MASARYKOVA UNIVERZITA
Přírodovědecká fakulta
Ústav teoretické fyziky a astrofyziky

BAKALÁŘSKÁ PRÁCE

Brno 2017 Matej Kosiba

MASARYKOVA UNIVERZITA

Přírodovědecká fakulta

Ústav teoretické fyziky a astrofyziky

Klasifikace galaxií pomocí strojového učení

Bakalářská práce

Matej Kosiba

Vedoucí bakalářské práce:
doc. Mgr. Norbert Werner, Ph.D.

Konzultant:
doc. Ing. Lukáš Burget, Ph.D.

Brno 2017

Bibliografický záznam

Autor:	Matej Kosiba Přírodovědecká fakulta, Masarykova univerzita Ústav teoretické fyziky a astrofyziky
Název práce:	Klasifikace galaxií pomocí strojového učení
Studijní program:	Fyzika
Obor:	Astrofyzika
Vedoucí práce:	doc. Mgr. Norbert Werner, Ph.D.
Konzultant:	doc. Ing. Lukáš Burget, Ph.D.
Akademický rok:	2016/2017
Počet stran:	viii + 58
Klíčová slova:	galaxie, morfologie, průzkum, klasifikace obrázků, neuronová síť

Bibliographic record

Author:	Matej Kosiba Faculty of Science, Masaryk University Department of Theoretical Physics and Astrophysics
Title of Thesis:	Classification of galaxies with machine learning application
Degree Programme:	Physics
Field of study:	Astrophysics
Supervisor:	doc. Mgr. Norbert Werner, Ph.D.
Advisor:	doc. Ing. Lukáš Burget, Ph.D.
Academic Year:	2016/2017
Number of Pages:	viii + 58
Keywords:	galaxy, morphology, survey, image classification, neural network

Abstrakt

Morfologie galaxií je rozhodující informací při studiu jejich formování a evoluce. Množství dat, které bylo pořízeno celými prohlídkami před deseti lety je v současnosti pořizováno v průběhu jediné noci. Očekávané veliké prohlídky, které budou dokončeny v blízké budoucnosti, budou produkovat ješte mnohem více dat mnohem lepšího rozlišení. Toto nesmírné množství dat již nemůže být klasifikováno lidmi, takže vývoj a použití automatického zpracování je logickým a nevyhnutelným krokem. Naše práce vychází z výherního řešení *„Galaxy Challenge"*, kterým byla konvoluční neuronová síť. Tuto síť jsme modifikovali, aby byla lépe přizpůsobena klasifikování galaxií po jejím natrénování na větším množství anotovaných galaxií, než které bylo poskytnuto v *„Galaxy Challenge"*. Naše síť správně klasifikovala více než 85 % eliptických a 87 % spirálních galaxií z našeho testovacího datasetu. Největší vědecký přínos naší práce vidíme v budoucím zkoumání závislostí mezi vztahem morfologie galaxií s prostředím v kterém se nachádzejí a červeným posunem, které bude potřebovat ohromné množství klasifikovaných objektů.

Abstract

Morphology of galaxies is a crucial information for studying their formation and evolution. Data volumes of entire survey's decade ago are being produced in a single night. Large surveys which are due to be released in near future are expected to produce even more data and of much higher quality. This vast amount of data can no longer be classified by humans, so development and usage of artificial programs is logical and necessary step. We present convolution neural network as a solution to galaxy classification. Our work is based on winning solution of Galaxy Challenge, which was a convolutional neural network. We modified this network to be more suitable for future training on larger dataset of classified galaxies, than that provided in Galaxy Challenge. Our neural network correctly classified more than 85 % elliptical and 87 % spiral galaxies of our testing dataset. The biggest scientific impact of our work lies in future determination of the galaxy morphology density relation as a function of redshift, which requires vast numbers of objects being classified.

MASARYKOVA UNIVERZITA
Přírodovědecká fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Akademický rok: 2016/2017

Ústav: Ústav teoretické fyziky a astrofyziky

Student: Matej Kosiba

Program: Fyzika

Obor: Astrofyzika

Ředitel *Ústavu teoretické fyziky a astrofyziky* PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje bakalářskou práci s názvem:

Název práce: Klasifikácia galaxií pomocou strojového učenia

Název práce anglicky: Galaxy Classification using Machine Learning

Oficiální zadání:
Large sky surveys such as the ongoing Dark Energy Survey (DES) and the upcoming Large Synoptic Sky Survey (LSST) will generate enormous amounts of high quality imaging data of galaxies. These data, in many different filters, will provide invaluable information on galaxy evolution. The exploitation of these vast datasets will require advanced machine learning techniques. The student will build and test a convolutional neural networks based technique to classify the morphological types of galaxies in sky survey images automatically. The neural networks will be trained on images classified in the galaxy ZOO citizen science project and the technique will be tested on SDSS, Pan-Starrs, and/or DES images.

Jazyk závěrečné práce: anglický

Vedoucí práce: doc. Mgr. Norbert Werner, Ph.D.

Datum zadání práce: 7. 11. 2016

V Brně dne: 24. 11. 2016

Souhlasím se zadáním (podpis, datum):

Kosiba 7.12.2016
Matej Kosiba
student

doc. Mgr. Norbert Werner, Ph.D.
vedoucí práce

prof. Rikard von Unge, Ph.D.
ředitel Ústavu teoretické fyziky a astrofyziky

Poděkování

V prvom rade by som sa chcel poďakovať mojim vedúcim, Norbertovi Wernerovi a Lukášovi Burgetovi. Norbertovi Wernerovi najmä za dôveru, ktorú do mňa vložil, keď sa so mnou rozhodol spolupracovať na bakalárskej práci na tému, o ktorej som z počiatku nemal žiadne bližšie informácie. Za vytvorenie tohto projektu, veľkú motiváciu, veľmi priateľský prístup, odborné rady a v neposlednej rade za jeho vynikajúce nápady. Lukášovi Burgetovi sa chcem poďakovať hlavne za nesmierne množstvo času, ktorý mi venoval, za to, že ma učil programovať a pracovať v operačnom systéme Linux v podstate od úplných začiatkov, taktiež za veľmi priateľský prístup a veľmi podrobnú pomoc s postupom písania práce. Rád by som sa poďakoval Michalovi Hradišovi za vedenie semináru o neurónových sietiach, kde mi poskytol cenné rady a priestor na diskusiu ohľadom mojej bakalárskej práce.

Veľké poďakovanie patrí mojej spolužiačke Anne Juráňovej, ktorá mi nesmierne pomáha počas celého môjho štúdia a v časoch najhorších ma vie vždy motivovať. Vždy bola pre mňa nesmiernou oporou. Ďalej sa chcem poďakovať mojej priateľke Eve Šikutovej najmä za jej obrovskú psychickú podporu, vytvorenie príjemného domáceho prostredia a pochopenia, ktoré rozhodne mnohokrát nebolo jednoduché. Xene Hamas sa chcem poďakovať za jej nesmiernu psychickú podporu a možnosť vyrozprávať všetko, čo som mal na jazyku. V neposlednom rade sa chcem poďakovať aj môjmu bratovi Tomášovi Kosibovi a celej mojej rodine za ich obrovskú podporu a dôveru, ktorú do mňa vkladali.

Prohlášení

Prohlašuji, že jsem svoji bakalářskou práci vypracoval samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 16. května 2017

Podpis autora

Contents

Chapter 1

Introduction

1.1 Morphology of galaxies

1.1.1 Galaxy morphology impact

Fraction of morphological distorted galaxies increases with their redshift which means, that morphology of galaxies has strong correlation with redshift. Morphology of galaxies in high redshifts is also a key element for cosmology and its studying of our Universe development. For example, if we live in an accelerating universe, most elliptical galaxies would form at redshifts $z > 1$, however, most elliptical galaxies in decelerating universe would form at $z < 1$ (Abraham and van den Bergh (2001)).

There has been whole study made about galaxy correlations as a function of morphological type (Davis and Geller (1976)) whose results indicate, that differences in galactic morphology are not just results of environmental interactions which are significant long after era of galaxy formation. Even the correlation of nature and evolution of galaxy clustering with morphological characteristics of galaxies in big clusters and small groups has been discussed for many years (Spitzer and Baade (1951), Field and Saslaw (1971), Oemler (1974), Gunn and Gott (1972), Karachentsev and Karachentseva (1975)). Elliptical and lenticular galaxies are dominant in central regions of rich compact clusters while containing just few spirals and irregular galaxies. In comparison, less dense clusters contain many spiral galaxies, similarly distributed as galaxies in field. More elliptical galaxies are also found in small groups which have short crossing times compared with looser groups where are not that many elliptical galaxies as in small groups.

Automatic galaxy classification applied on near-future surveys will provide tremendous amount of morphology classified galaxies enabling us to study evolution of galaxies in much better detail than hitherto possible. We will be able to measure ratio of different morphological types of galaxies depending on redshift, on local spacial density (galaxies being in clusters, near clusters or in field) and much more.

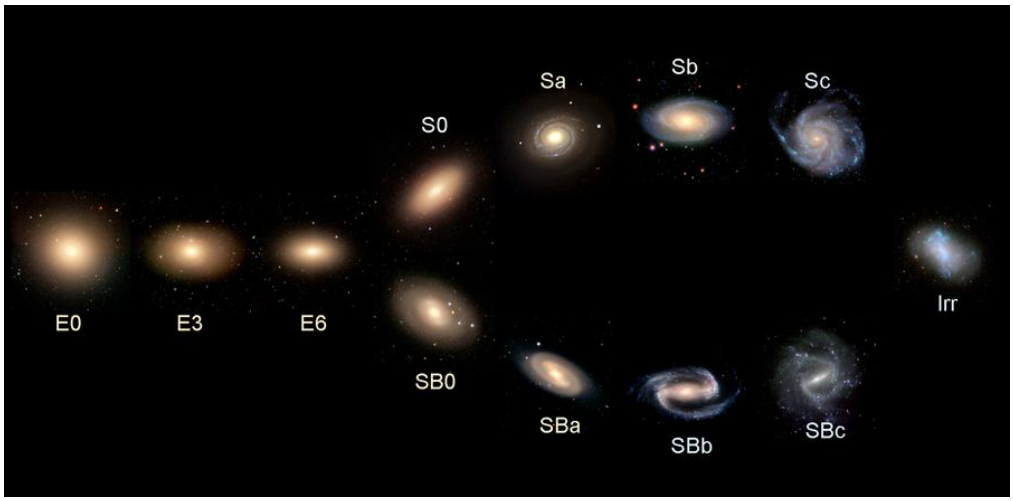


Figure 1.1: Hubble galaxy classification scheme (Hubble (1926)).

1.1.2 Hubble galaxy classification scheme

Hubble galaxy classification scheme (Figure 1.1) divides galaxies according to their morphology into 4 categories.

Elliptical galaxies are smooth, featureless objects appearing as ellipses in images. They are marked with letter E and number representing their degree of ellipticity. For example, E0 are spheroid elliptical galaxies and E7 (which are not shown in the Figure 1.1) are at the edge of their ellipticity.

Spiral galaxies have disk-like structure with usually 2 spiral arms. They are characterized by massive star formation inside of those spiral arms. Their central bulge is similar to elliptical galaxies, it is elliptical, without big star formation and with larger concentration of stars. Spiral galaxies are further divided into spirals with a bar structure in the centre and without it. Spiral galaxies without a bar and with a bar are marked with letter S, SB respectively. Both sub-categories of spiral galaxies have also a small latter (a, b, c) indicating looseness of their spiral arms. For instance, Sa is a spiral galaxy without a bar and with tighter spiral arms than Sc which is also without a bar but with loose spiral arms.

Lenticular galaxies are the third category of Hubble galaxy classification scheme. They are characterized with the central bulge and disk-like structure as spiral galaxies, but without visible spiral arms and they do not have active significant star formation. Lenticular galaxies are further divided into two types, S0 without a bar structure and SB0 with a bar-like structure.

Last category is including irregular galaxies marked as Irr. Small Magellanic Cloud is an example of such galaxy. Irregular galaxies have very distorted shapes without spiral arms or galactic bulge as spiral galaxies have.

1.1.3 Hubble-de Vaucouleurs galaxy classification scheme

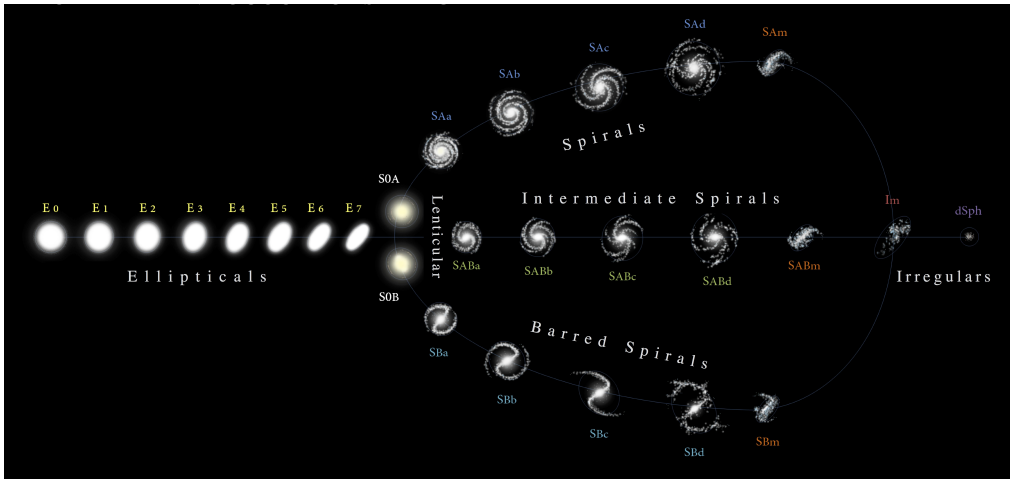


Figure 1.2: Hubble-de Vaucouleurs galaxy classification scheme (de Vaucouleurs (1959)).

Hubble-de Vaucouleurs galaxy classification (de Vaucouleurs (1959)) is an extension of original Hubble galaxy classification (Hubble (1926)) first described by Gérard de Vaucouleur in 1959. De Vaucouleur was not satisfied with Hubble's classification of spiral galaxies only according to tightness of their spiral arms and presence or absence of a bar rising from galactic nuclei. He was arguing that it does not adequately describe full range of observed galaxy morphologies. He retained Hubble's basic division of galaxies into ellipticals, spirals, lenticulars and irregulars but argued that rings and lenses are very important morphology structures of spiral galaxies, thus he improved Hubble's galaxy classification of more complex classification of spiral galaxies based on three morphological characteristics.

All spiral galaxies mark start with S following by few other letters. First additional marking letter describes properties of galactic bar. Those which have bars are marked as SB, without bars as SA and galaxies which are having slight indication of a bar, but it is not prominent as in SB galaxies, have marking SAB.

Second criterion determines whether galaxies posses ring-like structures (additionally marked with (r)), or not (additionally marked with (s)). Transition galaxies have additional mark (s). Those types are not in the Figure 1.2.

Third criterion determines tightness of spiral arms. It is an extension of basic Hubble's spiral arm tightness classification having 5 categories (a, b, c, d, m) instead of Hubble's tree (a, b, c). This distribution describes better the crossing of spirals into irregulars. Large Magellanic Cloud with its classification in Hubble-de Vaucouleurs system as SB(s)m is a representation for the last category of tightness criterion called m.

Category Im is out of spirals and its representative is a Small Magellanic Cloud.

1.1.4 The ATLAS^{3D} comb diagram

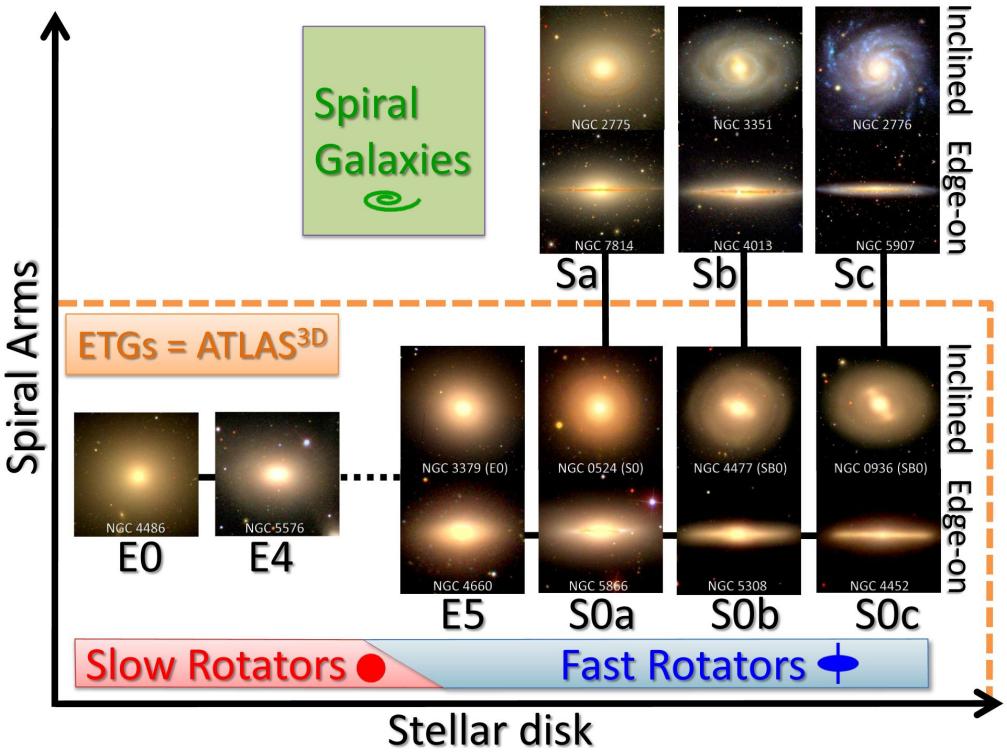


Figure 1.3: New comb diagram proposed by ATLAS^{3D} team (Cappellari et al. (2011)).

E. Hubble though, that the elliptical galaxies evolve into the lenticular galaxies, which further evolve into the spiral galaxies. This opinion gave rise to terms "early types", which are elliptical and lenticular galaxies, and "late types", which are spiral galaxies (we will not use those terms in order to avoid miss interpretation). However, nowadays opinion is exactly opposite. It seems more likely that the spiral galaxies further evolve into the lenticular galaxies, which further evolve into the elliptical galaxies due to processes such as ram-pressure stripping (Book and Benson (2010)) and galactic merging.

The elliptical and lenticular galaxies seem to be much more interesting than we though. It is nearly impossible to visually distinguish the gas and dust lacking face-on disks of stars from much rounder spheroids of elliptical and lenticular galaxies. This is a reason for debates which continue for decades about a fraction of hidden disks-like systems in the elliptical and lenticular galaxies.

We though that elliptical and lenticular galaxies rotate very slowly in comparison with spiral galaxies. It is explained as a reason of chaotic star rotation in elliptical and lenticular galaxies which nearly nulls angular momentum of those galaxies compared to massive rotation speed of stars in spiral galaxies, where nearly all stars rotate in similar direction inside of the spiral disk. However,

according to study of Cappellari et al. (2011) who studied star velocities of 260 elliptical and lenticular galaxies, around 66 % of those galaxies rotate extraordinary fast, due to fact, that they possess disk-like structure. This disk does not have gas, dust and rapid star formation as disks of spiral galaxies, which also makes it less visible.

ATLAS$^{\mathbf{3D}}$ (Cappellari et al. (2011)) team divide elliptical and lenticular galaxies into two groups called "slow rotators" and "fast rotators". Slow rotators are spheroidal elliptical or lenticular galaxies without disk-like structure. Fast rotators exhibit disk-like structure similar to those of spiral galaxies only without gas, dust and rapid star formation. They also point to strong evolutional relation of spiral galaxies with those fast rotators. Fast rotators seem to be old spiral galaxies which already consumed majority of their gas and dust for star formation. Fast and slow rotators can not be optically distinguished (as they have been classified as same morphological objects until now). However, by star velocity measurements all inclinations of those galaxies can be strictly classified as fast and slow rotators. Those are reasons why ATLAS$^{\mathbf{3D}}$ team considers Hubble galaxy classification inadequate and propose new scheme for galaxy morphology classification (1.3) based on star velocities.

Measurement of star velocities requires spectroscopic observations. Star velocities can not be measured just based on morphology. However, for first, we need to have morphology classified galaxies in order to find candidates (elliptical and lenticular galaxies) for spectroscopic measurements of star velocities to further distinguish those candidates to fast and slow rotators.

1.1.5 Morphology density relation

Morphology density relation (Figure 1.4) describes the relation between galaxy morphology shapes and environment they are located in.

Morphology density relation shows, that field environment consists mainly of spiral galaxies with just a few elliptical and lenticular galaxies. As environment density rises, representation of spiral galaxies decreases while representation of elliptical and lenticular galaxies increases. Elliptical and lenticular galaxies are mainly located in clusters of high densities.

Morphology density relation shows that galaxy evolution is strongly connected with environment. Future surveys with automatic image classification techniques will allow us to study such relationships in much greater detail.

1.2 Galaxy Zoo projects

1.2.1 First Galaxy Zoo project

The First Galaxy Zoo project started in 2007 (Lintott et al. (2008)). People who participated at this project classified images of galaxies which were extracted from the Sloan Digital Sky Survey (SDSS) (York et al. (2000)). Then they had to classify those galaxies according to the decision tree of this Galaxy Zoo project,

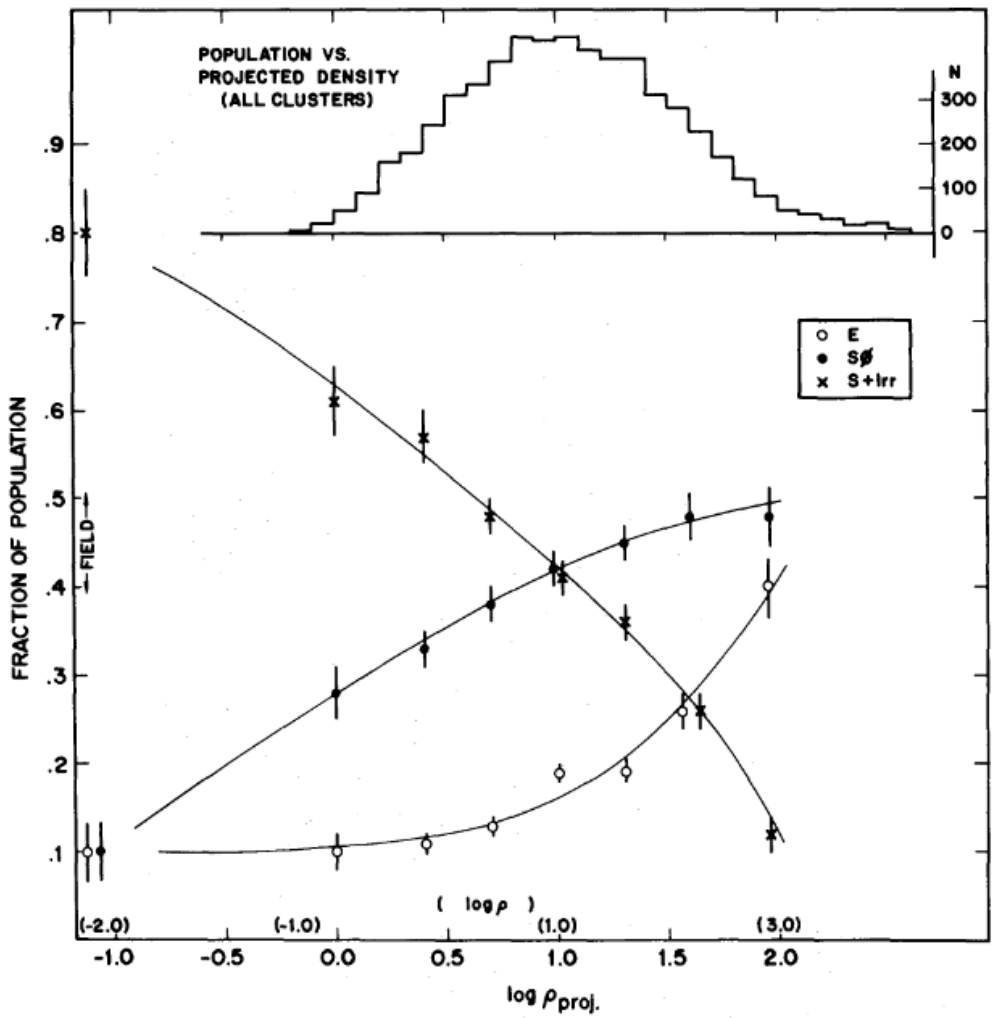


Figure 1.4: Morphology density relation of galaxies from 55 rich clusters studied by Dressler (1980). It shows the fraction of galaxy population as a function of log of projected density in galaxies Mpc^{-2}. Distribution of the galaxies over the bins of projected density is shown in upper histogram. E are elliptical galaxies, S0 lenticular galaxies and S + Irr are spiral and irregular galaxies.

which included only the most basic questions like if the galaxy is elliptical, spiral or mergers. For those, that were classified as spiral galaxies, classifiers were further asked to choose if the spiral arms were rotating clockwise, anticlockwise or if they were edge-on. This has a particular meaning for determining the rotation of galaxy.

Every galaxy was presented to multiple people, which helped to estimate probabilities of morphological types. It is vital to keep in mind, that even people do not agree in types of galaxies, which means that automatic classification techniques can not give results without deviations. According to Lintott et al. (2008) nearly one million of galaxies were classified in this first project of Galaxy Zoo.

1.2.2 Second Galaxy Zoo project

Galaxy Zoo 2, which is the second project of Galaxy Zoo, resulted in more than 16 million classifications of galaxy morphology of more than 300 000 galaxies (Willett et al. (2013)). Like in the previous project, this project took images from SDSS (York et al. (2000)). The main difference is that people had to describe galaxies according to more sophisticated decision tree (image of decision tree (Figure 1.5), decision tree in words (Table 2.7)) than in first Galaxy Zoo. Those innovations were, for example, in defining the relative strengths of galactic bulges and spiral arms, determining the shapes of edge-on disks, as well as whether bars or bulges are included or not. Citizen scientists provided great results. Compared with professional astronomers, their accuracy was more than 90 % (Willett et al. (2013)).

1.2.3 The Galaxy Challenge

The Galaxy challenge was an international online challenge distributed on Kaggle platform and supported by Winton Capital. Its goal was developing an algorithm for automatic computer classification of galaxies according to their morphology. Galaxy Challenge used labeled data from Galaxy Zo 2 crowd-sourcing project. Its winner is Sander Dieleman with his solution Dieleman et al. (2015). He built and trained convolutional neural networks for galaxy classification based on their morphology.

1.3 Increasing data volumes

The same data volumes, which were produced by entire surveys a decade ago, are nowadays possible to acquire during a single night. Moreover, a real-time data analysis is usually desired. This enormous and growing amount of data must be analysed in an automatic and sophisticated way. Crowd-sourcing projects can not be applied to such large data volumes that will be provided by surveys in near future (Figure 1.6).

Task	Questions	Responses	Next
01	Is the galaxy simply smooth and rounded with no sign of a disk?	smooth	02
		features or disk	07
		star or artifact	**end**
02	Could this be a disk viewed edge-on?	yes	09
		no	03
03	Is there a sign of a bar feature through the centre of the galaxy?	yes	04
		no	04
04	Is there any sign of a spiral arm pattern?	yes	10
		no	05
05	How prominent is the central bulge, compared with the rest of the galaxy?	no bulge	06
		just noticeable	06
		obvious	06
		dominant	06
06	Is there anything odd?	yes	08
		no	**end**
07	How rounded is it?	completely round	06
		in between	06
		cigar-shaped	06
08	Is the odd feature a ring, or is the galaxy disturbed or irregular?	ring	**end**
		lens or are	**end**
		disturbed	**end**
		irregular	**end**
		other	**end**
		merger	**end**
		dust lane	**end**
09	Does the galaxy have a bulge at its centre? If so, what shape?	rounded	06
		boxy	06
		no bulge	06
10	How tightly wound do the spiral arms appear?	tight	11
		medium	11
		loose	11
11	How many spiral arms are there?	1	05
		2	05
		3	05
		4	05
		more than four	05
		can't tell	05

Table 1.1: The Galaxy Zoo decision tree in words. It consists of 11 tasks and 37 possible responses. The numbers of tasks does not represent their order in decision tree. Texts in "Questions" and "Responses" were displayed to participants together with the icons in Figure 1.5.

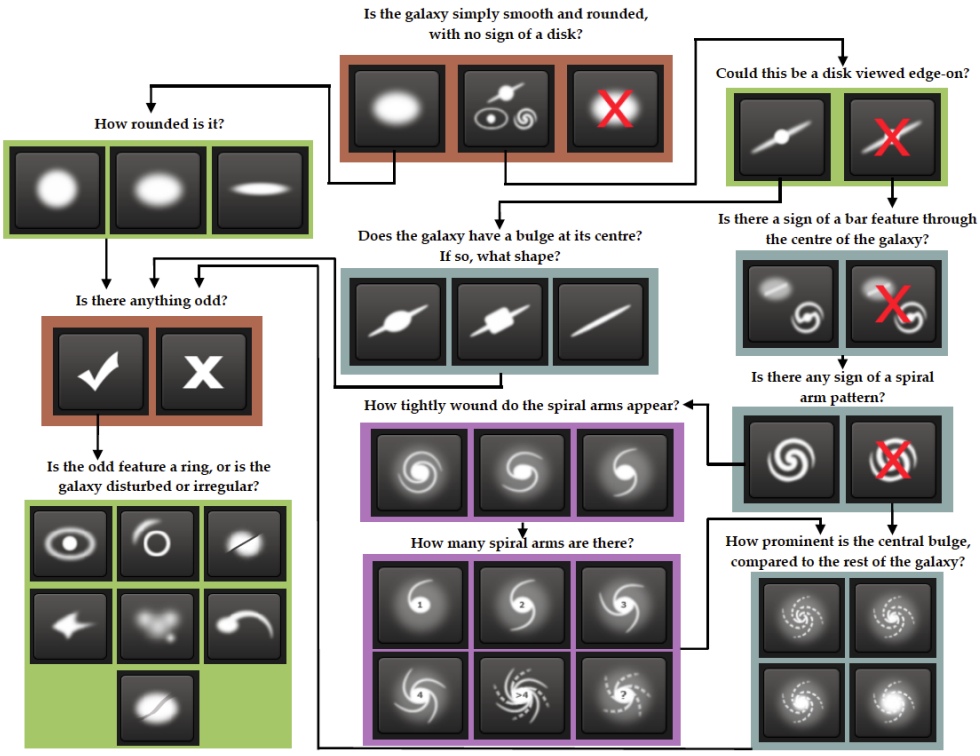


Figure 1.5: This image represents the flowchart of decision tree of Galaxy Zoo 2 [2]. It begins in the centre at the top. The colour represents the relative depth of questions in the decision tree. Those outlined in brown are asked for every galaxy. Green colour represents one step, blue two steps and purple three steps below branching points in decision three.

1.4 Large sky surveys

1.4.1 Introduction

There is enormous amount of a high quality image data of galaxies, which are not sufficiently analysed and even more are about to be made. Literally, tens of billions "*superb*" (as LSST developers say) images are yet to come. The future of analysing such tremendous amount of data lies in automatic processing and machine learning techniques such as convolutional neural networks that are arguably one of nowadays best tools for classification of such amount of image data. We would like to introduce some of currently working large surveys and the LSST which is right now under construction.

1.4.2 Sloan Digital Sky Survey (SDSS)

Creators of Sloan Digital Sky Survey (SDSS) developed their own filter system called *u' g' r' i' z'*, which is practically same as Gunn griz filter system (Oke

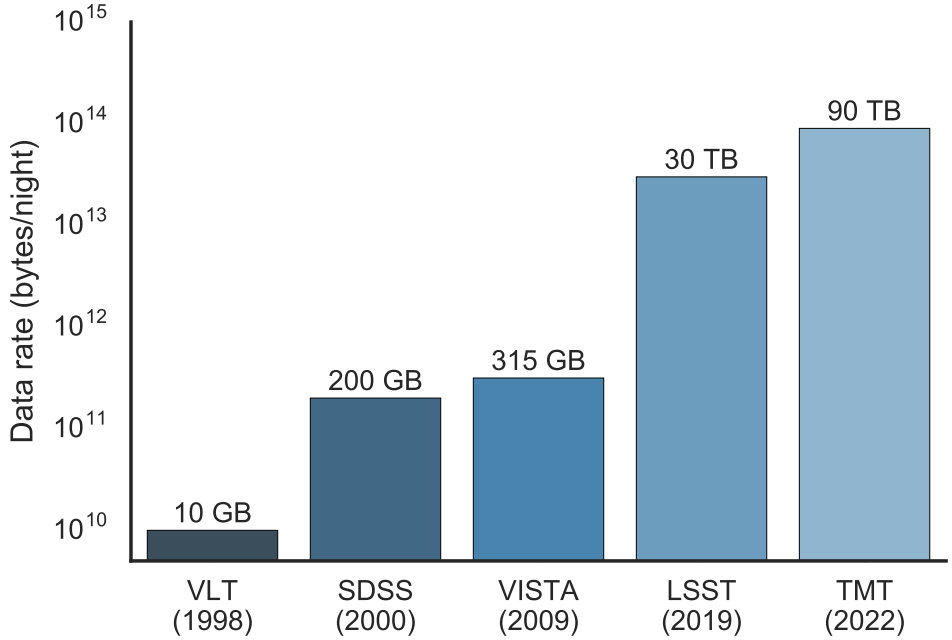


Figure 1.6: "*Increasing data volumes of existing and upcoming telescopes: Very Large Telescope (VLT), Sloan Digital Sky Survey (SDSS), Visible and In- frared Telescope for Astronomy (VISTA), Large Synoptic Survey Telescope (LSST) and Thirty Meter Telescope (TMT).*" (Kremer et al. (2017))

and Gunn (1983), Schild (1984)). The CCD chip of SDSS camera is divided into six columns, one for each filter. As the telescope scans the sky, pictures are being taken sequentially on every filter with time steps equal to the motion of telescopes in order to take a picture of exactly same field on sky in each of the six filters. (York et al. (2000))

SDSS made many major discoveries in the field of astrophysics. For example, its images of quasars allowed us to look back to the times when our universe was old only few billion years and showed that black holes had a stage of rapid early growth.

1.4.3 Dark Energy Survey (DES)

The Dark energy Survey (DES) is an international collaborative project. Its main goals are to map hundreds of millions of galaxies, to detect thousands of supernovae and to find patterns of cosmic structure, which may help to reveal properties of mysterious dark energy that is believed to accelerate the expansion of our Universe. DES is searching southern skies and it began on August 31, 2013. (The Dark Energy Survey Collaboration (2005))

Over the time period of five years, from 2013 to 2018, DES is recording

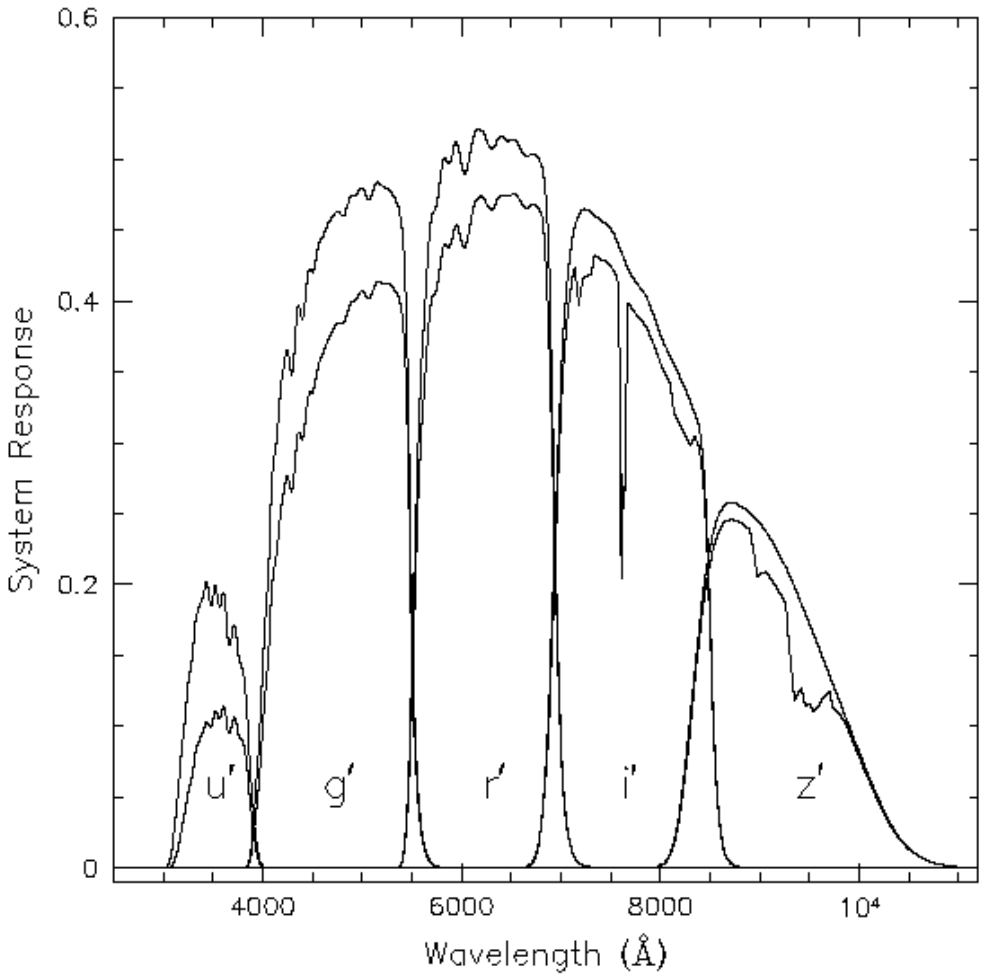


Figure 1.7: According to [3] this picture represents *"The SDSS system response curves. The responses are shown without atmospheric extinction (upper curves) and as modified by the extinction at 1.2 airmasses (lower curves). The curves represent expected total quantum efficiencies of the camera plus telescope on the sky."*

information from 300 million galaxies. It also uses a fraction of its time to discover thousands of supernovae by observing small patches of sky almost once a week. [5]

The DES uses a CCD camera mounted on a Blanco 4-m telescope at Cerro Telolo Inter-American Observatory (CTIO). The survey is carried out using 5 filters, where 4 are those from SDSS (g, r, i and z). Fifth filter is Y (400 - 1050 nm) plus it has 3 additional slots for possible installation of other filters. (Honscheid et al. (2008))

The accuracy of determining photometric redshifts (photo-z's) of galaxies will be dependent also on the galaxy type. (The Dark Energy Survey Collaboration

(2005))

1.4.4 The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)

The first project of the PAN-STARRS projects, the Pan-STARRS observatory, embarked in May 2010. The observations were carried using 5 filters, g_{P1}, r_{P1}, i_{P1}, z_{P1} and y_{P1} seeing the sky in visible and near infrared spectrum. It was also the first project that intended to find asteroids that could threaten the Earth. For this purpose, continuous and repeating scanning of the sky was carried out 12 times in five years. (Magnier et al. (2013))

During four years of observing, pan-STARRS collected information of extensive quality about over 3 billion stars, galaxies and other sources [6].

1.4.5 The intermediate Palomar Transient Factory (iPTF)

The iPTF is a project built upon a previous Palomar Transient Factory (PTF), which started in March 2009. It uses the 48inch Schmidt telescope (P48) at the Palomar Observatory equipped with a camera, which has two filter options R and g. The iPTFs main goal is to search for young supernovae and fast transients. (Cao et al. (2016))

The iPTF will changeover to the Zwicky Transient Factory (ZFT) in 2017. The ZFT will be using reworked version of the same telescope as the iPTF, but it will use a new camera, which will enable an every night full scan of a visible sky and will directly lead to the LSST era.

1.4.6 Large Synoptic Sky Survey (LSST)

The LSST is currently under construction in Chile. In its ten-year survey LSST will provide more than 37 billion images of galaxies [7] with its 3.2 billion-pixel camera. It is due to start observing in 2019. The LSST observations will be carried out using 6 filters u, g, r, i, z and y. Every image will be of the size equivalent to 40 full moons (almost 10 square degrees of the sky). The LSST will be able to map the whole sky in just a few days (Ivezic et al. (2008)). One of its main objectives is discovering new transients. Those are objects which change brightness over time. Some may have period of tens of years, while others just of a few seconds. Some of those changes will be caused by extremely rare events. It is expected that the LSST will see millions of transients per night where real-time data analysis is needed for follow-up observations. The LSST will provide about 30 terabytes of images per night, which is approximately 60 times more than for the SDSS.

The LSST will provide images with such resolution that it is expected to discover many faint objects, which are impossible to observe today. It will be able to detect objects 10 million times fainter than visible with human eye. This is very important mainly for cosmology, because fainter galaxies are further away

not just in space, but also in time. The distribution of morphological types of those faint galaxies will be crucial for better understanding the evolution of our Universe itself.

The LSST data will be used to make the three-dimensional cosmic map with unprecedented depth and detail. This map will serve to various purposes. For example, for locating the dark matter, characterizing properties of the dark energy, tracking the transient objects and studying our Galaxy into much greater details. It will be also used to protect the Earth by detecting and tracking asteroids that might impact the Earth.

The LSST will use its observations of several billion galaxies to study their masses and influence on the distortion of space-time. Those measurements will be also used to further understand the dark matter and dark energy. Particularly the influence of dark matter on the development of structure of the Universe on a cosmic scale and how dark energy behaves with cosmic time or with redshift.

Redshift z is crucial for study of the dark energy and physics related with it. The spectroscopic measurement of redshift requires huge amount of observing time, which makes it impossible to be done for billions of galaxies that will be observed by the LSST. This is a reason why the developers of the LSST want to use photometric redshift measurement instead. One of possible technique to estimate the photometric redshift is a machine-learning based on the neural networks. (LSST Dark Energy Science Collaboration (2012))

1.5 Related work

Machine learning techniques, but mainly artificial neural networks, have been used in astronomy research for more than two decades starting with their application on star-galaxy discrimination (Odewahn et al. (1992), Bertin (1993)) and classification of galaxy spectra by Folkes et al (1996). More recent example is their usage for photometric redshift estimation (Collister and Lahav (2004)).

Chapter 2

Neural networks as a solution to image recognition

2.1 Introduction

In this chapter, we describe the neural networks. It starts with comparison of biological and mathematical neuron, continues with explanation of feed-forward neural networks and their components and ends with description of convolutional neural networks and their specifics.

2.2 Comparison with biological brain

Neural networks took an inspiration in biological neural systems. Neurons are the basic computation units of the brain (Figure 2.1). They are connected with synapses. Mathematical neurons, which are used in artificial neural networks, mimic the behaviour of the biological neurons.

The biological neurons receive input signals (electric impulses and chemicals) through their dendrites and produce an output signal via an axon, where the neuron connects to other neurons with synapses to other dendrites of other neurons. A neuron is triggered and active only if the combination of its input signals reaches some threshold condition resulting in transferring its information to the successor neurons.

Each connection between mathematical neurons has its weight. The mathematical model of neuron (Figure 2.1) calculates weighted sum of its inputs and activates only if this sum reaches its treshold condition. If the treshold is reached, mathematical neuron applies the non-linear function on the result of the weighted sum and sends this information to other neurons.

2.3 Single-hidden-layer neural networks

The single-hidden-layer neural networks consists of an input layer followed by one hidden layer, which is followed by an output layer. The hidden layer is any layer between the input layer and the output layer. Following equations show an output of the hidden layer h and an output of the output layer y.

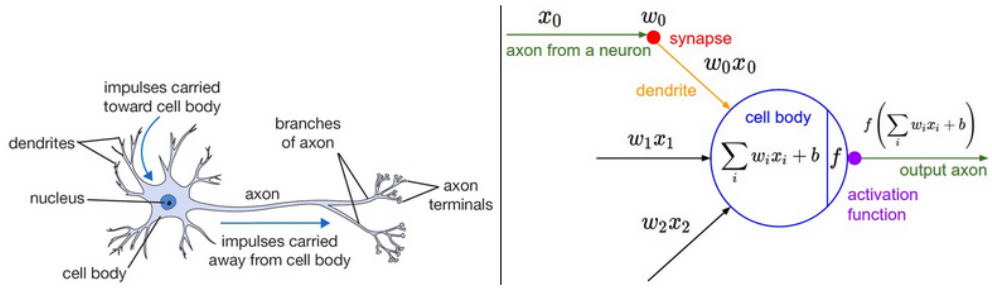


Figure 2.1: Comparison of biological neuron on the left and mathematical neuron on the right [8].

$$\boldsymbol{h} = f_1(\mathbf{W}_1\mathbf{x}_1 + \mathbf{b}_1) \tag{2.1}$$

$$\boldsymbol{y} = f_2(\mathbf{W}_2\mathbf{x}_2 + \mathbf{b}_2) \tag{2.2}$$

f_1 is the non-linear function applied on the input layer and f_2 on the hidden layer. $\mathbf{x}_1$ is an input and $\mathbf{x}_2$ is an output vector of the hidden layer, which is an input of an output layer. $\mathbf{W}_1$ and $\mathbf{W}_2$ are the matrixes of weights, which will be learned by the neural network together with the bias parameters $\mathbf{b}_1$ and $\mathbf{b}_2$. Bias enables to shift the activation function to the left or to the right. For example, lets consider an input of the mathematical neuron to be a number 2, adding a bias would allow an output of 0, which would not be possible otherwise. This makes neural networks more flexible and usually improves their performance. Rectified linear unit (ReLU) is nowadays most commonly used lon-linear function. Sigmoid function od hyperbolic tangent are also in use.

Sigmoid function

$$\sigma(\boldsymbol{z}) = \frac{1}{1 + \mathrm{e}^{-\boldsymbol{z}}}. \tag{2.3}$$

Hyperbolic tangent

$$\sigma(\boldsymbol{z}) = \frac{\mathrm{e}^{2\boldsymbol{z}} - 1}{\mathrm{e}^{2\boldsymbol{z}} + 1}. \tag{2.4}$$

Rectified linear unit (ReLU)

$$\sigma(\boldsymbol{z}) = \max(0, \boldsymbol{z}). \tag{2.5}$$

Where $\boldsymbol{z} = \boldsymbol{w}\boldsymbol{x} + b$, $\boldsymbol{w}$ is a matrix of weights, b is a bias parameter and x represents input data.

2.4 Deep Neural Networks

What distinguishes the deep neural networks from the single-hidden-layer neural networks is their depth. The number of hidden layers defines the neural networks

depth. Traditional machine learning uses simple neural networks composed of one input layer and one output layer, at most with one hidden layer between. So word deep is a strictly defined technical term for neural networks with more than one hidden layer (those will be further explained).

In deep neural networks, each layer of neurons is being trained on different features. Those features are based on previous layer's output.

Neurons in deeper layers exhibit higher level of abstraction and respond to more complex patterns in the input images as can be seen in Figure (2.2).

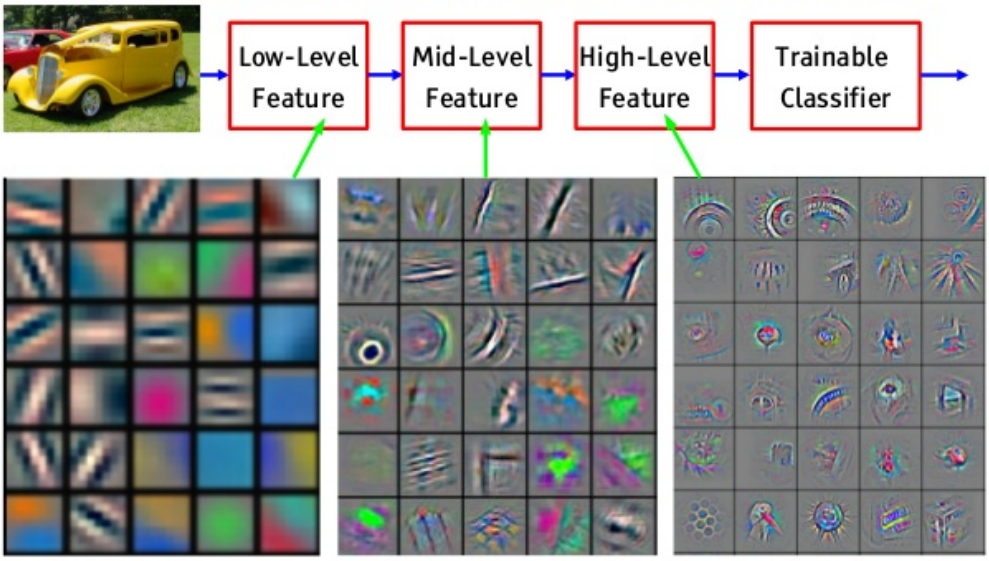


Figure 2.2: Feature visualization of CNN trained on ImageNet (Zeiler and Fergus (2013).

2.5 Architecture of Standard Feed-forward Neural Network

The information flow in a standard feed-forward neural network, as its name indicates, continuous straightforwardly from the input to the output, without any loops. Information is always fed forward and never returns back as, for example, in recurrent neural networks, which will not be further discussed in this work.

The feed-forward neural networks architecture (Figure 2.3) consists of neurons and layers. The first layer is called input layer. It represents an information entry to the network. Then there is a number of hidden layers and finally the output layer, which calculates the output of neural network. The number of hidden layers can not be declared by some strict rules as well as an amount of

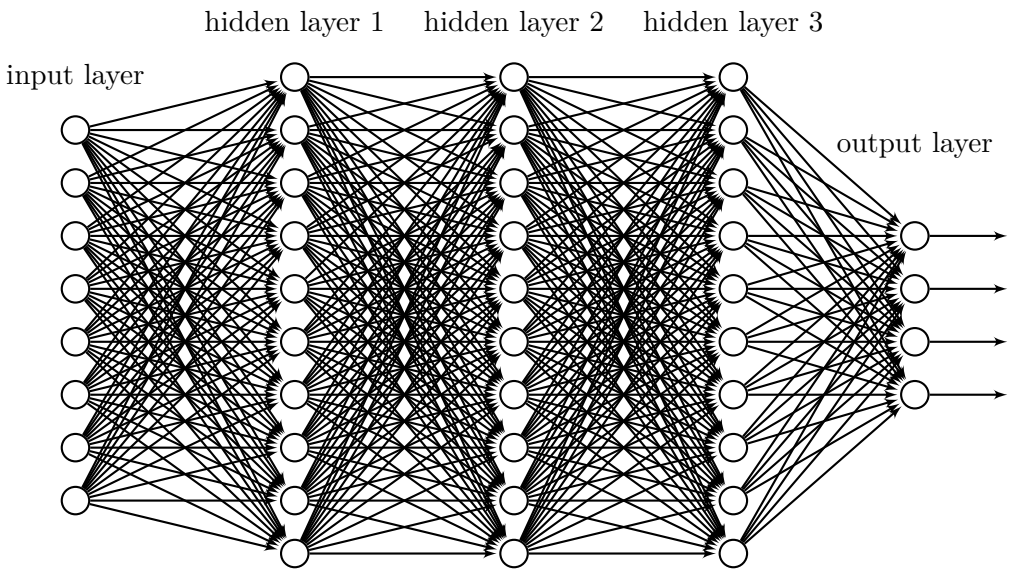


Figure 2.3: Image of feed-forward neural network consisting of one input layer, three hidden layers and one output layer [4].

neurons for each hidden layer. Different architectures were proposed for different purposes. When creating neural network for a new task, it is recommended to start with an architecture that works for a similar problem, and than to try adjust its properties. The feed-forward architecture is often of the first choice for its simplicity. An important property of this architecture is that all the layers are fully connected to the adjacent layers. Every single neuron of the layer connects to all the neurons in adjacent layers, but does not have any connection with neurons in the other layers. [4]

This results in very large number of parameters, which is the main difference and disadvantage in comparison with the convolutional neural networks.

The output of layer j is a vector and it is calculated by following formula

$$\mathbf{x}_j = f_j(\mathbf{W}_j \mathbf{x}_{j-1} + \mathbf{b}_j), \tag{2.6}$$

where $\mathbf{x}_{j-1}$ is the input to j-th layer with matrix of weights $\mathbf{W}_j$, vector of biases $\mathbf{b}_j$ and non-linear activation function f_j

2.6 Loss function

Any neural network can be seen as a nonlinear function

$$\mathbf{y} = f(\mathbf{x}; w),$$

where $\mathbf{x}$ is an input (an image of galaxy in our work), $\mathbf{y}$ is an output vector (probabilities of different galaxy morphology types) and w is a set of trainable parameters. For example, for the single hidden layer feed-forward neural network

from equations (2.1) and (2.2) of Table 2.7. The set of trainable parameters would be $w = \{\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2\}$. We will use w_i to denote single parameter from this set (e.g. one coefficient of one of the weight matrices or bias vectors). To train the neural network, we have a set of N training examples consisting of inputs $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N,\}$ and corresponding desired (ground true) outputs $\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N,\}$ (e.g. the human annotations of galaxy morphologies from the Galaxy Challenge). We need to define a loss function measuring the quality of particular parameter set w based on computing the error between predicted output vectors $\mathbf{y}_n$ and ground true vectors $\mathbf{t}_n$. There are many possible loss function. To be consistent with the metric used for the Galaxy Challenge, we have used *Mean Square Error* as the loss function in our work

$$L(w) = \frac{1}{2N} \sum_{n}^{N} |\mathbf{t}_n - \mathbf{y}_n|^2 = \frac{1}{2N} \sum_{n}^{N} \sum_{d}^{D} (t_{nd} - y_{nd}), \tag{2.7}$$

where y_{nd} is d-th element of the D-dimensional vector $\mathbf{y}_n$ (and similarly for t_{nd}).

2.7 Gradient Descent (GD)

The learning process of neural network is based on searching for the combination of learnable parameters providing the lowest error of the loss function. We will use the gradient of the loss function to find a direction along which we should change our weight vector. This direction will be mathematically guaranteed to be the direction of the fastest descent of the loss function (LeCun et al. (1999)).

$$w_i := w_i - \alpha \frac{\partial}{\partial w_i} L \tag{2.8}$$

This update of weights is simultaneously performed for all values of weights (i = 1, ... , number of weights). Parameter α is called the learning rate. Its function is to manually change the gradient update. it is often used in actual training because weights tend to increase far too much in each iteration, which would make them diverge an “over correct”.

The gradient descent is an optimization method, which updates all the weights at once after running through all samples in the training dataset once (this is called an epoch). However, its alternative, the stochastic gradient descent (SGD), updates the weights progressively after a subset of the training sample from the training dataset.

2.8 Back-propagation

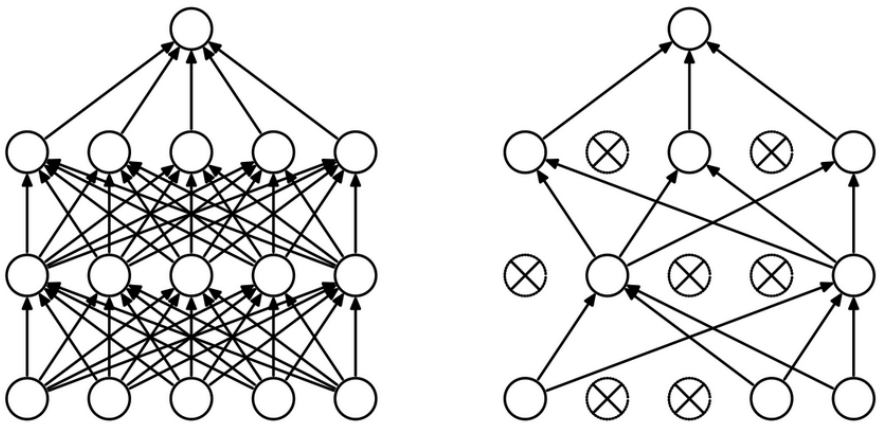
The back-propagation is a learning process of neural network. It uses the gradient descent method to find the minimum of a loss function. The solution of learning problem is a combination of weights with minimum loss function. The back-propagation could be divided into two repeating phases. The network is firstly

given a vector as its input. It is propagated through the whole network resulting in an output. The networks output is compared with the desired output by using a loss function, which calculates error values for every single neuron in the network starting at the output layer and propagating through the whole network. The phase consists of updating all of the weights by desired optimization function.

2.8.1 Dropout and Vanishing gradient problem

Traditional neural networks used just few hidden layers. It was hard to learn deeper layers with many layers applying non-linearities. Gradient information would vanish during propagation through layers, which made it difficult to learn parameters of lower layers (Hochreiter et al. (2001), Glorot et al. (2011)).

The dropout is a regularization technique, which reduces the vanishing gradient problem. When applied on a layer, the dropout randomly sets output values of previous layer to zero with probability p. p is typically chosen to be 0.5. The dropout is applied in every single sample presented to the network, creating almost certainly new, unique neural network (Figure 2.4(b)) for every sample. Those dropped out neurons are not participating in the learning process. The neurons can not rely on the active presence of the other neurons, which makes them more robust and less like to memorize the data (overfitting).



(a) Standard Neural Net (b) After applying dropout.

Figure 2.4: Illustration of dropout.

2.9 Convolutional neural networks

The convolutional neural networks make an assumption, that their input is an image. This assumption allows great simplifications, which results into reducing the amount of parameters in the network. Convolutional neural networks have their inputs and outputs of each layer organized as 3-dimensional matrix, where

the locality of the neighbouring elements in the matrix is significant. ([1], LeCun et al. (1999))

2.9.1 Convolutional layer

The input of convolutional layer is a stack of feature maps (for example the colour rgb channels of an image). The convolutional layer takes this input and convolves it with its features resulting in an output of different feature maps. Filters of deeper convolutional layers have a shape of 3-dimensional matrix, so they could operate across all feature maps and provide 3-dimensional convolution.

This can be implemented by replacing the matrix-vector product $f(\mathbf{W}_j\mathbf{x}_{j-1})$ in equation (2.6) with a sum of convolutions. The input of layer j is represented as a set of K matrices $\mathbf{X}_{j-1}^{(k)}$ with $k = 1 \ldots K$. Each of this matrices represents different feature map. Representation of the output feature maps $\mathbf{X}_j^{(l)}$, $l = 1 \ldots L$ are represented by following formula (Dieleman et al. (2015))

$$\mathbf{X}_j^{(l)} = f\Big(\sum_{k=1}^{K} \mathbf{W}_j^{(k,l)} * \mathbf{X}_{j-1}^{(k)} + v_j^{(l)}\Big). \tag{2.9}$$

Matrices $\mathbf{W}_j^{(k,l)}$ represent the filters of layer j and $b_j^{(l)}$ represents the bias of feature map l. Mark * represents the 2-dimensional convolution, which has following formula

$$\mathbf{X}[i,j] * \mathbf{H}[i,j] = \sum_k \sum_l \mathbf{X}[k,l]\mathbf{H}[i-k, j-l], \tag{2.10}$$

where i and j are filter width and height, $\mathbf{X}$ is the input and H is the filter.

The neurons in a convolutional layer are connected only to a small region of previous layer in comparison with fully connected layers in feed-forward neural networks (Figure 2.5) ([1], LeCun et al. (1999)).

The parameters of convolutional layer consist of set of learnable filters (LeCun et al. (1999)). In this case, the network is learning filters that become active when seeing some feature. The first convolutional layer learns to recognize just the most basic features like for example differently rotated edges and coloured blotches. Deeper layers learn more complex features. The interpretation of those features is much harder as they do not have classic 3 spacial dimensions of width, height and depth.

Each convolutional filter has a set of filters and each of them will produce a separate 2 dimensional depth slice. The output volume is produced by stacking the depth slices along the depth dimension. Each neuron in the convolutional layer is connected just locally with an input in the height and width dimensions, but always fully in depth dimension. This region of connection is called the receptive field of neuron, which is equivalent with a filter size.

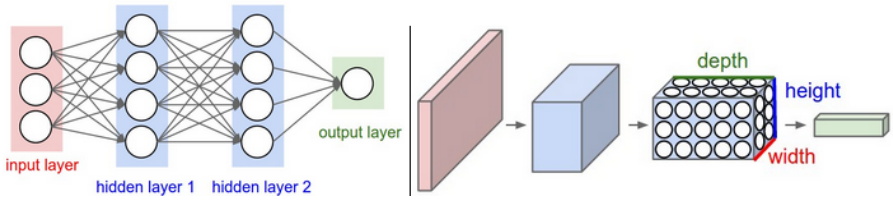


Figure 2.5: Comparison of architectures of fully connected neural network (left) and convolutional neural network (right) [1].

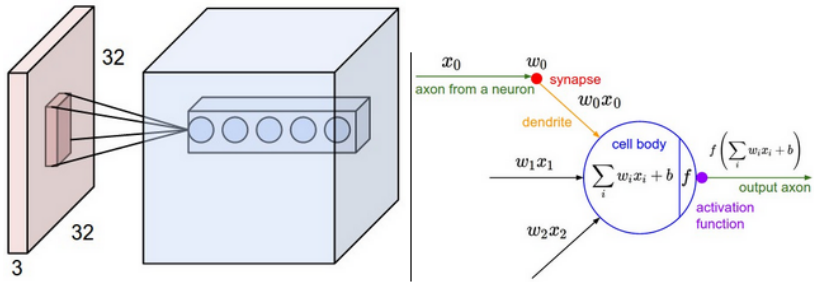


Figure 2.6: The left showss an example of an input image of volume 32×32×3 (where 3 is the rgb colour dimension) and the first convolutional layer with neurons. The left side of the picture depicts a receptive field, which shows the local connection of neurons in the spatial dimension, but full connection in depth. All of those 5 neurons in the example are looking at the same receptive field. On the right side is a scheme of mathematical neuron. Those neurons are similar to those in the regular neural networks, producing dot product of their weights with the input, followed by non-linearity. The difference is in their connectivity, which is now restricted in the width and height dimensions. ([1], LeCun et al. (1999))

Output of convolutional layer

The output of a convolutional layer is specified by 3 hyper-parameters, filter size, stride and padding. Number of feature maps determines the depth dimension of output.

Stride defines how the filter is moving on the picture, how many pixels does it jump in its movement. It will be better explained on an example. Suppose that we have image of dimensions 32×32×3 and filter 5×5×3. With stride 1 the output of this convolutional layer would have sizes 27×27×3.

Padding (Figure 2.7) is adding zeros to the spacial dimension of the image. Adding padding +2 to the example above will give us an image of dimensions 36×36×3. Applying the convolution with filter of sizes 5×5×3 and stride 1 will give output of 32×32×3.

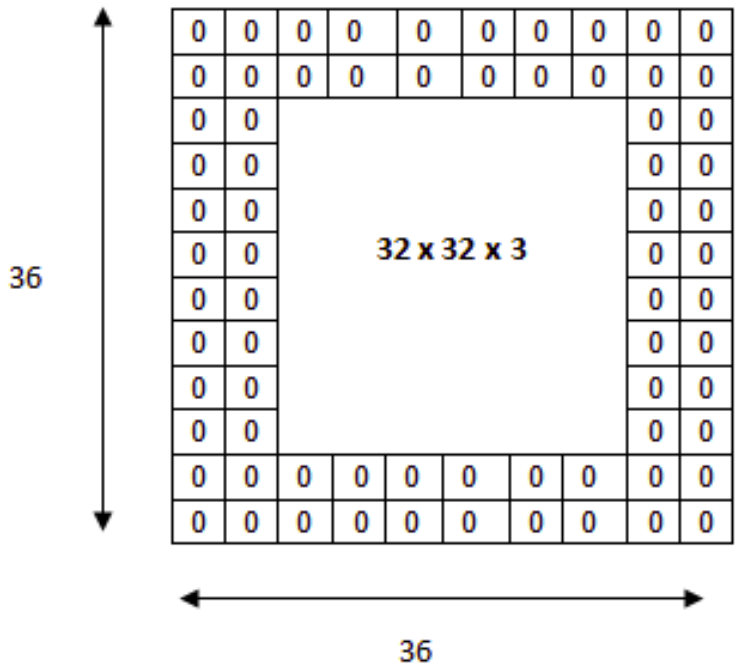


Figure 2.7: Illustration of padding +2 applied on image of size 32×32×3 (depth dimension is not illustrated)

Parameter sharing

Number of parameters can be dramatically reduced by parameter sharing. It is based on the idea, that if one feature is useful for computation at some spatial position, than it should be useful for computation at a different position. In other words, this makes constraint for the neurons in the same depth slice to use the same weights and bias. Depth slice is a single two-dimensional slice of depth of convolutional layer. (For example, volume of size 42×42×69 has 66 depth slices, each of size 42×42). Practical usage of this is in back-propagation, every neuron will still compute the gradient for its own weights, but these gradients will be added up across each depth slice and updating only single set of weights per slice. [1], LeCun et al. (1999))

2.9.2 Pooling layer

The pooling layer reduces the spacial dimension of the representation given by a convolutional layer, which results to decrease of the amount of parameters and the number of computations in a network. Pooling works independently on every depth slice of its input. It has a stride parameter similar to that of a filter of convolutional layer. Pooling usually applies MAX operation. For instance, lets consider pooling of filter size 2×2 applied with stride 2 (Figure 2.9). This pooling would apply MAX operation on 4 numbers in each of its steps, returning

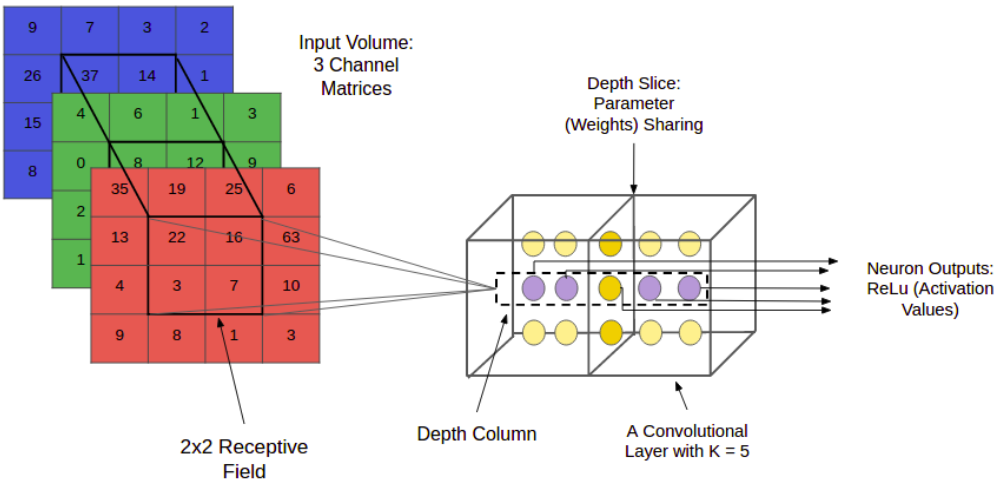


Figure 2.8: An illustration of a convolution with filter size 3×3×5, where 5 is filters depth, applied on 3 rgb colour dimensional image, for example classical JPEG image [7].

just the maximum value of those four. It downsamples every depth dimension by 2, which reduces the data volume to 25 %. (Ranzato et al. (2007))

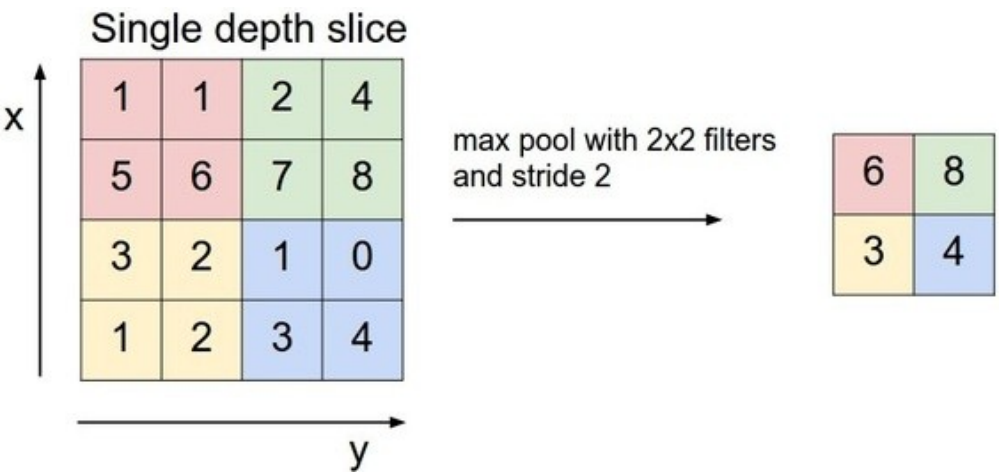


Figure 2.9: Illustration of max pooling with filter size 2×2 and stride 2.

2.9.3 Non-linearity as an activation function

Convolutional neural networks use non-linear activation functions similar to the regular neural networks. Nowadays standard non-linearity is ReLU. It has several computational advantages in comparison with other non-linearities like, for example, the sigmoid function or the hyperbolic tangent. Using ReLU activation decreases the problem with vanishing gradient (Hochreiter (1998),

Glorot et al. (2011)). Another thing is that neurons that use ReLU increase sparsity in the hidden layers (Glorot et al. (2011)). Another benefit is the increase of effectiveness of gradient descent during back-propagation due to ReLU's piecewise linear nature. For instance, according to Krizhevsky et al. (2012) implementation of their convolutional neural network with ReLU non-linearity made their network six times faster than equivalent networks with hyperbolic tangent non-linearity.

2.9.4 Fully connected layer

The fully connected layers used in a convolutional neural networks are similar to the one, used in regular neural networks (Figure 2.3). Each neuron is connected with every single neuron in the previous layer.Their activations are also computed using a matrix multiplication followed by adding a bias offset.

2.9.5 Architecture of Convolutional Neural Network (CNN)

Figure 2.10 is an example of architecture of convolutional neural network. It shows a CNN which is made to classify pictures of digits. Size of the input layer is 32×32, which is also size of training images. Second layer is a convolutional layer which has 28×28 feature map size and convolutional filter (also called kernel) of size 5×5. Output of this convolutional layer is of 28×28 size. Third layer is a sub-sampling layer. Its output is 14×14 and its filter dimension is 2×2 which means, that it sub-samples its input dimension (which is an output of second layer, convolutional layer) 2 times in width and height dimensions. Fourth layer is again a convolutional layer with feature map dimension 10×10 and filter dimension 5×5. It is again followed by sub-sampling layer. The spacial dimension of fifth layers filter is 2×2 and output dimension 5×5, so it again sub-samples 2 times. Following layer is sixth layer, which is a fully connected layer followed by non-linearity. This layer classifies its inputs to specific classes, giving them probabilities of being certain digit.

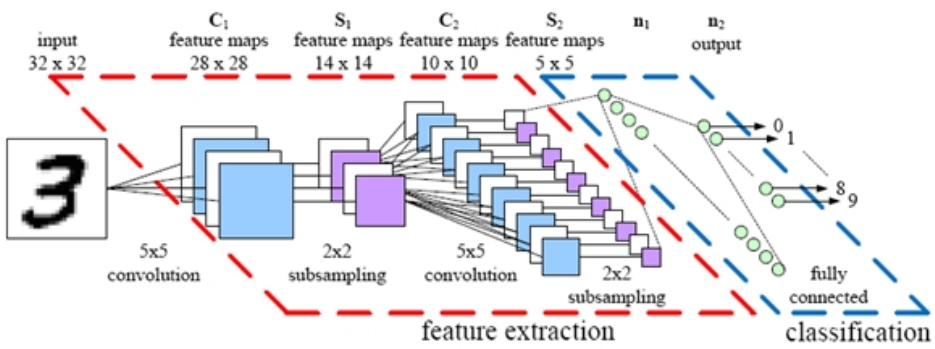


Figure 2.10: An example of architecture of a convolutional neural network [9].

Chapter 3

Application of neural networks to images of galaxies

3.1 Dataset

Our dataset consists of 61 578 424×424×3 JPEG rgb colour images of galaxies from SDSS. All galaxies are centred in images. All the images have been labeled by citizen scientists participating in Galaxy Challenge resulting. The labels consist of 37 numbers corresponding to answers given to the participants. The participants were first given already classified galaxies to determine their accuracies and to give their classifications credibility. Because every galaxy was classified by several people (between 40 to 50), in order to improve classification accuracy, these numbers can be represented as probabilities of a galaxy being of a particular morphological type.

3.2 Data augmentation

Our data augmentation and preprocessing done in our work is based on the work of Dieleman et al. (2015). We used python (van Rossum (1995)), sci-kit image (van der Walt et al. (2014)) and numpy module (Dubois et al. (1996)) for all data augmentation and preprocessing.

We make random augmentations of images in every epoch. Epoch is a period in which neural network trains once on all training images. Images are supplied to the network by a generator. The generator takes raw images as the input, makes all transformations to make sixteen 45×45×3 images of every raw image (those transformations are further described in section "Data preprocessing") and yields them in batches as the input to neural network. A random augmentation is applied on the raw images before all those transformations and it consists of random shift in range (-4, 4) pixels, random rotation in range (0, 360) degrees and random zoom in range (1/1.3, 1.3). Those augmentations are small enough to preserve all the morphological shapes of galaxies, but big enough to decrease the neural networks overfitting.

3.3 Data preprocessing

Neural networks highly benefit from larger training datasets.

For creating a larger dataset, we make use of the fact that galaxies are invariant to rotations and translations. If we flip the image of galaxy, rotate or translate it, it will always be the same galaxy, which means that the same label can be used for all of the new versions of the galaxy.

All the additional augmentations we make are done to as described in previous section already randomly augmented raw images. The raw images (after augmentation) are firstly flipped. Then both versions, (flipped and not flipped), are rotated by 45 degrees in anti-clockwise direction, which makes 4 different views of 1 raw image. Figure 3.1 shows an example of those transformations on image with id 448630. All those 4 views still have dimensionality $424\times424\times3$.

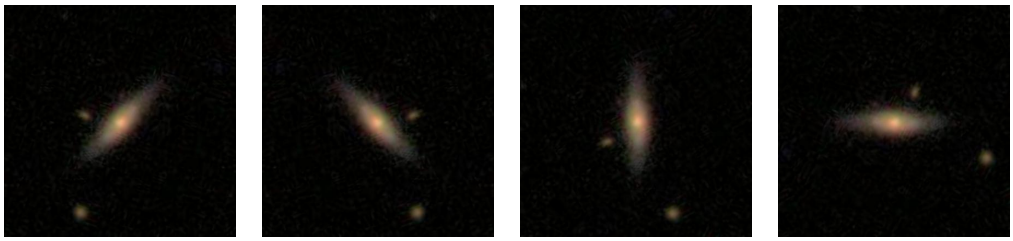
The next step is cropping those images to $207\times207\times3$ and downsampling by factor of 3 into $69\times69\times3$ dimensions. This transformation is shown in Figure 3.2.

The last step creates 4 new versions of every of those four $69\times69\times3$ images. Every $69\times69\times3$ image is cut to four $45\times45\times3$ partially overlapping images. The overlapping part is in the centre of the image in order to gain more information about the galactic nuclei, as more of the questions that the neural network is trained to answer are about the features in the galactic nuclei then further away.

All of those transformations of raw images into 16 $45\times45\times3$ images needs to be done to classification as well.

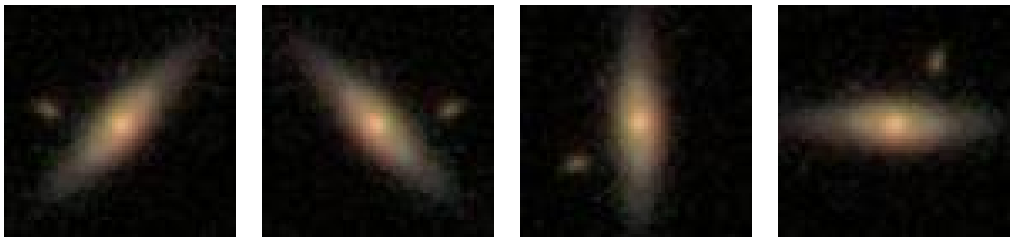
This process creates exactly 16 times larger dataset and significantly improved the performance. Another benefit from the increased amount of training data lies in the reduction of the overfitting effect. The reason is very logical and simple: it is more difficult for a neural network to simply memorize 16 times more images.

We decided to keep the first 6400 images of the dataset for the testing. Those images were not participating in any form in the actual training. The last 6 170 images from the dataset were selected for validation purposes so the neural network was trained only on the remaining 49 008 images.



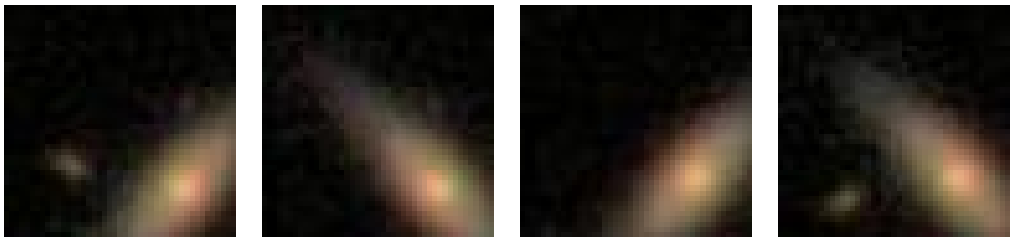
(a) (b) (c) (d)

Figure 3.1: (a) raw image, (b) flipped raw image, (c) rotated raw image by 45 degrees, (d) flipped and rotated raw image by 45 degrees. Rotations are made in anti-clockwise direction.



(a) (b) (c) (d)

Figure 3.2: Downsampling and cropping transformation performed at images from Figure 3.1 resulting in images of 69×69×3 dimensionality.



(a) (b) (c) (d)

Figure 3.3: Four 45×45×3 dimensional images. They are cropped versions of the image from Figure 3.2a and rotated to have their galactic nuclei in bottom right corner. The same transformations are performed on the other three images from Figure 3.2 as well, which results in 16 times larger dataset.

3.4 Presented Neural network

We present an innovation (Table 3.2) of the S. Dielemans neural network (Dieleman et al. (2015), Figure 3.4, Table 3.1). It is based on a combination of ideas, which gave rise to the inception neural network module (Szegedy et al. (2014))

and the "bottleneck" architecture. We build our neural network using Keras (Chollet et al. (2015)) environment with Theano (Theano Development Team (2016)) backend.

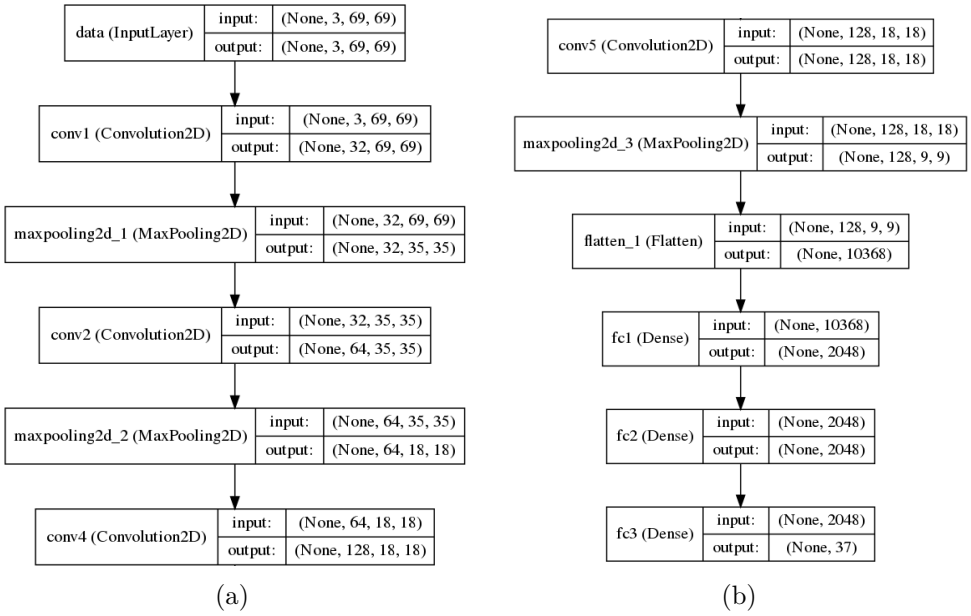


Figure 3.4: The architecture of the S. Dielemans (Dieleman et al. (2015)) neural network split to 2 parts for clarity ((b) continues below (a)).

	type	**features**	**filter size**	**non-linearity**	**initial weights**
1	convolution	32	6×6	ReLU	$N(0,0.01)$
1p	max-pooling	-	2×2	-	-
2	convolution	64	5×5	ReLU	$N(0,0.01)$
2p	max-pooling	-	2×2	-	-
3	convolution	128	3×3	ReLU	$N(0,0.01)$
4	convolution	128	3×3	ReLU	$N(0,0.1)$
4p	max-pooling	-	2×2	-	-
5	dense	2048	-	maxout (2)	$N(0,0.001)$
6	dense	2048	-	maxout (2)	$N(0,0.001)$
7	dense	37	-	constrains	$N(0,0.01)$

Table 3.1: The architecture of the S. Dielemans neural network (Dieleman et al. (2015)). Convolutional layers **1**, **2** and **4** are followed by max-pooling layers. All max-pooling layers (**1p**, **2p** and **4p**) have 2×2 filter size and stride 2. Initial biases were everywhere 0.1 except of the dense layers **5** and **6** with initial biases 0.01.

	type	features	filter size	non-linearity
1	convolution	32	6×6	ReLU
1p	max-pooling	-	2×2	-
2	convolution	64	5×5	ReLU
2p	max-pooling	-	2×2	-
3a	convolution	128	1×1	ReLU
3b	convolution	128	1×1	ReLU
3bb	convolution	128	3×3	ReLU
3c	convolution	128	1×1	ReLU
3cc	convolution	128	5×5	ReLU
3d	max-pooling	-	2×2	-
3dd	convolution	128	1×1	ReLU
3merger	merger	512	-	ReLU
4	convolution	128	3×3	ReLU
4p	max-pooling	-	2×2	-
5	dense	2048	-	ReLU
6	dense	2048	-	ReLU
7	dense	37	-	constrains

Table 3.2: The architecture of the presented neural network. The inception module consists of layers starting with number **3**. Layer **3merger** concatenates the filters of layers **3a**, **3bb**, **3cc** and **3dd**. This merged layer is an input of convolutional layer **4**, which is a "bottleneck" of our architecture. All layers are initialized with glorot-uniform initialization (Glorot and Bengio (2010)) (except of max-pooling layers which do not have any weights.

3.4.1 Inception module

The basic idea of the inception module is to connect the information from further away of observed point with information very close to observed point (Szegedy et al. (2014)). This is exactly what concatenation of convolutions with filter sizes 5×5(**3c**), 3×3(**3bb**) and 1×1(**3a**, **3dd**) does. Because galaxy morphology is not of a homogeneous distribution, but is characterized by different distribution of matter in space (mainly spiral, merger and irregular galaxies) we found this idea perfectly fitting for our task.

Original inception module used in GoogleNet (Szegedy et al. (2014)) does not have similar numbers of feature maps (numbers in column **features** in Table 3.2) for all of its convolutional layers, but they decreased for convolutional layers with bigger kernel sizes. It was made like this because the creators of the inception module worked on a task where the spacial correlation between points very close close to each other was much stronger than further away. This is why they originally decreased numbers of feature maps with increasing the filter sizes in the convolutional layers in the inception module. We found that the original architecture of inception module did not gave us as good results, as if all layers

had same number of features, which is in our opinion due to the fact that the morphology of galaxies has strong correlation even between points further away.

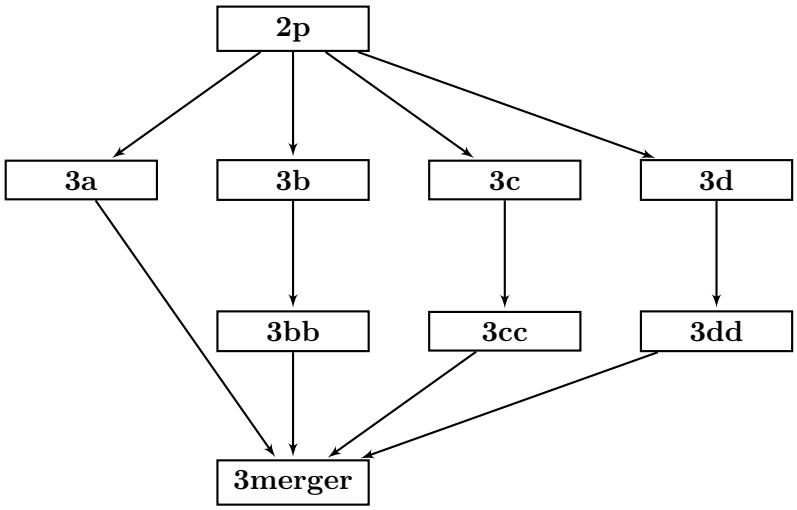


Figure 3.5: Implementation of inception module in our network (Table 3.2).

3.4.2 Bottleneck

Deep neural networks have a problem with non-learning neurons they might be more than 20 % in a very well working neural networks). Bottleneck layer is a layer which has significantly lower dimensionality than its neighbouring layers. Decrease of the dimensionality is helpful because it decreases the learning time, but the main reason we use the bottleneck is that it increases the networks robustness and reduces the overfitting. The neurons with higher activations tend to pass through bottleneck more than those with lower activations (Tishby et al. (2000)). It can be also seen as a layer reducing the sensitivity to the noise in the data. Bottleneck layer in our networks is convolutional layer **4** shown in Table 3.2.

3.4.3 How does our convolutional neural network see the world

In Figures 3.6 and 3.7, of the first convolutional layer **1** in our neural network (Table 3.4) are shown the filters. In general, first layers usually detect the most basic shapes. For example, filters 3.6k and 3.7c seem to detect edges. Filters 3.6l and 3.7g looks like curve detectors. Neural networks do not use only one filter for classification, but combination of all.

All of this gets much more complicated when the information passes to the deeper layers as they do not have classic 3 rgb spacial dimensionality as images do. Those layers are learning more complex features with increasing depth.

Figure 3.6: Filters 1-16 of first convolutional layer of our architecture.

Figure 3.7: Filters 17-32 of first convolutional layer of our architecture.

Chapter 4

Results

4.1 Classification of selected galaxies

We tested our neural network on 6 400 images of SDSS and 5 additional images. Those were images of galaxies M51 (Figure 4.1, 4.3), NGC 1365 (Figure 4.2), M51 (Figure 4.4) and UGC 12336 (Figure 4.5). We preprocessed those images manually into 424×424×3 spacial dimensionality, so they could be classified by our neural network.

Figure 4.1: Image of galaxy M51 captured by students of Faculty of Science, Masaryk University Department of Theoretical Physics and Astrophysics in observatory of city Vyškov. This image is a collage made of multiple images in different bands combined together to classical rgb JPEG image by Mgr. Filip Hroch, PhD. Our neural network made following predictions $P(elliptical) = 0.2\,\%$; $P(spiral) = 94.6\,\%$; $P(star\ or\ artifact) = 5.3\,\%$.

Figure 4.2: Figure of galaxy NGC 1365 taken by DES (The Dark Energy Survey Collaboration (2005)). Our neural network made following predictions $P(elliptical) = 3.1\,\%$; $P(spiral) = 79.7\,\%$; $P(star\ or\ artifact) = 17.2\,\%$.

4.2 ROC curves

The receiver operating characteristic curve (ROC) characterizes quality of a detection system as a trade-off between the probability of miss ($Miss$) against probability of false alarm (FA). The closer the curve copies the left vertical axis and the bottom horizontal axis, the more accurate the detection system is. The closer the curve comes to the shape of the diagonal curve connecting points [0, 100] and [100, 0], the less accurate it is. We show the ROC curves as our results according to the tasks (Table 2.7) of the Galaxy Challenge.

Figure 4.3: Figure of galaxy M31 1365 taken by Pan-STARRs (Magnier et al. (2013)). Our neural network made following predictions $P(elliptical) = 33.2\,\%$; $P(spiral) = 64.5\,\%$; $P(star\ or\ artifact) = 0.2\,\%$.

Figure 4.4: Figure of galaxy M51 1365 taken by Pan-STARRs (Magnier et al. (2013)). Our neural network made following predictions $P(elliptical) = 0.2\,\%$; $P(spiral) = 97.8\,\%$; $P(star\ or\ artifact) = 1.9\,\%$.

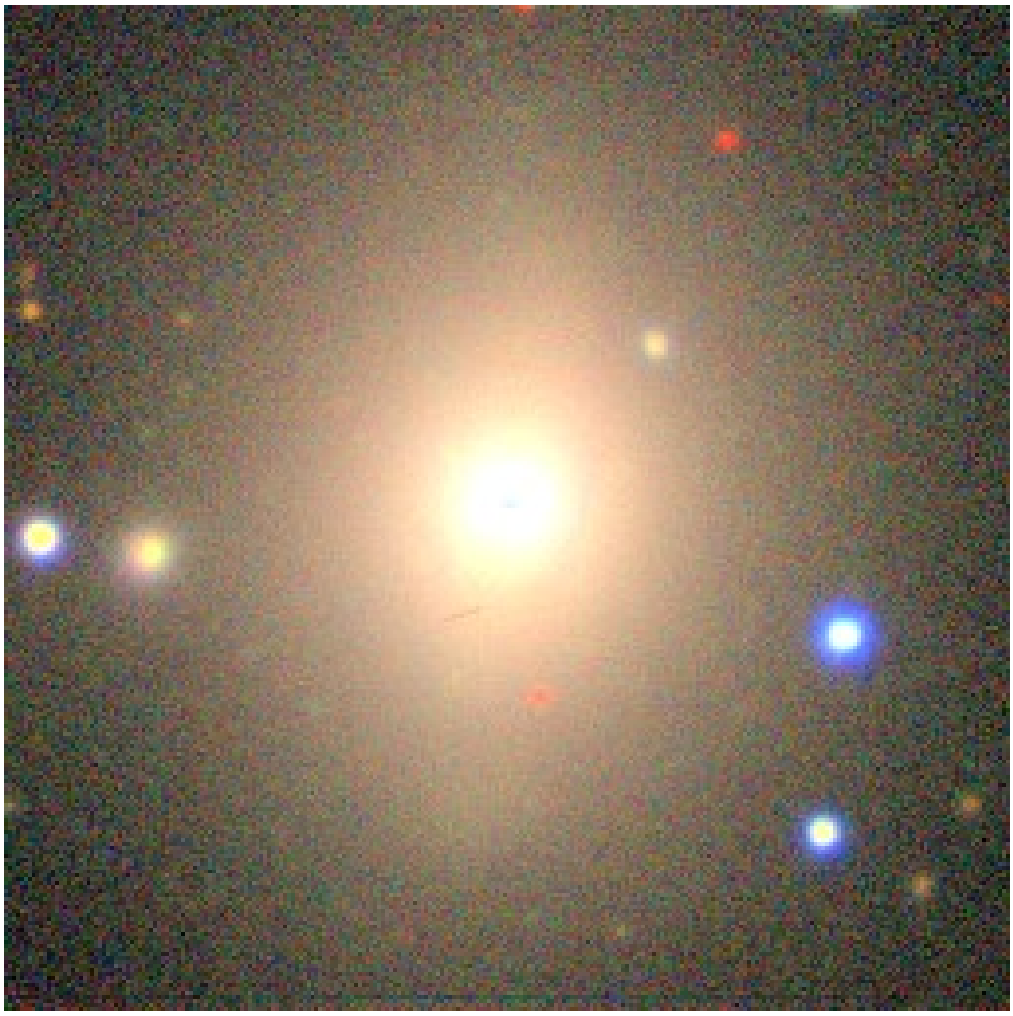


Figure 4.5: Figure of galaxy UGC 12336 taken by DES (The Dark Energy Survey Collaboration (2005)). Our neural network made following predictions $P(elliptical) = 75.8\,\%$; $P(spiral) = 1.2\,\%$; $P(star\ or\ artifact) = 24.0\,\%$.

Task	Questions	P_1 [%]	P_2 [%]	P_3 [%]	P_4 [%]	P_5 [%]
01	smooth	0.2	3.1	33.2	0.2	75.8
	features or disk	94.6	79.7	64.5	97.8	1.2
	star or artifact	5.3	17.2	0.2	1.9	24.0
02	edged-on – yes	3.6	8.3	23.7	1.7	0.0
	edged-on – no	91.0	71.4	40.8	96.0	0.1
03	bar – yes	16.3	42.5	12.0	35.3	0.0
	bar – no	74.7	28.9	28.7	60.6	0.0
04	spiral arm pattern – yes	89.3	52.6	25.7	95.8	0.0
	spiral arm pattern – no	1.7	18.8	15.1	0.3	0.1
05	no bulge	0.8	11.2	5.1	4.8	0.0
	just noticeable	18.8	15.0	14.8	28.7	0.0
	obvious	52.5	27.5	14.4	44.7	0.0
	dominant	18.9	17.7	6.4	17.8	0.0
06	anything odd – yes	63.7	99.3	62.4	64.0	36.8
	anything odd – no	36.2	0.7	37.6	36.0	63.2
07	completely round	0.1	0.9	0.0	0.1	49.7
	in between	0.0	1.0	14.1	0.1	26.1
	cigar-shaped	0.0	1.1	19.1	0.0	0.0
08	ring	15.4	19.5	6.1	12.6	1.2
	lens or arc	8.9	0.8	0.0	4.6	2.8
	disturbed	10.7	3.0	14.2	7.9	5.4
	irregular	4.0	0.0	17.7	0.3	3.5
	other	17.7	45.0	12.0	18.2	19.7
	merger	7.1	30.1	12.4	20.3	0.4
	dust lane	0.0	0.0	0.0	0.0	0.0
09	rounded	2.3	4.4	15.9	1.1	0.0
	boxy	0.3	0.9	2.6	0.2	0.0
	no bulge	0.9	2.9	5.2	0.4	0.0
10	tight	34.7	2.3	7.8	41.7	0.0
	medium	37.5	16.2	8.5	41.1	0.0
	loose	17.2	34.0	9.4	13.0	0.0
11	1	16.3	3.8	3.8	8.6	0.0
	2	22.3	40	9.5	43.1	0.0
	3	9.3	0.1	0.6	13.5	0.0
	4	0.2	0.0	0.0	0.7	0.0
	more than four	19.3	5.8	0.4	14.8	0.0
	can't tell	21.8	2.8	11.5	15.1	0.0

Table 4.1: Table of classifications provided by our neural network. P_1 corresponds to Figure 4.1, P_2 to Figure 4.2, P_3 to Figure 4.3, P_4 to Figure 4.4 and P_5 to Figure 4.5.

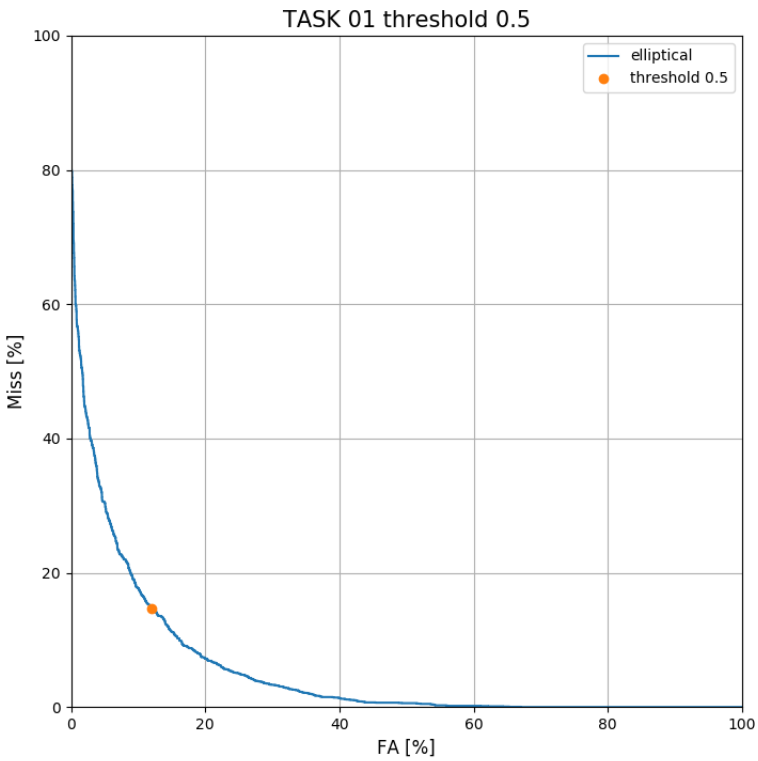


Figure 4.6: Figure of ROC curves for TASK 01, category “elliptical” (Table 2.7) with treshold at 0.5.

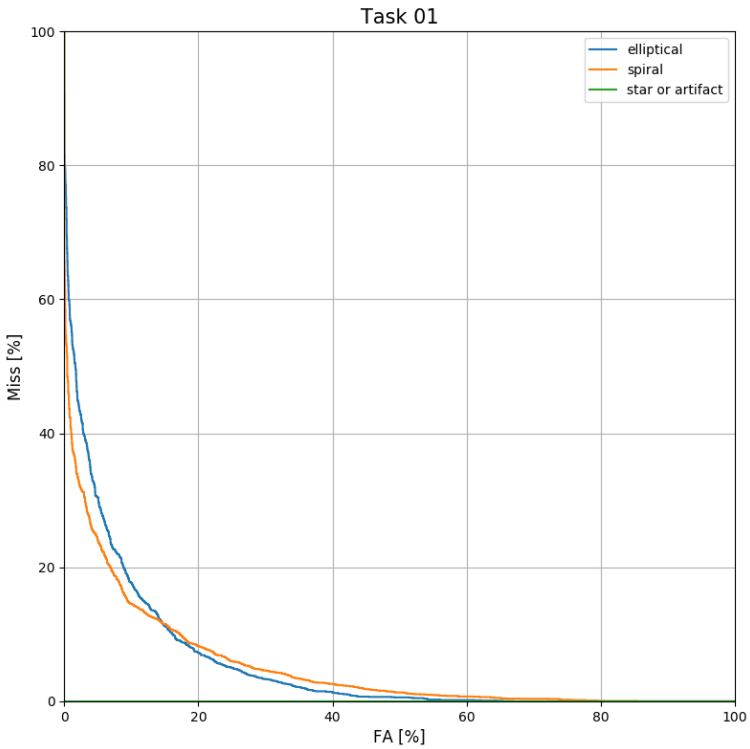


Figure 4.7: Figure of ROC curves for TASK 01, where our neural network classified the object as elliptical or spiral galaxy, star or artifact (Table 2.7).

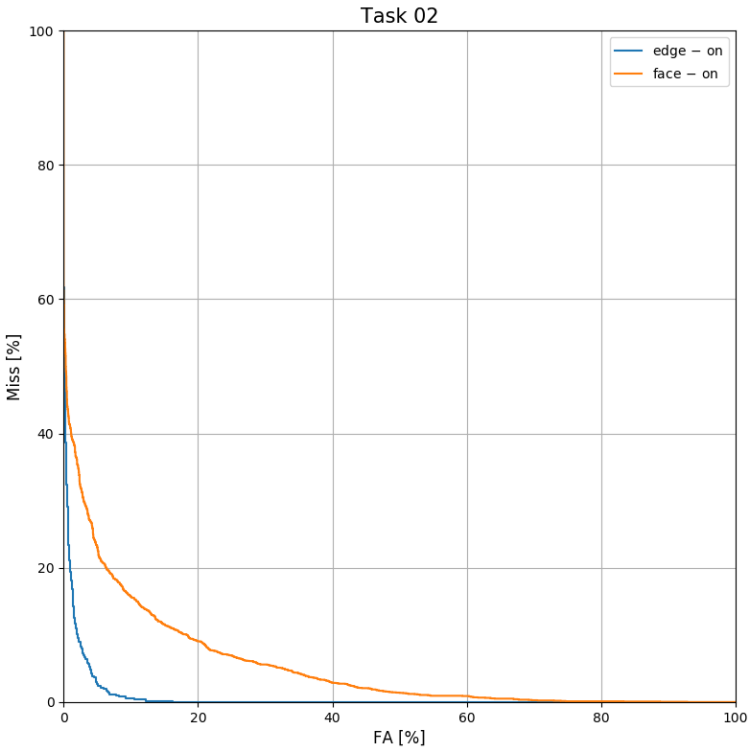


Figure 4.8: Figure of ROC curves for TASK 02, where our neural network had to choose if is the galaxy edged-on od faced-on (Table 2.7).

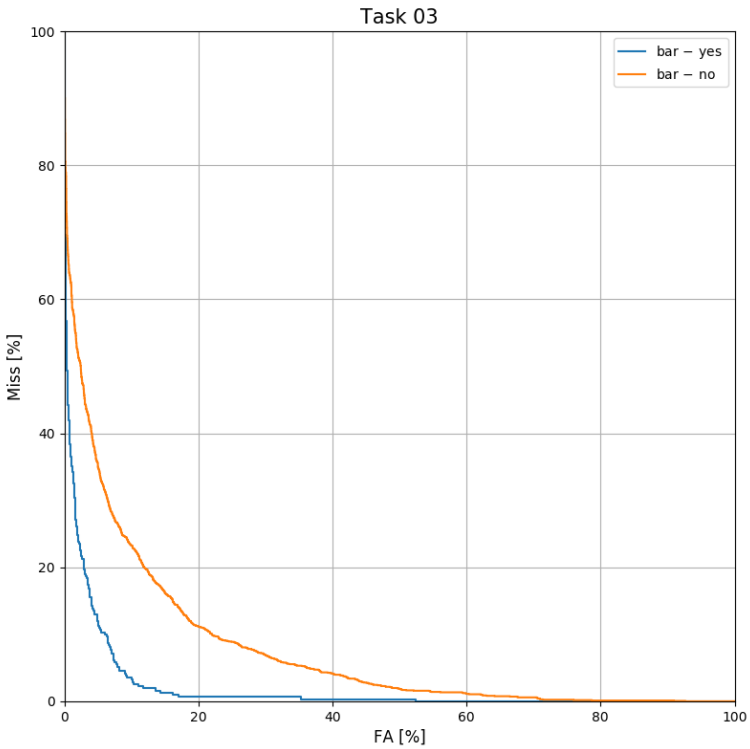


Figure 4.9: Figure of ROC curves for TASK 03, in which our neural network classified the galaxy according to presence of a bar (Table 2.7).

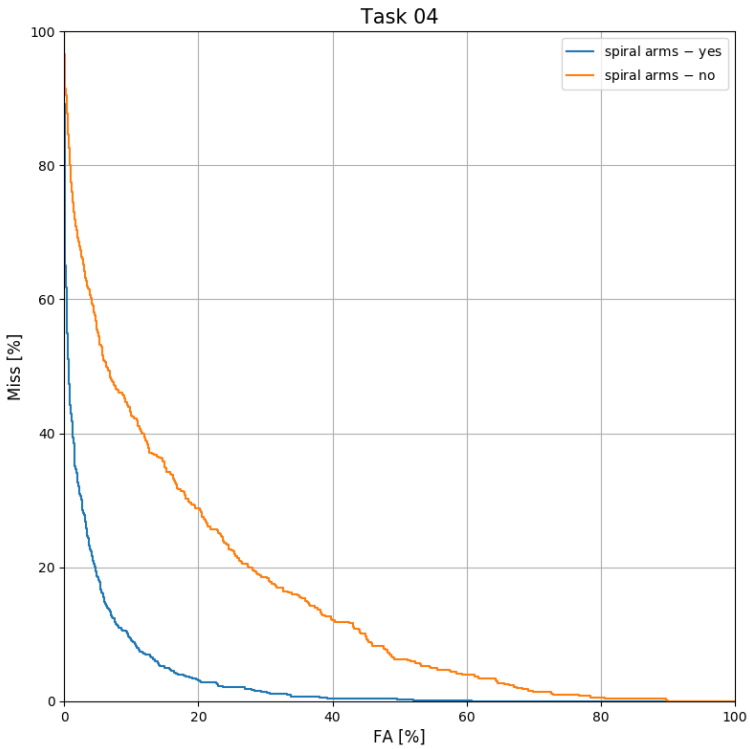


Figure 4.10: Figure of ROC curves for TASK 04, whether galaxy consists of spiral arms or not (Table 2.7).

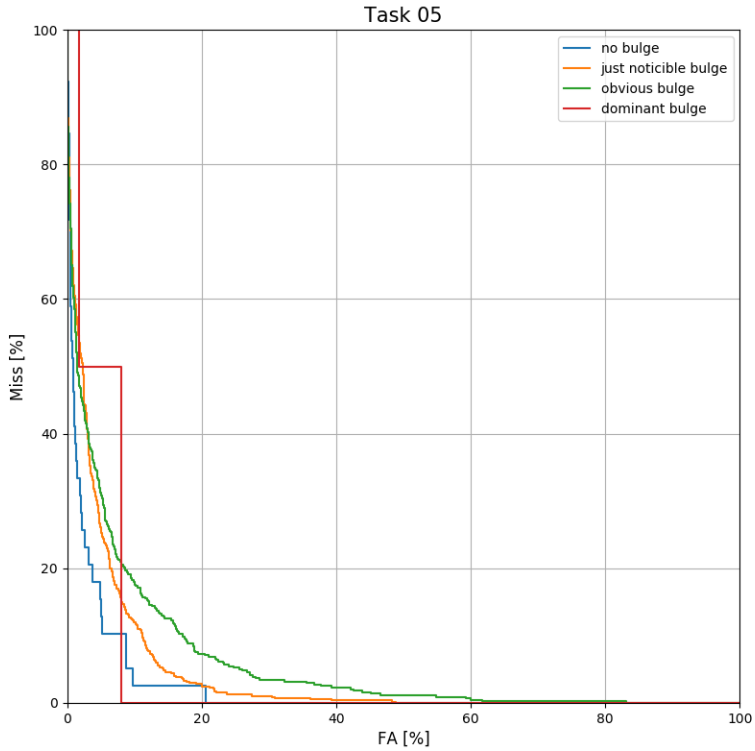


Figure 4.11: Figure of ROC curves for TASK 05, about classification whether galaxy has no bulge, just noticeable, obvious or dominant bulge (Table 2.7).

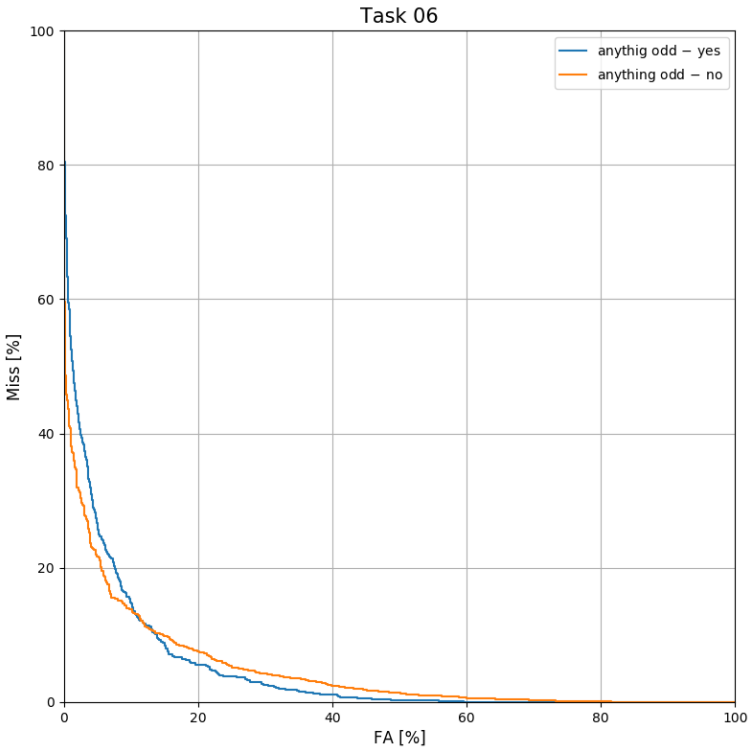


Figure 4.12: Figure of ROC curves for TASK 06, in which the network determines whether there is anything odd (Table 2.7).

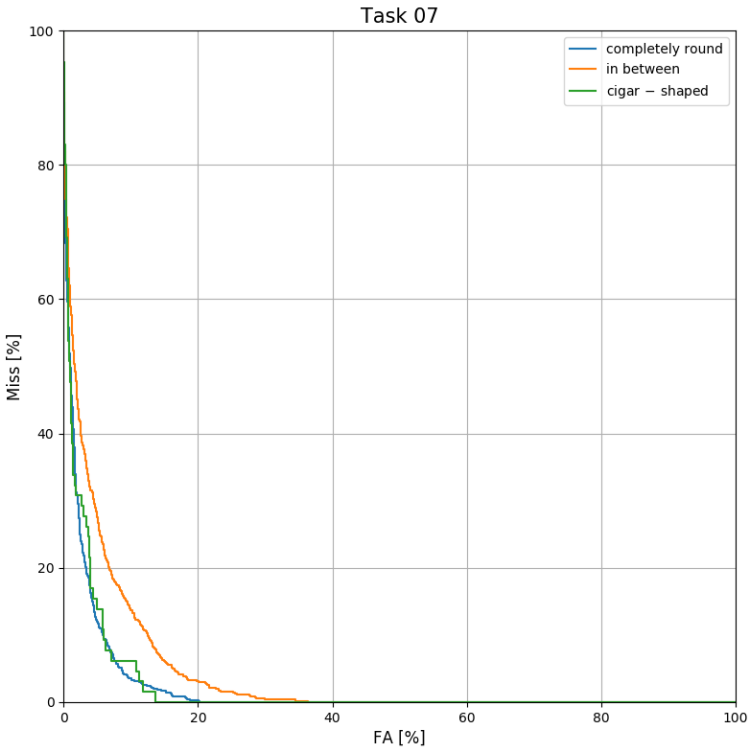


Figure 4.13: Figure of ROC curves for TASK 07, where our neural network classifies elliptical galaxies into round, cigar-shaped or in between (Table 2.7).

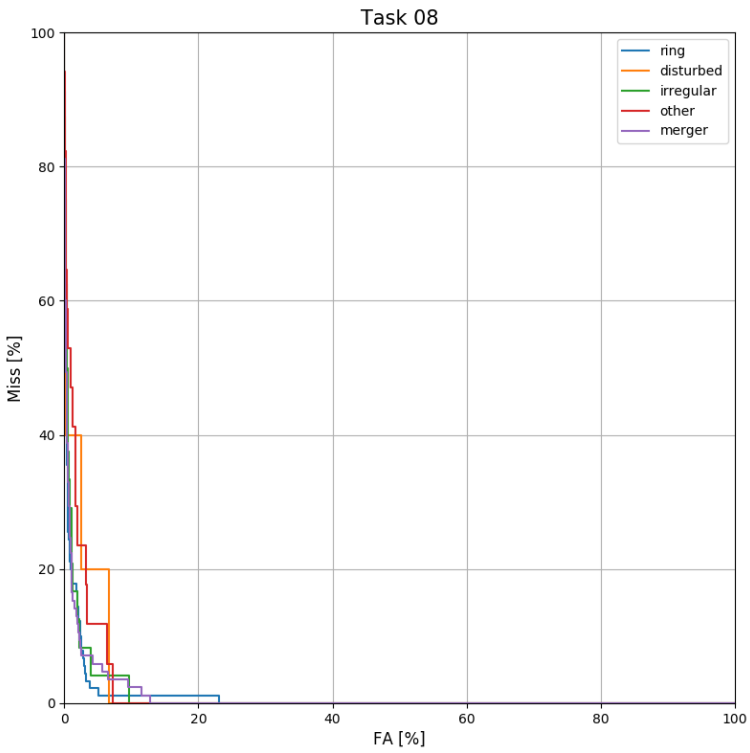


Figure 4.14: Figure of ROC curves for TASK 08, where odd miscellaneous features like rings, dust lanes, lenses, arcs, disturbances, irregularities, mergers or other are being classified (Table 2.7). 2.7).

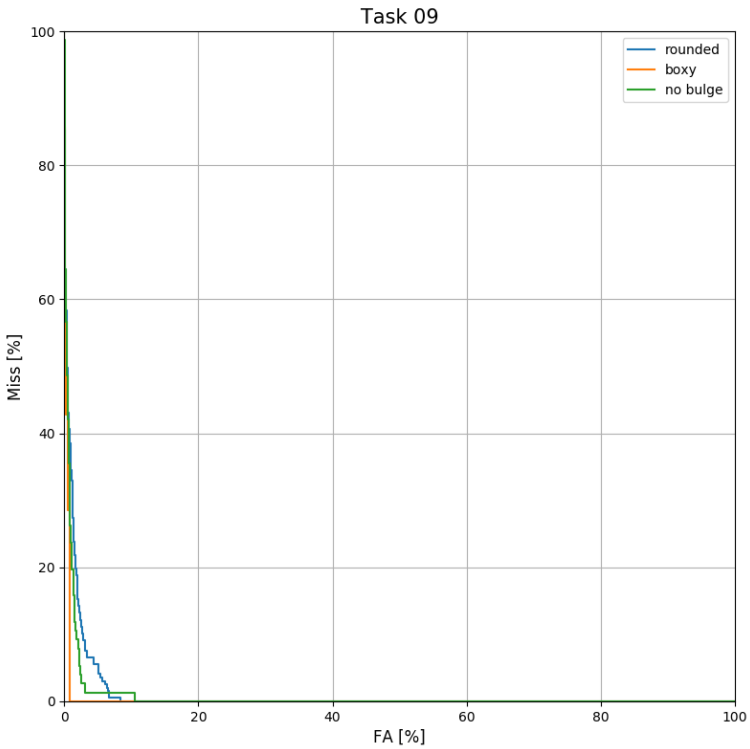


Figure 4.15: Figure of ROC curves for TASK 09, in which the bulge is classified as either rounded, boxy, or missing (Table 2.7).

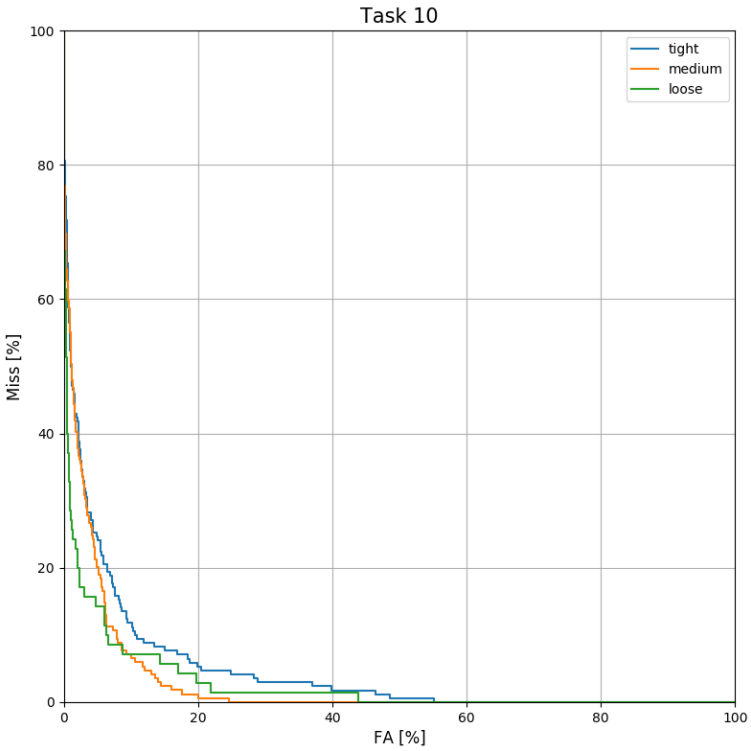


Figure 4.16: Figure of ROC curves for TASK 10, where our neural network decides whether the spiral arms are wound tightly, loosely, or in-between (medium)."

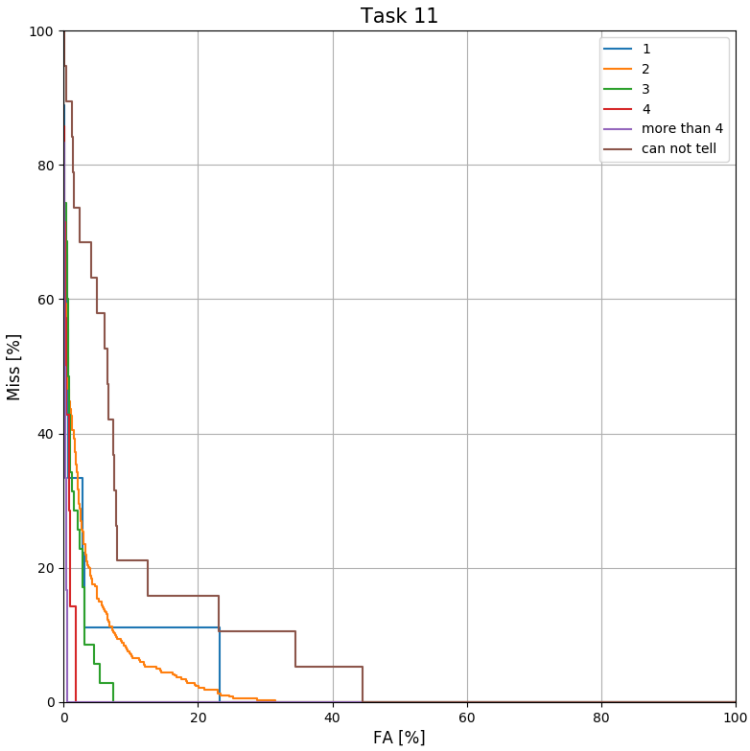


Figure 4.17: Figure of ROC curves for TASK 11, where our neural network decides how many spiral arms the galaxy has (Table 2.7).

Chapter 5

Discussion and Future plans

5.1 Discussion

5.1.1 Classification of selected galaxies

The classification accuracy of the people participating in the Galaxy Challenge decreased with the more complex questions deeper in the decision tree (Figure 1.5). This is the reason why our neural networks accuracy decreases with the increasing complexity of morphology classification. We can not expect our network to produce perfect results for questions that even people disagree about. For example, classification of both pictures of galaxy M51 (Figure 4.1 and 4.4) of tasks 01 (probabilities of 94.6 % (Figure 4.1) and 97.8 % (Figure 4.4) of spiral category), task 02 (probabilities of 91.0 % (Figure 4.1) and 96.0 % (Figure 4.4) of edge-on – no category) and task 04 (probabilities of 89.3 % (Figure 4.1) and 95.8 % (Figure 4.4) of spiral arm pattern – yes category) proved great results. On the other hand, classification of task 03 for NGC 1365 (Figure 4.2) correctly gave bigger value for bar – yes ccategory, but it is not as good, as in the example with M51. This is exactly due to fact, that even people participating in Galaxy Challenge performed worse at classification of more complex tasks.

We want to stress, that the probabilities for the categories deeper in the decision tree (Figure 1.5) do not sum to 100 %, instead they must sum to their "parent" category. For example, the probabilities of the categories of task 03 (bar – yes and bar – no) must sum to the probability given to the category edge-on – no of task 02.

The classification of UGC 12336 (Figure 4.5) nicely excluded the possibility of the spiral galaxy category by giving it only 1.2 % probability. However, the network gave 24.0 % probability for the star or artifact category and 75.8 % probability for the elliptical category. It happened because of the presence of a parasite stars near the centre of the image.

5.1.2 ROC curves

Figure 4.6 shows ROC curve for TASK 01 category "elliptical" with chosen threshold of 0.5. All points of ROC curves are defined by certain threshold value. Different threshold values corresponds to the different combination of

the probability of false alarm FA [%] and probability of miss $Miss$ [%]. FA for this example of TASK 01 "elliptical" category would happen, if networked classified spiral galaxy as an elliptical. $Miss$ would happen for classification of elliptical galaxy as a spiral galaxy. Our networks results are $FA = 12.0\,\%$ and $Miss = 14.7\,\%$. It means, that our network classified 12.0 % of spiral galaxies in our testing dataset as elliptical galaxies and 14.7 % of elliptical galaxies in our testing dataset as spiral galaxies for threshold value = 0.5. TASK 01 is basically not a binary problem, there is also the category "star or artifact", but we do not consider its share as relevant because our test dataset consists of 3493 spiral galaxies, 2745 elliptical galaxies and only 2 stars or artifacts.

Figure 4.7 of ROC curves of TASK 01 (2.7) shows, that the curve for the category "star or artifact" copies exactly the left vertical and bottom horizontal axis. In this case, this does not mean perfect, 100 % result. However, it indicates that our 6 400 test images included just very few answers for the "star or artifact" category. It is vital to understand that stars or artifacts are really easily distinguishable from galaxies, which is why those few examples in our dataset were perfectly classified.

Figure 4.8 shows that the classification of "edge-on – yes" spiral galaxies was much easier than the classification of "edge-on – no" spiral galaxies for our neural network. We interpret this as the fact that the "edge-on – no" spiral galaxies can be miss classified as elliptical galaxies more likely than the "edge- – yes" spirals.

Figure 4.9 shows a similar situation as Figure 4.8. The detection of barred spiral galaxies was much easier due to the fact that the bar structure is kind of unique structure, which makes it easier for a neural network to classify.

The classification of the spiral galaxies with arm patterns was very good in comparison with the classification of the spiral galaxies without spiral arm patterns (Figure 4.10). This may be mainly result of the similarity between the elliptical galaxies and the spiral galaxies lacking the spiral arm patterns.

The ROC curve corresponding to the detection of spiral galaxies with dominant bulge (Figure 4.11) can not be reliable as it consists only of a few points. This is a consequence of poor representation of this morphological type in our test dataset. The same applies to ROC curve of galaxies without bulge.

Figure 4.12 shows two ROC curves for TASK 06 (Table 2.7).

Figure 4.13 shows ROC curves for TASK 07 (Table 2.7). It shows that the classification of completely round elliptical galaxies is easier than classification of the other types.

The ROC curves showd in Figure 4.14 have nice tendency to copy the left vertical and the bottom horizontal axis, but they are made only of few points, so they can not be considered reliable. The ROC curves for the categories "lens or arcs" and "dust lane" could not be showd as there were no data for them in our test dataset.

Figure 4.15 shows that most of the spiral galaxies in our test dataset consists of rounded bulges because the ROC curve for "rounded" category consists of most points.

Figure 4.16 shows that the classification of the spiral galaxies with medium tight spiral arms resulted with better performance than the classification of spiral galaxies with tight spiral arms. It also shows that the representation of spiral galaxies with loose spiral arms is much poorer that the other two types.

Figure 4.17 nicely shows that the majority of the spiral galaxies in our test dataset consists of the spiral galaxies with 2 spiral arms because the ROC curve for this category consists of most points (most answers). The other categories do not even have enough points to be reliably described.

5.2 Future plans

5.2.1 Discovering and classifying galaxies

There are several ways we can continue our work. Our neural network was trained on 424×424×3 JPEG rgb colour images of galaxies which were in the centre of the images. Therefore, our network can only classify images similar to the training images (images with centred galaxy and same dimensionality and format). One possible future improvement of our network is to training our network so that it could find all galaxies in its input images and classify them.

5.2.2 Gravitational Lenses

A very interesting field of study is discovering gravitational lenses. For example, gravitational lenses are used to understand the nature of mysterious dark matter. The light is bend around the massive galaxy between us and the distorted galaxy much further away due to the gravitational influence. The light passing from the distant galaxy is not bended only due to the influence of the gravity of the ordinary matter of galaxy that is lensing the light, but also because of the gravity of the mass of the dark matter around the galaxy. The gravitational lenses can be used to measure perturbations in fluctuations of the dark matter particles (Moustakas et al. (2009)). Stage 1 of the crowd-sourcing Zooniverse project called Space warps (Marshall et al. (2016)) already made a great observations. Over 37 000 participants made more than 11 million classifications in a period of 8 months. Stage 1 resulted in 3381 candidates for gravitational lenses. Those were further examined in Space warps Stage 2 (More et al. (2016)) resulting in 29 promising of 59 total gravitational lenses candidates. The goal of this search was to identify the gravitational lens candidates possibly missed by the robots, which previously searched for the gravitational lenses in Canada France Hawaii Telescope (Kalirai et al. (2001)) images. Those new classifications can be used to re-train the existing robots, or, what we consider as our possible future goal, to make even better classifier.

5.2.3 Supernovae

Discovering supernovae as soon as possible is crucial for analysing the properties of stars exploded to form the supernovae and processed of explosions start up and proceeding. Supernovae (particularly type Ia) are also important for astronomy and cosmology for calculating distances in the universe. The Ia supernovae are used as "standard candles" because they always explode under very similar conditions having same absolute magnitude. Other supernovae types can be used to calculate distances as well. Zooniverse project Snapshot Supernova is a crowd-sourcing project aimed to discover new supernovae using images from Sky Mapper telescope (Keller et al. (2007)) and Public ESO Spectroscopic Survey of Transient Objects (Smartt et al. (2015)). More than 40 000 participating volunteers provided more than 1.9 million classifications resulting in another dataset, which can be used to train neural networks for discovering supernovae.

Chapter 6

Summary and Conclusion

6.1 Summary

We present convolutional neural network as a solution to galaxy classification. Our neural network correctly classified more than 85 % elliptical and 87 % spiral galaxies of our testing dataset. It also proved to be capable of detecting bars, spiral patterns, and bulges in spiral galaxies. Even though its success rate was lower for the progressively more detailed tasks, such as identifying the shape of the bulge or finding dust-lanes, future training on larger dataset is expected to substantially improve our network. The biggest scientific impact of our work lies in future determination of the galaxy morphology density relation as a function of redshift, which requires vast numbers of objects being classified.

6.2 Conclusion

Invention of artificial algorithms for data processing is inevitable and logical step forward together with tremendous enlargement of vast datasets, which is, and will not be possibly processed otherwise.

Our implementation of convolutional neural network proved great results for morphology classification of galaxies. It can be used right now for various astrophysical purposes. For example, if we want to select galaxies of precise morphological type from an unclassified sample, we can use our network to make predictions for all of them. Afterwards, we would select just those, for example, with 95 % probability of needed category. On the other hand, if we needed to select as much, for example elliptical galaxies, as possible, we could look for those, with probabilities of being an elliptical galaxy greater, than some still acceptable value. This approach would lead to finding more objects, but with greater amount of miss classified spiral galaxies as an elliptical galaxies.

Our network can be also further retrained on larger training dataset of classified galaxies, which will increase its performance.

Bibliography

[1] cs231n: Convolutional neural networks for visual recognition. http://cs231n.github.io/convolutional-networks/.

[2] the galaxy zoo decission tree. https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/details/the-galaxy-zoo-decision-tree.

[3] the photometric camera and the ccds. http://www.astro.princeton.edu/PBOOK/camera/camera.htm.

[4] neural networks and deep learning. http://neuralnetworksanddeeplearning.com/index.html.

[5] the dark energy survey overwiev. http://www.darkenergysurvey.org/the-des-project/overview/.

[6] pan-starrs releases catalogue of 3 billion astronomical sources. https://phys.org/news/2016-12-pan-starrs-catalogue-billion-astronomical-sources.html.

[7] convolutional neural networks (cnns): An illustrated explanation. http://xrds.acm.org/blog/2016/06/convolutional-neural-networks-cnns-illustrated-explanation/.

[8] "neural networks part 1: Setting up the architecture." notes for cs231n convolutional neural networks for visual recognition, stanford university. http://cs231n.github.io/neural-networks-1/.

[9] when does deep learning work better than svms or random forests? http://www.kdnuggets.com/2016/04/deep-learning-vs-svm-random-forest.html.

R. G. Abraham and S. van den Bergh. The Morphological Evolution of Galaxies. *Science*, 293:1273–1278, August 2001. doi: 10.1126/science.1060855.

E. Bertin. *Science with Astronomical Near-Infrared Sky Surveys: Proceedings of the Les Houches School, Centre de Physique des Houches, Les Houches, France, 20–24 September, 1993.* Springer Netherlands, 1993. ISBN 9789401109468.

L. G. Book and A. J. Benson. The Role of Ram Pressure Stripping in the Quenching of Cluster Star Formation. ApJ, 716:810–818, June 2010. doi: 10.1088/0004-637X/716/1/810.

Y. Cao, P. E. Nugent, and M. M. Kasliwal. Intermediate Palomar Transient Factory: Realtime Image Subtraction Pipeline. PASP, 128(11):114502, November 2016. doi: 10.1088/1538-3873/128/969/114502.

M. Cappellari, E. Emsellem, D. Krajnović, R. M. McDermid, and P. et al. Serra. The ATLAS^{3D} project - VII. A new look at the morphology of nearby galaxies: the kinematic morphology-density relation. MNRAS, 416:1680–1696, September 2011. doi: 10.1111/j.1365-2966.2011.18600.x.

François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

A. A. Collister and O. Lahav. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. PASP, 116:345–351, April 2004. doi: 10.1086/383254.

M. Davis and M. J. Geller. Galaxy Correlations as a Function of Morphological Type. ApJ, 208:13–19, August 1976. doi: 10.1086/154575.

G. de Vaucouleurs. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53:275, 1959.

S. Dieleman, K. W. Willett, and J. Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. MNRAS, 450:1441–1459, June 2015. doi: 10.1093/mnras/stv632.

A. Dressler. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. ApJ, 236:351–365, March 1980. doi: 10.1086/157753.

Paul F. Dubois, Konrad Hinsen, and James Hugunin. Numerical python. *Computers in Physics*, 10(3), May/June 1996.

G. B. Field and W. C. Saslaw. Groups of Galaxies: Hidden Mass or Quick Disintegration? ApJ, 170:199, December 1971. doi: 10.1086/151203.

Folkes et al. *Highlights of Spanish Astrophysics II*. Springer, 1996. ISBN 9780792369745.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, volume 9, pages 249–256, May 2010.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of*

the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11), volume 15, pages 315–323. Journal of Machine Learning Research - Workshop and Conference Proceedings, 2011.

J. E. Gunn and J. R. Gott, III. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. ApJ, 176:1, August 1972. doi: 10.1086/151605.

G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv e-prints*, July 2012.

Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 6(2):107–116, April 1998. ISSN 0218-4885. doi: 10.1142/S0218488598000094.

Sepp Hochreiter, A. Steven Younger, and Peter R. Conwell. *Learning to Learn Using Gradient Descent*, pages 87–94. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. ISBN 978-3-540-44668-2. doi: 10.1007/3-540-44668-0_13.

K. Honscheid, D. L. DePoy, and for the DES Collaboration. The Dark Energy Camera (DECam). *ArXiv e-prints*, October 2008.

E. Hubble. No. 324. Extra-galactic nebulae. *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 324:1–49, 1926.

Z. Ivezic, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, and AlSayyad et al. LSST: from Science Drivers to Reference Design and Anticipated Data Products. *ArXiv e-prints*, May 2008.

J. S. Kalirai, H. B. Richer, G. G. Fahlman, J.-C. Cuillandre, and P. et al. Ventura. The CFHT Open Star Cluster Survey. I. Cluster Selection and Data Reduction. AJ, 122:257–265, July 2001. doi: 10.1086/321140.

I. D. Karachentsev and V. E. Karachentseva. Type and shape differences for isolated and paired galaxies. Soviet Ast., 18:428, February 1975.

S. C. Keller, B. P. Schmidt, M. S. Bessell, P. G. Conroy, and P. et al. Francis. The SkyMapper Telescope and The Southern Sky Survey. PASA, 24:1–12, May 2007. doi: 10.1071/AS07001.

J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. Steenstrup Pedersen, and C. Igel. Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy. *ArXiv e-prints*, April 2017.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. 2012.

Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*, pages 319–, London, UK, UK, 1999. Springer-Verlag. ISBN 3-540-66722-9.

C. J. Lintott, K. Schawinski, A. Slosar, K. Land, and Bamford et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. MNRAS, 389:1179–1189, September 2008. doi: 10.1111/j.1365-2966.2008.13689.x.

LSST Dark Energy Science Collaboration. Large Synoptic Survey Telescope: Dark Energy Science Collaboration. *ArXiv e-prints*, November 2012.

E. A. Magnier, E. Schlafly, D. Finkbeiner, M. Juric, and Tonry et al. The Pan-STARRS 1 Photometric Reference Ladder, Release 12.01. ApJS, 205:20, April 2013. doi: 10.1088/0067-0049/205/2/20.

P. J. Marshall, A. Verma, A. More, C. P. Davis, and S. et al. More. SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses. MNRAS, 455:1171–1190, January 2016. doi: 10.1093/mnras/stv2009.

A. More, A. Verma, P. J. Marshall, S. More, and E. et al. Baeten. SPACE WARPS- II. New gravitational lens candidates from the CFHTLS discovered through citizen science. MNRAS, 455:1191–1210, January 2016. doi: 10.1093/mnras/stv1965.

L. A. Moustakas, K. Abazajian, A. Benson, A. S. Bolton, and J. S. et al. Bullock. Strong gravitational lensing probes of the particle nature of dark matter. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, 2009.

S. C. Odewahn, E. B. Stockwell, R. L. Pennington, R. M. Humphreys, and W. A. Zumach. Automated star/galaxy discrimination with neural networks. AJ, 103:318–331, January 1992. doi: 10.1086/116063.

A. Oemler, Jr. The Systematic Properties of Clusters of Galaxies. Photometry of 15 Clusters. ApJ, 194:1–20, November 1974. doi: 10.1086/153216.

J. B. Oke and J. E. Gunn. Secondary standard stars for absolute spectrophotometry. ApJ, 266:713–717, March 1983. doi: 10.1086/160817.

M. Ranzato, F. J. Huang, Y. L. Boureau, and Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007. doi: 10.1109/CVPR.2007.383157.

R. Schild. CCD observations of galaxies in clusters at high redshift. ApJ, 286:450–463, November 1984. doi: 10.1086/162620.

S. J. Smartt, S. Valenti, M. Fraser, C. Inserra, and D. R. et al. Young. PESSTO: survey description and products from the first data release by the Public ESO Spectroscopic Survey of Transient Objects. A&A, 579:A40, July 2015. doi: 10.1051/0004-6361/201425237.

L. Spitzer, Jr. and W. Baade. Stellar Populations and Collisions of Galaxies. ApJ, 113:413, March 1951. doi: 10.1086/145406.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and S. et al. Reed. Going Deeper with Convolutions. *ArXiv e-prints*, September 2014.

The Dark Energy Survey Collaboration. The Dark Energy Survey. *ArXiv Astrophysics e-prints*, October 2005.

Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016.

N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. *ArXiv Physics e-prints*, April 2000.

S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, and J. D. et al. Warner. scikit-image: Image processing in python. *PeerJ*, 2:e453, 6 2014. ISSN 2167-8359. doi: 10.7717/peerj.453.

Guido van Rossum. Python tutorial. Report CS-R9526, April 1995.

K. W. Willett, C. J. Lintott, S. P. Bamford, K. L. Masters, and Simmons et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. MNRAS, 435:2835–2860, November 2013. doi: 10.1093/mnras/stt1458.

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, and Annis et al. The Sloan Digital Sky Survey: Technical Summary. AJ, 120:1579–1587, September 2000. doi: 10.1086/301513.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

All internet sources ([1] - [9]) were up to date 18.5.2017.