НЕВРОННИ АРХИТЕКТУРИ С ДЪЛБОКО СТРУКТУРИРАНО ОБУЧЕНИЕ, ИЗПОЛЗВАНИ В КОМПЮТЪРНОТО ЗРЕНИЕ

1. Използване на конволюционни невронни мрежи.

Конволюционните невронни мрежи са водеща архитектура, в теорията на компютърното зрение, за задачи свързани с разпознаване, класификация, локализация, сегментиране и откриване на обекти от изображения или видео. CNNs архитектури са дълбоки невронни мрежи с право разпространение на сигнал (Feed-Forward Neural Network – FFNN), проектирани да приемат директно за вход изображение под формата на масив или на вектор с фиксиран размер, съдържащи интензитета на пикселите в 1 канал (монохроматично изображение - бинарно или в сивата скала) или 3 канала (хроматично изображение - цветно).

Прилагането на конволюционни филтри осигурява средство за улавяне на пространствени характеристики на наблюдаваните обекти в сцената, основани на класически методи използвани при обработката на изображения. Филтрите на конволюция и дълбочината на мрежата, осигуряват извличане на локални характеристики, които позволяват на изкуствената невронна мрежа да се справя, в сценарии, при които обектът е изместен от очакваната позиция в кадъра. Това е пряко свързано със свойствата на конволюционните мрежи по отношение на афинните трансформации - транслационна еквивариантност и транслационна инвариантност. CNN подрежда своите неврони в три размерности – ширина, височина и дълбочина, и всеки слой трансформира тримерен входен обем от активации, в тримерен изходен обем. В основата на CNN са заложени четири ключови идеи, които се възползват от свойствата на сигналите – локални връзки, споделени тегла, обединяване и използване на много слоеве. Те свойства осигуряват и основните предимства пред напълно свързаните многослойни перцептрони (Fully Connected Multilayer Perceptron – FC MLP). Дълбоките CNN са йерархично структурирани, в които характеристиките от по-високо ниво се получават от такива от по-ниско ниво. Подробно проучване на най-новите архитектури на дълбоки невронни конволюционни мрежи е направено в следващите раздели.

www.eufunds.bg
Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.

2. Използване на генеративна състезателни мрежи.

Генеративна състезателна (съревнователна) мрежа (Generative Adversarial Network - GAN) е вид мрежа за дълбоко обучение, която може да генерира нови данни със сходни характеристики като входните реални данни от обучаемия набор. Те се състоят от две мрежи, които се обучават заедно.

Първата се нарича генератор и подава вектор от произволни стойности като вход (латентни входове), създавайки данни със същата структура като данните на обучение.

Втората мрежа е дискриминатор, който подава данните под формата на партиди, съдържащи наблюдения както от обучителните данни от генератора, така и от генерираните данни и ги оценява. Дискриминаторът се опитва да класифицира наблюденията като „реални“ и „генерирани“. Генераторът генерира реалистични данни, а дискриминаторът се обучава от силни представяния на характеристики, които са характерни за данните на обучение. Приложенията на GANs са много и разнообразни – генериране на музика, мода, изобразително изкуство, дигитална криминалистика. Намират голяма реализация в науката за подобряване на астрономически изображения, за симулация на експерименти по физика на елементарните частици, за ранна медицинска диагностика и много други, и разнообразни приложения.

3. Използване на автоенкодери (Autoencoders).

Автоенкодерите (Autoencoders) са друг вид дълбоки невронни архитектури, които се използват за разпознаване на изображения. Те са вид генеративни модели за машинно обучение и използват основно алгоритми за неконтролирано обучение. Автоенкодерите са мрежи с право разпространение на сигнала и се състоят от енкодер и декодер блокове. За задачи като класификация на изображения могат да приемат като вход и трансформирани изображения, като на изхода на невронната мрежа реконструират оригиналното добро изображение. Входният и изходния слоеве имат еднакъв брой неврони, което цели на изхода на невронната мрежа реконструкция на входа, вместо прогнозиране на целеви изход. Автоенкодерът научава представяния от набора от данни в енкодер блока. Енкодерът кодира входните данни, обикновено с намаляване на размерността им. Декодерът реконструира приблизително

оригиналното входно изображение от енкодера, минимизирайки разликата между входа и изхода, като се учи да игнорира незначителни данни като шум. В контекста на ISAR изображенията, интерес биха имали автоенкодер невронните мрежи за намаляване на шума (Denoising Autoencoder – DAE). Те се опитват да постигнат добро представяне, приемайки частично повреден вход и възстановявайки в голяма степен оригиналното незашумено изображение на изхода на мрежата. Вариационните автоенкодер невронни мрежи (Variational Autoencoder - VAEs) се състоят от енкодер и декодер блокове, и функция на загубата (Loss Function). Тя цели да минимизира грешката при реконструкция на изображението в декодер блока. Обучаемият алгоритъм, във вариационните автоенкодери, оптимизира загубата по отношение на параметрите в двата блока с помощта на метода на градиентното спускане. Основните характеристика на автоенкодер мрежите е намаляване на размерността на входните изображения и извличане на информация под формата на представяния от входните данни.

4. Използване на мрежи с дългосрочна-краткосрочна памет.

Мрежите с дългосрочна-краткосрочна памет (Long Short-Term Memory - LSTM) са специален вид рекурентни невронни мрежи (Recurrent Neural Networks – RNN) използвани за дълбоко обучение. Те могат да обработват както точки от данни на изображения, така и да намират връзки при поредици от данни, включващи време – видео поток, естествен език, звук и др. LSTM невронната архитектура е водеща в дълбокото обучение и се използва за задачи свързани с обработка на последователности - като разпознаване и генериране на ръкописен текст, разпознаване на реч, многоезична езикова обработка, езиково моделиране, разпознаване и композиране на музика, оптично разпознаване на символи, машинен превод, синтактичен анализ на естествен език, генериране надпис на изображения, видео към текстово описание, адаптивна роботика и управление, внимателно зрение, видео поток, дизайн на лекарства и др. Рекурентните невронни мрежи имат верижна топология съставена от повтарящи се модули (единици). Тяхното предимство пред другите стандартни невронни мрежи, е че могат да оперират с поредица от вектори. Повтарящият се модул на стандартна RNN има опростена структура от един слой, докато LSTM модула се състои от четири слоя, което позволява на мрежата да проявява временно динамично

поведение. LSTM мрежите са способни да запомнят информация за дълги периоди от време. Архитектурата на LSTM мрежата е изрично проектирана да се справя с фундаменталния проблем с изчезващия градиент в дълбокото обучение, и предотвратява изчезването или експлозията на обратното разпространение на грешката. Общата LSTM единица се състои от клетка, две входни порти (input and input gate), изходна порта (output gate) и порта за забравяне (forget gate). Архитектурата на LSTM мрежата се характеризира с постоянна линейност на състоянието на клетката, заобиколена от четири нелинейни слоя на състоянията на портите. Порталните единици имат сигмоидна нелинейност и определят сами своите стойности на работа въз основа на текущо състояние и на входните данни, като балансират текущия вход и предишни състояния. Портите сами контролират състоянието на клетката като премахват или добавят информация към LSTM модула. По този начин LSTM клетката е самоадаптивна и има вътрешна повторяемост, в допълнение към външното повторение на рекурентните мрежи.

Обучаемият модел създава много дълбока абстракция, нарастваща итеративно. Двете входни порти на LSTM единицата запомнят стойности през произволни интервали от време, използвайки в общия случай различни функции за активация. Те работят заедно и решават какво да добавят към състоянието на клетката в зависимост от входа. Портата за забравяне избира кои стойности да премахне от старото състояние на клетката въз основа на текущите входни данни. Изходната клетка определя кои стойности от състоянието на клетката трябва да бъдат предадени на изхода. Входът на LSTM единицата се състои от изхода на предишното състояние и всички нови данни, предоставени в текущата времевата стъпка. LSTM невронните мрежи са изключително подходящи за класифициране, обработка и изготвяне на прогнози от времеви серии.

Разработени са различни хибридни варианти на LSTM мрежи, от които по-известни са: LSTM автоенкодери (LSTM Autoencoders), LSTM с внимание (LSTM with Attention), мултипликативна LSTM (Multiplicative LSTM - mLSTMs), LSTM шпионска връзка (LSTM Peephole Connection), Затворена повтаряща се единица (Gated Recurrent Unit).

5. Използване на невронни трансформаторни мрежи.

Най-новите модели за дълбоко обучение, които се основават на дълбоки невронни архитектури и привличат все по-голям изследователски интерес в областта на компютърното зрение са невронните мрежи трансформатори (Transformers) и подредените капсулни автоенкодер невронни мрежи (Stacked Capsule Autoencoders - SCAE). Трансформаторите са въведени през 2017 г. от екип на Google Brain и са дълбоки невронни архитектури, базирани на механизъм на самовниманието (Self-Attention). Проектирани са да обработват последователни входни данни, претегляйки значението на всяка част от входа различно, а впоследствие обработвайки целия вход наведнъж, за разлика от рекурентните мрежи. Това намалява времето за обучение и позволява извършването на повече паралелни изчисления. Използват се предимно в областта на NLP, но намират и приложение в компютърното зрение, като заменят моделите на RNNs, изместваики LSTM мрежите.

Моделът на трансформаторите използва архитектура последователност към последователност (Sequence-to-Sequence – Seq2Seq), състояща се от енкодер-декодер в класическия си вариант. Кодиращите и декодиращите елементи могат да бъдат по-няколко, подредени един върху друг. Кодиращият елемент енкодер, се състои от две части - механизъм за самовнимание и невронна мрежа с право разпространение на сигнала. Енкодерите в невронната мрежа имат една и съща архитектура и са много сходни помежду си. Енкодерът приема входната последователност и я картографира в пространство с по- високо измерение. Самовниманието се изчислява с помощта на три вектора. За всяка част от изображението, самовниманието създава вектор на заявка, вектор ключ и вектор стойност. Тези вектори позволяват матрично изчисление на самовниманието. Създаденият вектор от последователности (кодировки) се подава в декодера, който го превръща, трансформира в друга изходна последователност.

Всеки слой от енкодера съдържа кодировки, за това кои части от входовете са свързани една с друга. Механизмът на самовниманието приема входните кодировки от предишния енкодер и претегля тяхната относимост един към друг, за да генерира изходни кодировки. Модулът за внимание с много глави (Multi-head attention module) свързва енкодер и декодер елементите.

Декодерът се състои от три основни компонента – механизъм за самовнимание, механизъм за внимание върху кодирането и невронна мрежа с право разпространение на сигнала.

Първо, декодерът приема информацията за позициите на частите на входните данни и вгражда изходната последователност като свой вход с механизъм на самовниманието.

Вторият елемент в декодера - механизъм на вниманието, помага на декодера да извлича съответната информация от входните данни.

Третият елемент е мрежата с право разпространение, която обработва допълнително изходите с помощта на нормализиране на данните (Data Normalization).

Специализирани невронни мрежи трансформатори са - зрителните трансформатори (Vision Transformer - ViT), базирани изцяло на самовнимание без CNN. Те се прилагат директно към последователност от отделните части на изображения (кръпки/patches) и се справят отлично с задачи като класификация. На входа на невронната мрежа, изображението се разделя на отделни фрагменти с фиксиран размер, подавайки в следващите слоеве линейните проекции на тези части заедно с позицията им. С помощта на определяне на позицията на отделните кръпки от изображението, се запазва пространствената/позиционна информация. Входните изображения се изравняват до множество малки части на изображението, което е изчислително много по-ефективно за голям набор от данни.

Съществуват и хибридни архитектури CNN + Transformer, т. нар. откриващи трансформатори (Detection Transformer - DETR). DETR невронните мрежи се справят със семантична сегментация (Semantic Segmentation) - присвояват етикет на класа за всеки пиксел и сегментиране на инстанция (Instance Segmentation) - откриване и сегментиране на всеки обект, като двете задачи се обединяват под наименованието паноптично сегментиране (Panoptic Segmentation). Модела се справя с глобалното разбиране на изображението с помощта на самовниманието и има способността да разграничава припокриващи се обекти.

Съществуват и други модели на невронни мрежи базирани на трансформатори, които се справят с не анотирани данни, използвайки неконтролирано обучение.

Подредените капсулни автоенкодери играят все по-голяма роля в компютърното зрение и машинното обучение. Те са версия на капсулираните мрежи, но се обучават неконтролирано с помощта на самоконтролирано обучение. За първи път са представени през 2019 г. от екип от учени от Applied AI Lab – Институт по роботика на Университета в Оксфорд, Департамент по статистика на Университета в Оксфорд, Google Brain и DeepMind [86]. SCAE невронните мрежи се състоят от автоенкодери и капсули като специализирана част от модел, който описва абстрактен обект. Автоенкодер невронните мрежи с подредени капсули се състоят от енкодери с капсули на частите (Part Capsule Autoencoder - PCAE), последвани от енкодери с капсула на обектите (Object Capsule Autoencoder). По този начин се дефинира генеративния процес, който се състои от два етапа – PCAE и OCAE. PCAE откриват отделни малки части и техните пози от изображението, и използват тези части за сглобяване на обектите в OCAE. Постига се извличане и улавяне на всички възможни конфигурации на частите на обекта, което помага на капсулите да научават еквивариантни представяния за обекта. OCAE използват частите и позите за да разсъждават за обектите. Предвижданото на позите в наблюдаваната сцена се извършва с помощта на афинни трансформации - транслация, мащабиране и ротация. За всяка капсула на обекта се научават матрици на трансформации, представляващи геометричната връзка между обект и неговите части. PCAE използва базиран на CNN енкодер, предвиждащ една карта на характеристиките за всеки параметър на капсулата, последвана от обединяване, основано на вниманието (attention-based pooling). Активирането на капсулите на частите, описва части, а не пиксели, които могат да имат произволни позиции в изображението. Енкодерите с капсули на обектите предсказват различните позиции на отделните части като рядък набор от обекти, където всеки настоящ обект предвижда няколко части.

В енкодерите с капсули на обектите се използват наборен трансформаторен енкодер (Set Transformer). Процесът на декодиране на капсулите на обектите, към капсулите на частите, се извършва от отделни многослойни перцептрони (Multilayer Perceptrons - MLPs). По един за всяка

капсула на обекта, който предсказва параметрите на капсулата от изходите на наборния трансформаторен енкодер. PCAE декодера, шаблона за частите, по един за всяка капсула на частите. SCAE са дълбоки невронни архитектури, при които произволен енкодер научава еквивариантни представяния от извличане на части и техните пози и групирането им в обекти.

Описаните дълбоки невронни мрежи заемат централно място в изследванията за машинно обучение и при приложенията в областта на компютърното зрение. Разгледаните сложни архитектури се справят с голям спектър от задачи, свързани не само с приложното поле на компютърното зрение. Всички разгледани невронни мрежи могат да обработват сензорен вход под формата на изображение или видео поток. При направения обзор на DNNs, са разгледани мрежи използващи различни подходи на машинното обучение – контролирано, неконтролирано и самоконтролирано обучение. Тенденцията в развитието на сложните невронни архитектури е да се обучават от малък набор от не анотирани входни данни, чрез оптимизиране на загубата. Тези модели също могат да бъдат предварително обучени върху големи немаркирани набори от данни във фаза на самоконтрол, и след това фино настроени за изпълнение на конкретна задача. Голямото предизвикателство пред DL, DNNs и CV е да се справят със задачи като предсказване на липсващи части от изображения или да прогнозират липсващи, или бъдещи видео кадри, от текущи кадри, т.е. да се представи несигурността в прогнозата за изображения или видео поток.