

Добар дан свима. Данас ћу вам представити свој дипломски рад на тему Реализација модела за класификацију слика хране уз коришћење трансферног учења. Фокус овог рада је на конволуционим неуронским мрежама и њиховој примени у класификацији слика, где сам се у свом пројекту фокусирао на класификацију слика хране користећи доступан скуп података сликама и класам хране. Проћи ћемо кроз структуру ових мрежа, изазове у тренирању оваквих модела, као и практичне примене у стварном свету.

---

Конволуционе неуронске мреже, или CNN-ови, су један од најважнијих алата који користимо данас у обради слика и видеа. У суштини, неуронске мреже функционишу тако што покушавају да имитирају начин на који ради наш мозак – оне уче кроз искуство. Свака неуронска мрежа је састављена од више слојева "неурона", где сваки слој узима информације, обрађује их и затим их шаље следећем слоју. Код CNN-ова, ови слојеви су специјално дизајнирани да препознају различите обрасце у сликама, попут ивица, облика и текстура.

Пре него што су CNN-ови постали популарни, потребно је било ручно дефинисати шта ће рачунар тражити на слици – на пример, морали бисмо унапред одредити како изгледа ивица или одређени облик. Сада, захваљујући CNN-овима, они сами уче да препознају ове карактеристике директно из података. То значи да, ако желимо да наш систем препознаје различите животиње, не морамо ми унапред да кажемо шта чини мачку мачком или пса псом – CNN ће то сам научити гледајући много примера.

На пример, ако имамо мрежу која треба да класификује слике аутомобила, она ће кроз различите слојеве научити да препознаје гуме, светла, па чак и различите углове и боје аутомобила. Што више података дајемо мрежи, она ће боље разумети различите карактеристике и моћи ће да прави тачније предикције.

Конволуционе неуронске мреже се развијају уз помоћ дубоког учења, које омогућава овим моделима да буду "паметнији" и да боље разумеју комплексније обрасце у подацима. То је разлог зашто су данас широко коришћене у стварима као што су препознавање лица, медицинска дијагностика на основу слика или чак у аутомобилима који се сами возе. Све ово постаје могуће јер CNN-ови могу да се прилагоде различитим задацима и уче директно из података које им дајемо, без потребе да их ручно програмирамо за сваку нову ситуацију.

---

Основни слојеви CNN-а функционишу тако што корак по корак анализирају слике и извлаче информације које су битне за решавање одређеног задатка. Први на реду су **конволуциони слојеви**. Можемо их замислити као мале филтере који „прелазе“ преко слике и хватају различите детаље, као што су ивице, облици или текстуре. На пример, на слици пса, ови слојеви ће научити да препознају његове контуре, крзно или облике попут њушке и ушију. Како се слојеви нижу један на други, мрежа добија све прецизније информације о слици, идентификујући све сложеније карактеристике.

Након тога долазе **pooling слојеви**, који у суштини поједностављују информације тако што смањују величину података. Ово звучи као да губимо нешто, али заправо нам помаже да фокусирамо само на битне детаље, чиме се смањује количина информација коју модел обрађује и смањује вероватноћа да ће се „заглавити“ на небитним детаљима. Замисли то као да од велике, детаљне слике направимо мању верзију која задржава најважније делове. Ово такође чини модел бржим и ефикаснијим.

На крају, имамо **потпуно повезане слојеве**. Ови слојеви раде тако што узимају све те научене карактеристике – као што су облици, текстуре и обрасци – и комбинују их како би направили коначну предикцију. Ови слојеви су, у неку руку, „мозак“ CNN-а који доноси одлуку на основу свих информација које је раније прикупио. Ако је реч о класификацији слика, овај слој би на крају одредио да ли је на слици пас, мачка, ауто или нешто друго.

Да би цела ова прича функционисала, ту је и **активциони слој**, попут **ReLU**. Он уводи неопходну „нелинеарност“ у модел, што значи да омогућава мрежи да учи сложеније односе између података. Уместо да се ослања само на једноставне математичке односе, модел може да препозна и решава сложеније обрасце који су често присутни у стварним сликама. ReLU је важан јер омогућава да мрежа не „заглави“ на негативним вредностима података, што доприноси бољим резултатима и бржем учењу.

Дакле, сви ови слојеви раде заједно како би CNN био способан да научи од основних до веома сложених облика и детаља, што га чини моћним алатом у обради слика и видеа.

---

Кључни моменти у развоју конволуционих неуронских мрежа обележени су неким од најважнијих модела у историји машинског учења. Један од првих, **LeNet-5**, био је развијен 1998. године за препознавање руком писаних бројева, што је тада био огроман пробој. Замисли само колико је важно било да аутоматски препознамо бројеве на чековима или формуларима, и то без људске интервенције! То је био један од првих реалних примера употребе неуронских мрежа у пракси.

Међутим, прави револуционарни тренутак десио се 2012. године са моделом **AlexNet**. Овај модел је освојио такмичење ImageNet, и тада је свет схватио колико су дубоке мреже, односно „дубоко учење“, моћне. AlexNet је први модел који је показао да је могуће тренирати велике и комплексне мреже користећи графичке процесоре (GPU) уместо стандардних процесора, што је убрзало цео процес и довело до бољих резултата.

Каснији модели, као што су **VGGNet** и **ResNet**, додатно су побољшали тачност и способност мрежа да уче. VGGNet је био познат по томе што је користио једноставну и елегантну архитектуру, док је **ResNet** увео иновацију звану резидуално учење. Овај концепт је омогућио да мреже постану још дубље, решавајући проблеме који настају када модели постану сувише сложени – проблеме са којима се до тада нисмо знали носити.

---

С друге стране, један од главних изазова код тренирања CNN-ова је феномен зван **преобучавање**. У суштини, ово се дешава када модел постане предобар у учењу одређеног скупа података, до те мере да буквално "учи напамет" или "буба" податке. Он постаје одличан у препознавању слика на којима је обучен, али када се суочи са новим, непознатим сликама, модел се изгуби јер не зна како да примени своје научено знање. Замисли то као ученика који је научио све одговоре за тест напамет, али када добије неко питање које је мало другачије, не може да се снађе.

Овај проблем преобучавања може бити веома фрустрирајућ, јер значи да модел није довољно "флексибилан" да се избори са стварним светом, где подаци нису увек исти као они на којима је трениран. Како бисмо се борили са тим, користимо разне технике попут аугментације података, регуларизације, или раног заустављања тренинга када видимо да се резултати погоршавају на тестним подацима.

Још један изазов је сама сложеност тренирања ових дубоких модела. Они захтевају изузетно много рачунарских ресурса, посебно времена и снаге. Зато су графички процесори (GPU) постали толико важни. Они могу да обрађују огромне количине података паралелно, што знатно убрзава процес тренирања у поређењу са класичним процесорима (CPU).

У ствари, модели су толико захтевни да су се развили читави онлине сервиси који изнајмљују графичке картице програмерима и истраживачима. Овако не мораш куповати скупу опрему, већ једноставно користиш рачунарску снагу са удаљених сервера да би тренирао своје моделе. Ова решења омогућавају да тренирање, које би иначе трајало данима или чак недељама, буде завршено за много краће време, што је огроман напредак у овој области.

Уз коришћење унапред тренираних модела, као што су ResNet или Inception, програмери могу брзо прилагодити већ постојеће моделе на своје задатке, уместо да све граде од почетка. Ово штеди и време и ресурсе, а омогућава много бржу примену CNN-ова у различитим индустријама.

---

Пре него што слика уђе у CNN, морамо да је припремимо, што подразумева нормализацију и скалирање како би се слике ускладиле са очекивањима модела. Још један важан корак је аугментација података – рецимо, ротирање или "flipовање" слика. Овим техникама правимо више варијација исте слике и тиме смањујемо ризик да рачунар „учи напамет“, док уједно смањујемо потребу за већим скупом података.

Трансферно учење нам омогућава да користимо већ трениране моделе, као што је InceptionV3, који је већ научио одређене карактеристике на другом скупу података. Овај приступ нам значајно скраћује време тренинга јер не морамо тренирати модел од почетка. Додају се само додатни слојеви који се прилагођавају специфичном проблему. Технике оптимизације као што су SGD и RMSprop помажу да модел брже конвергира ка решењу.

Што се тиче евалуације, меримо перформансе модела кроз метрике као што су тачност и губитак. Тачност нам показује колико је модел дао исправних предикција, док је губитак метрика која нам помаже у оптимизацији током тренирања. Матрица конфузије је корисна јер показује где модел најчешће греши и које класе меша. Циљ је, наравно, повећати тачност и смањити губитак, иако модел никада неће бити савршен и погодиће сваки податак.

Током првих 10 епоха, примећујемо смањење губитка на тренинг сету, али након треће епохе губитак на валидационом сету почиње да расте, што је знак преобучавања. Тачност на тренинг сету наставља да расте, али на валидационом сету опада после четврте епохе. У другом пролазу од 100 епоха, тачност на тренинг сету досеже 90%, али се на валидационом сету стабилизује на 65%. Ово значи да модел има проблема са генерализацијом, што захтева примену техника као што је рано заустављање или додатна аугментација података.

Закључно, иако CNN-ови представљају револуцију у анализи слика, суочавамо се са изазовима попут преобучавања и потребе за бољом интерпретацијом модела. У будућности, циљ је развој ефикаснијих и скалабилнијих модела за примену у што ширем спектру индустрија.