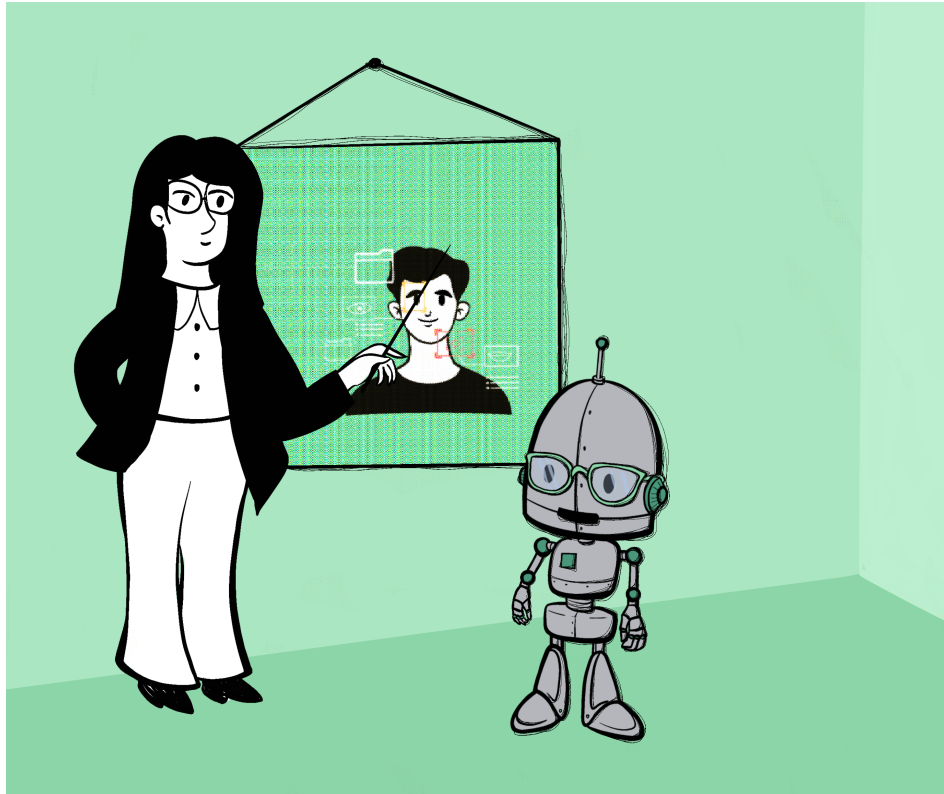


Module 6

Computer Vision

*"**Computer Vision** ist ein komplexer Bereich, sodass es noch **viel Forschung** bedarf, bis all die bekannten - und noch bisher unbekannten - Problemstellungen gelöst werden können."*



Über das Modul

In diesem Modul wird der Hauptfokus auf Computer Vision (**CV**) gelegt. Dabei sollen die bereits erlernten Fähigkeiten und Wissensbereiche aus den vorherigen Kapiteln erweitert und zur kritischen Auseinandersetzung mit dem Thema **CV** anregen. Vor allem das Modul **Supervised Learning** ist für ein vertieftes Verständnis der kommenden Inhalte von Vorteil, ist jedoch nicht verpflichtend zu absolvieren. Auch ein Quereinstieg ist kein Problem und vermittelt die wichtigsten Lernziele.

Das Ziel dieses Modules ist ein grundlegendes Verständnis dafür zu schaffen, inwiefern sich die Art und Weise zu "sehen" zwischen Menschen und Computern unterscheidet, welche Techniken und Algorithmen das Erfassen von Objekten ermöglichen, aber auch eine kritische Auseinandersetzung und Diskussionsgrundlage zu schaffen. Diese Ziele werden vor allem durch praktische "Hands-On" und "CS-Unplugged" Beispiele erreicht und mit einem Praxisbeispiel, in welchem die Schüler*innen ein eigenes Machine-Learning Face-Lock Programm in Scratch entwickeln, abgeschlossen. Vertieftes theoretisches Wissen bezüglich Künstlicher Intelligenz sowie mathematisches Hintergrundwissen ist hierbei nicht notwendig.

Ziele

Die Schüler*innen sind in der Lage...

- ...Gemeinsamkeiten und Unterschiede in der menschlichen und computerbasierten visuellen Informationserfassung zu nennen
- ...zu erklären, wie ein Computer visuelle Informationen speichert und verarbeitet
- ...grundlegende Algorithmen zur Objekterkennung zu nennen und können diese anhand einfacher Beispiele erklären
- ...einfache Machine-Learning **CV**-Algorithmen selbst praktisch anwenden
- ...die Grenzen und Möglichkeiten von **CV** aufzeigen

Agenda

Zeit	Inhalt
10 min	Einführung

Zeit	Inhalt
15 min	Theorie - Verarbeitung von Bildern
15 min	Theorie - Klassische CV Verfahren
30 min	Theorie + Arbeitsblatt - Viola Jones
40 min	Praxis - ML Gesichtserkennung mit Scratch
10 min	Diskussion - Möglichkeiten und Grenzen

Einführung - Was ist Computer Vision (CV)?

Bei der Einführung in **CV** geht es darum, dass sich die Schüler*innen mit grundsätzlichen Konzepten des "Sehens" und der "Objekterkennung" aktiv im Rahmen einer Plenumsdiskussion auseinandersetzen. Als Basis und Leitfaden dieses Einführungsprozesses dienen die **Präsentationsfolien**.

The introduction to **CV** is about the students actively dealing with the basic concepts of "seeing" and "object recognition" in a plenary discussion. The presentation slides serve as the basis and guide for this introductory process.

Vergleich Mensch - Computer

(Folien 2 - 5)

In den Folien 2-3 werden die Schüler*innen im Rahmen einer Plenumsdiskussion an das Thema, wie Menschen beziehungsweise Computer ihre Umwelt wahrnehmen können, herangeführt. Wichtig ist hier vor allem, dass die Ähnlichkeiten des grundlegenden Prozesses zwischen Menschen und Computer ersichtlich sind. Hierbei können diverse Methoden (Tafel, aber auch diverse Online-Tools) zur Ideensammlung etc. verwendet werden.

Sowohl Menschen als auch Computer arbeiten nach dem EVA-Prinzip (Eingabe - Verarbeitung - Ausgabe). Wir sehen etwas mit unseren Augen, verarbeiten dies mit unserem Gehirn und handeln dementsprechend. Bevor wir jedoch nach diesem Schema arbeiten können, müssen Menschen (vom Kleinkindalter aufwärts), als auch ein Computer (z.B. Supervised Learning) gewisse Fähigkeiten erlernen. Es wird iterativ geübt, bis man beispielsweise einen Satz sprechen kann, oder einen Apfel als Apfel erkennt. Dieses anfängliche Lernen wird in dem ersten Schritt auf Folie 5 dargestellt. Nach dem Prozess des Lernens kann man sich der Herausforderung stellen, und Konversationen führen, oder aber auch Äpfel von Birnen unterscheiden. Im Falle des Computers, müssen wir jedoch Bildmaterial zur Verfügung stellen, welches der Computer verarbeiten soll. Das würde dem E des EVA-Prinzips entsprechen. Danach folgt die Verarbeitung (V) durch den Computer mittels speziellen Computer Vision Algorithmen. Als letzter Schritt folgt die Ausgabe (A), in der beispielsweise die Objekte auf dem Bild markiert werden oder auch das Handy durch die Erkennung des Gesichtes entsperrt wird.

Auf Folie 6 werden exemplarisch noch einige typische Aufgaben von **CV** dargestellt.

Classification

Klassifizierung des Objektes auf dem Foto als eine Katze

Classification + Localization

Hierbei wird die Katze auf diesem Bild zusätzlich noch lokalisiert und mit einem roten Rechteck markiert

Object Detection

Bei diesem Beispiel werden verschiedene Objekte innerhalb des Bildes klassifiziert und auch lokalisiert

Material

-  CV - Introduction.pdf

Grundlagen digitaler Bilder

Im letzten Abschnitt wurde mittels Diskussion und Beispielen in das Thema grundlegend eingeführt, was nun im folgenden Abschnitt durch einige theoretische Grundlagen digitaler Bilder vertieft werden soll. Fokus soll hierbei auf die interne Verarbeitung und Speicherung von Bildern gelegt werden.

Wie werden Bilder in einem Computer gespeichert?

(Folien 8 – 9)

Die Informationsverarbeitung- und Speicherung funktioniert binär. D.h. der Computer kennt nur die Zustände 0 und 1. 0 bedeutet kein Strom und 1 bedeutet Strom. Nun muss man sich überlegen, wie man Unmengen an Bildinformationen in 0en und 1en darstellen kann. Man muss hierbei etwas um die Ecke denken. Um es möglichst einfach zu halten, sehen wir uns ein Bild in Graustufen an. Bei einem Graustufenbild werden verschiedene Bildbereiche dunkler oder heller dargestellt. Je dunkler ein Bildpunkt ist, desto niedriger ist sein repräsentativer Zahlenwert. Bei helleren Bildpunkten ist der Wert dementsprechend größer. Der Zahlenbereich reicht von 0 bis 255, somit stellt 0 sehr dunkel und 255 sehr hell dar. Wandelt man diese Dezimalzahl anschließend noch in eine Binärzahl um ($4 \Rightarrow 1000$, $3 \Rightarrow 11$) kann man diese in einem Computer abspeichern.

Auf Folie 9 ist eine Gegenüberstellung zwischen dem menschlichen Sehen und jenes eines Computers. Durch die vorher bereits kurz erläuterte Art und Weise wie Bilder in einem Computer abgespeichert werden, erhält man eine regelrechte Flut an Daten. Spannend ist jedoch, sobald man ein wenig aus dieser Unmenge an Zahlen herauszoomt, können wir mit unserem Auge sogar den Inhalt des im Computer gespeicherten Bildes der Schildkröte nur anhand der Pixelwerte erkennen.

Welche Informationen könnte man bereits aus diesen Pixelwerten herauslesen? Wenn ein helles Feld neben einem dunklen Feld liegt, kann man durch einfache Subtraktion einen Unterschied der Werte feststellen. Da Objekte in Bildern meist irgendwelche markanten Kanten haben, welche sich farblich von der Umgebung unterscheiden, kann ein Computer zum Beispiel durch einfache Mathematik erkennen, dass es sich hier um die Kante eines Tisches oder der Linie auf einem Fußballfeld handelt (mehr dazu später im Kapitel Edge-Detection).

Ähnlich funktioniert es bei RGB-Bildern, nur dass hier viel mehr Informationen in Form von Zahlen übereinandergelegt werden. Dadurch kann sich der Computer eine Farbe errechnen/speichern und diese dann darstellen, aus Gründen der Einfachheit gehen wir in diesem Modul immer von Bildern in Graustufen aus.

Wichtig für uns ist aktuell nur, dass wir durch spezielle Algorithmen und diverse mathematische Methoden, aus diesen Zahlenmengen, sehr viel Information extrahieren können. Ein einfacher Einstieg soll durch das gemeinsame Untersuchen der Funktionsweise des Flood-Fill Algorithmus gegeben werden.

Flood Fill

(Folie 10)

Auf Folie 10 soll mittels einer Plenumsdiskussion die Funktionsweise des Flood Fill Algorithmus erarbeitet werden. Dabei wird den Schüler*innen Zeit gelassen, um sich selbst die Funktionsweise des Algorithmus zu überlegen und die einzelnen Schritte zu notieren. Nach dieser Phase werden die Lösungsansätze im Plenum besprochen und verglichen.

Generelle Funktionsweise des Algorithmus: Wie es sich bereits aus dem Namen vermuten lässt, wird das Bild "geflutet" oder auch ausgemalt. Zwecks einfacher Illustration bieten sich hier Zahlen gut an. Das Bild wird so weit ausgemalt, bis man auf eine andere Farbe stößt. Je nachdem welche Zahl gezeichnet wurde, wird mehr oder weniger Fläche ausgemalt. Jede Zahl hat eine bestimmte freie, nicht bemalte Fläche und kann somit vom Computer durch mathematische Berechnungen identifiziert werden.

Schritte: (Beispiel)

1. Nimm zwei Farben (Hintergrund und Schrift)
2. Setze den Pinsel auf einen Startpunkt
3. Male so lange bis der Punkt unter dem Pinsel eine andere Farbe als der Hintergrund hat
4. Wenn alles ausgemalt, überprüfe Füllstand der Pinsel-Farbe

Dieser einfache Algorithmus bringt auch einige Einschränkungen mit sich. Zum Beispiel, was wenn die Zahlen unsauber geschrieben sind und zum Beispiel eine Null eine offene Stelle hat und dadurch der innere Bereich auch ausgefüllt wird?

Bekannte Beispiele für die Anwendung dieses Algorithmus wären zum Beispiel das legendäre Minesweeper-Game oder auch die Flächenfüllfunktion in MS-Paint.

Material

-  CV - Introduction.pdf

Klassische CV Algorithmen

Zu diesem Zeitpunkt sollten sich die Schüler*innen ein grobes Bild machen können, wie ein Computer grundsätzlich einfache Dinge in digitalen Bildern erkennen kann. Aber was, wenn die Objekte komplizierter werden, die Farbunterschiede undeutlicher und die Komplexität allgemein zunimmt?

Objekterkennung arbeitet häufig auf Basis von Machine- und Deep Learning Algorithmen, welche im Laufe der Zeit perfektioniert wurden. Oft können diese sehr komplexen Algorithmen, durch ihre einfacheren, klassischen Vorgänger ersetzt werden. Das Stichwort ist Reduzierung der Komplexität durch Abstraktion. Auf den Folien 14 und 15 wird dies mittels eines einfachen Beispiels dargestellt. Selbst nach dem Entfernen jeglicher Bildeigenschaften, bis auf die Kontur des Elefanten, bleibt der Elefant als Elefant erkennbar.

Klassische Kantenerkennungs-Algorithmen können zum Beispiel der Sobel oder Canny Algorithmus sein. Diese Algorithmen basieren sehr stark auf den mathematischen Grundlagen der Matrizenrechnung, jedoch können nach Teilschritten des Algorithmus, bereits klare Kanten der Objekte erkannt werden. (Folie 16-24)

Erinnern wir uns zurück an das Kapitel der internen Bilddarstellung. Auf Folie 16 wird noch einmal exemplarisch diese interne Darstellung aufgezeigt. So ist zum Beispiel sehr gut zu erkennen, wie dieser Roboter aus einzelnen Pixeln mit bestimmten Zahlenwerten besteht. Aufgrund der einfacheren Handhabung wird die RGB-Darstellung der Pixelwerte gegen jener der Graustufendarstellung ausgetauscht, das Prinzip bleibt jedoch dasselbe. Betrachtet man die Zahlenwerte aus einer gewissen Entfernung, lässt sich für das menschliche Auge bereits eine ungefähre Form erkennen. Aber wie könnte nun ein Computer mit diesen Zahlenwerten eine "Kante" dieses Objekts erkennen? Richtig, durch mathematische Berechnungen, vorwiegend Matrizenrechnungen.

Auf der Folie 17 wird nun das Ergebnis der horizontalen Kantenerkennung gezeigt. Das linke Bild wurde mittels eines Python Skriptes selbst erstellt und kann z.B. für Interessierte Schüler*innen als Exkurs bereitgestellt werden. In der rechten Grafik sieht man die entsprechenden Zahlenwerte, nachdem die horizontale Kantenberechnung durchgeführt wurde. Auf Folie 18-19 wird nun die mathematische Technik dahinter genauer erläutert. So kompliziert diese

Abbildungen auch aussehen, so einfach ist deren Algorithmus zur Wertberechnung. Es werden lediglich einfache Rechenregeln angewandt. Man subtrahiert z.B. Zelle Ab von Zelle Aa und erhält somit einen Zahlenwert. Je höher dieser Zahlenwert, desto höher ist der farbliche Unterschied der einzelnen Pixel. Ergo, wir haben höchstwahrscheinlich einen Pixel einer Kante erkannt. Macht man diese Vorgehensweise für alle Pixelwerte, erhalten wir nicht nur einzelne Werte, sondern ganze Linien, in dem Fall Kanten, welche als solches erkannt werden können. Nun sollen gemeinsam mit den Schüler*innen und Schülern weitere Werte dieser Matrix berechnet werden (die Lösung findet sich im Notiz-Bereich der jeweiligen Folien)

Das Prinzip der vertikalen Kantenerkennung funktioniert auf dieselbe Weise, daher wird nicht weiter darauf eingegangen. Hat man nun vertikale und horizontale Kanten berechnet, kann man diese beiden Ergebnisse kombinieren und erhält somit eine Matrix bzw. ein Bild, welches nun alle Kanten beinhaltet. Dies wird auf Folie 23, 24 nochmals visualisiert und auf Folie 25 mit dem originalen Bild verglichen.

In aller Kürze zusammengefasst ergeben sich für den Algorithmus folgende drei Schritte:

- Berechnung der horizontalen Kanten
- Berechnung der vertikalen Kanten
- Kombination der beiden Ergebnisse

Material

-  CV - Introduction.pdf

Gesichtserkennung mit dem Viola Jones Algorithmus

Im Rahmen einer Plenumsdiskussion sollen die Folien 27 bis 30 mit den Schüler*innen besprochen werden. Die notwendigen Informationen befinden sich zur Gänze auf den Folien.

Auf Folie 31 werden erste technische Grundlagen des Viola Jones Algorithmus dargestellt. Ein wichtiger Bestandteil des Algorithmus ist das "Sliding-Window", mit dem das Bild schrittweise von links nach rechts und von oben nach unten abgetastet wird und nach möglichen (auch mehreren) Gesichtern und deren Merkmalen gesucht wird. Die Erkennung der einzelnen Gesichtsabschnitte (Auge, Nase, Mund, ...) erfolgt mit sogenannten "Haar-Like Features", welche auf Folie 32 dargestellt sind.

Die wichtigsten Haar-Features sind:





- Edge Features
- Line Features
- Four-Rectangle Features

Diese Features und auch das Sliding-Window haben keine fixe Größe, sondern werden dynamisch nach Bedarf angepasst. In Folie 33 wird diese Vorgehensweise nochmals dargestellt und auf Folie 34 noch zusätzlich ein **Video**, welches den Ablauf des Viola Jones Algorithmus zusammenfasst und visualisiert, verlinkt.

Der Viola Jones Algorithmus hat einige Einschränkungen: Mit ihm kann man nur frontale Gesichter erkennen, was in gewissen Szenarien und Anwendungsfällen problematisch sein kann und mit der Notwendigkeit, andere Computer Vision Verfahren einzusetzen, einhergehen. Weiters ist wichtig zu betonen, dass Viola Jones zwar Gesichter detektieren, nicht jedoch verschiedene Gesichter unterscheiden kann.

Als nächstes (Folien 35 - 36) folgt das Hands-On Beispiel zum Viola Jones Algorithmus. Die benötigten Materialien und auch die Erklärung der Aufgabe findet sich in der Arbeitsblatt-Datei und auf den Folien. Zusätzlich gibt es für interessierte bzw. fortgeschrittene Schüler*innen die Möglichkeit, den Viola Jones Algorithmus anhand eines bereits Python Skriptes auszuprobieren (die Aufgabenstellung und Musterlösung findet sich auf Folie 37).

Material





-  CV - Introduction.pdf
-  <https://youtu.be/hPCTwxFOqf4>
-  CV - Viola Jones.pdf
- Für Fortgeschrittene:  CV - Viola Jones Python.zip

Machine Learning in CV

Auf Folie 39 findet sich nochmals eine kurze Zusammenfassung. Computer Vision (**CV**) und Machine Learning (ML) sind zwei Bereiche von KI. Computer Vision beschäftigt sich mit dem computerbasierten Verarbeiten von visuellen Informationen, welches durch Ansätze von Machine Learning noch weiter verbessert werden kann (z.B. Erkennung von Bewegungen, Folgen von Objekten in Echtzeit, ...). Da Machine Learning – speziell Supervised Learning – bereits in einem anderen Modul aufgegriffen und besprochen wurde, fokussieren wir uns nun zum Abschluss auf "Hands-On" Beispiele, in denen ML und **CV** kombiniert werden.

Die Schüler*innen und Schüler sollen dabei ein virtuelles Smartphone mittels eines vorher erstellten Machine Learning Modells programmieren. Die notwendigen Unterlagen und Erklärungen finden sich wiederum in der entsprechenden Arbeitsblatt-Datei.

Material

-  CV - Introduction.pdf
-  <https://machinelearningforkids.co.uk>
-  CV - Face Lock.pdf
-  CV - Face Lock Solutions.pdf

Möglichkeiten und Grenzen

Computer Vision ist ein komplexer Bereich, sodass es noch viel Forschung bedarf, bis all die bekannten – und noch bisher unbekannten – Problemstellungen gelöst werden können. Zudem haben wir auch gesehen, dass selbst so einfache Algorithmen wie der Viola Jones Algorithmus, praktisch überall seit Jahrzehnten Anwendung finden. Als Abschluss dieses Moduls soll nun auf etwaige Fragen, Anmerkungen etc. der Schüler*innen eingegangen werden und ein Austausch stattfinden. Auf Folie 42 finden sich Leitfragen, welche als Anregung benutzt werden können, jedoch sollte sich ein dynamischer Gesprächsverlauf entwickeln.

Falls an diesem Punkt noch Zeit zur Verfügung steht, kann je nach Bedarf und Interesse noch auf Folie 43 eingegangen werden. Hierbei wird ein von einem Computer klassifiziertes Bild einer Ampel, durch einfache Änderungen von Pixeln, so verändert, dass es den **CV** Algorithmus nicht mehr gelingt, die Ampel richtig zu klassifizieren. Weitere Informationen bezüglich der Zuverlässigkeit von **CV** Algorithmen finden sich auf der Homepage, welche auf Folie 43 im Notizbereich verlinkt wurde.

Material

-  CV - Introduction.pdf