simply confused with a noise when it is low. It also results in release records with different forms and sizes. Noises and artifacts are constant throughout an experiment recording. Both do not have a pattern of form, quantity, or intensity. In addition to making it difficult to visualize the releases, they may even be confused with substances.

Finally, since there was no similar dataset available to the community, images with DA releases were generated, and all labels were made manually for the proposed experiments.

1.3 Objectives

The main objective of this research was to develop and evaluate the performance of a classification system capable of identifying the phasic release of dopamine, based on visual characteristics extracted from images generated with FSCV data. The system explored the use of different descriptors as well as classifiers.

To achieve this objective, it was necessary to obtain a new dataset with dopamine release images which, in addition to being adequately labeled, it was preprocessed. Patches were extracted from the original images, which were explored different parameters. Subsequently, its features were extracted and, finally, a classifier was used to generate the results. Such results were also combined with those obtained using other features descriptors before generating a final decision.

The datasets created were made available to the scientific community to be used in future works.

Chapter 2

Theoretical Background

To make this document self-contained, this chapter reviews the theoretical foundations that support the work proposal. First we present the basic concepts of dopamine and the fast-scan cyclic voltammetry. In sequence, the concept and stages of pattern recognition are reviewed, highlighting the representations and classification.

2.1 Dopamine

The neurotransmitter dopamine (DA; 3,4-dihydroxyphenethylamine), as well as epinephrine and norepinephrine, is a biogenic amine from the catecholamines group (Jackowska and Kryszynski, 2013). These neurotransmitters synthesis occurs from the amino acid tyrosine (S)-2-amino-3-(4-hydroxyphenyl)-propanoic acid, which through the enzymatic action of tyrosine hydroxylase (TH), will convert tyrosine into L-DOPA (L-3-4-dihydroxyphenylalanine). So, the L-DOPA is converted to DA by the enzyme DOPA decarboxylase (DDC).

The DA is transported to synaptic vesicles by the vesicular dopamine transporter (VMAT) and released by exocytosis when the synaptic terminal is depolarized (Jones et al., 2014). This release may occur in Tonic or Phasic way. The amount of neurotransmitter that will be released depends on the intensity of the stimulus, and also its predictive capacity and the salience associated (Grace, 1995).

The Tonic release occurs when dopaminergic fibers exhibit a low frequency of action potentials that result in the release of small amounts of dopamine continuously. The Phasic release occurs when dopaminergic fibers exhibit high frequencies of short-acting action potentials, resulting in a large increase in short-acting dopamine synaptic concentration (Robinson et al., 2009). Each of these is responsible for distinct types of behaviors, because, depending on the concentration of DA, distinct neural pathways can be activated (Grace, 1995). To study these release processes of DA, there are techniques to monitor the extracellular concentration of this neurotransmitter such as microdialysis and fast-scan cyclic voltammetry (FSCV).

The microdialysis provides a good chemical selectivity, however, it has a poor temporal resolution of the order of minutes to hours and it is more adequate to observe tonic dopamine release (Robinson et al., 2009). The FSCV performs a high temporal resolution sampling, tenths of second. Fast-scan cyclic voltammetry is more adequate to phasic dopamine release, and it is also used for real-time observations associated with behavior studies.

2.2 Fast-Scan Cyclic Voltammetry

The fast-scan cyclic voltammetry is an electrochemical detection technique. The current generated by the electrons that dopamine donate to or get from the electrode is proportional to the dopamine concentration. FSCV is able to measure variations in dopamine concentration with high temporal and spatial resolution (Yorgason et al., 2011). Therefore, it provides information about how fast dopamine is released from and reuptaked to synaptic terminals, which is very important to study effects of drugs of abuse (Gomez-A et al., 2017; Fawaz et al., 2009).

In the FSCV, a potentiostat is used to pass a voltage ramp rapidly through a carbon fiber electrode measuring the current. During this voltage ramp, nearby electroactive chemical species, such as dopamine, adenosine or monoamines, are oxidized and/or reduced resulting in alterations in current amplitude. This amplitude is proportional to the concentration of the species (Swamy and Venton, 2006).

The current read also is referred to as a voltammogram that is collected and compared across time to verify changes in concentrations of electroactive chemical species. The voltammetry has a high temporal resolution and it is an ideal technique for measuring rapid presynaptic signaling events in the brain. The measure of the DA release and transporter activity in limbic brain regions is a common application of voltammetry (Yorgason et al., 2011).

All informations are stored in a numerical matrix, and each information collected in the experiments is represented by the applied potential on the y-axis, the x-axis is the cycle (time), and the color is current. The colors are based on a false color palette agreed by researchers and used by FSCV analysis softwares.

2.3 Pattern Recognition

In pattern recognition the classification task can be understood as the assignment of a class to a feature vector, extracted from a sample to be classified, that is called pattern. Generally in this type of problem the patterns received are organized into a predefined number of classes with supervised learning, in which the input and the desired output data are provided. Each sample is a pair consisting of an input feature vector and a desired output class. An optimal scenario will allow the correct classification for unseen instances. However, there are cases in which the available attributes to characterize the samples do not obviously differentiate each class.

Classic examples in the application of pattern recognition are: writing recognition, fingerprint recognition, speech recognition and face recognition. The classical approach to the development of pattern recognition systems foresees three well-defined steps: preprocessing, feature extraction and classification (Duda et al., 2012). These steps are represented in the Figure 2.1.

The initial step of sample preparation usually applies segmentation trying to isolate parts of interest. In addition, filtering tasks such as noise reduction are part of the preprocessing stage. In many cases, in the feature extraction stage we extract features of visual attributes such as texture and color. Finally, classification algorithms are used on the extracted descriptors in order to assign a class to each standard submitted to the system.

2.4 Representation

The representation or feature extraction stage is quite important in the development of pattern recognition systems. Among the visual features as well as shapes and colors, the texture is

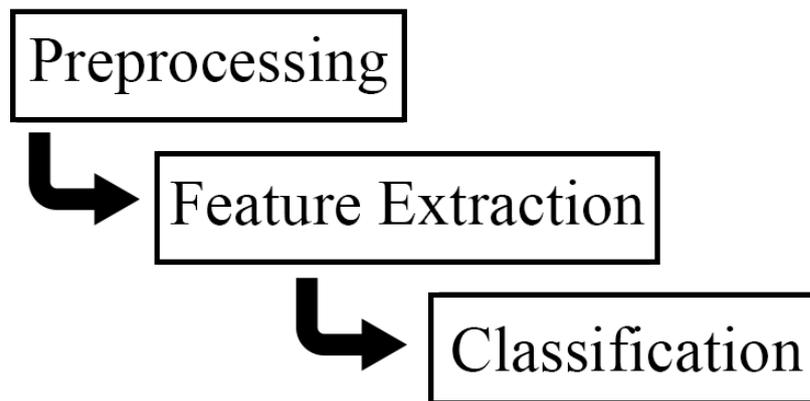


Figure 2.1: Pattern recognition system steps.

perceived easily by a person, contributing to the identification of objects in a given scene. Texture corresponds to a visual pattern which is usually related to the pixel distribution in a region and properties of the image object such as color, brightness, and size. Thus, this attribute contains significant information about the content of the image, being much explored in applications of computer vision. According to Gonzalez and Woods (2010), the texture may be a set of statistical features or other local image properties that are constant, periodic or with little variation.

2.4.1 Local Binary Pattern (LBP)

A widely used texture descriptor is the Local Binary Pattern, which has been used in several application domains: face recognition (Ahonen et al., 2006), musical genre recognition (Costa et al., 2012), handwriting identification (Bertolini et al., 2013) and classification of bird species (Zottesso et al., 2016).

Originally used as a complementary descriptor for local image contrast, the LBP was adapted and has become a good structural approach to texture (Ojala et al., 2002). Certain local binary pattern to a pixel neighborhood are considered fundamental image texture properties and the histogram of occurrence of these characteristics is used as features.

According to Ojala et al. (2002), the LBP operates on an image pixel and its adjacent ones to find a histogram of local binary pattern. To be able to operate textures of different scales, it can create patterns establishing different quantities of neighbors for its operation. Such variations are identified by $LBP_{P,R}$ in which P is the number of neighboring pixels existing in a region of radius R around the central pixel C (Figure 2.2).

The LBP values are computed by comparing the intensity of the central pixel with its neighbors. According to Mäenpää (2003), the LBP works as a threshold where the value 1 is taken if the neighbor intensity is equal to or higher than C , otherwise the value is 0. Then, these values are multiplied by weights of neighboring pixels, which correspond to powers of two with exponent equal to their order number in the neighbors sequence, the first one being equal to zero. Finally, the central pixel is assigned the sum of this calculation.

The total of LBP features calculated depends on the variation of adjacent pixels in relation to the central one. For example, in $LBP_{8,2}$ there are 2^8 possible patterns. However not all patterns need to be used as distinct feature. There is a concept of uniformity for the local binary patterns, which is based on the number of transitions between one and zero present in the sequence associated with the pattern. A binary code is considered uniform if the number of transitions is less than or equal to two, and the code is treated as a circular list.

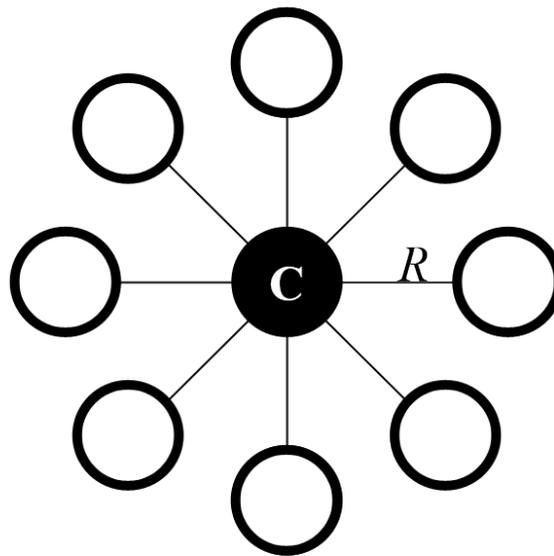


Figure 2.2: Central pixel C with P neighboring ones.

Thus it is possible to use values associated with uniform patterns generating a smaller vector. Ojala et al. (2002) established a histogram composed of 58 possible uniform combinations, in addition to all nonuniform patterns found in an additional vector position. These 59 characteristics can be obtained using the LBP which uses 8 neighbors and a value of R equal to 2 (Figure 2.3), instead of 2^8 patterns.

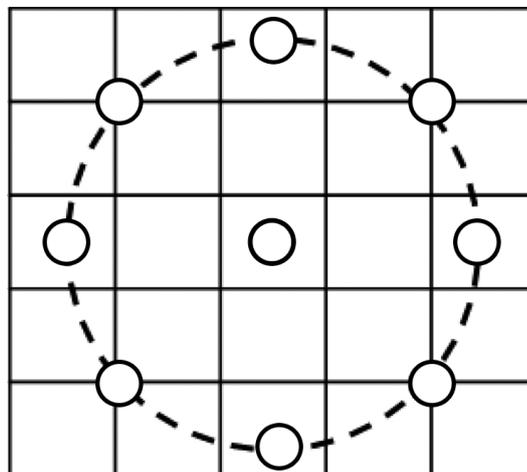


Figure 2.3: Example of a neighborhood with $LBP_{8,2}$.

2.4.2 Local Phase Quantization (LPQ)

Originally developed to identify the texture in blurred images (Ojansivu and Heikkilä, 2008), the Local Phase Quantization has shown good performance in both blurred and clear images. This descriptor is based on blur invariance property and uses the extracted local phase information using the 2D Discrete Fourier Transform (DFT) computed on a rectangular neighborhood, that is a local window for each pixel of the image.

The local phase information of an image of size $N \times N$ is given by the Short-Time Fourier Transform (STFT) presented in Equation 2.1.

$$\hat{f}_{u_i}(x) = (f \times \phi_{u_i})x \quad (2.1)$$

in which the filter ϕ_{u_i} is result of the Equation 2.2.

$$\phi_{u_i} = e^{-j2\pi u_i^T y} \mid y \in \mathbb{Z}^2 \mid \|y\|_\infty \leq r \quad (2.2)$$

where $r = (m - 1)/2$ is the size of the window and u_i is a 2D frequencies vector.

This descriptor considers only four complex coefficients that correspond to the 2D frequencies: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, $u_4 = [a, -a]^T$, where $a = 1/m$. The STFT is expressed through the notation vector according to Equation 2.3.

$$\hat{f}_{u_i}(x) = w_{u_i}^T f(x) \quad (2.3)$$

$F = [f(x_1), f(x_2), \dots, f(x_{x_2})]$ is denoted as a matrix $m^2 \times N^2$ which comprises neighborhood of all pixels of the image and $w = [w_R, w_I]$, where $w_R = Re[W_{u1}, W_{u2}, W_{u3}, W_{u4}]$ and $w_I = Im[W_{u1}, W_{u2}, W_{u3}, W_{u4}]$.

The Re and Im represent respectively the real and imaginary parts of a complex number, and the transformation matrix ($8 \times N^2$) is given by $\hat{F} = wF$.

According to Ojansivu and Heikkilä (2008), the function $f(x)$ is a result of the first order Markov process. The coefficient of correlation between two pixels x_i and x_j is exponentially related to its distance L^2 . A covariance matrix C of size $m^2 \times m^2$ is defined for the vector f , according to the Equation 2.4.

This covariance matrix of the coefficients is obtained by $DwCw^T$. D is not a diagonal matrix, the coefficients are correlated and may be no longer through $E = C^T \hat{F}$, where V is an orthogonal matrix derived from the singular value decomposition (SVD) of matrix D , with $D' = V^T D V$.

$$C_{i,j} = \sigma^{\|x_i - x_j\|} \quad (2.4)$$

The Equation 2.5 shows how the coefficients are quantized, where e_{ij} are components of E .

$$q_{ij} = \begin{cases} 1 & \text{if } e_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

Finally such binary elements are transformed to decimal (Equation 2.6), comprising integer values between 0 and 255. Thus the LPQ histogram is composed of the vector of 256 positions, which represents all the positions of the input image.

$$b_j = \sum_{i=0}^7 q_{ij} 2^i \quad (2.6)$$

2.5 Representation Learning

A classification system performance is heavily dependent on the choice of data representation or features used. And the inability to extract and organize discriminative informations from the data impacts poor results that can be obtained (Bengio et al., 2013).

Automatic learning representation can make it easier to extract important information to build a classification system. Among the ways of learning representations there are the deep learning methods, like convolutional neural networks (CNN).

A convolutional neural network is a variation of a multi-layer perceptrons network and consists of layers with different functions. Figure 2.4 shows an example of a CNN and the number of each layer can vary from network to network. Initially it is common to apply the data to input layers known as convolutional layers (Vargas, 2016). These layers are composed of neurons, and each neuron is responsible for applying a trainable filter to a specific image area. Basically a neuron is being connected to a set of pixels of the previous layer and for each connections is applied a weight. The respective weights of its connections, produces an output passed to the next layer. The weights assigned to the connections of a neuron can be interpreted as a matrix representing the filter of a convolution of images in the spatial domain (kernel). The weights are shared across neurons from a same layer, leading the filters to learn patterns which occur in any part of the image.

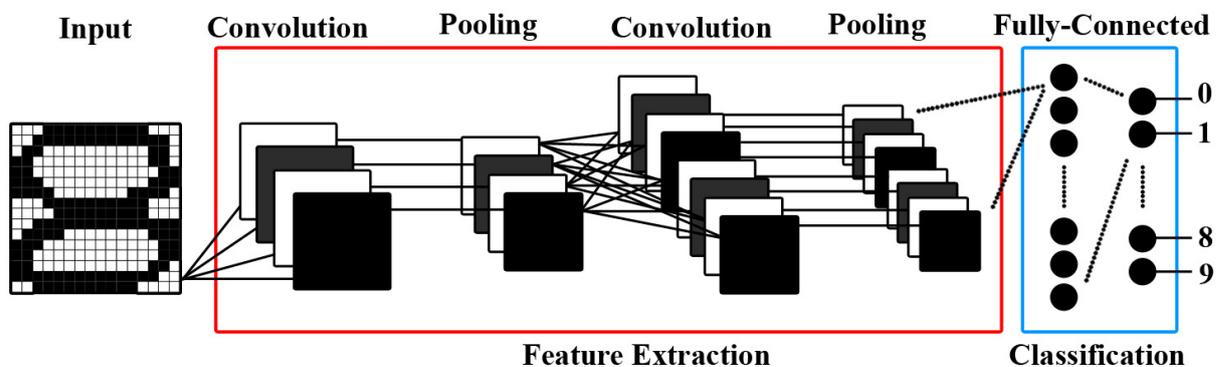


Figure 2.4: Example of a convolutional neural network layers.

In the CNN convolutional layers, it is not necessary to specify which filters or features to be used. It is defined only the architecture of the filters: sizes, stride and quantity per layer. The learning process of the network changes the weights throughout the training, searching automatically for the best values for the input dataset.

A very important layer commonly used after the convolutions is the pooling layer. The function of this layer is to reduce the dimensionality of the data in the network. This reduction is important for training faster, but also to create spatial invariance. The pooling layer only reduces the height and width of a map.

When it is desired to perform a classification, it is appended after the set of the convolutional and pooling layers at least one fully-connected layer. This fully-connected layer is responsible for tracing a decision path to each class based on results of the filters from previous layers.

After the fully-connected layer the last step is the classification function. It is fundamental in training, since it influences the learning of the filters and consequently the result of the network. However, it is not necessary to classify the input dataset using only CNN, it is possible to use the features extracted using representation learning with another classification algorithm.

There is also the transfer learning approach, which can assist in tasks with a small number of input images to train and learn parameters. A CNN often contains a huge number of parameters, which is necessary a huge number of images for training. Even a direct learning using thousands examples is problematic, thus a transfer learning can use the internal representation learned from another task, and re-use it to extract features from a new target problem.

Figure 2.5 shows an example of transfer learning based on the Oquab et al. (2014) proposal. A CNN is trained in the source task, using its images. The parameters learned are transferred to the target task, except for the last layer. The target task contains a smaller dataset and to compensate the different images distribution, two additional fully-connected layers are added at the end of the network, training it by using the target dataset and then classifying it.

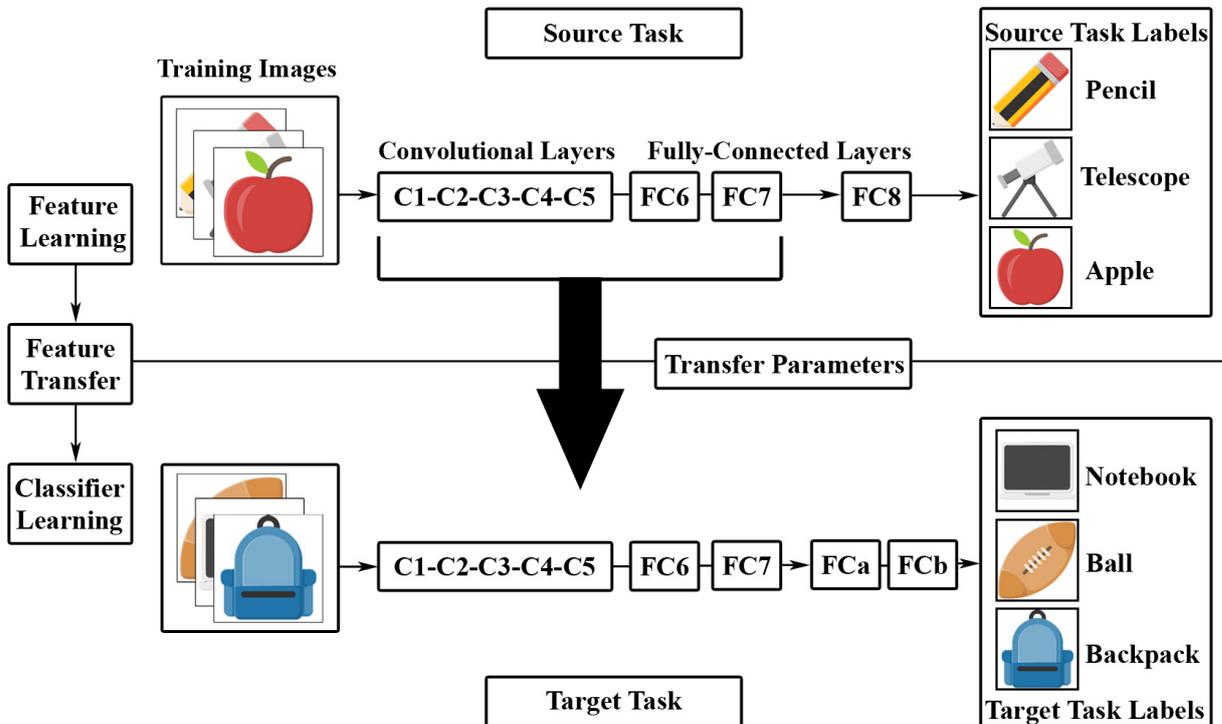


Figure 2.5: Example of transfer learning.

2.5.1 Inception

Inception is a deep convolutional neural network architecture with a design that allows increasing the depth and also the width of the network, while keep the computational cost constant (Szegedy et al., 2015).

Images from a same class may have huge variation in the location and size of the information. Choosing the best kernel size for the convolution is not so easy. So the Inception uses multiple sizes filters operating on the same level.

The Figure 2.6 shows the inception modules. The naïve module performs convolution on an input using 3 different sizes of filters: 1×1 , 3×3 , 5×5 . Max pooling is also performed and the results are concatenated before the next module. But even a few number of 5×5 is computationally expensive. Szegedy et al. (2015) added an extra 1×1 convolution before the 3×3 and 5×5 convolutions to limit the number of input channels.

Thus, an Inception is a neural network consisted of inception modules stacked upon each other, with occasional max-pooling layers to reduce the resolution of the grid. Using this dimension reduced inception module, a neural network known as GoogLeNet (Inception v1) was built.

Through Inception v2 and v3, several improvements were applied in the network model (Szegedy et al., 2016), such as smart factorization methods. A filter $n \times n$ can be factored into a

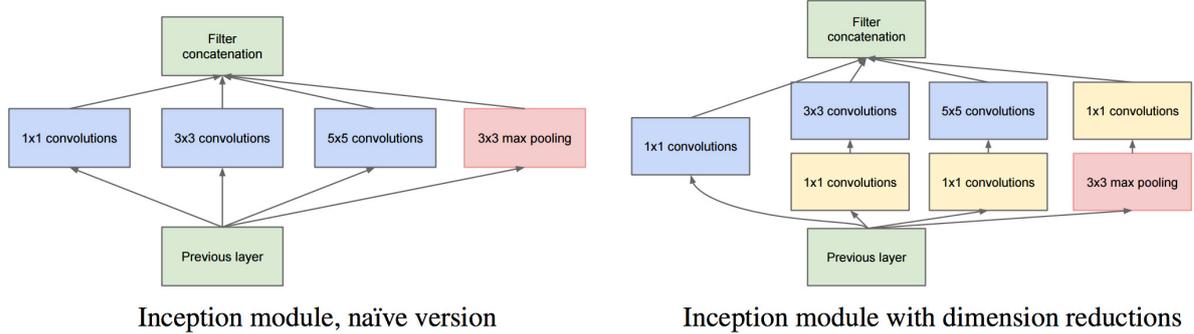


Figure 2.6: Inception module. Extracted from Szegedy et al. (2015).

combination of $1 \times n$ and $n \times 1$ convolutions to improve computational speed. Szegedy et al. (2016) also altered the auxiliary classifiers, which contribute more near the end of the training step. The Inception v3 network model can be seen in Figure 2.7.

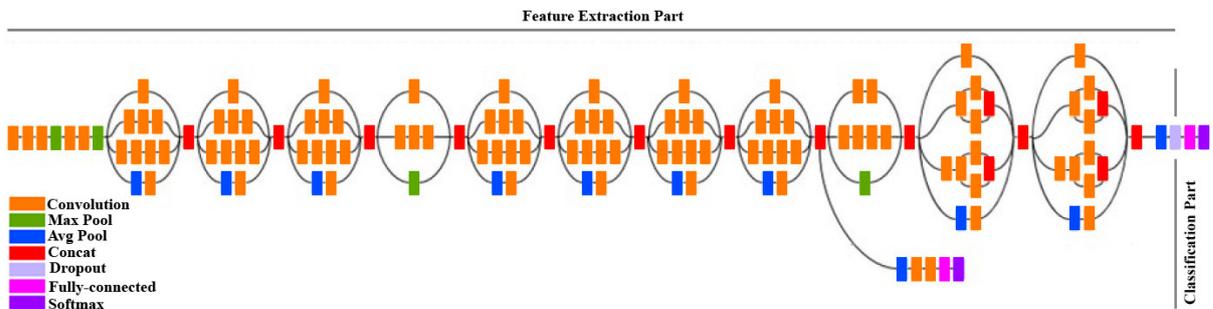


Figure 2.7: Overview of Inception v3 network model.

2.5.2 You Only Look Once (YOLO)

YOLO is an accurate and fast approach to object detection (Redmon and Farhadi, 2017). A single neural network evaluation predicts bounding boxes and class probabilities directly from full images (Figure 2.8). You Only Look Once (YOLO) at a sample to predict the classes and their locations (Redmon et al., 2016).

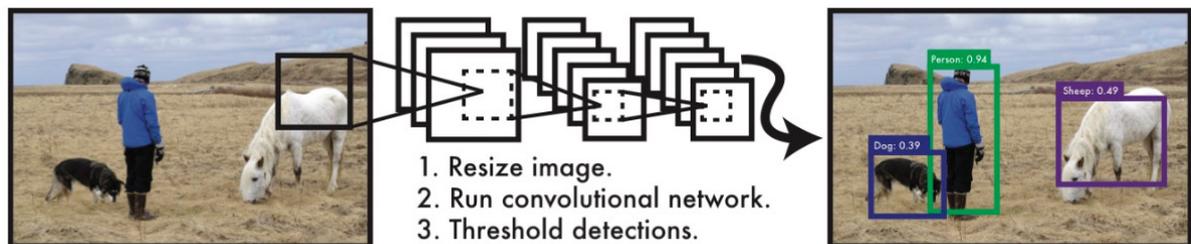


Figure 2.8: The YOLO Detection System. Extracted from Redmon et al. (2016).

The YOLO model initially divides the input sample into an $S \times S$ grid (Figure 2.9). Each grid cell predicts a fixed number bounding boxes. If the center of an object falls into the cell, it is responsible for predicting that object.

Each cell predicts B bounding boxes and each box has a confidence score. This score shows how confident the YOLO is that there is an object in the box and how accurate it is (Redmon et al., 2016). The confidence score should be zero if there is no object in that grid cell.

Each bounding box also consists of: (x, y, w, h) . The box width w and height h are normalized, x and y are offsets to the corresponding cell. x, y, w, h are between 0 and 1. Each cell also predicts C conditional class probabilities. And it is the probability that the detected object belongs to a specific class. All these predictions are encoded as an $(S, S, B \times 5 + C)$ tensor.

YOLOv1 predicted the bounding boxes using fully-connected layers, which were removed since its second version and now uses anchor boxes. So instead of directly predicting a bounding box, YOLOv2 and v3 predict offsets from a predetermined set of predetermined boxes.

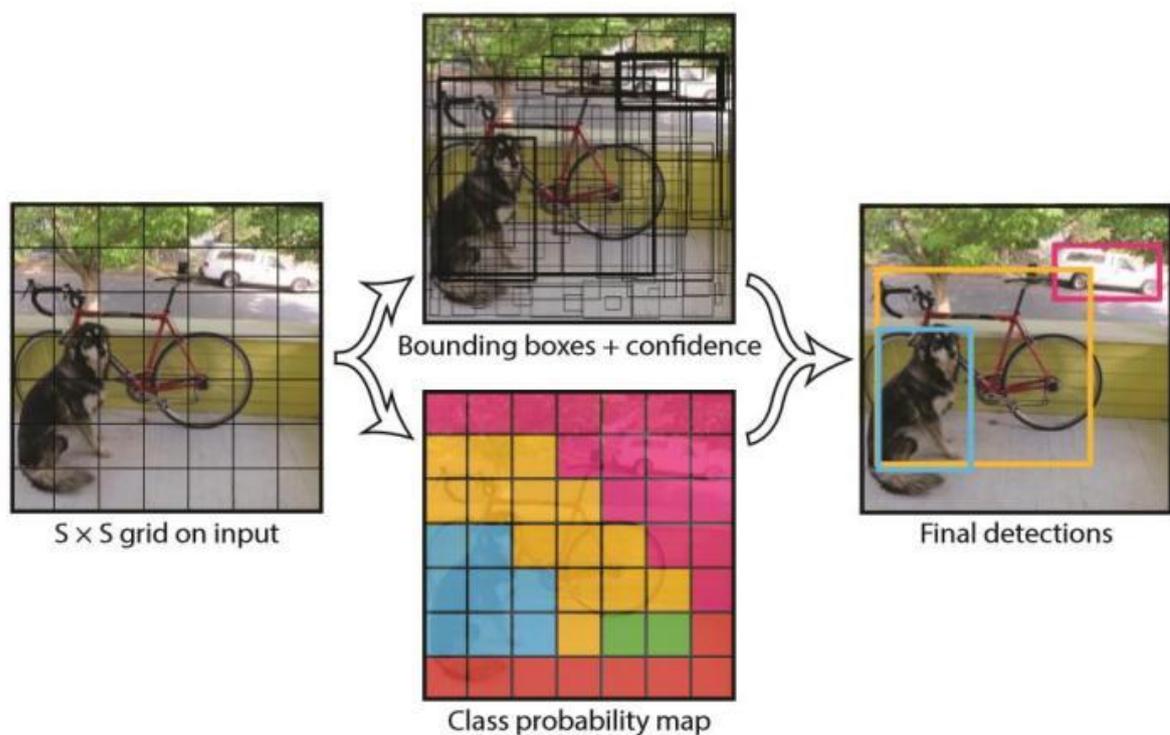


Figure 2.9: The model detection. Extracted from Redmon et al. (2016).

The YOLOv3 network for performing feature extraction uses successive 3×3 and 1×1 convolutional layers (Redmon and Farhadi, 2018). The model is significantly larger than other versions (v1, v2, fast and tiny) with 53 convolutional layers, as shown in Table 2.1

Table 2.1: YOLOv3 model (Darknet-53)

Type	Filter	Size	Output	
Convolutional	32	3×3	256×256	
Convolutional	64	$3 \times 3 / 2$	128×128	
Convolutional	32	1×1		x1
Convolutional	64	3×3		
Residual			128×128	
Convolutional	128	$3 \times 3 / 2$	64×64	
Convolutional	64	1×1		x2
Convolutional	128	3×3		
Residual			64×64	
Convolutional	256	$3 \times 3 / 2$	32×32	
Convolutional	128	1×1		x8
Convolutional	256	3×3		
Residual			32×32	
Convolutional	512	$3 \times 3 / 2$	16×16	
Convolutional	256	1×1		x8
Convolutional	512	3×3		
Residual			16×16	
Convolutional	1024	$3 \times 3 / 2$	8×8	
Convolutional	512	1×1		x4
Convolutional	1024	3×3		
Residual			8×8	
AvgPool		Global		
Fully-Connected		1000		
Softmax				

2.6 Classification

The support vector machine (SVM) is a set of supervised learning techniques able to analyze data, recognize patterns and classify them. This classifier, introduced by Vapnik (1995), has been widely used with success. It has presenting competitive results in the most diverse applications, such as classification of bird species (Zotesso et al., 2016), music genre recognition (Costa et al., 2011) and identification of handwriting (Bertolini et al., 2013).

The standard SVM is defined as a non-probabilistic binary linear classifier, that has as input a set of data, and is able to predict for each input, which of the possible classes it is part of. Initially, with a training algorithm and set of examples already defined to which category each belongs, the SVM constructs a model that assigns the new examples to one category or another.

2.6.1 Classifiers Combination

When the classifiers outputs present for each sample an estimate of probability for each class in the classification system, it is possible to perform a combination of classifiers. Kittler et al. (1998) proposed some merging rules to combine the predictions of different classifiers. Among these rules we have:

- **Maximum:** among all classifiers, the class with the highest prediction score is chosen.
- **Average:** it calculates the mean between the prediction values of each class present in all classifiers, and chooses the class with the highest average.
- **Sum:** it sums all the predictions values of each class in all classifiers, and chooses the class with the highest final value.
- **Product:** similar to the Sum Rule, but it calculates the product of the values instead of adding up.

2.7 Evaluation Measures

An automatic classification system should be evaluated on the basis of evaluation measures. Accuracy is one of the most common measures used in several areas, in which it is measured how efficient the system is from the point of view of classification. It is basically the percentage of final score that the system obtained taking into account the total number of correctly classified samples, and the total number of samples classified (correct or not). Other criteria commonly used in evaluating the efficiency of a system are Precision, Recall, and F-measure.

- **Precision:** It is the result of the number of True Positives (T_p) samples over the number of True Positives plus the number of False Positives (F_p). If the classification system contains a large number of False Positives (F_p), the precision decreases. Its formula is expressed by Equation 2.7.

$$Precision = \frac{T_p}{T_p + F_p} \quad (2.7)$$

- **Recall:** It is the result of the number of True Positives (T_p) samples over the number of True Positives plus the number of False Negatives (F_n). If the classification system contains a large number of False Negatives (F_n), the precision decreases. Its formula is expressed by Equation 2.8.

$$Recall = \frac{T_p}{T_p + F_n} \quad (2.8)$$

- **F-measure or F-score:** It is defined as the harmonic mean of Precision and Recall. Its formula is expressed by Equation 2.9.

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.9)$$

2.8 Receiver Operating Characteristic Curve

The performance of a binary classifier can be evaluated using a graphical plot called Receiver Operating Characteristic (ROC) curve (Metz, 1978). By plotting the true positive rate (TPR) against the false positive rate (FPR) at different threshold settings, it is possible to generate a ROC curve.

Figure 2.10 shows an ROC space, which is defined by TPR as y-axis and FPR as x-axis. The diagonal line represents a test with no discriminating ability (DeLong et al., 1988), which is the same as guess a class by flipping coins.

Points below the diagonal represent worse results than a random classifier. Points above it represent better classifications. The ideal point is (0,1). And the closer an ROC curve comes to it, the better its discriminating.

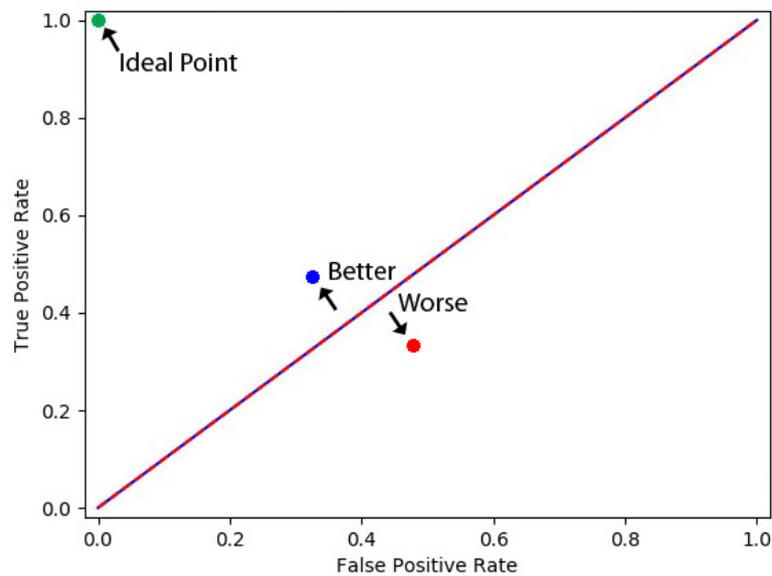


Figure 2.10: Example of ROC space.

Chapter 3

State-of-the-Art

The identification and analysis of dopamine release is a well-known and studied problem in the area of neuroscience and medicine (Da Cunha et al., 2012, 2015; Grace, 1995). However, when we are dealing specifically with systems in the computer science area, such as pattern recognition systems, we have an original problem to be explored and tested. To the best of our knowledge, no similar work has been found in the literature on automatic identification of images with phasic dopamine release, generated from data obtained by FSCV.

Borman et al. (2017) presented an algorithm for detection of transient adenosine release. However, an analysis of current (nA) data is performed at two specific voltages (1.4 V and 1.0 V) which are well-defined parameters for the oxidation of this substance. No features of visual image information are extracted. In this case, it is verified peak moments at these two points and, once they matched the proposed rules, an adenosine release is identified. The rules are based on the fact that a peak must be present at both oxidation voltages, the secondary peak (1.0 V) must lag the primary one (1.4 V), the ratio of secondary to primary peak currents must be in an acceptable range consistent with adenosine, and the duration of secondary peak must be longer than the primary one. The f-measure obtained with this system had a mean of 90%, being tested in four different datasets.

Yorgason et al. (2011) created a Demon Voltammetry and Analysis software suit, which is compatible with existing hardware and is capable of detecting and quantifying electrically evoked DA release and uptake. It has been developed for data collection, analysis, and figure making. This software also contains stimulation settings for performing pulse and burst stimulations. The analysis tools contains features for examining data and figure making that allows for creation of publication quality figures which can be exported. They also used it to compare the sensitivity of multiple kinetic measures of release and uptake to cocaine-induced alterations in electrically evoked dopamine.

Nguyen and Venton (2015) presented how FSCV can be used to adenosine release detection, and also compared this technique with others used to measure adenosine. The other methods are: Microdialysis which is one of the most general techniques for monitoring neurochemical changes; Electrophysiology studies monitor the firing of neurons and can be used to examine the effects of adenosine release; and Amperometric Biosensors that directly measure adenosine at platinum electrodes coated with enzymes. They concluded there are several advantages using fast-scan cyclic voltammetry and it is the best method for measuring rapid changes. The carbon-fiber FSCV microelectrode is the small size of the electrodes, which are less invasive than microdialysis probes, and are also cheap and easy to make compared to biosensors. But the biggest advantage is the time resolution of the measurements, which is the fastest method

for measures. Biosensors take too long to respond to a change in adenosine and thus are not fast enough to respond to all changes in spontaneous adenosine releases.

Cyclic Voltammetry is not only used to analyze dopamine or adenosine. It is widely used to analyze the most diverse substances. Park et al. (2014) examined zirconium electrochemical redox behaviors based on cyclic voltammetry results. Masek et al. (2014) investigated using cyclic and differential pulse voltammetry the process and the kinetics of the electrochemical oxidation of morin in an anhydrous electrolyte. Kumar et al. (2016) examined therapeutic potential of selective serotonin and norepinephrine reuptake inhibitor in an animal model of comorbidity between epilepsy, depression- and impulsive-like behavior, using FSCV to measure the strength of serotonergic and noradrenergic tones.

Many advances have occurred in the area of pattern recognition, which addresses the classification of items according to certain classes, or categories provided in a domain of a problem (Duda et al., 2012). Despite the lack of equivalent publications on automatic identification of dopamine images, recognition techniques involving preprocessing, feature extraction and the elaboration of a classification system, have been successfully applied in recognition tasks in the most diverse domains. Cavalin and Oliveira (2017) presented a survey of texture classification methods, which were already applied in writer identification (Bertolini et al., 2013), forest species classification (Tou et al., 2007; Nasirzadeh et al., 2010), medical diagnosis (Sutton and Hall, 1972; Harms et al., 1985; Khademi and Krishnan, 2008), geo-processing (Haralick et al., 1973) and agriculture (Jhuria et al., 2013; Pujari et al., 2015).

Haralick et al. (1973) used texture features for classification of photomicrograph (five kinds of sandstones), aerial photograph (eight land-use categories) and satellite images (eight land-use categories). In this paper the features were based on gray-level spatial-dependence matrices, also known as gray-level co-occurrence matrices (GLCM), which were extracted from these three different datasets. Each dataset was divided into two parts: a training and a testing set. The best accuracy was 89% for the photomicrographs images, 82% for the aerial photographic images, and 83% for the satellite images. They also concluded that the texture features probably have a general applicability for a wide variety of images classification.

Cavalin et al. (2013) investigated the extraction of multiple feature vectors based on image segmentation and multiple feature sets. Texture descriptors as gray-level co-occurrence matrix (GLCM), local binary patterns (LBP) and linear phase quantization (LPQ) were used to extract features from a 112 species database containing microscopic images of wood. Support vector machines (SVM) were used for training and testing, and the results demonstrated that the proposed approach increased the recognition rates. The best recognition rate was 93.2% as a result of LPQ and GLCM features combined.

Bertolini et al. (2013) showed a classification scheme based on dissimilarity representation, in which were used texture descriptors to perform a writer identification and verification. Two different databases were used: BFL database which is composed of 315 writers and a total of 945 images; and IAM database with a total of 650 different writers. Their experiments using LBP and LPQ features, and SVM as classifier were able to surpass previous results in the literature, and the best accuracies were using LPQ features: 96.7% on the BFL database and 99.2% on the IAM database.

Pujari et al. (2015) presented different image processing techniques used to identify and classify fungal diseases which affect different agriculture and horticulture crops. They focused on visual symptoms and extracted features from fungal disease in plants images. The datasets used were composed of different crops, for example: fruit, vegetable and commercial crops images. And for each one was used a different approach. The identification of fruit crops disease was performed extracting features as block-wise, GLCM and gray-level run-length matrix (GLRLM).

And to classify these images into normal or severely, moderately and partially affected, the nearest neighbor (NN) classifier was used considering the Euclidean distance. The accuracies with block-wise, GLCM and GLRM features were respectively: 94.08%, 91.37% and 86.71%. The identification of fungal disease affected on vegetable crops was performed extracting local binary patterns (LBP) features and classifying them using a proposed neuro-kNN, which reached the accuracy of 91.54%. Using discrete wavelet transform (DWT) features and a probabilistic neural network (PNN) classifier, the identification on commercial crops dataset accuracy was 86.58%.

Lucio and Costa (2015) presented a bird species identification approach using sounds, but not using acoustic features. The classification system was based on textural features of spectrogram images. The dataset used was composed of 2814 audio samples of 46 different species, which were converted to spectrogram. From those images, features were extracted using local binary patterns (LBP), local phase quantization (LPQ) and Gabor filters. The classification step was also performed using SVM and the best f-measures obtained were: 76.39%, 69.09% and 75.67% for LBP, LPQ and Gabor filters features respectively.

In addition to all the researches using texture descriptors and SVM, since the first works introducing convolutional neural networks (LeCun et al., 1989), it has also been applied in several domains. Character recognition which is a problem well known in the literature was investigated by LeCun et al. (1998). They applied deep neural network to a handwritten digit database known as MNIST. There were also work using convolutional networks for traffic sign recognition (Sermanet and LeCun, 2011) and real-world house numbers classification (Sermanet et al., 2012).

Until Krizhevsky et al. (2012) achieved a significantly higher accuracy using CNN on ImageNet dataset, winning the ILSVRC 2012 competition. And since then, the researches just show more and more advance on convolutional neural networks over other classification methods in the most varied databases.

Redmon et al. (2016) presented a unified architecture model for real-time object detection, called YOLO. It runs a single convolutional network on the input image, identifying objects and classifying them. This approach has been highlighted in several applications, such as the vehicle detection and classification presented by Zhou et al. (2016). Laroca et al. (2018) developed a real-time automatic license plate recognition system based on the YOLO detector and Severo et al. (2018) published an iris location study also using a fine-tuned YOLO model.

Chapter 4

Methodology

This chapter describes the techniques, datasets, and parameters used for the creation of the proposed classification system. From a high-level, the process follows the classic steps presented by Duda et al. (2012) and it was explored three mainly approaches: training and testing with original images; training and testing with extracted image patches; and a combined approach.

4.1 Datasets

Fast-scan cyclic voltammograms (FSCV) color plot images were obtained from the Laboratory of Central Nervous System of the Federal University of Parana (UFPR) at Curitiba, Brazil and from D. Robinson's Laboratory of the University of North Carolina (UNC) at Chapel Hill, United States of America. Data were recorded from a carbon fiber electrode placed in the nucleus accumbens of anesthetized rats. Dopamine release was evoked by electrical stimulation of the ventral tegmental area (20 pulses, 0.5 ms per pulse).

FSCV measurements were taken with a Wireless Instantaneous Neurotransmitter Concentration Sensor system (WINCS, Mayo Clinic, Rochester, Minnesota, USA) and processed with WINCSware with MINCS software (version 2.10.4, Mayo Clinic, Rochester, MN, USA). Every 100 ms, a triangular wave form potential of -0.4V to +1.3V to -0.4V was applied at a rate of 300 V/s to the carbon-fiber recording electrode versus the Ag/AgCl- reference electrode. Oxidative and reductive currents were continuously sampled at 100,000 samples/s and 944 samples/scan.

The images were generated from 30 different experimental records with a total of 1005 electrically evoked dopamine release, resulting in two datasets. Each record has dopamine release evoked with different magnitude of electrical stimulation (200-600 μ A, 50-60 Hz).

4.1.1 Dataset I

The first dataset of cyclic voltammogram images (Matsushita et al., 2018) was obtained exclusively from the Laboratory of the Central Nervous System of the Federal University of Parana (UFPR). This initial dataset consists of 9 different experimental recordings with a total of 285 phasic DA releases, which was processed and transformed into images with resolution of 1200×900 pixels. Each record has releases with different concentration, that results in images with different intensities and sizes of these events.

Taking into account that the phasic dopamine has a random nature of release over time and a certain potential (voltage) variation, these moments were labeled manually and with this information it was possible to classify them. These labels saved in text files contain information

of the approximated x-axis release interval and release peak of each image. They also contain information from which experimental recording the image belongs to and the original recording cycle (time) from which it was extracted. Figure 4.1 shows red lines plotting this release interval labeled.

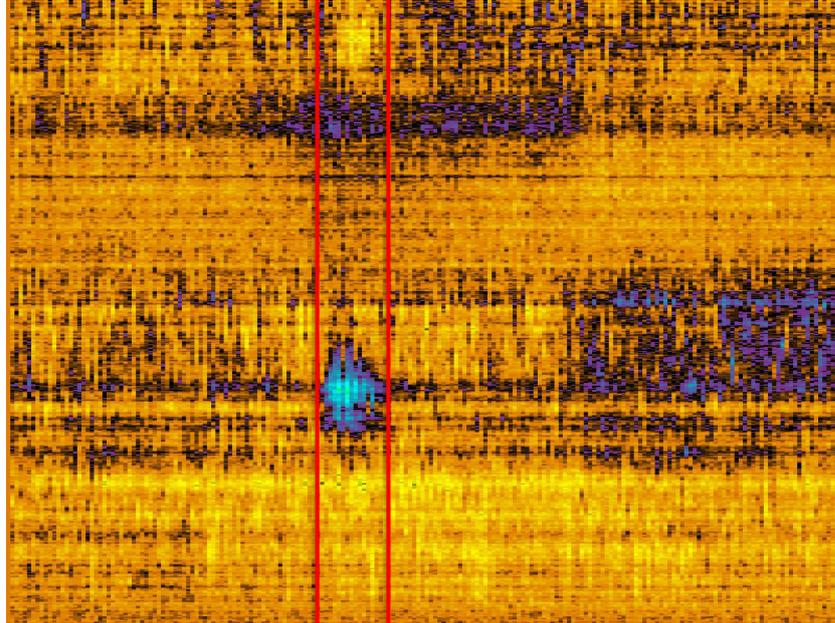


Figure 4.1: DA release interval.

Image patches with size of 100×100 , 120×120 and 150×150 pixels were manually created using the release moments labeled information. The central column (x-axis) of these patches were based on the release peak. Despite of the voltage variation it was possible notice a common area in all images, which is approximated between the pixel 515 and 615 of the y-axis. This area was used to set the height of the patch, that can be increased or decreased proportionally depending on the resolution used.

Peaks can be observed by green lines in the Figure 4.2, while the common release region is limited by red ones.

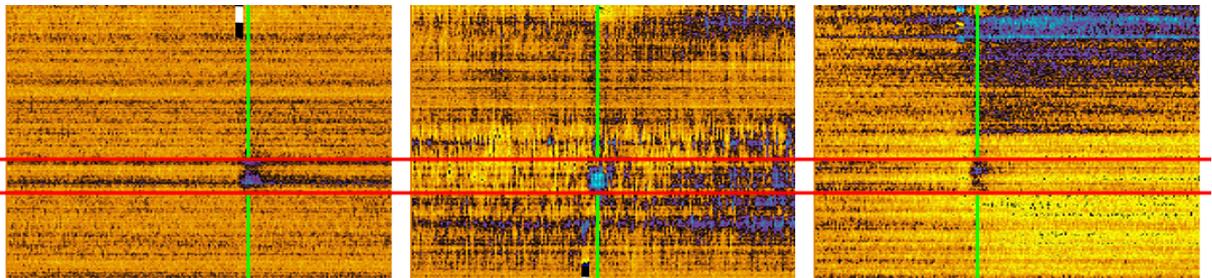


Figure 4.2: Peaks and common dopamine release region.

Aiming at a balanced dataset, 285 random patches of areas without dopamine release were extracted, but they keep using the common area to delimit the height. The 570 images patches of the two classes were randomly divided into 3 folds with 190 images each (95 DA release and 95 Non-release). These manually extracted patches folds constitutes the training dataset.

Also a testing dataset was created using automatic image patches extraction, that is closer to real life situation. A 120×120 sliding window was applied over the entire images, moving horizontally from 60 to 60 pixel inside the common area. Since the labeled information contains the approximated release interval, it was possible to separate them into two classes: (1) phasic DA release patches, and (2) non-release patches. Figure 4.3 shows an example of automatically extracted patches.

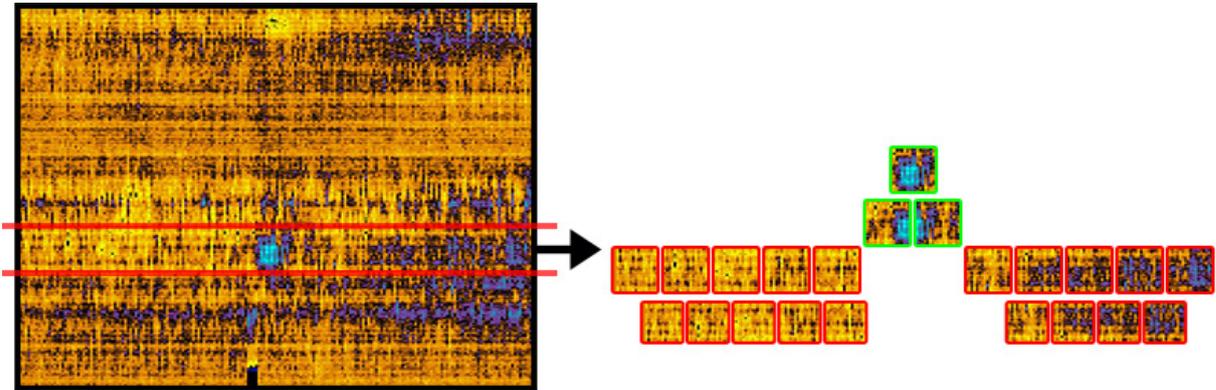


Figure 4.3: Example of automatically extracted patches.

The automatic patches were divided into 3 folds as well. It is important to note that the image distribution were the same as the manual one, but now with more than one patch extracted from each original image. Patches from the same image are in the same fold, ensuring there were no different training and testing folds with samples from the same source.

Since a complete experiment has more recording time with no dopamine release, more 912 samples from 48 images without DA complemented this dataset, 304 for each fold. And for the use of texture descriptors, all images were converted to grayscale.

4.1.2 Dataset II

The second version of the dataset is composed of images generated from all the 30 experimental records, including those in Dataset I. Unlike the first version, this one has not only phasic dopamine release images, but also full images with no DA release. In total there are 2010 images, 1005 of each of these classes, with resolution of 875×656 pixels. Despite the resolution difference, the images on both datasets represent a 20 seconds recording.

During the generation of these FSCV images, it is common to use a background subtraction before applying a fake color palette. Normally for each image, one column is selected to subtract the values from the others. In the case of the first dataset, this process was done manually during its generation. In the dataset II, this process is done automatically, choosing 3 different background positions: the Background A was selected from the beginning of each image (0.5 seconds), the Background B from the middle of each image (10 seconds), and the Background C from the end of each image (19.5 seconds).

These images with different background end up generating different results, as it is possible to be observed in Figure 4.4. Thus it is possible to explore different approaches of training and testing, since for each DA release 3 images were generated.

Like the first version, it also has a common region now between the pixel 320 and 520 of the y-axis. And all processes from labeling to patches extraction have also been applied to this dataset. The patches size used was 200×200 and all images were divided into 3 folds as well.

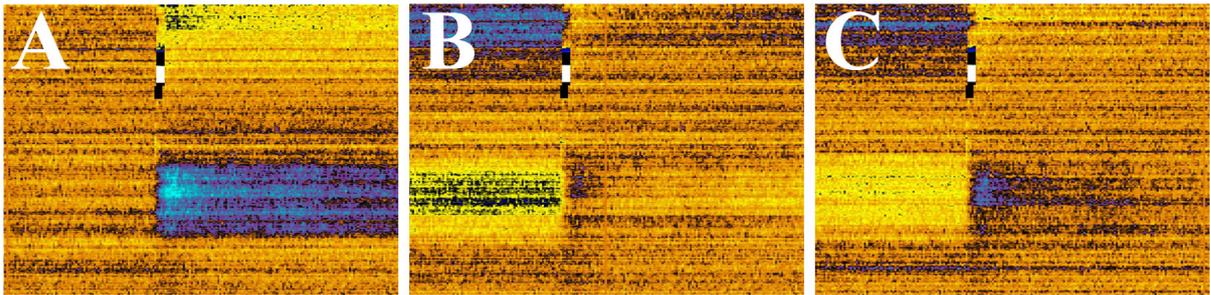


Figure 4.4: Images generated using different background positions.

The sliding window was applied over the images, moving horizontally from 135 to 135 pixel inside the common area.

4.2 Image Patches Approach

The first approach used only the extracted image patches and it is illustrated in Figure 4.5: texture descriptors were applied over the labeled image patches to extract features, then the classification step was performed using the manually extracted patches features as training model and the automatically extracted ones as testing. It was also performed tests using convolutional neural networks replacing these two steps. Each sample was classified as Phasic Dopamine Release (Class 1) or Non-Release (Class 2), and an evaluation metric will be applied to calculate the accuracy and f-measure.

4.2.1 Texture Descriptors

The feature extraction was performed using texture descriptors presented in Chapter 2.4 with grayscale input images. $LBP_{8,2}$ was applied considering only the so-called uniform patterns, which produces a histogram of 59 values as a feature vector. It was also tested a variant of LBP with the same parameters, known as Robust Local Binary Pattern (RLBP) (Chen et al., 2013), that was proposed to be more robust in noisy images. The LPQ generates a vector with 256 characteristics and was applied using different window sizes in order to find the best one.

4.2.2 Classification

Different classifiers were used in the experiments, but the best results were always obtained using Support Vector Machines. Thus, the classification was performed according to the Chapter 2.6, using SVM with RBF (radial basis function) kernel, and parameters C and γ optimized by grid-search. The SVM implementation used was proposed by Chang and Lin (2011) and it is known as LIBSVM.

It was used a cross-validation, and when one among the 3 folds was used as testing set, the other two were used as training set. Thus, it is guaranteed that no classified sample were used in the training. All extracted patches experiments were performed using manual extracted patches as training set, which has the same number of samples for both classes.

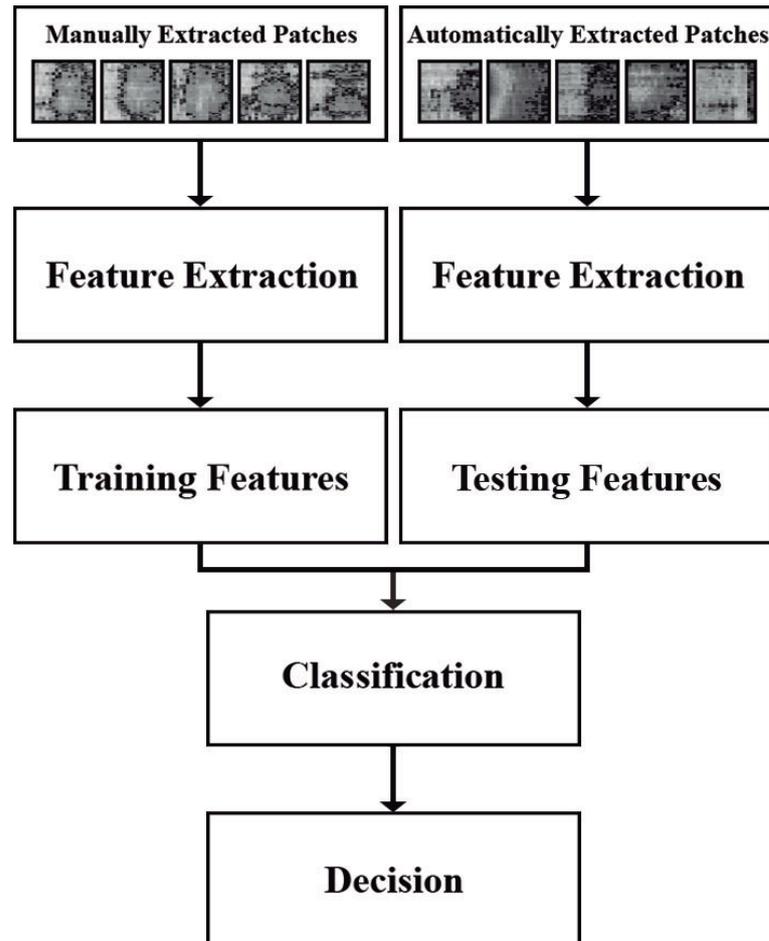


Figure 4.5: An overview of the image patches approach.

4.2.2.1 Early and Late Fusion

The use of SVM generates as a result an estimate of probability for each class existing in the classification system. When we extract different features and classify with support vector machine, this resulting predictions can be combined before having a final decision, as can be seen in the Figure 4.6. The identified patterns, even if incorrectly and using the same classifier, do not necessarily overlap. Thus, for some cases combinations of classifiers predictions were performed, and it is called Late Fusion. The merging rules used are: Maximum, Product and Sum (Kittler et al., 1998).

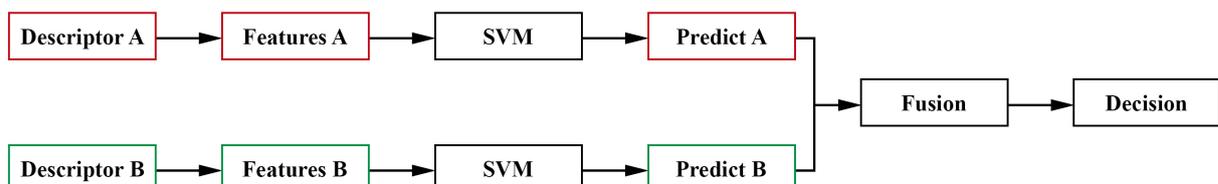


Figure 4.6: Overview of Late Fusion.

For some experiments, instead of using the late fusion, it is possible to concatenate the feature vectors before the classification. The features extracted with different descriptors were used as one larger vector, for example: the 59 LBP features and the 256 LPQ features are fused

as a single array with 315 positions (Figure 4.7). It is called Early Fusion and the Figure 4.8 represents the overview of this technique.

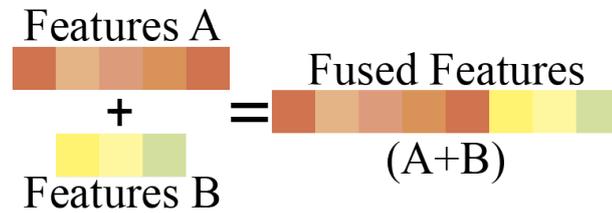


Figure 4.7: Representation of Early Fusion.

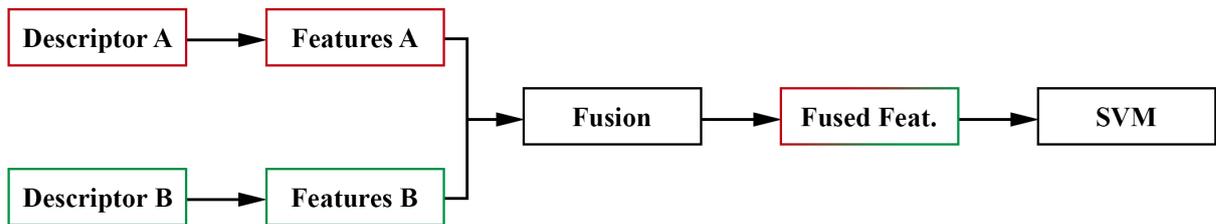


Figure 4.8: Overview of Early Fusion.

4.2.3 Convolutional Neural Networks

The use of neural networks was applied only in tests performed with Dataset II. This second version has already a significantly larger amount of samples than the first one, even though a data augmentation technique was applied for some tests.

The technique consisted of adding up to three random variation in training images including: rotation, translation, brightness, blur, saturation, sharpening. Figure 4.9 shows examples after the data augmentation, in which an original training sample generates 14 new samples increasing the training set.

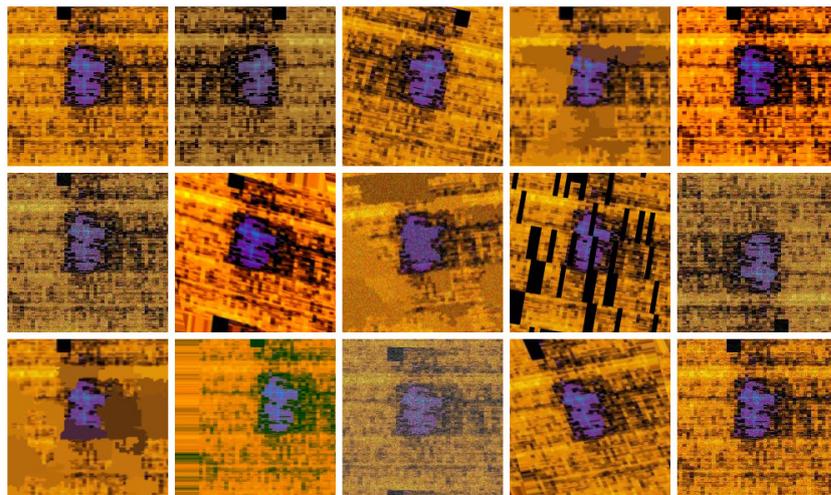


Figure 4.9: Examples of Data Augmentation.

The convolutional neural network model used in this Image Patches Approach was proposed by Roecker et al. (2018). It was designed with principles to simplify the model and use low-resolution images, useful to development of systems with limited resources.

Table 4.1 describes the model architecture which can also be observed in Figure 4.10. It received as input an RGB image, in which the input passed through a stack of convolutional layer with variable number of filters 3×3 . The convolution output was set up with a leaky rectifier activation function (LReLU). Then a spacial pooling performed a maximum-value subsampling with a 2×2 window and stride of 2.

Fully-connected layers have structure similar to multilayer perceptron (MLP) receiving the previously stages results as input (Roecker et al., 2018). The only existing difference of the model used in this work for the one presented by Roecker et al. (2018), is the last layer which does the classification and has 2 units. In this output a softmax (normalized exponential function) was applied to result into probabilities. The best results were using a learning rate of 0.0001 and the batch size was 50. These tests were run for 100 epochs.

Table 4.1: Architecture of Roecker et al. (2018) model

#	Layer	Parameters	Stride (x, y)
1	Convolutional	$3 \times 3 \times 32$	(1, 1)
2	Convolutional	$3 \times 3 \times 32$	(1, 1)
3	Pooling	2×2	(2, 2)
4	Convolutional	$3 \times 3 \times 32$	(1, 1)
5	Convolutional	$3 \times 3 \times 32$	(1, 1)
6	Pooling	2×2	(2, 2)
7	Fully-connected	512	-
8	Fully-connected	512	-
9	Fully-connected	2	-
10	Softmax	2	-

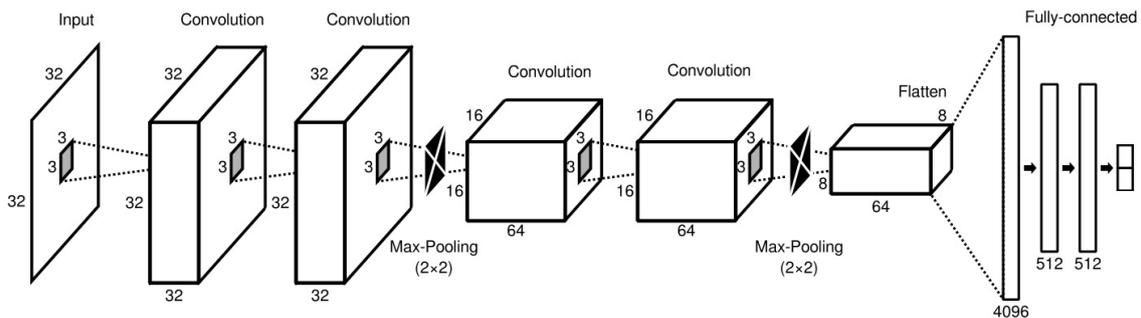


Figure 4.10: Overview of Roecker et al. (2018) model.

4.2.4 Proposed Metrics

Using the idea of automatically generated patches of an image, the same release may be present in more than one patch according to its size and the amount of DA released. Since it is important to identify this release within the labeled interval, regardless of whether you have one or more patches at that moment, they were all considered as only one sample. Therefore, for example, even if there were 10 patches, and all 10 were correctly classified, it was considered as 1 hit. It is enough that one of the patches of this interval is correctly classified. If the 10 are incorrectly classified, it was considered one miss classification.

However, false positives are not desirable. Thus, all patches that do not contemplate a DA release were analyzed as individual samples. It was considered a hit if it was classified as Non-release (TN). Using this proposed strategy, the correct answers were counted and the f-measure were calculated. The f-measure is an important metric to analyze this type of problem, since the automatic dataset will not have both balanced classes.

4.3 Combined Approach

The second approach follows the same methodology of representation and classification used in the extracted patches approach, however it also used full images. Since a training set containing complete samples with both dopamine release and non-release was required, this approach was investigated only with the more complete dataset.

Figure 4.11 shows an overview of this approach, which contains two main steps: the original images approach and the image patches approach. In the first main step, features was extracted from the full images and then classified. In addition to the original samples, two zoning variations were tested (Figure 4.12). The first one uses only the common region of phasic dopamine release, and the second one also adds a region of the top of the original sample (between the pixel 0 and 90 of the y-axis), which in some releases has visual information that could be important for extracting features.

In this part of the approach, besides the techniques of representation of the previous approach, some experiments were also carried out using the Inception v3 network described in the Chapter 2.5. The RGB input images were resized to $299 \times 299 \times 3$ which is the default value of the Inception v3 model used. For these tests, the transfer learning technique was applied, using the pre-trained parameters on ImageNet dataset (Russakovsky et al., 2015). The only altered layer of the original network was the end of the fully-connected layers, so that it had the output for our two classes.

In transfer learning, we reused these transferred weights for the feature extraction layers, which is the most complex part of the model. Only the classification part (fully-connected layers) was re-trained using our dataset as input. The best results were obtained using a learning rate of 0.01 and the batch size was 100. These tests were run for 15,000 iterations.

In the second main step, once it was decided if an input contains a DA release or not, patches were extracted from those images classified as DA Release (Class 1). The patches will be classified exactly as the first approach, with the exception of metrics, and this allows even the wrong decisions to have the classification reviewed.

As it is assumed that each image contains only one release of dopamine, the classification of these patches allows a more precise decision of its location. If all patches from a single image are classified as Class 2 (non-release), the image from which they were extracted and reclassified to that other class. Otherwise, it means that at least one patch has been identified as containing the DA release and, if correct, the image itself was considered as a hit. With this metric it was possible to calculate hits and errors per image, keeping the dataset analysis balanced and providing results of possible dopamine release images as well as most precise regions within them.

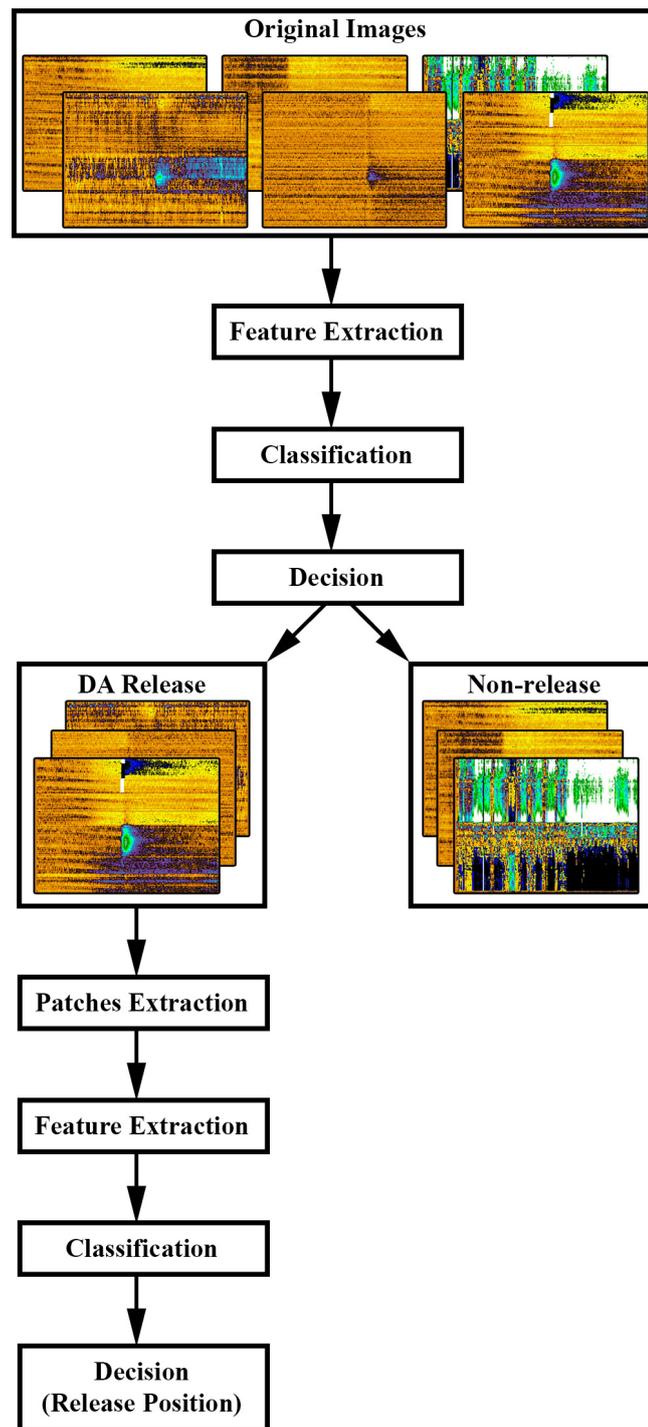


Figure 4.11: An overview of the combined approach.

4.4 Object Detection

Finally, some experiments based on the Combined Approach were performed using YOLOv3. YOLO allows the creation of an end-to-end system, in which a single architecture allows the detection and classification of objects. In this way it was possible to use original images as input and provide an accurate result of the location of the dopamine release in the sample, without the necessity of the two steps with patches extraction (Figure 4.13).

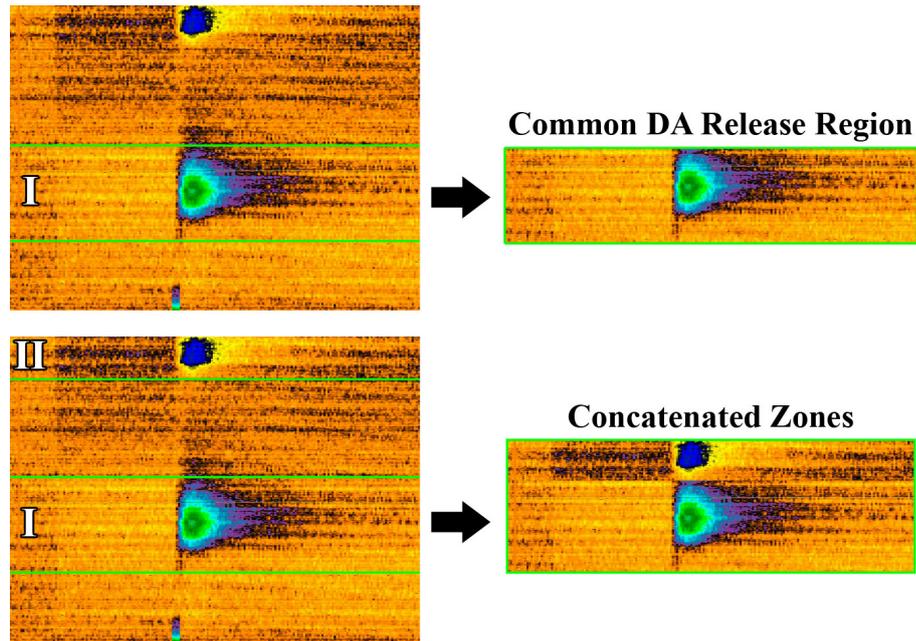


Figure 4.12: Examples of the original image and two zoning variation.

This approach is based on object detection, which has been used successfully in different applications (Zhou et al., 2016; Laroca et al., 2018; Severo et al., 2018). YOLO identifies the Class 1 (phasic dopamine release) in the input samples. And, unlike all other approaches, there will be only one class to be identified. When it is not detected, then it is assumed that there is no release of dopamine (Class 2).

The DA release labels were converted to the YOLO pattern of bounding boxes. Some training sessions were carried out with the files of the DA images, and others with all the images. The evaluation metrics are similar to the previous one. If a dopamine release is identified in the correct location, compared with the original labels, it is already considered a hit. If there is any identification in an image of Class 2, it will be considered a miss classification.

Two models were used: Yolo and Tiny Yolo with weights pre-trained on ImageNet, both in version 3. From its default settings only the input resolution size and the output classes has been changed. Different tests were performed, some using the default input size of 416×416 and others using its variable of random resize. Since the version 2 model, the convolutional and pooling layers can be resized on the fly, it means that instead of using only one input size, for each 10 iterations, the network size will be randomly resized (input and output) to size between 320×320 and 608×608 . The best results were obtained using a learning rate of 0.001 and the batch size was 64 and 8 subdivision. These tests were run for 25,000 iterations.

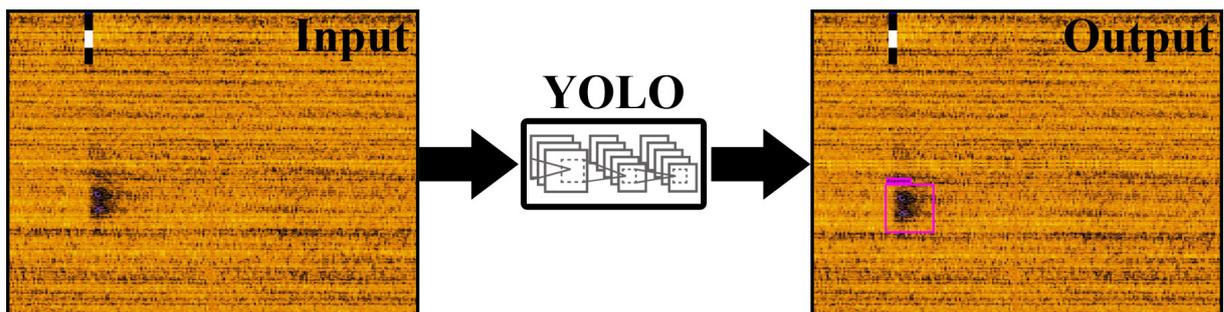


Figure 4.13: Simplified YOLO overview.

Chapter 5

Results

This chapter presents the best results obtained in tests, in which different approaches were used for the construction of classifiers. Some experiments were also performed with a combination of predictions or approaches allowing comparisons and analysis of these results. The results of tests using a balanced dataset, with the same number of images in the two classes, are represented by Accuracy, and the tests with different numbers by the F-measure.

5.1 Dataset I

The experiments using the initial version of the dataset were performed only following the image patches approach, and initially manual images were tested to analyze the best parameters.

5.1.1 Manually Extracted Patches

The training and testing set used was composed of manually generated patches from the original images. As described in the Chapter 4 for each testing fold, the other two were used as training.

5.1.1.1 Texture Descriptors

Table 5.1 summarizes the best results obtained using the features extracted with $LBP_{8,2}$ and $RLBP_{8,2}$ descriptors and SVM as classifier. On the other hand, the Table 5.2 presents the results using different window size parameters of the LPQ.

It is possible to notice that the LBP and RLBP results are quite similar, but the tests using LPQ features showed a better performance. The best results were achieved with LPQ (window size 9) and 120×120 patches, the accuracy obtained in the best case was 95.96%.

Table 5.1: Results obtained with LBP/RLBP and SVM using manually extracted patches

Patch Size	Descriptor	Accuracy (%)
100×100	LBP	92.63 ± 2.27
100×100	RLBP	93.68 ± 0.86
120×120	LBP	92.63 ± 2.58
120×120	RLBP	92.63 ± 1.72
150×150	LBP	91.40 ± 1.08
150×150	RLBP	91.23 ± 1.63

Table 5.2: Results obtained with LPQ and SVM using manually extracted patches

Patch Size	Window Size	Accuracy (%)
100 × 100	3	92.28 ± 1.08
100 × 100	5	94.74 ± 0.43
100 × 100	7	95.44 ± 0.50
100 × 100	9	94.91 ± 0.50
100 × 100	11	94.21 ± 1.14
120 × 120	3	94.91 ± 1.08
120 × 120	5	94.91 ± 0.50
120 × 120	7	95.44 ± 0.66
120 × 120	9	95.96 ± 0.25
120 × 120	11	95.61 ± 0.66
150 × 150	3	93.68 ± 0.43
150 × 150	5	95.09 ± 0.25
150 × 150	7	93.33 ± 2.03
150 × 150	9	95.26 ± 2.27
150 × 150	11	94.56 ± 1.08

5.1.1.2 Early Fusion

The Early Fusion method was used for SVM classifications presented in the Table 5.3. The classification was performed using features of the best LPQ parameter, for each patch size, concatenated with LBP or RLBP features, according the methodology presented in the subsection 4.2.1. The results got a little improvement and the best accuracy (96.67%) was a result of the LBP and LPQ features extracted from 120 × 120 patches.

Table 5.3: Results obtained with Early Fusion using manually extracted patches

Patch Size	Descriptor 1	Descriptor 2	Accuracy (%)
100 × 100	LBP _{8,2}	LPQ ₇	96.14 ± 0.66
100 × 100	RLBP _{8,2}	LPQ ₇	95.79 ± 0.43
120 × 120	LBP_{8,2}	LPQ₉	96.67 ± 0.50
120 × 120	RLBP _{8,2}	LPQ ₉	96.32 ± 0.43
150 × 150	LBP _{8,2}	LPQ ₉	94.74 ± 1.14
150 × 150	RLBP _{8,2}	LPQ ₉	94.91 ± 0.99

5.1.1.3 Late Fusion

Finally, a Late Fusion of SVM predictions was also performed. The set with patches with resolution of 120 × 120 pixels presented good performances in all previously experiments and it was chosen for these tests. Using either the maximum, sum or product rule the accuracy and f-measure was the same for this specific problem, which has only two classes. The Table 5.4 summarizes the results of the predictions fusion using sum rule.

These results show that the combination of even weaker classifiers can provide better results. Such a fusion of predictions provided the combination of complementary information, even when errors occur. The accuracy of 97.37% was the best obtained in all preliminary

Table 5.4: Results obtained with Late Fusion using manually extracted patches

Prediction 1	Prediction 2	Accuracy (%)
LBP _{8,2}	LPQ ₃	97.02 ± 0.50
RLBP_{8,2}	LPQ₃	97.37 ± 1.14
LBP _{8,2}	LPQ ₅	94.91 ± 1.38
RLBP _{8,2}	LPQ ₅	95.79 ± 1.14
LBP _{8,2}	LPQ ₇	95.79 ± 0.74
RLBP _{8,2}	LPQ ₇	96.49 ± 0.89
LBP _{8,2}	LPQ ₉	95.61 ± 0.25
RLBP _{8,2}	LPQ ₉	96.49 ± 0.89
LBP _{8,2}	LPQ ₁₁	96.49 ± 0.50
RLBP _{8,2}	LPQ ₁₁	97.19 ± 0.50

experiments using manually extracted patches, being the result of the combination of a SVM classifier with RLBP features and the classifier with LPQ (window size 3).

5.1.2 Automatically Extracted Patches

Since the preliminary tests with 120×120 manually extracted patches showed good results, the following experiments were performed using 120×120 automatically extracted patches as testing set. Table 5.5 summarizes the results obtained using texture descriptors, while Table 5.6 shows the results using early and late fusion (sum rule).

Table 5.5: Results with texture descriptors and SVM using automatically extracted patches

Patch Size	Descriptor	F-Measure(%)
120×120	LBP _{8,2}	62.32 ± 0.43
120×120	RLBP _{8,2}	69.85 ± 0.40
120×120	LPQ₉	77.23 ± 1.20

Table 5.6: Results obtained with Early and Late Fusion using automatically extracted patches

Fusion	Descriptor 1	Descriptor 2	F-Measure (%)
Early Fusion	LBP _{8,2}	LPQ ₉	72.69 ± 0.50
Early Fusion	RLBP _{8,2}	LPQ ₉	72.46 ± 0.84
Late Fusion	LBP _{8,2}	LPQ ₉	72.55 ± 0.37
Late Fusion	RLBP_{8,2}	LPQ₉	75.47 ± 0.69

One can note that in these cases the LPQ performance was superior to the other descriptors, and even the combinations of feature vectors or predictions did not improve the f-measure. The concatenation of two features array not necessary will improve the results, since the complexity will rise too.

The best f-measure using automatic patches was 77.23%, that is justified by the different numbers of samples between both classes, there are significantly more patches without DA release. Despite this low f-measure rate the Dopamine Release Class accuracy rate is about 98%, and it is very important not to miss releases identification on the recordings.

Figure 5.1 shows the ROC curves of the best results of each approach, where we can see how the ideal curve would be if all the patches could be extracted exactly in the most correct position. Such ROC curves will serve as a comparison for the next experiments.

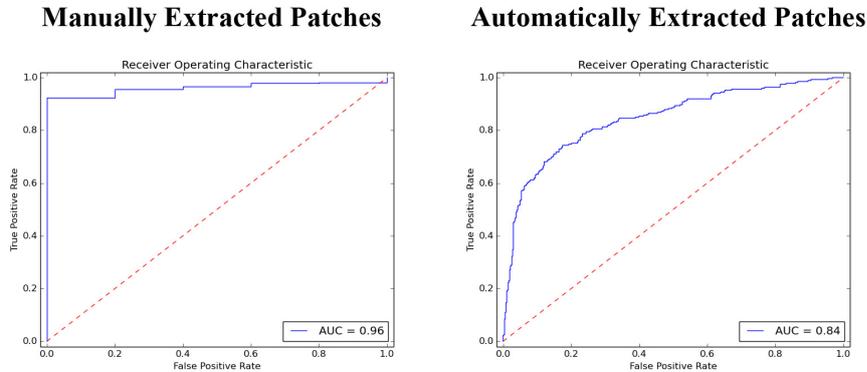


Figure 5.1: ROC curves of the best Dataset I tests.

5.2 Dataset II

The second and more complete version of the dataset allowed exploring new approaches. As in the previous version, the initial tests were exploring texture descriptors and their best parameters. However, the samples used as training and testing set were the original images, without patches extraction.

5.2.1 Texture Descriptors

There are three different training sets, each generated with different background as presented in Chapter 4.1.2. Table 5.7 shows the results obtained by training and testing full images with a specific background and different texture descriptors, but for each dataset tests were also performed with their two zoning variations: the first containing only the common DA release region and the second containing two regions of the image concatenated. The results obtained with these tests are shown in Table 5.8.

It was possible to notice that of all the varied images, the best results were always using the background position A and C. The best result in these preliminary tests was using the common DA release region, the background position C dataset and LPQ as descriptor, with an accuracy of 75.72%.

In addition to the experiments performed separately, tests were also done combining the training sets of different background positions. Thus, the respective folds of positions A, B and C were grouped, keeping the same original images together. The testing sets were explored individually as shown in the Table 5.9. The best result was 73.88% and it was not better than the previously one (75.72%) besides having a higher cost for the training.

Using the best results with background position A and C, late fusion tests were also performed according to the Table 5.10. As before, the fusions showed the same results for the different fusion rules, so only the results with the sum rule were presented. The accuracy of 79.15% was resulted from the combination of LPQ and SVM predictions, and it was the best result obtained with texture descriptors and these samples.

Table 5.7: Results from Dataset II obtained with texture descriptors

Background Position	Descriptor	Accuracy(%)
Background A	LBP_{8,2}	73.77 ± 0.31
Background A	RLBP _{8,2}	73.28 ± 0.64
Background A	LPQ ₃	73.44 ± 0.78
Background A	LPQ ₅	72.39 ± 1.39
Background A	LPQ ₇	71.19 ± 1.29
Background A	LPQ ₉	72.09 ± 0.56
Background B	LBP _{8,2}	69.20 ± 0.58
Background B	RLBP _{8,2}	69.10 ± 1.89
Background B	LPQ ₃	72.04 ± 2.72
Background B	LPQ₅	72.39 ± 0.86
Background B	LPQ ₇	70.35 ± 0.74
Background B	LPQ ₉	69.15 ± 0.88
Background C	LBP _{8,2}	70.15 ± 1.52
Background C	RLBP _{8,2}	70.70 ± 1.85
Background C	LPQ₃	73.93 ± 0.31
Background C	LPQ ₅	73.13 ± 1.29
Background C	LPQ ₇	71.09 ± 0.73
Background C	LPQ ₉	68.36 ± 2.76

Table 5.8: Tests using different zoned images

Zoning Type	Background	Descriptor	Accuracy(%)
DA Release Region	A	LBP _{8,2}	74.43 ± 0.74
DA Release Region	A	RLBP _{8,2}	73.43 ± 1.69
DA Release Region	A	LPQ ₃	75.32 ± 1.65
DA Release Region	B	LBP _{8,2}	67.16 ± 1.64
DA Release Region	B	RLBP _{8,2}	66.67 ± 1.22
DA Release Region	B	LPQ ₃	69.75 ± 0.51
DA Release Region	C	LBP _{8,2}	70.40 ± 0.51
DA Release Region	C	RLBP _{8,2}	70.30 ± 0.88
DA Release Region	C	LPQ₃	75.72 ± 0.60
Concatenated Zones	A	LBP _{8,2}	73.38 ± 1.11
Concatenated Zones	A	RLBP _{8,2}	73.08 ± 0.88
Concatenated Zones	A	LPQ ₃	75.32 ± 0.61
Concatenated Zones	B	LBP _{8,2}	67.95 ± 1.78
Concatenated Zones	B	RLBP _{8,2}	64.79 ± 1.30
Concatenated Zones	B	LPQ ₃	68.36 ± 2.76
Concatenated Zones	C	LBP _{8,2}	70.95 ± 0.37
Concatenated Zones	C	RLBP _{8,2}	70.20 ± 0.14
Concatenated Zones	C	LPQ ₃	75.52 ± 0.56

To finalize the investigations with texture descriptors, the same tests of Dataset I automatically extracted patches were performed. In Table 5.11 it is possible to observe that the best Dataset II automatically extracted patches result was of 74.78%, inferior to the 77.23% of the first one. And as with Dataset I, early and late fusions did not improve results (Table 5.12).

Table 5.9: Results obtained using a combined training set

Background Position	Descriptor	Accuracy(%)
Background A	LBP _{8,2}	73.63 ± 2.44
Background A	RLBP _{8,2}	73.73 ± 0.76
Background A	LPQ ₃	73.03 ± 1.01
Background B	LBP _{8,2}	68.16 ± 1.62
Background B	RLBP _{8,2}	68.31 ± 1.34
Background B	LPQ ₃	70.05 ± 1.04
Background C	LBP _{8,2}	69.20 ± 0.61
Background C	RLBP _{8,2}	70.40 ± 0.86
Background C	LPQ₃	73.88 ± 1.71

Table 5.10: Results obtained using Late Fusion of background position A and C predictions

Fusion Rule	Descriptor A	Descriptor C	Accuracy(%)
Sum Rule	LBP _{8,2}	LBP _{8,2}	76.97 ± 0.92
Sum Rule	LBP _{8,2}	RLBP _{8,2}	76.32 ± 0.70
Sum Rule	LBP _{8,2}	LPQ ₃	79.10 ± 1.44
Sum Rule	RLBP _{8,2}	LBP _{8,2}	76.87 ± 0.49
Sum Rule	RLBP _{8,2}	RLBP _{8,2}	76.07 ± 1.06
Sum Rule	RLBP _{8,2}	LPQ ₃	79.05 ± 0.91
Sum Rule	LPQ ₃	LBP _{8,2}	76.42 ± 1.93
Sum Rule	LPQ ₃	RLBP _{8,2}	76.77 ± 0.63
Sum Rule	LPQ₃	LPQ₃	79.15 ± 0.67

Table 5.11: Results using texture descriptors and automatically extracted patches from Dataset II

Patch Size	Descriptor	F-Measure(%)
200 × 200	LBP_{8,2}	74.78 ± 0.93
200 × 200	RLBP _{8,2}	72.42 ± 0.85
200 × 200	LPQ ₃	71.86 ± 1.09

Table 5.12: Results obtained using Early and Late Fusion and automatically extracted patches from Dataset II

Fusion	Descriptor 1	Descriptor 2	F-Measure (%)
Early Fusion	LBP_{8,2}	LPQ₃	69.05 ± 0.41
Early Fusion	RLBP _{8,2}	LPQ ₃	68.09 ± 1.21
Late Fusion	LBP _{8,2}	LPQ ₃	67.54 ± 1.16
Late Fusion	RLBP _{8,2}	LPQ ₃	67.32 ± 1.29

5.2.2 Convolutional Neural Networks

Dataset II brought greater complexity, and many new images without dopamine release. In these images there are the most diverse variations and noises. Despite the greater challenge, it was possible to explore new approaches such as the use of Convolutional Neural Networks.

The first experiments were using the Inception v3 and Roecker et al. (2018) models. As shown in the previous Chapter 4.3, the Inception tests were performed using RGB input images (resized to $299 \times 299 \times 3$) and pre-trained weights on ImageNet. Table 5.13 shows the results

using learning rate of 0.01 and the batch size of 100. These tests were run for 15,000 iterations and the best accuracy was 95.72% using data augmentation in the original images training set, which was already much better than all results obtained with texture descriptors.

Table 5.13: Results obtained using Inception v3

Zoning Type	Data Augmentation	Accuracy(%)
Original Samples	No	95.67 \pm 0.53
Original Samples	Yes	95.72 \pm 0.51
DA Release Region	No	95.07 \pm 0.85
DA Release Region	Yes	95.17 \pm 0.51
Concatenated Zones	No	94.53 \pm 0.19
Concatenated Zones	Yes	94.98 \pm 0.39

Unlike the Inception tests, no pre-trained parameters were used for the Roecker et al. (2018) model. Each of the training sets were performed for 100 epochs and initially only images generated using the background position A were used. These initial tests without data augmentation expressed in Table 5.14 served to test different input sizes. From these results, only the best parameters were used for the tests with the other training sets and using data augmentation (Table 5.15). An accuracy of 97.31% was the best result obtained, using a training set of zoned images in the common DA release region and generated by background B.

In spite of the excellent result obtained, when we analyzed the predictions, it was possible to observe that both false positives and false negatives had very high values, and this inflicted in the ROC curve obtained (Figure 5.2) that can be compared with a more optimal result of Figure 5.1. Perhaps for this reason, no predictions fusion tests showed improvements, no longer being an interesting approach to be used.

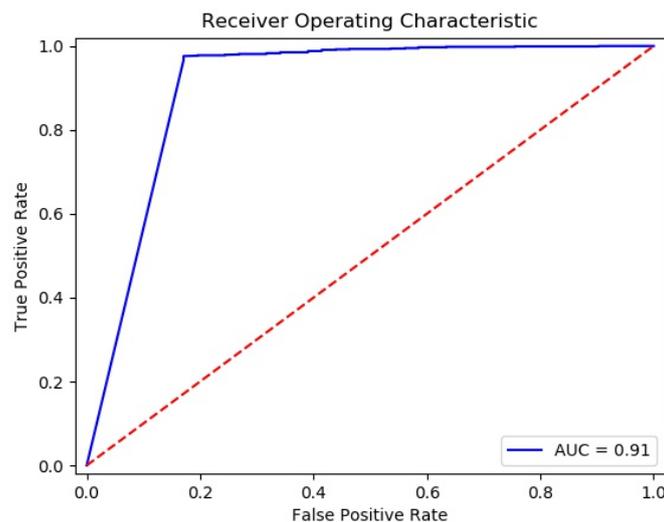


Figure 5.2: ROC curve of the best test using entire images approach.

Tests using automatically extracted patches approach and this model with data augmentation also got good results. Table 5.16 shows that the best result had an f-measure of 95.64%. However, for 1005 DA release, there are more than 9000 patches without release, generating hundreds of false positives.

Table 5.14: Results obtained using Roecker et al. (2018) CNN model

Input Size	Zoning Type	Accuracy(%)
350 × 262	Original Samples	85.87 ± 2.24
350 × 80	DA Release Region	92.64 ± 0.14
350 × 116	Concatenated Zones	84.28 ± 5.21
262 × 197	Original Samples	83.68 ± 3.05
262 × 60	DA Release Region	94.13 ± 1.99
262 × 87	Concatenated Zones	85.07 ± 4.23
175 × 131	Original Samples	91.14 ± 1.49
175 × 40	DA Release Region	94.73 ± 0.51
175 × 58	Concatenated Zones	89.35 ± 2.31
88 × 66	Original Samples	91.01 ± 1.75
88 × 20	DA Release Region	88.86 ± 4.58
88 × 29	Concatenated Zones	85.97 ± 3.92

Table 5.15: Results obtained with Data Augmentation using Roecker et al. (2018) CNN model

Background Position	Zoning Type	Accuracy(%)
Background A	Original Samples (175 × 131)	96.47 ± 0.61
Background A	DA Release Region (175 × 40)	96.47 ± 0.67
Background A	Concatenated Zones (175 × 58)	96.22 ± 1.42
Background B	Original Samples (175 × 131)	96.72 ± 0.12
Background B	DA Release Region (175 × 40)	97.31 ± 0.64
Background B	Concatenated Zones (175 × 58)	96.62 ± 0.55
Background C	Original Samples (175 × 131)	94.78 ± 1.06
Background C	DA Release Region (175 × 40)	96.22 ± 0.14
Background C	Concatenated Zones (175 × 58)	95.77 ± 0.78

Table 5.16: Results obtained using automatically extracted patches and CNN

Background Position	Patch Size	F-measure(%)
Background A	200 × 200	94.44 ± 1.28
Background B	200 × 200	85.25 ± 1.34
Background C	200 × 200	95.64 ± 1.04

The alternative to this problem was to extract the patches directly from the images classified as positive of the best experiment from Table 5.15, and then classify these patches. Following the methodology presented in the previous chapter, this approach allowed even a revision of the false positives: Considering the 41 FP there were only 17 now, and the final accuracy of the complete process was **98.36%** ± 2.39.

This classifier was very versatile in identifying phasic dopamine release in different situations: in small amounts or in the middle of noises. Figure 5.3 shows examples classified as Positive, where a few samples containing noise/artifacts from recording are still missclassified (FP).

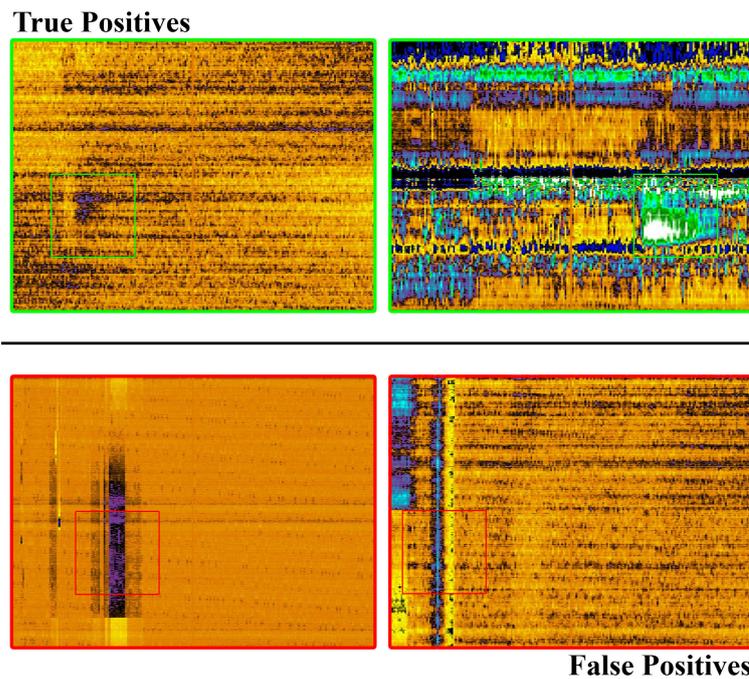


Figure 5.3: Example of True Positives and False Positives.

5.2.2.1 YOLO

Despite excellent results already obtained, the previous approach requires two different steps of training and classification with neural networks. Using YOLO, it was possible to directly identify the original image without the need to extract patches.

Table 5.17 shows the accuracy obtained using the Tiny model and RGB inputs with size of 416×416 . The best result was 96.82% with a training composed only of phasic dopamine release samples, and background position A. These same parameters were replicated to the complete YOLOv3 model, which is larger and has a higher computational cost. However, the accuracy did not improve, being $96.57\% \pm 0.68$.

The last two tests performed were obtained again using the best parameters, but using the random resize. New anchors were calculated using kmeans in the Dataset II to result in the same default number of the model: 6. The result was $97.56\% \pm 0.49$, but the best accuracy was **$97.66\% \pm 0.67$** using the default anchors.

YOLO only identifies the phasic dopamine release class, but it is possible to analyze true and false positives predictions. As in the results of the other model, false positives also have a high value, resulting in the ROC curve of Figure 5.4. These high values present in CNN also made it impossible to apply a good threshold for rejection, and thus try to improve the final decision. Like the other classifier, YOLO also makes correct DA identification in the most varied situations, and in Figure 5.5 we can observe some examples of True Positives and False Positives.

Table 5.17: Results obtained using Tiny YOLOv3 model

Training Set	Background Position	Accuracy(%)
Only DA release	Background A	96.82 ± 1.13
Only DA release	Background B	95.37 ± 1.08
Only DA release	Background C	96.27 ± 0.21
Complete	Background A	95.82 ± 1.52
Complete	Background B	94.63 ± 1.84
Complete	Background C	92.19 ± 0.79

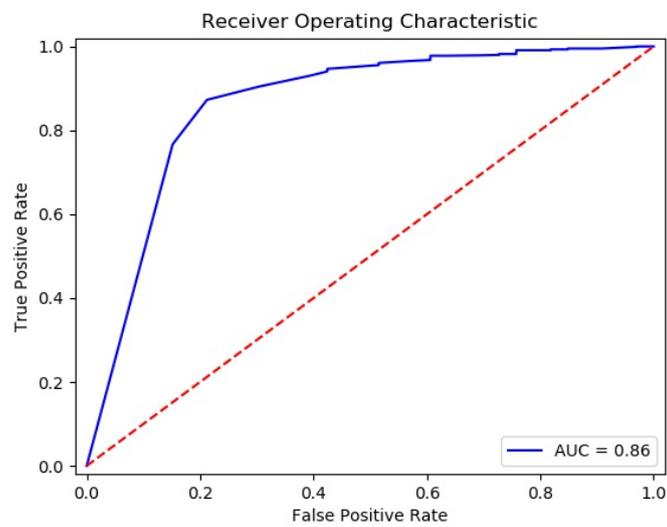
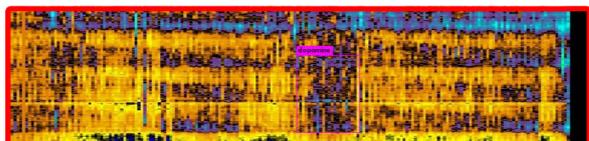
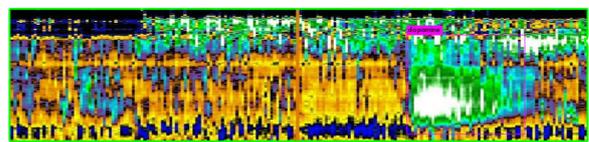
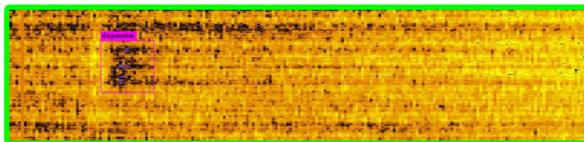


Figure 5.4: ROC curve of the best test using YOLO.

True Positives



False Positives

Figure 5.5: Example of True Positives and False Positives using YOLO.

Chapter 6

Final Considerations and Future Works

In current literature, pattern recognition systems using dopamine images generated from data collected with fast scan cyclic voltammetry are something new. There are many methods and parameters to explore. Thus, this work explored mainly the use of visual texture features and convolutional neural networks for the development of a classification system. The proposed methodology uses original images and patches extracted from them to construct an automatic identification system. However it was necessary to create two new dataset, which were made available to the scientific community.

Preliminary experiments were performed using a dataset composed of patches. The best manually extracted patches accuracy obtained was 97.37% using the fusion of classifiers with different texture features. However there are much more patches without DA release using automatically extracted patches, what really happens in a real life recording, and it results in a challenging problem to get high f-measure rates and minimize the false positives. The best f-measure was 77.23% using LPQ as texture descriptor.

The second version of the dataset not only brought new approaches to explore but also a way to avoid the problem of the unbalanced classes. The tests in this dataset, which contains greater diversity in phasic dopamine releases and also complete images without any release, resulted in similar values using texture descriptors and automatically extracted patches. The f-measure with automatically extracted patches from Dataset II was 74.78% using LBP. And the accuracy with entire images approach using Late Fusion of different background positions predictions was 79.15%. But by far the best results were using Convolutional Neural Networks.

The Roecker et al. (2018) model was able to classify entire images with an accuracy of up to 97.31%. By itself, this result is already very good by classifying DA release within 20 seconds of an experimental recording. But using a combined approach, extracting patches from images classified as Class 1, it was also possible to provide a more accurate result within each of the images with an accuracy of 98.31%.

The combined approach provides an excellent result, but YOLO tests resulted in an accuracy of 97.66%. Even though this value is somewhat lower, the YOLO allows the identification of phasic DA release directly from the original samples, without the need for two different training and classification steps.

This method also results in identifications in varied positions and sizes, being more versatile and accurate than patches, which are always restricted to a defined size. Thus, in a problem of identifying a substance that varies many patterns, YOLO also becomes an excellent choice for an automatic classification. Although there were visual variations in the images, there was no definitive background position that was always better than the others. It was possible to achieve great results in all of them.

In future work, we intend to continue to build a larger and more varied dataset, including other substances, to better exploit the problem and to favor the development of new classification systems. In this way, we can create classifiers for other substances, as well as different classes of dopamine release by quantity. We also intend to experiment different neural networks models and descriptors.

Bibliography

- Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041.
- Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798 – 1828.
- Bertolini, D., Oliveira, L. S., Justino, E., and Sabourin, R. (2013). Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, 40(6):2069–2080.
- Borman, R. P., Wang, Y., Nguyen, M. D., Ganesana, M., Lee, S. T., and Venton, B. J. (2017). Automated algorithm for detection of transient adenosine release. *ACS Chemical Neuroscience*, 8(2):386–393.
- Cavalin, P. and Oliveira, L. S. (2017). A review of texture classification methods and databases. *Energy (θ , D)*, 1:1.
- Cavalin, P. R., Kapp, M. N., Martins, J., and Oliveira, L. E. (2013). A multiple feature vector framework for forest species recognition. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 16–20. ACM.
- Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.
- Chen, J., Kellokumpu, V., Zhao, G., and Pietikäinen, M. (2013). RLBP: Robust local binary pattern. In *Proceedings of the British Machine Vision Conference*, pages 1–12.
- Costa, Y. M. G., Oliveira, L. S., Koerich, A. L., and Gouyon, F. (2011). Music genre recognition using spectrograms. In *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*, pages 1–4. IEEE.
- Costa, Y. M. G., Oliveira, L. S., Koerich, A. L., Gouyon, F., and Martins, J. G. (2012). Music genre classification using lbp textural features. *Signal Processing*, 92(11):2723–2737.
- Da Cunha, C., Boschen, S., Gómez-A, A., Ross, E. K., Gibson, W., Min, H. K., Lee, K. H., and Blaha, C. D. (2015). Toward sophisticated basal ganglia neuromodulation: Review on basal ganglia deep brain stimulation. *Neuroscience and Biobehavioral Reviews*, 58:186–210.
- Da Cunha, C., Gómez-A, A., and Blaha, C. D. (2012). The role of the basal ganglia in motivated behaviour. *Reviews in the Neurosciences*, 28(5-6):747–767.

- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.
- Fawaz, C., Martel, P., Leo, D., and Trudeau, L.-E. (2009). Presynaptic action of neurotensin on dopamine release through inhibition of D(2) receptor function. *BMC Neuroscience*, 10:96.
- Gomez-A, A., Fiorenza, A., Boschen, S., Sugi, A., Beckman, D., Ferreira, S., Lee, K., Blaha, C., and Da Cunha, C. (2017). Diazepam inhibits electrically evoked and tonic dopamine release in the nucleus accumbens and reverses the effect of amphetamine. *ACS Chem. Neurosci.* 2017, 8(2):300–309.
- Gonzalez, R. C. and Woods, R. E. (2010). *Processamento digital de imagens*. Pearson Prentice Hall, São Paulo.
- Grace, A. A. (1995). The tonic/phasic model of dopamine system regulation: its relevance for understanding how stimulant abuse can alter basal ganglia function. *Drug and Alcohol Dependence*, 37(2):111–129.
- Haralick, R. M., Shanmugam, K., et al. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621.
- Harms, H., Gunzer, U., and Aus, H. (1985). Combined local color and texture analysis of stained cells. *Computer Vision, Graphics, and Image Processing*, 32(3):414.
- Jackowska, K. and Krysinski, P. (2013). New trends in the electrochemical sensing of dopamine. *Analytical and Bioanalytical Chemistry*, 405(11):3753–3771.
- Jhuria, M., Kumar, A., and Borse, R. (2013). Image processing for smart farming: Detection of disease and fruit grading. In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, pages 521–526. IEEE.
- Jones, D. R., Moussaoud, S., and McLean, P. (2014). Targeting heat shock proteins to modulate alpha-synuclein toxicity. *Analytical and Bioanalytical Chemistry*, 7(1):33–51.
- Khademi, A. and Krishnan, S. (2008). Medical image texture analysis: A case study with small bowel, retinal and mammogram images. In *Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on*, pages 001949–001954. IEEE.
- Kittler, J., Hatef, M., Duin, R. P., and Matas, J. (1998). On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Kumar, U., Medel-Matus, J.-S., Redwine, H. M., Shin, D., Hensler, J. G., Sankar, R., and Mazarati, A. (2016). Effects of selective serotonin and norepinephrine reuptake inhibitors on depressive- and impulsive-like behaviors and on monoamine transmission in experimental temporal lobe epilepsy. *Epilepsia*, 57(3):506–515.

- Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2018). A robust real-time automatic license plate recognition based on the yolo detector. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lucio, D. R. and Costa, Y. M. G. (2015). Bird species classification using spectrograms. In *Computing Conference (CLEI), 2015 Latin American*, pages 1–11. IEEE.
- Mäenpää, T. (2003). *The Local binary pattern approach to texture analysis: Extensions and applications*. Oulun yliopisto.
- Masek, A., Chrzescijanska, E., and Zaborski, M. (2014). Electrooxidation of morin hydrate at a pt electrode studied by cyclic voltammetry. *Food Chemistry*, 148:18–23.
- Matsushita, G. H. G., Oliveira, L. E. S., Sugi, A. H., da Cunha, C., and Costa, Y. M. G. (2018). Automatic identification of phasic dopamine release. In *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5. IEEE.
- Metz, C. E. (1978). Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier.
- Nasirzadeh, M., Khazael, A. A., and bin Khalid, M. (2010). Woods recognition system based on local binary pattern. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2010 Second International Conference on*, pages 308–313. IEEE.
- Nguyen, M. D. and Venton, B. J. (2015). Fast-scan cyclic voltammetry for the characterization of rapid adenosine release. *Computational and Structural Biotechnology Journal*, 13:47–54.
- Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.
- Ojansivu, V. and Heikkilä, J. (2008). Blur insensitive texture classification using local phase quantization. In *Image and signal processing*, pages 236–243. Springer.
- Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014:1717–1724.
- Park, J., Choi, S., Sohn, S., Kim, K.-R., and Hwang, I. S. (2014). Cyclic voltammetry on zirconium redox reactions in licl-kcl-zrcl4 at 500° c for electrorefining contaminated zircaloy-4 cladding. *Journal of The Electrochemical Society*, 161(3):H97–H104.
- Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M., and Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature*, 422(6932):614–618.
- Pujari, J. D., Yakkundimath, R., and Byadgi, A. S. (2015). Image processing based detection of fungal diseases in plants. *Procedia Computer Science*, 46:1802–1808.

- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525.
- Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *CoRR*, abs/1804.02767.
- Robinson, D. L., Howard, E. C., McConnell, S., Gonzales, R. A., and Wightman, R. M. (2009). Disparity between tonic and phasic ethanol-induced dopamine increases in the nucleus accumbens of rats. *Alcoholism: Clinical and Experimental Research*, 33(7):1187–1196.
- Roecker, M. N., Costa, Y. M. G., Almeida, J. L. R., and Matsushita, G. H. G. (2018). Automatic vehicle type classification with convolutional neural networks. In *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5. IEEE.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Sermanet, P., Chintala, S., and LeCun, Y. (2012). Convolutional neural networks applied to house numbers digit classification. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3288–3291. IEEE.
- Sermanet, P. and LeCun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2809–2813. IEEE.
- Severo, E., Laroca, R., Bezerra, C. S., Zanlorensi, L. A., Weingaertner, D., Moreira, G., and Menotti, D. (2018). A benchmark for iris location and a deep learning detector evaluation. *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7.
- Sutton, R. N. and Hall, E. L. (1972). Texture measures for automatic classification of pulmonary disease. *IEEE Transactions on Computers*, 100(7):667–676.
- Swamy, B. E. K. and Venton, B. J. (2006). Subsecond detection of physiological adenosine concentrations using fast-scan cyclic voltammetry. *Analytical Chemistry* 2007, 79(2):744–750.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.
- Tou, J. Y., Lau, P. Y., and Tay, Y. H. (2007). Computer vision-based wood recognition system. In *Proceedings of International workshop on advanced image technology*. Citeseer.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag.
- Vargas, A. C. G.; Paes, A. V. C. N. (2016). Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. *Conference on Graphics, Patterns and Images*, 29:1–4.

- Wanat, M. J., Willuhn, I., Clark, J. J., and Phillips, P. E. M. (2009). Phasic dopamine release in appetitive behaviors and drug addiction. *Current drug abuse reviews*, 2(2):195–213.
- Yorgason, J. T., España, R. A., and R.Jones, S. (2011). Demon voltammetry and analysis software: analysis of cocaine-induced alterations in dopamine signaling using multiple kinetic measures. *Journal of Neuroscience Methods*, 202(2):158–164.
- Zhou, Y., Nejati, H., Do, T.-T., Cheung, N.-M., and Cheah, L. (2016). Image-based vehicle analysis using deep neural network: A systematic study. In *Digital Signal Processing (DSP), 2016 IEEE International Conference on*, pages 276–280. IEEE.
- Zottesso, R. H. D., Matsushita, G. H. G., Lucio, D. R., and Costa, Y. M. G. (2016). Automatic segmentation of audio signal in bird species identification. In *Computer Science Society (SCCC), 2016 35th International Conference of the Chilean*, pages 1–11. IEEE.