

Tasvirdagi matnlarni tanib olish uchun neyron tarmoqlari tashabbuskorliklari

Mavzuna Xayrullo qizi Karimova
mavzunakarimova71@gmail.com
Muqaddas Baxtiyor qizi Madayeva
muqaddasmadayeva@gmail.com

Annotatsiya: Matni aniqlash, shuningdek, optik belgilarni aniqlash sifatida ham tanilgan, bosma yoki qo'lda yozilgan matni tahrirlash, qidirish va tahlil qilish oson bo'lgan raqamli formatga aylantiradi. Bu matn tasvirlarini tahlil qilishni va ulardagi belgilar va so'zlarni tanib olishni o'z ichiga oladi. Optik belgilarni aniqlash - bu hujjatlarni skanerlashga bo'lgan talab ortib borayotgani va ma'lumotlarni samarali va aniq yozib olish zarurati tufayli tez rivojlanayapti. Optik belgilarni aniqlash ko'plab sohalarda, jumladan bank, sog'liqni saqlash, hukumat va ta'limda muhim texnologiyaga aylandi. Optik belgilarni aniqlash bozoridagi ba'zi imkoniyatlarga quyidagilar kiradi: Katta ma'lumotlar tahlilining yuksalishi: Har kuni hosil bo'ladigan raqamli ma'lumotlar ortib borayotganligi sababli, optik belgilarni aniqlash tasvir va hujjatlar kabi tuzilmagan ma'lumotlar manbalaridan ma'lumotlarni skanerlashi va olishi mumkin. Mashinani o'rganish va chuqur o'rganish yutuqlari: optik belgilarni aniqlash ilg'or mashinani o'rganish algoritmlari va chuqur neyron tarmoqlar yordamida sezilarli darajada yaxshilanishi mumkin, bu aniqlik va samaradorlikni oshiradi.

Kalit so'zlar: neyron tarmoq, algoritim, rasm, tasvir, CNN, RNN, model arxitekturasi

Neural network initiatives for image text recognition

Mavzuna Xayrullo kizi Karimova
mavzunakarimova71@gmail.com
Muqaddas Baxtiyor kizi Madayeva
muqaddasmadayeva@gmail.com

Abstract: Text recognition, also known as optical character recognition, converts printed or handwritten text into a digital format that is easy to edit, search, and analyze. It involves analyzing text images and recognizing characters and words in them. Optical character recognition is rapidly developing due to the increasing demand for document scanning and the need to capture data efficiently and accurately. Optical

character recognition has become an important technology in many industries, including banking, healthcare, government, and education. Some of the opportunities in the optical character recognition market include: The rise of big data analytics: Digital data generated every day as data increases, optical character recognition can scan and retrieve data from unstructured data sources such as images and documents. Advances in machine learning and deep learning: Optical character recognition can be greatly improved with advanced machine learning algorithms and deep neural networks, increasing accuracy and efficiency.

Keywords: neural network, algorithm, picture, image, CNN, RNN, model architecture

Deep Learning yordamida matnni aniqlash

Chuqur o'rganish matnni aniqlash sohasida inqilob qildi, natijada aniqlik va ishlashning keskin yaxshilanishiga olib keldi. Chuqur o'rganishga asoslangan matnni aniqlashning bir nechta yondashuvlari mavjud, jumladan.

- Konvolyutsion neyron tarmoqlari (CNN): CNN ko'pincha tasvirga asoslangan matnni aniqlash uchun ishlatiladi. Kirish tasviri xususiyatlarni ajratib oladigan va matn tasvirini o'rganadigan konvolyutsion qatlamlar tomonidan quvvatlanadi. Keyin CNN chiqishi keyingi ishlov berish va matnni tanib olish uchun takroriy neyron tarmog'iga (RNN) uzatiladi.

- Takroriy neyron tarmoqlari (RNN): RNN qo'l yozuvi va nutqni aniqlash kabi ketma-ketlikka asoslangan matnni aniqlashda keng qo'llaniladi. RNNlar ketma-ket ma'lumotlarni qayta ishlash uchun qayta aloqa zanjirlaridan foydalanadilar, bu ularga uzoq muddatli bog'liqliklar va kontekstli ma'lumotlarni olish imkonini beradi.

- Enkoder-dekoder tarmoqlari: Kodlovchi-dekoder tarmoqlari matnni oxirigacha aniqlash uchun ishlatiladi. Kirish tasviri avval xususiyat vektoriga kodlanadi, so'ngra belgilar yoki so'zlar ketma-ketligiga dekodlanadi. Ushbu tarmoqlar samaradorlik va aniqlikni oshirib, oxirigacha o'qitilishi mumkin.

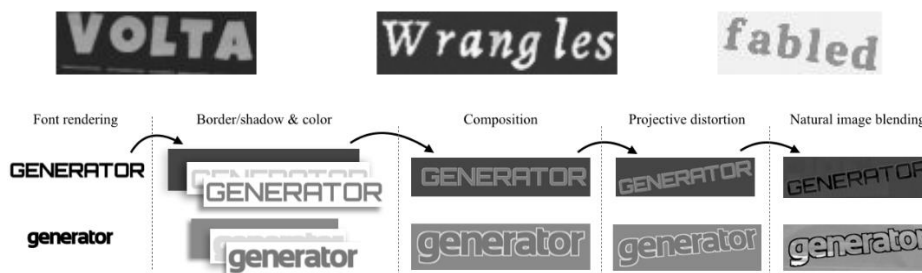
Matnni aniqlash modelimizni yaratish uchun quyidagi bosqichlardan fodalaniadi:

1. Ma'lumotlar to'plamini yig'ish
2. Ma'lumotlarni oldindan qayta ishlash
3. Tarmoq arxitekturasini yaratish
4. Yo'qotish funksiyasini aniqlash
5. Trening modeli
6. Bashoratdan olingan natijalarni dekodlash

Ma'lumotlar to'plamini yig'ish

Biz Visual Geometry Group tomonidan taqdim etilgan ma'lumotlardan foydalanildi. Bu jami 10 GB hajmdagi katta ma'lumotlar to'plami. Bu yerda men tekshirish ma'lumotlar to'plami uchun 0,08% tasvirlar bilan mashg'ulot to'plami

uchun atigi 40000 ta rasmdan foydalandim. Bu ma'lumotlar quyida ko'rsatilgan rasmlarga o'xshash matnli tasvir segmentlarini o'z ichiga oladi:



Ma'lumotlarni oldindan qayta ishlash

Kirish tasvirini ham, chiqish yorlig'ini ham oldindan qayta ishlashimiz kerak. Kirish tasvirimizni oldindan qayta ishlash uchun quyidagilarni ishlatamiz:

1. Tasvirni o'qing va kulrang o'lchamdagi tasvirga aylantirish;
2. To'ldirishdan foydalanib, har bir rasmni (128,32) o'lchamda qilish;
3. Arxitekturaning kirish shakliga mos kelishi uchun tasvir o'lchamini (128,32,1) sifatida kengaytirish.
4. Rasm pikseli qiymatlarini 255 ga bo'lish orqali normallashtiring.

Chiqish yorliqlarini oldindan qayta ishlash uchun quyidagilardan foydalaning:

1. Tasvir nomidagi matnni tasvir nomidan o'qish mumkin, chunki tasvirning ichida yozilgan matn mavjud.
2. Funksiya yaratish orqali so'zning har bir belgisini qandaydir raqamli qiymatga kodlang ("a":0, "b":1 'z':26 va hokazo). Aytaylik, bizda "abab" so'zi mavjud bo'lsa, bizning kodlangan yorlig'imiz [0,1,0,1] bo'ladi.

1. So'zlardan maksimal uzunlikni hisoblang va uni maksimal uzunlik bilan bir xil o'lchamda qilish uchun har bir chiqish yorlig'ini to'ldiradi. Bu RNN arxitekturasi chiqish shakliga mos kelishi uchun amalga oshiriladi.

Oldindan ishlov berish bosqichida ikkita ro'yxatni yaratishimiz kerak: biri yorliq uzunligi, ikkinchisi esa RNN ga kirish uzunligi. Ushbu ikkita ro'yxat bizning CTC yo'qotish uchun muhim. Yorliq uzunligi - har bir chiqish matni yorlig'ining uzunligi va kirish uzunligi arxitekturada 31 bo'lgan LSTM qatlamiga har bir kirish uchun bir xil bo'ladi.

Model arxitekturasi

Model arxitekturasi yaratish va uni oldindan ishlangan ma'lumotlar bilan o'rnatish.

$$Model = CNN + RNN + CTC \text{ yo'qolishi}$$

Modelimiz uch qismdan iborat:

1. Konvolyutsion neyron tarmoq qatlami

Konvolyutsion neyron tarmog'i (CNN yoki ConvNet) - bu vizual tasvirlarni tahlil qilishda muvaffaqiyatli qo'llanilgan, sun'iy neyron tarmoqlarni ta'minlaydigan chuqur o'rganish sinfidir. CNN har qanday tasvirni parcha-parcha solishtiradi va uni aniqlash

vaqtida qidiradigan qismlar xususiyat deb ataladi. Tasvirdan xususiyatlarni ajratib olish uchun konvolyutsion neyron tarmoqdir. CNNda uchta asosiy operatsiya mavjud: a) Konvolyutsiya. b) ReLU. c) Pooling yoki Sub

a) konvolyutsiya qatlami

CNN uchun Convolutionning asosiy maqsadi kirish tasviridan xususiyatlarni olishdir. Har bir konvolyutsiya qatlami to'rt o'lchamli $N \times Rangli \text{ kanal} \times kenglik \times balandlik$ dagi to'plamli kirish sifatida tasvirni oladi. Yadrolar yoki filtrlar, shuningdek, to'rt o'lchovli (xususiyatlar xaritalari soni, xususiyat xaritalari soni, filtr kengligi va filtr balandligi) o'rganilishi mumkin bo'lgan parametrlar (og'irliklar va noaniqliklar) to'plamidir. Har bir konvolyutsiya qatlamida to'rt o'lchovli konvolyutsiya tasvirlar to'plami va xususiyat xaritalari o'rtasida ular orasidagi p nuqta bilan hisoblanadi. Konvolyutsiyadan so'ng o'zgarishlar faqat tasvirning kengligi va balandligi bo'ladi.

b) Rektifikatsiya qilingan chiziqli birlik

ReLU deb nomlangan qo'shimcha operatsiya har bir Convolution operatsiyasidan so'ng ishlatilgan. Rektifikatsiyalangan chiziqli birlik (ReLU) neyron tarmog'ining xujayrasi bo'lib, uning x ga berilgan chiqishini hisoblash uchun quyidagi faollashtirish funksiyasidan foydalanadi:

$$R(x) = \text{Maks}(0, x)$$

c) Pooling yoki Sub

Ushbu qatlamda parametrlar va hisob-kitoblarni kamaytiradigan kichraytirilgan xaritalarni olish uchun xususiyat xaritasining o'lchami kamayadi. Birlashtirish Maks, O'rtacha yoki Sum bo'lishi mumkin. Konvolyutsiya qatlamidagi filtrlar soni birlashtirishdan olingan chiqish xaritalari soni bilan bir xil. Pooling tuzatilgan xususiyat xaritalaridan ma'lumot oladi va keyin algoritmgga muvofiq uni qisqartiradi.

2. Takrorlanuvchi neyron tarmoq

Vaqt bosqichida ketma-ket chiqishni bashorat qilish uchun takroriy neyron tarmoq. Takroriy neyron tarmog'i (RNN) neyron tarmog'ining bir turi bo'lib, oldingi bosqichdagi chiqish joriy bosqichga kirish sifatida beriladi. gapning keyingi so'zini bashorat qilish kerak bo'lganda, oldingi so'zlar talab qilinadi va shuning uchun oldingi so'zlarni eslab qolish zarurati tug'iladi. RNN ning asosiy va eng muhim xususiyati - bu ketma-ketlik haqidagi ba'zi ma'lumotlarni eslab turadigan yashirin holat. RNNda hisoblangan narsalar haqidagi barcha ma'lumotlarni eslab qoladigan "xotira" mavjud.

Uzoq qisqa muddatli xotira (LSTM) tarmoqlari ketma-ketlikni bashorat qilish muammolarida tartibga bog'liqlikni o'rganishga qodir bo'lgan takrorlanuvchi neyron tarmoq turidir. Bu mashina tarjimasini, nutqni aniqlash va boshqalar kabi murakkab muammoli sohalarda talab qilinadigan xatti-harakatlardir. LSTMlar chuqur o'rganishning murakkab sohasidir. LSTMlar nima ekanligini va ikki tomonlama va ketma-ketlik kabi atamalar sohaga qanday bog'liqligini bilish qiyin bo'lishi mumkin. LSTMlar vaqt va qatlamlar bo'yicha qaytarilishi mumkin bo'lgan xatoni saqlab

qolishga yordam beradi. Doimiyroq xatolikni saqlab, ular takrorlanuvchi tarmoqlarga ko'p vaqt davomida (1000 dan ortiq) o'rganishni davom ettirishga imkon beradi va shu bilan sabab va oqibatlarini masofadan bog'lash uchun kanal ochadi. LSTM yacheykasida axborot oqimini tartibga soluvchi uchta turli eshiklari bor.

1. eshikni unutilish;
2. kirish eshigi;
3. chiqish eshigi.

CTC yo'qotish funksiyasi

Har bir vaqt bosqichi uchun chiqishni bashorat qilish uchun ishlatiladigan transkripsiya qatlami bo'lgan CTC yo'qotish funksiyasi. CTC LOSS Aleks Graves RNNni o'qitish uchun ishlatiladi, bu qo'lda yozilgan hizalanish muammosini bartaraf qiladi, chunki qo'lda yozilgan har bir yozuvchining moslashuvi har xil. Shunchaki tasvirda yozilgan narsalarni (Ground Truth Text) va BLSTM chiqishini beriladi, keyin u salbiy maksimal ehtimollik yo'lini minimallashtirish maqsadida yo'qotishni hisoblab chiqadi.

Foydalanilgan adabiyotlar

1. Chjan, L.; Quyosh, L.; Li, V.; Chjan, J.; Cai, V.; Cheng, C.; Ning, X. Barmoq tomirlarini aniqlash uchun qisman eng kichik kvadratlar diskriminant tahliliga asoslangan qo'shma bayesian ramka. IEEE Sens. J. 2021 , 22 , 785-794. [Google Scholar] [CrossRef]
2. Ning, X.; Tian, V.; Yu, Z.; Li, V.; Bay, X.; Vang, Y. Hcfnn: Tasvirni tasniflash uchun yuqori tartibli qamrov funksiyasi neyron tarmog'i. Shaklni tan olish. 2022 , 131 , 108873. [Google Scholar] [CrossRef]
3. Ning, X.; Tian, V.; U, F.; Bay, X.; Quyosh, L.; Li, W. Giper-kolbasa qamrovi funksiyasi neyron modeli va tasvirni tasniflash uchun o'rganish algoritmi. Shaklni tan olish. 2023 , 136 , 109216. [Google Scholar] [CrossRef]
4. Chen, Z.; Silvestri, F.; Vang, J.; Chju, X.; Ahn, H.; Tolomei, G. Relax: O'zboshimchalik bilan bashorat qilingan modellar uchun mustahkamlovchi o'rganish agenti tushuntiruvchisi. Axborot va bilimlarni boshqarish bo'yicha 31st ACM xalqaro konferentsiyasi materiallarida, Atlanta, GA, AQSh, 2022 yil 17–22 oktyabr; 252–261-betlar. [Google olimi]
5. Vey, X.; Saha, D. Bilgan: Simsiz kanallardan neyron tarmoqlardan foydalangan holda kalitlarni yaratish. Simsiz xavfsizlik va mashinalarni o'rganish bo'yicha 2022 yilgi ACM seminari materiallarida, San-Antonio, TX, AQSh, 2022 yil 19 may; 45–50-betlar. [Google olimi]