**ТEХНИЧECКИ УНИВEРCИТEТ – COФИЯ**

Фaкултeт пo тeлeкoмуникaции

Cпeциaлнocт: Тeлeкoмуникaции

**Курсов проект**

**по „Основи на Аудио и Видео Технологиите“**

*Тeмa:*  
Основни характеристики на дълбоко машинно обучение за компютърно зрение

|  |  |  |
| --- | --- | --- |
| **Студенти:** |  |  |
| Михаил Александров Александров | Фaк. №: 111217168 | Група: 51 |
| Живко Руменов Йорданов | Фaк. №: 111217169 | Група: 51 |
|  |  |  |
| **Преподавател:** | **доц. д-р инж. Павлина Колева** | |
|  |  | |
| Дата: ................. | Подпис: .................... | |
|  |  | |
| София, 2021 | | |

# Съдържание

[Съдържание - 2 -](#_Toc93959484)

[1. ЗАДАНИЕ - 3 -](#_Toc93959485)

[2. Увод - 4 -](#_Toc93959486)

[3. Анализ на текущото състояние на проблема по литературни данни - 6 -](#_Toc93959487)

[3.1 Какво е машинно обучение - 6 -](#_Toc93959488)

[3.2 Какво е дълбоко обучение - 7 -](#_Toc93959489)

[3.3 Какво представляват невронните мрежи - 8 -](#_Toc93959490)

[3.4 Какво са конволюционни слоеве - 9 -](#_Toc93959491)

**ТЕХНИЧЕСКИ УНИВЕРСИТЕТ**

**–**

**СОФИЯ**



Катедра: Радиокомуникации и видеотехнологии

Дата на задаване: 28.09.2020

Краен срок на предаване: 23.01.2020

Образователна степен: бакалавър, специалност: ТК

# ЗАДАНИЕ

**ЗА КУРСОВ ПРОЕКТ**

**ПО „ОСНОВИ НА ВИДЕО И АУДИО ТЕХНОЛОГИИТЕ“**

на студентите:

Живко Руменов Йорданов, гр. 51, Фак. номер: 111217169

Михаил Александров Александров, гр. 51, Фак. номер: 111217168

**Тема: 35. Основни характеристики на дълбоко машинно обучение за компютърно зрение**

Съдържание на курсовата работа *(8-10 стр. при изпълнение от един студент и 15 стр.*

*при изпълнение от двама студенти, Размер на шрифта - 12, междуредие - 1,5)*

1. Заглавна страница/ задание
2. Увод
3. Анализ на текущото състояние на проблема по литературни данни *(цитиране на източниците в края на проекта)*
4. Описание на проблема *(матет. формули, фигури, таблици и др.)*
5. Реализация *(описание на алгоритми, хардуер, програмен код на симулации, блокови схеми и т.н.)*
6. Примери за приложение
7. Използвана литература *(цитиране на източниците)*

Студент: .........................................

(Живко Йорданов)

Студент: ......................................... Научен ръководител: ....................

(Михаил Александров) (гл. ас д-р инж. Н Нешов)

София 1000, бул. “Св. Климент Охридски” 8, тел.: 02/ 965 21 11, факс: 02/ 868 32 15, www.tu-sofia.bg

# Увод

Технологията за компютърно зрение е една от най-обещаващите области за научни изследвания в рамките на изкуственият интелект и компютърните науки, и предлага огромни предимства за бизнеса в нашето съвремие. В сърцевината си, областта на компютърното зрение се фокусира върху проектиране на компютърни системи, които притежават способността да улавят, да разбират, и да интерпретират важна визуална информация, която се съдържа в изображенията и видео данните. След това компютърните визуални системи превеждат тези данни, използвайки контекстуални знания, предоставени от хора, използвани за задвижване на вземането на решения. Превръщането на едно изображение от необработен вид в концепция от по-високо ниво, така че хората или компютрите да могат да интерпретират и действат по него, е основната цел на технологията за компютърното виждане. Трябва обаче, да се направи важно разграничение между компютърното зрение и обработката на изображения. Обработката на изображения е науката за извършване на промени на един образ, така че да се произведе нов образ с определени (засилени) характеристики. Тези промени може да включват – увеличаването на резолюцията, нормализиране на яркостта и контраста, изрязване, замъгляване, или някаква друга дигитална трансформация, която би била необходима за конкретната цел. Дигиталната обработка на изображения не взема в предвид действителното съдържание на изображението - това е просто поредица от механични трансформации, предприети да променят изображението за някаква определена цел.

Системите за човешкото зрение имат огромното предимство да бъдат информирани за знания за цял живот, които помагат да се контекстуализират данните в нашето зрително поле. Нашите очни ябълки улавят визуалната информация например - образа на една котка, и нашият предишен опит интерпретира това събиране на отразена светлина и се свързва с понятието за котка. Сложността на нашата система за визуално възприятие и тясната ѝ връзка с нашата памет и по-високи способности за разсъждение дава на тези визуални данни необходимият контекст, за да осигури качество в ежедневните активности. Тези човешки способности, докато са недостъпни за компютрите, могат да бъдат имитирани ефективно чрез алгоритми за машинно обучение. Но както се оказва, обучаването на машини, които да имитират тази основна човешка фунцкия, която е гордо демонстрирана от 5 годишни деца по цял свят, е изключително трудно. Решаването на този проблем непрекъснато заема най-ярките умове в изследванията на изкуственият интелект.

# Анализ на текущото състояние на проблема по литературни данни

## Какво е машинно обучение

Машинното обучение е знанието, че компютрите се учат и действат като хора. Машинното обучение е компютърната парадигма, която води до растежа на "големите данни" и изкуствения интелект. То се основава на развитието на невронни мрежи и дълбокото обучение. Обикновено това се описано като имитация на начина, по който хората се учат, но това е малко погрешно определение. Машинното обучение всъщност се отнася до статистически анализ и итеративно обучение. Има много различни видове алгоритми за машинно обучение. Всеки ден се публикуват стотици. Те обикновено са групирани стил на учене (контролирано обучение, неконтролирано учене, обучение с полу-контролиран достъп) или по-късно споразумения във форма или функция (напр. класификация, регресия, дърво на решенията, групиране, задълбочено обучение и др.). Независимо от стила или функцията на обучение, всички комбинации се състоят от следното:

* представителство (набор от класификатори или езика, който компютърът разбира)
* оценка (известен също като функция за обективно / отчитане)
* оптимизация (методът за търсене, често най-добрият класификатор за оценяване, се използват както методите за оптимизация,

Основната цел на алгоритмите за машинно обучение е да се обобщава допълнително, т.е. да се интерпретират успешно данни, които никога преди не са били представени. Има различни начини да оставим машините да се учат. От използването на обикновени дървета за вземане на решения до групиране на слоеве от изкуствени невронни мрежи (последното е дало възможност за дълбоко учене). Докато фокусът често е върху избора на най-добрия алгоритъм за обучение, изследователите са открили, че някои от най-интересните въпроси не са отговорени от наличните ресурси. Важен момент е, че машинното обучение не е просто, а дори и автоматизация. Ако мислите така, вие със сигурност пропускате ценните прозрения и възможности, които машините могат да ви предложат.

Машините, които учат са полезни, тъй като с цялата си обработваща мощност те могат по-бързо да намират модели в големи данни. Иначе тези части биха били пропуснати от хората. Машинното обучение е инструмент, който може да се използва за подобряване на решаването на проблеми. Той може да направи информирани заключения в по-широк спектър от проблеми като диагностициране на болести и съставяне на решения за борба с глобалното изменение на климата. Срещат се и обаче предизвикателства и ограничения при машинното обучение. Една от най-честите грешки при начинаещите е успешното тестване на данните за обучение с илюзията, че сте успешни. Част от събирането на данни не казва нищо за целия набор от данни. Когато алгоритъмът за обучение (т.е. ученикът) не работи, понякога той успее, като доставя повече данни на машината. Това обаче може да доведе до проблеми със мащабируемостта, където имаме повече данни, отколкото времето за учене. Машинното обучение не е самоцел. Освен това не е полезно да се използва като общо решение. Веднъж обучен, алгоритъмът за машинно обучение е в състояние да сортира чисто нови входове през мрежата с голяма скорост и точност в реално време. Това го прави основна технология за компютърно зрение, гласово разпознаване, езикова обработка, както и научноизследователски проекти.

## Какво е дълбоко обучение

Дълбокото обучение или йерархичното обучение е подмножество на машинното обучение в изкуственият интелект, който може да имитира функцията за обработка на данни на човешкия мозък и да създава подобни модели, които мозъкът използва за вземане на решения. Противно на алгоритмите, базирани на задачи, системите за дълбоко обучение се учат от представянето на данни - те могат да се учат от неструктурирани или немаркирани данни. Архитектурите за дълбоко обучение като дълбоките невронни мрежи, мрежи от вероятности(бейсови мрежи), повтарящи се невронни мрежи, и конволюционните невронни мрежи са намерили приложения в областта на компютърното зрение, аудио/говорното разпознаване, машинният превод, филтрирането в социалните мрежи, биоинформатиката, дизайнът за лекарства и много други.

## Какво представляват невронните мрежи

Невронните мрежи са съставени от набор от алгоритми, които са моделирани спрямо структурата на човешкият мозък. Тези алгоритми могат да интерпретират сензорни данни чрез машинно възприятие и надписвайки или групирайки необработените данни. Те са предназначени да разпознават числови модели, които се съдържат във вектори, в които са всички данни от реалния свят (изображения, звуци, текстове, времеви редове и т.н) трябва да бъдат преведени. В основата си, основната задача на невронните мрежи е да групира и класифицира необработените данни - те групират немаркираните данни въз основа на приликите, които са намерени във входните данни и след това се класифицират данните въз основа на етикетираният набор от данни за обучение. Невронните мрежи могат автоматично да се адаптират към променящият се вход. Така че, не е необходимо наново да проектираме изходните критерии всеки път, когато входът се промени, за да генерира възможно най-добрия резултат.

Невронните мрежи понастоящем са най-популярният начин да се направи Deep Learning, но съществуват и други начини за постигане на машинно обучение. Дълбокото обучение в компютърното зрение стана възможно чрез изобилието от данни за изображения в съвременния свят, плюс намаляването на разходите за изчислителната мощност, необходима за обработването му. Големи мащабни комплекти от изображения като ImageNet, CityScapes и CIFAR10 събраха милиони изображения с точно обозначени функции за алгоритми за дълбоко обучение. Привидно за една нощ, изпълнението на алгоритмите за дълбоко обучение надмина тридесет години работа по ръчни детектори на функции. Допълването на достатъчен брой добре обозначени изображения към дълбока визуална система, която дава възможност да се разберат точните нюанси на ниво пиксел и определя отделните компоненти на по-голямото изображение. Тя автоматично ще научи къде са ръбовете, и как особени комбинации от ръбове, които се различават по цвят и контраст един от друг и от фона се комбинират, за да образуват определени функции. Изображението на Фиг. 1 показва как една система за дълбоко обучение може да идентифицира характеристиките на котка.

A picture containing calendar

Description automatically generated

Фиг. 1

## Какво са конволюционни слоеве

Конволюционните слоеве от по-висок ред на невронната мрежа ще започнат да разбират, че ако на изображението има 4 крака, глава, опашка и тяло, следователно въпросното изображение може да съдържа котка. От необработени визуални данни на ниво пиксел, машината връща концепция от по-висок ред – „котка“ – въз основа на последователното добавяне и класифициране на тези отделни компоненти. Изображението на Фиг. 2 показва това прогресивно разбиране в контекста на човешкото разпознаване на лицето.

A picture containing website

Description automatically generated

Фиг. 2

Разликата е, че традиционните системи за зрение включват човека, който казва на машината какво трябва да има, вместо обратният случай, където алгоритъма за дълбоко обучение автоматично извлича характеристиките за това какво има на изображението. Подходът "отдолу нагоре" е значително по-ефективен за определени видове проблеми с анализа на изображенията, много от които използваме често в ежедневието си. При диагностициране на тъканна проба обаче тези разлики стават ценни. Малките, незабележими колебания в плътността на пикселите могат да означават ранното начало на болестта рак - подробности, които дори експерт патолозите е възможно да пропуснат. Човешкото визуално възприятие удря бариера на разделителната способност на изображението при около 2290 пиксела на инч и удря максималност(таван) за възприемане на движението при около 30 кадъра в секунда. Способността на системите за компютърно зрение да работят с точност на ниво пиксел, се повтаря бързо, и изпълнението последователно във времето предлага невероятен потенциал, за увеличаване или надминаване на човешкото възприятие.

CNN или конволюционната невронна мрежа (CNN - convolutional neural network) е клас невронни мрежи за дълбоко обучение. Накратко, можем да си представим CNN като алгоритъм за машинно обучение, който може да приеме входно изображение, да присвои важност (усвоими тегла и отклонения) на различни аспекти/обекти в изображението и да може да ги различава един от друг. Този тип мрежа работи, като извлича характеристики от изображенията. Всяка конволюционна невронна мрежа се състои от следното:

1. Входният слой, който е изображение в сива скала

2. Изходният слой, който е двоичен или мултикласов етикет

3. Скрити слоеве, състоящи се от конволюционни слоеве, слоеве ReLU (ректифицирана линейна единица), обединяващите слоеве и напълно свързана невронна мрежа

Много е важно да се разбере, изкуствените невронни мрежи, съставени от множество неврони, не могат да извличат характеристики от изображението. Това е мястото, където се появява комбинация от навиващи и обединяващи слоеве. По същия начин, конволюционните и обединяващите слоеве не могат да извършат класификация, следователно се нуждаем от напълно свързана невронна мрежа. Преди да преминем към концепциите, нека се опитаме да разберем тези отделни сегменти поотделно. Ето един пример от Фиг. 3 който илюстрира процеса на конволюционната невронна мрежа от входни до изходни данни.

Diagram, engineering drawing

Description automatically generated

Фиг. 3

Например, ако си представим, че имаме достъп до множество изображения на различни превозни средства, всяко обозначено като камион, кола, микробус, велосипед и т.н. Тогава идеята е да вземем тези предварително етикетирани/класифицирани изображения и да разработим алгоритъм за машинно обучение, който е в състояние за приемане на ново изображение на превозно средство и да го класифицира в правилната му категория или етикет. Преди да започнем да изграждаме невронна мрежа, трябва да разберем, че повечето изображения се преобразуват в сива форма, преди да бъдат обработени. Има значение и какви скали ще използваме – дали ще бъдат сиви скали или RGB/цветни изображения.

Diagram

Description automatically generated

Фиг. 4 - RGB цветни канали на изображение

Всяко цветно изображение има три канала, т.е. червено, зелено и синьо, както е показано на Фиг. 4. Има няколко такива цветови пространства като сивата скала, CMYK, HSV, в които може да съществува изображение. Предизвикателството с изображенията с множество цветни канали е, че имаме огромни обеми от данни, с които да работим, което прави процеса изчислително интензивен. В други светове мислете за това като за сложен процес, при който невронната мрежа или всеки алгоритъм за машинно обучение трябва да работи с три различни данни (RGB стойности в този случай), за да извлече характеристики на изображенията и да ги класифицира в съответните им категории. С други думи може да погледнем на това като на сложен процес, при който невронната мрежа или всеки алгоритъм за машинно обучение трябва да работи с три различни данни (RGB стойности в този случай), за да извлече характеристики на изображенията и да ги класифицира в съответните им категории. Ролята на CNN е да намали изображенията във форма, която е по-лесна за обработка, без да губи функции, критични за доброто прогнозиране. Това е важно, когато трябва да направим алгоритъма възможен за мащабиране до масивни набори от данни.

A picture containing table

Description automatically generated

Фиг. 5

Изображението на Фиг. 5 илюстрира стойността на сгънатата характеристика, когато ядрото се прилага към входното изображение. Разбираме, че данните за обучение се състоят от изображения в сивата скала, които ще бъдат вход за слоя на свъртане за извличане на характеристики. Конволюционният слой се състои от едно или повече ядра с различни тежести, които се използват за извличане на характеристики от входното изображение. Да кажем, че в примера по-горе работим с ядро (K) с размер 3 x 3 x 1 (x 1, защото имаме един цветен канал във входното изображение), с тегла, очертани по-долу. Когато плъзгаме ядрото върху входното изображение (да речем, че стойностите във входното изображение са интензитети на сивата скала) въз основа на теглата на ядрото, в крайна сметка изчисляваме характеристики за различни пиксели въз основа на техните околни/съседни пикселни стойности. Например, когато ядрото се прилага върху изображението за първи път, както е илюстрирано на фигурата по-горе, получаваме стойност на характеристика, равна на 4 в сгънатата матрица на характеристиките. Процесът на изместване на ядрото се нарича stride и в нашият пример ядрото ще се измести 9 пъти. Когато използваме операция със стойност на стъпка 1 ( Non-Strided ), имаме нужда от 9 итерации, за да покрием цялото изображение. Конволюционната невронна мрежа научава тежестта на тези ядра самостоятелно. Резултатът от тази операция е карта на характеристиките, която основно открива характеристики от изображенията, вместо да разглежда всяка една стойност на пиксела. Резултатът от тази операция е карта на характеристиките, която основно открива характеристики от изображенията, вместо да разглежда всяка една стойност на пиксела.

# Описание на проблема

В този курсов проект е разгледан класически проблем в машинното обучение и е реализирано решене което се основава на прилагане на техники от дълбокото обучение разгледани в точка 3 от проекта.

## Дефиниране на разглеждания проблем

Класическият проблем разгледан във този курсов проект е класификационен проблем при който трябва да се синтезира система, която чрез прилагане на техники от дълбокото обучение, да може, с голяма точност, да групира подадената информация на входа й.

В конкретният казус, системата ще бъде обучена да класифицира изображения на кучета и котки. За тази цел са подбрани 3000 фотографски изображения на кучета и котки, заснети в разнообразни ъгли, пози и при различна светлина. Тези изображения ще се използват за обучаване и в последствие, верификация на работоспособността на синтезираната система.

## Модел на решението

Системата която ще се предложи за решение на този проблем ще представлява синтез и в посредством реализация на конволюционна невронна мрежа.

Diagram

Description automatically generated

Фиг. 6

Обобщена блок схема на невронната мрежа е представена на Фиг. 6. На входа на мрежата се поддава изображение. На изхода резултата е представен чрез вероятности. Вероятността изображението да принадлежи към класът на кучетата и вероятността изображението да принадлежи към класът на котките Ако се съберат двете вероятности, то резултатът им винаги трябва да 1-ца.

Моделът на конволюционната мрежа показан на Фиг. 6 може условно да се раздели на два блока „Конволиционна част“ и „Класическа част“. Блокът който е наречен „Конволиционна част“ съдържа слоеве които са присъщи за обработката на изображения във една невронна мрежа. Блокът който е наречен „Класическа част“ представя невронната мрежа в класическият си вид. Тук информацията извлечена от „Конволиционна част“ се обработва и решението се взима.

Схемата на блокът „Конволиционна част“ от Фиг. 6, е представена на Фиг. 7.

Diagram

Description automatically generated

Фиг. 7