6.15 DEEP LEARNING: UNVEILING THE COSMOS THROUGH THE SLOAN DIGITAL SKY SURVEY*

I am sure that you, like every other human being who has ever done so, have been fascinated by the universe every time you have stared at the night sky. This fascination with the night sky is what has led us to build huge telescopes in many different wavelength domains in order to study the objects and structures present in the cosmos.

Let us say that you do end up building a large optical telescope to observe different kinds of objects – planets, stars and galaxies. The planets in our Solar System are easily identifiable, based on the huge number of easily available previous images available... well, everywhere!

But when it comes to stars and galaxies, they can be extremely far away compared to their size. In other words, the angle subtended by both stars and galaxies is so small that it is not possible to tell them apart immediately. This can only be done by measuring what is known as the SPECTRUM (see Fig. (6.11) and Fig. (6.12) and read the footnote about a Blackbody[2]). In other words, you have to observe these objects at many different wavelengths, and figure out how the intensity changes as a function of wavelength.

In practice, of course, it is NOT possible to observe any SINGLE wavelength of light using any telescope. Instead, here is the best one can do:

1. Make sensors that are sensitive to the optical range of wavelengths
2. Record the response of the sensor over the entire wavelength range – this is done to correct for effects in the spectrum due to the sensors
3. Use filters that only allow certain wavelength ranges to go through, e.g., red, blue, green; record the response of the filters to incoming light as well
4. Observe the same object through different filters

2 A black–body is a perfect absorber-emitter – basically, let in some radiation into a darkened closed sphere and allow it to reach equilibrium – the radiation that this body emits is known as blackbody radiation. Planck derived the formula for a spectrum of such a body which agreed with experiment, by famously assuming that light was emitted in ‘packets’ - thus was born the field of quantum physics.

5. Correct for the response of the filters and the sensors – this yields the spectrum from the object

This is what the SDSS does too – and then compares the absorption/ emission spectra to known ones. And this, in a nutshell, is how to tell apart different objects in the universe.

Naturally, a lot of detail has been swept under the carpet here – details that you could learn as a student of Astronomy. If you are interested, write to us and watch this space – you never know when we might drop a book on Machine Learning for Astronomy!

A Cool Aside

Here is something you do not really think about every day – but you should!

Do you think sophisticated and large facilities like SDSS are the only ones that can get spectra from heavenly bodies? Think again...

You can use a prism to get the spectrum of sunlight – you can also use diffraction grating. The difficulty in getting the solar spectrum is not the availability of these tools (diffraction gratings are available for less than Rs. 2000/- nowadays) but the fact that you cannot look at the spectrum directly without burning a hole through your cornea!

So then, if you cannot look at the sun to look at its spectrum produced by a diffraction grating, what can you do instead?

Turns out, the answer is quite simple: reflect the sunlight off, say, a small needle (so that the reflected light DOES NOT make you blind!), then photograph it using your cell phone –place an appropriate (You may very well ask why we said **appropriate** – the reason is that all diffraction gratings deviate incoming light to a certain extent, depending on the distance between the slits in the grating. If the deviation is too small, the source and the spectrum will be too close, and you will, therefore, not be able to see the spectrum at all. On the other hand, if the deviation is too large, you would need a very wide-angle lens in order to photograph a spectrum whose angular separation from the source is large – these extra-wide-angle lenses would easily distort the image of the spectrum, introducing a new problem. Hence the need for balance) diffraction grating in front of the cell phone camera and ensure that the spectrum and source are both in the frame of the photograph. See Fig. (6.13) and Fig.(6.14).

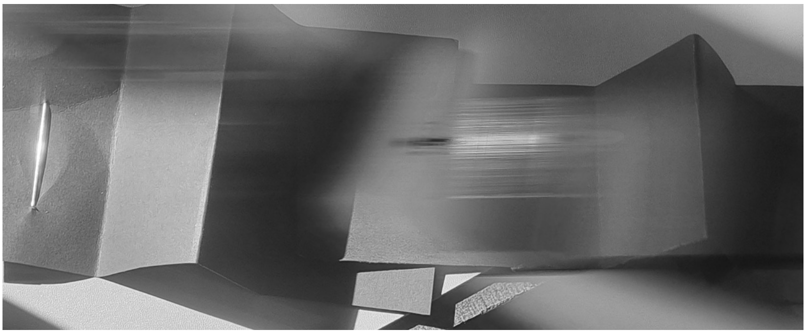

Fig. 6.13: *A 'needle spectrum' of the sun, taken by one of the authors on 06th January 2023 as a demonstration of how easy it is to obtain spectra of celestial objects. Sunlight was reflected off a needle, and then photographed using a cell phone camera – with one small addition… a diffraction grating was placed in front of the cell phone camera. Black paper was used as a background, for two reasons: 1. To ensure that there is only one reflecting surface, otherwise there will be multiple spectra that might even overlap, producing a lot of confusion and 2. To ensure visibility of the solar spectrum – the spectrum looks much brighter with a black rather than a light background.*

In principle, you can use a diffraction grating with any spacing between the successive slits, but it is advantageous to adjust to a spacing that does not require you to use a wide-angle lens to ensure that the spectrum, as well as the source, are visible in the resulting photograph.

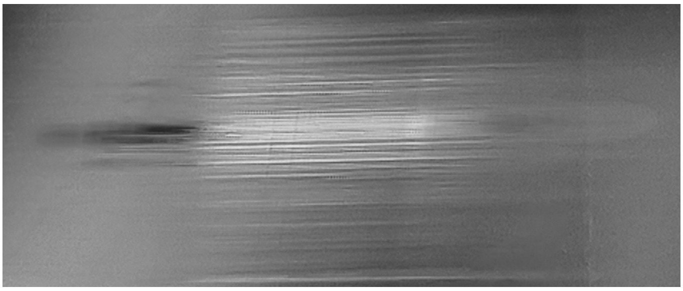

Fig. 6.14: *A 'needle spectrum' of the sun, taken by one of the authors on 06th January 2023 as a demonstration of how easy it is to obtain spectra of celestial objects. Sunlight was reflected off a needle, and then photographed using a cell phone camera – with one small addition… a diffraction grating was placed in front of the cell phone camera. Black paper was used as a background, for two reasons: 1. To ensure that there is only one reflecting surface, otherwise there will be multiple spectra that might even overlap, producing much confusion and 2. To ensure visibility of the solar spectrum – the spectrum looks much brighter with a black rather than a light background.*

In principle, you can use a diffraction grating with any spacing between the successive slits; however, we chose a grating with just the right spacing such that both the spectrum and the source are visible in the photograph, without any distortions – this way, we would be able to make meaningful measurements of the spectrum.

6.15.1 Introducing the Sloan Digital Sky Survey (SDSS)

The Sloan Digital Sky Survey (SDSS), akin to a modern-day celestial cartographer, has systematically mapped the heavens, offering us an unprecedentedly detailed view of the universe. This monumental survey has captured the essence of the cosmos, spanning across stars, galaxies, and quasars (QSOs), transforming the canvas of the night sky into a rich tapestry of data.

In the pursuit of understanding these celestial entities, we turn to the avant-garde field of Deep Learning, a subset of Machine Learning, poised at the intersection of data science and astronomy. The SDSS dataset serves as the perfect crucible for applying Deep Learning techniques, enabling us to not only discern the mysteries of the universe but also to revolutionise our approach to astronomical data analysis.

The Sloan Digital Sky Survey (SDSS) dataset, a cosmic tapestry, unveils a spectrum of parameters that delineate the characteristics of stars, galaxies, and quasars (QSOs). These celestial bodies, each unique in their cosmic narratives, are differentiated by a suite of attributes captured in the SDSS dataset. Our exploration here is aimed at demystifying these differences in a manner that is technical yet accessible, casting light on the nuanced distinctions that separate stars from galaxies and QSOs.

Stars, Galaxies, and Quasars: A Comparative Analysis

Stars

Stars, those shimmering jewels of the night sky, are characterised in the SDSS dataset by specific parameters that reflect their inherent properties. Key among these is:

1. **Magnitude Parameters (u, g, r, i, z):** These parameters, spanning the ultraviolet (u) to the near infrared (z), signify the star's brightness as seen through different filters. Stars, with their relatively smaller sizes and closer proximity, often exhibit distinctive brightness patterns across these filters.
2. **Redshift:** For stars, the redshift is typically close to zero, indicating their relative proximity within our galaxy. This parameter, a measure of the extent to which light has been stretched as it travels through the cosmos, is pivotal in distinguishing stars from more distant galaxies and QSOs.
3. **Point Spread Function (PSF) Magnitudes:** The PSF magnitudes (psfMag_u, psfMag_g, etc.) are crucial in identifying stars. Since stars appear as point sources due to their vast distances, their light distribution can be closely approximated by a point spread function.

The Galactic Conclaves: Galaxies

Galaxies, vast islands of stars bound together by gravity, present a different set of characteristics:

1. **Magnitude Parameters:** Galaxies, composed of billions of stars, have a broader and more diffused spectrum across the u, g, r, i, z magnitudes. Their light, an amalgamation of countless stars, gas, and dust, imparts a distinct signature in these filters.
2. **Redshift:** Galaxies generally exhibit higher redshift values compared to stars. This is indicative of their much greater distances, with the redshift providing clues about their velocity and position in the expanding universe.
3. **Exponential Model Fit Parameters (expAB_u, expAB_g, etc.):** These parameters, derived from fitting an exponential model to the light profile of galaxies, help in distinguishing them from stars. Unlike stars, galaxies do not conform to a point spread function, necessitating more complex models to describe their extended structure.

The Enigmatic Quasars (Quasi–Stellar Objects): QSOs

Quasars or QSOs, the lighthouses of the distant universe, are supermassive black holes at the centers of galaxies. Their properties are marked by extreme luminosities and high redshifts.

1. **Magnitude Parameters:** The magnitudes of QSOs are generally high across all filters, a testament to their incredible luminosity. Their light, however, is often red- shifted into the infrared part of the spectrum, a characteristic captured in the SDSS magnitudes.
2. **Redshift:** QSOs typically have exceedingly high redshift values, reflecting their position at the very fringes of the observable universe. This is a key indicator in distinguishing them from stars and galaxies.
3. **Color Indices:** The color indices (e.g., u-g, g-r) of QSOs are distinct from those of stars and galaxies. These indices, calculated from the magnitudes in different filters, can reveal the presence of strong emission lines, a hallmark of quasars.

Understanding SDSS Data Through Visualisation

To illuminate these differences, let us consider plotting some of these parameters against each other. For instance, a color-magnitude diagram shown in Fig.(6.15) (plot- ting, say, u-g against g-r) can reveal clustering patterns that differentiate stars, galaxies, and QSOs. Similarly, a redshift histogram can visually separate these celestial entities based on their distances from us.

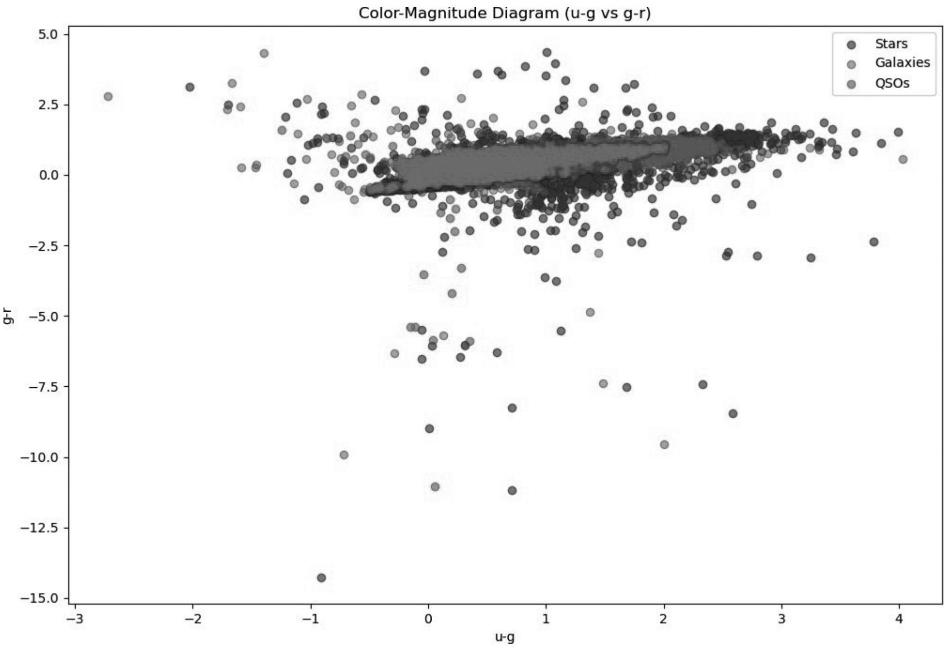

Fig. 6.15: *The Color–Magnitude diagram of the entire SDSS Data Release 18 dataset. This color-magnitude diagram (CMD) serves as a vibrant map charting the astronomical characteristics of various celestial objects, presenting a comparative analysis of stars (in blue), galaxies (in green), and quasars (QSOs, in red) based on their intrinsic luminosities and temperatures. The horizontal axis (u-g) represents the color index, which is a measure of the object's temperature, with lower values indicating hotter, more ultraviolet-rich objects, while the vertical axis (g-r) indicates the object's brightness in green light relative to red light, with lower values signifying brighter objects. Stars are predominantly clustered along a diagonal sequence, indicative of the main sequence where stars of different temperatures align according to their luminosity and temperature. Galaxies exhibit a wider spread in color and a shift towards dimmer magnitudes, reflective of their varied stellar populations and the influence of interstellar dust. Quasars are notably scattered throughout the diagram, with many having high u-g values, denoting the significant red- shifting of their light due to their rapid recession speeds and the expansion of the universe. The diagram encapsulates the diversity of the universe's constituents, with the distribution of objects illustrating the complex interplay between an object's luminosity, temperature and distance from Earth, providing astronomers with insights into the evolutionary stages of these celestial bodies.*

In the realm of astrophysics, this comparative analysis is more than merely academic. It forms the bedrock upon which further explorations and discoveries are built. By understanding the subtle and overt distinctions between stars, galaxies, and QSOs, astronomers can unravel the history of the universe, from the fiery births of stars to the enigmatic behavior of black holes at the centers of distant galaxies.

Let us proceed with some visualisations to bring these concepts to life. We will create plots that highlight the differences in magnitude parameters, redshift, and color indices for stars, galaxies, and QSOs within the SDSS dataset.

The color-magnitude diagram (CMD) shown in Fig. (6.15) is like a graph that shows the relationship between the true brightness of stars, galaxies, and quasars (QSOs) and their colors. The ′g-r′ on the vertical axis tells us about the color, where a bigger number means redder, and a smaller number means bluer. The ′u-g′ on the horizontal axis helps us compare how bright objects are in ultraviolet light compared to green light. The position of a star or galaxy on this graph gives us clues about what It is like: how hot it is, how old it might be, and how far away it is from us.

In this particular graph, we can see three different groups: stars are marked in blue, galaxies in green, and QSOs in red. Stars form a thick line going from the bottom left to the top right, showing that they can have a wide range of colors and brightnesses. Galaxies are spread out more, mostly above the line of stars, which means they are usually less brightly in ultraviolet light. QSOs are scattered around, but they tend to be further to the right, which means they are usually brighter in ultraviolet light compared to how they appear in green light. This graph helps astronomers quickly see the differences between these objects and learn more about them.

The redshift distribution graph shown in Fig. (6.16) is for stars, galaxies and quasars (QSOs). Redshift happens when the light from an object in space stretches out, usually because that object is moving away from us. It is like the sound of a siren on a moving fire truckit changes pitch as it moves past us. In this graph, the redshift is on the bottom (horizontal axis), and the number of objects is on the side (vertical axis). The higher the bar, the more objects there are with that amount of redshift.

For stars, which are the blue bars, we see they are all clumped up at the left end with redshifts close to zero. This means they are not moving away from us very much, they are in our own galaxy. Galaxies, shown in green, have a wide range of redshifts, which tells us they are at all different distances, some close and some far away, all moving away at different speeds. The red bars are the QSOs, and they are pretty spread out with high redshifts. This means they are really far away and moving away from us very quickly. By looking at this graph, scientists can get a quick snapshot of how far away different space objects are and how fast they are going, which helps them figure out how the universe is expanding.

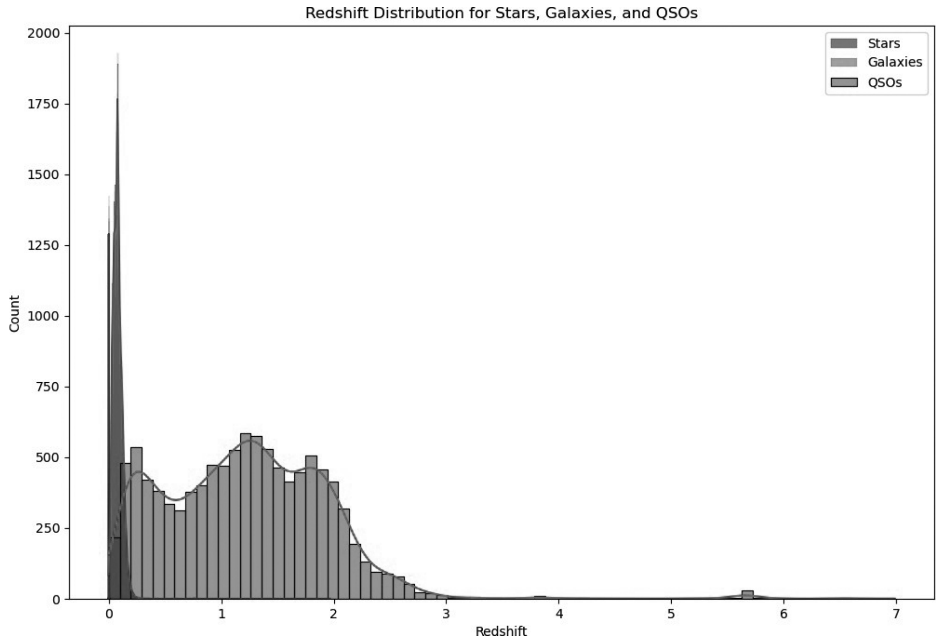

Fig. 6.16: *The Redshift Distribution of stars, galaxies and Quasars or QSOs on the SDSS DR18. This redshift distribution histogram provides a quantitative depiction of the cosmic recessional velocities for various celestial objects, categorised into stars, galaxies and quasars (QSOs), as inferred from their spectral redshifts. In the histogram, the x-axis represents redshift (z), which correlates to the velocity at which objects are moving away from the Earth, with higher redshifts indicating greater velocities and distances. Stars, predominantly within our own Milky Way galaxy, exhibit a narrow distribution concentrated near zero redshift, reflecting their relatively stationary position with respect to us. Galaxies display a broader, bell-shaped distribution, indicating a wide range of distances and recessional velocities, which is consistent with the expanding universe framework where more distant galaxies move away faster. Quasars, the most distant and energetic objects in the universe, are represented by a long tail extending to high redshifts, con- firming their cosmological distances and the profound expansion rates associated with the early universe. The histogram encapsulates the vast scale of the cosmos, from local stars to the most remote quasars, offering a visual narrative of the universe's dynamic and ever-evolving nature. Stars are at z = 0, as expected.*

The Sloan Digital Sky Survey and Its Profound Impact on Digital Photography

We are going to take an unusual route here. Instead of describing how the SDSS is star–spangled awesome when it comes to Astronomy, we are going to tell you **how the SDSS has caused the Digital Camera Revolution – which started in the 1990s**. Remember that Digital Cameras are part of a ≈ USD 5 billion industry, growing at about 5% per year – the fact that an esoteric field like Astronomy caused this industry to come into being and bloom is significant.

We hope that the discussion that follows will increase your understanding of the interplay between Science, Technology Development and Industry – this is necessary for all Machine Learning/ AI/ Data Science Professionals, and our hope in including this tangential description is to make you – the reader – aware of this interplay, so that you can leverage it in your career.

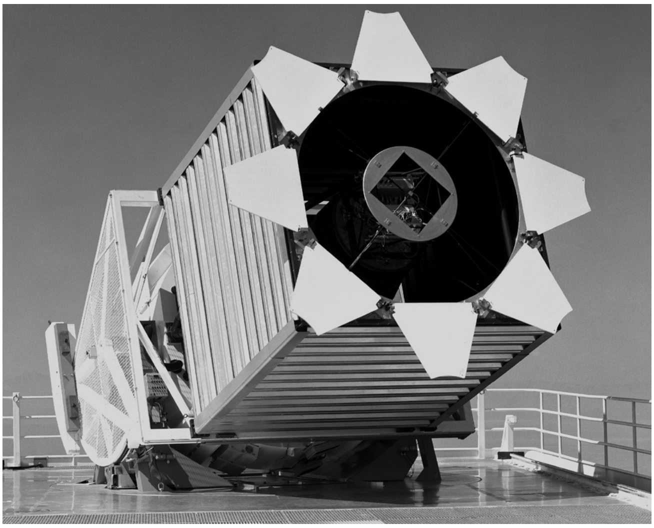

Fig. 6.17: *The SDSS 2.5 metre telescope. The image displays the Sloan Digital Sky Survey (SDSS) telescope in all its engineering splendor, a testament to the intersection of astronomical science and technological advancement. This 2.5-meter telescope, housed at the Apache Point Observatory in New Mexico, is a marvel of modern optics, featuring a distinctive, segmented, 120-megapixel digital camera capable of capturing wide-field images across five optical bands, from the ultraviolet to the near-infrared. The telescope's unique design, with its flower-like array of petal-shaped baffles, minimises stray light, enhancing its ability to record the faint light of distant celestial objects accurately. The SDSS telescope has been instrumental in mapping the large-scale structure of the universe, studying the distribution of luminous and non-luminous matter and contributing to the digital camera revolution with its pioneering imaging technology. This photo captures the telescope's intricate lattice of support and the open enclosure which allows it to peer deep into the night sky, collecting data that has fueled countless research projects and expanded our understanding of the cosmos. [Photo courtesy: Sloan Digital Sky Survey]*

The Dawn of a New Era in Astronomy: Introduction to SDSS The Sloan Digital Sky Survey (SDSS) heralded a transformative era in astronomical observation and imaging technology. Conceived and executed with unprecedented precision, SDSS was not merely an astronomical mission but a technological marvel that reshaped our understanding of the universe and the tools we use to perceive it.

- **Astronomical Aspirations:** SDSS aimed to create the most detailed three-dimensional maps of the Universe, capturing data across a significant portion of the sky with its advanced instrumentation and digital imaging capabilities.

- **Innovative Instrumentation:** At the core of SDSS's technological prowess was its 2.5-meter telescope, situated at the Apache Point Observatory, equipped with a highly advanced digital camera and spectrographs.

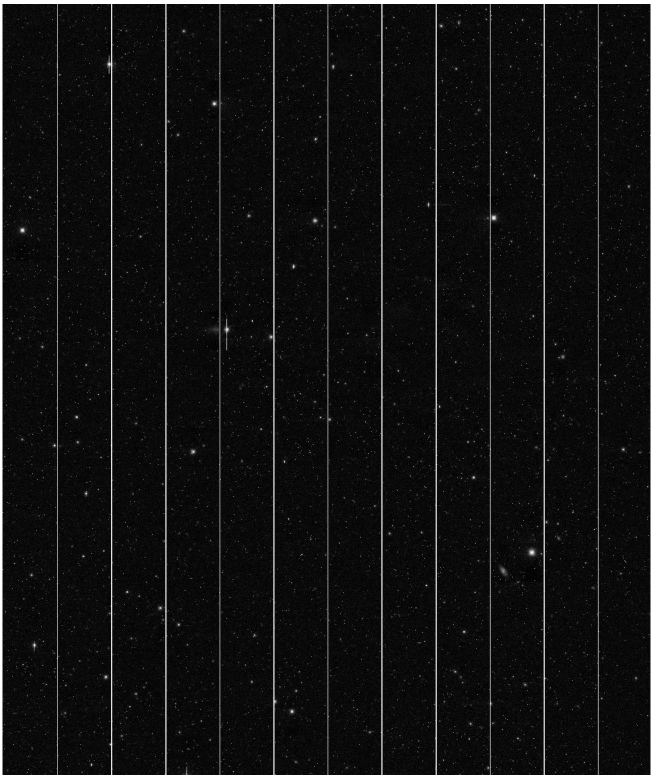

Fig. 6.18: *This image presents a meticulously assembled mosaic of a celestial expanse, measuring 2.5 degrees on each side The intricate tapestry of the cosmos was woven over two consecutive nights using the Sloan Digital Sky Survey's formidable 2.5-meter telescope paired with its intricate mosaic camera. The camera's configuration, composed of six columns of Charge-Coupled Devices (CCDs), captures six distinct scanlines of the sky. These scanlines, each separated by slight gaps, are the canvas upon which light from distant stars and galaxies is painted. To achieve a continuous celestial panorama, the telescope was slightly shifted between the two observational runs, allowing the separate scanlines to interlock perfectly. This subtle adjustment is a dance of precision, ensuring that the final image is devoid of discontinuities, offering observers a seamless view into the depths of the universe. The resulting image is a testament to the blend of human ingenuity and technological prowess, providing astronomers with a contiguous and unbroken map of a vast section of the night sky.*

[Photo courtesy: Sloan Digital Sky Survey]

Understanding the SDSS Setup The technical setup of SDSS was a confluence of optical and digital technologies, pushing the boundaries of what was possible in astronomical imaging.

1. **The Telescope and Camera:** The heart of SDSS's imaging capabilities was its 120-megapixel camera, one of the largest of its time, capturing detailed images across five optical bands. The camera's design, featuring Charge-Coupled Devices (CCDs), was a pioneering effort in high-precision astronomical imaging (Source: [astro-ph/9809085] Untitled Document).

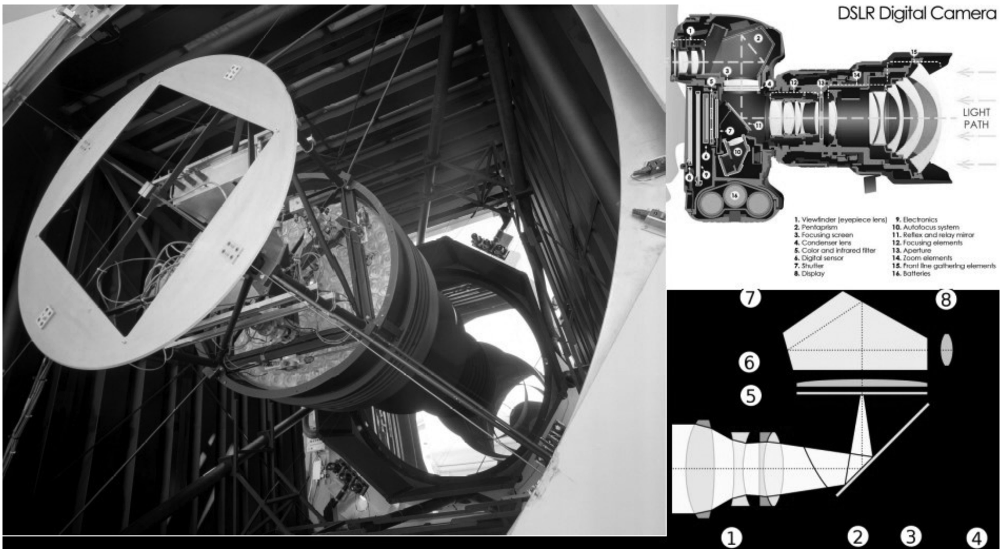

Fig. 6.19: *This composite image, a triptych of optical prowess, encapsulates the technological lineage from the Sloan Digital Sky Survey (SDSS) to the intricate mechanisms of modern digital single-lens reflex (DSLR) cameras. On the left, the SDSS telescope's interior reveals the large-scale, complex structure designed for mapping the cosmos, featuring its wide-field camera with an array of charge-coupled devices (CCDs) that capture the light of distant celestial bodies with unparalleled precision. These astronomical CCDs, with their ability to record faint light across a wide area of the sky, have profoundly influenced the development of imaging sensors in DSLR cameras, exemplified by the dissected Nikon DSLR shown in the center. This camera, much like the SDSS's camera, relies on a sophisticated array of optics and electronics to render high-resolution images. To the right, a schematic of a DSLR's optical path – from the lens (1), through the reflex mirror (2), pentaprism (6) and viewfinder (7), to the digital sensor (4) – demonstrates the similar principles of light capture and processing that unite telescopes and cameras. The SDSS's pioneering imaging technology not only advanced our understanding of the universe but also propelled the mass production capabilities and sensor designs that have made such intricate DSLR cameras widely available to both amateur and professional photographers, allowing the cosmos's grandeur to inspire the artistry of terrestrial snapshots.*
Left photo courtesy Sloan Digital Sky Survey
Right top photo courtesy ephotozine.com
Right bottom figure courtesy Wikipedia

2. **Optical Configuration and Image Processing:** SDSS's camera system utilised an intricate arrangement of filters and lenses, along with sophisticated image processing techniques, to produce high-quality astronomical data. This setup was instrumental in achieving the level

of detail and accuracy that SDSS is renowned for (Source: [astro-ph/0006396] The Sloan Digital Sky Survey: Technical Summary).

The Revolution in Digital Photography: Tracing SDSS's Impact

SDSS's influence extended significantly beyond the realm of astronomy, playing a pivotal role in the advancement of digital photography.

- **Pioneering Digital Imaging Technologies:** The development and implementation of CCD technology in SDSS's camera marked a turning point in digital imaging. This technology, essential for capturing high-resolution astronomical images, also found its way into the burgeoning field of digital photography, reshaping the landscape of photographic technology.
- **From Astronomy to Consumer Cameras:** The advancements in CCD technology and digital image processing pioneered by SDSS directly influenced the design and capabilities of consumer digital cameras. The transfer of this technology from professional astronomical instruments to everyday photographic devices was a significant leap in the evolution of digital cameras (Source: [astro-ph/0006396] The Sloan Digital Sky Survey: Technical Summary). See Fig.(6.19).

Deep Learning

Let us now turn to the analysis of the SDSS Data Release 18 dataset.

Deep Learning, a term that echoes through the corridors of modern computational science, is an advanced form of Machine Learning. It is distinguished from traditional neural networks by its profound architecture, consisting of multiple hidden layers. These layers, akin to the strata of geological formations, allow Deep Learning models to learn and extract increasingly abstract features from raw data, a process reminiscent of peeling the layers of an onion to reveal its core.

Distinguishing Deep Learning from Traditional Neural Networks While traditional neural networks are the bedrock upon which Deep Learning is built, they are but the first steps in a much larger journey. Traditional neural networks typically consist of a shallow architecture with one or two hidden layers. In contrast, Deep Learning networks delve deeper, with numerous hidden layers, each layer learning more complex and abstract representations of the data. This depth empowers Deep Learning models to capture intricate patterns that are often elusive to their shallower counterparts.