ALTERNATOR
Misliti znanost.

Kako sodobne naprave samodejno prepoznajo hrano s fotografij?

23. 7. 2020

Številka: 35/2020

Avtor:

- Simon Mezgec



Foto: Arne Hodalič

Samodejno razpoznavanje fotografij hrane z razvojem sodobnih pristopov v zadnjih letih postaja vedno bolj priljubljen računalniški problem. Uporabnost reševanja tega problema se kaže v avtomatiziranju beleženja vnosa živil, kar je pogoj za izboljšanje kakovosti prehrane posameznika. Tradicionalno so se za ta namen uporabljali pristopi, kot so prehranski dnevniki in vprašalniki, vendar ti pristopi terjajo ročno delo posameznika, saj je treba po vsaki zaužiti jedi ali pijači shraniti podatke o tem, katera vrsta hrane je bila zaužita, o načinu priprave, količini oziroma masi itd., kar ne zagotavlja natančnega beleženja vnosa hrane, saj je recimo ocenjevanje količine za večino težavno, tehtanje zamudno, posameznik lahko podatke nehote prireja ali pa izgubi začetno motivacijo. S samodejnim razpoznavanjem slik hrane se zmanjša težavnost in skrajša trajanje vnašanja za posameznika, pa tudi odstrani subjektivnost človeškega vnosa. Po zajemu fotografije hrane algoritem samodejno prepozna to hrano in shrani podatke o fotografirani hrani za oceno kakovosti prehranskega vnosa. Ta pristop se torej lahko uporabi pri aplikacijah za pametne naprave, ki omogočajo tako beleženje vnosa živil kot tudi analizo kakovosti prehrane. Ta prispevek predstavlja, kako poteka razpoznavanje slik hrane, kakšni so njegovi začetki in ta čas najbolj natančne metode za razpoznavanje.

Težavnost in začetki razpoznavanja slik hrane

Razpoznavanje slik hrane sodi med probleme računalniškega vida. Gre za področje, katerega začetki segajo v 60. leta prejšnjega stoletja in je zaradi dobrih rezultatov postalo priljubljeno v zadnjih letih. Cilj računalniškega vida je visokonivojsko razumevanje slik in videov, kar algoritem pridobi tako, da obdela vhodno sliko ali sličice videa z različnimi pristopi, kot so detekcija, segmentacija ali razpoznavanje slike, ter vrne uporabno informacijo. Pri razpoznavanju slike algoritem razvrsti predmete na sliki v izhodne kategorije – pri razpoznavanju hrane te kategorije predstavljajo posamezne vrste hrane in pijač. Razpoznavanje poteka tako, da se posameznim predmetom razpoznavanja dodeli *značilnosti*, s pomočjo katerih jih lahko zanesljivo razlikujemo med seboj. Ta postopek je bil tradicionalno največji izziv pri računalniškem vidu, saj je treba značilnosti (kakšne barve je predmet, kaj šteje kot rob predmeta, kako velik je predmet itd.) najprej definirati tako, da zajamejo vse vizualne variacije predmetov, potem pa te značilnosti interpretirati, tako da so predmeti klasificirani v pravilne kategorije.

Problem razpoznavanja slik hrane velja za enega bolj zapletenih, kar je posledica vizualnih lastnosti hrane. Oteževalne okoliščine so:

- enaka jed se lahko pojavi v veliko različnih oblikah, kar je posledica tako priprave kot načina zajema slike;
- različne jedi so si na slikah lahko zelo podobne, kar otežuje razlikovanje med njimi;
- velik delež vizualne informacije sestavin se izgubi pri pripravi jedi;
- za pijače je na voljo zelo malo vizualnih informacij – po navadi samo barva, količina in način serviranja;
- različnih vrst hrane in pijač je ogromno.

Pri večini ostalih problemov računalniškega vida, kot so razpoznavanje šarenice, zaznavanje avtomobilov, sledenje

športnikov na igrišču ali nadzor kakovosti proizvodnega procesa, teh težav ni ali pa jih je bistveno manj. Kako lahko torej algoritem sam prepozna, katera vrsta hrane je na sliki, in to brez človeškega posredovanja?

Prva ideja raziskovalcev na tem področju je bila *ročna definicija značilnosti* hrane. Pri tem pristopu raziskovalec sam določi, katere značilnosti naj algoritem išče na sliki. Če želi razpoznati recimo paradižnik, v kodo algoritma eksplicitno napiše, kakšne oblike naj bo paradižnik, kakšne barve, kakšne teksture, mogoče tudi, kje na sliki je ali s čim je obdan, algoritem pa potem glede na te definicije išče ujemajoče se predmete na sliki. To je nato treba ponoviti za čisto vsako vrsto hrane. Kot je omenjeno zgoraj, pa je lahko vizualna podoba hrane zelo drugačna od slike do slike, še zlasti, kadar so te narejene s pametnim telefonom ob neidealnih pogojih – kar je na eni sliki okroglo, je na drugi lahko podolgovato, kar je na eni sliki rdeče barve, je na drugi lahko oranžne, in kar ima na eni sliki gladko teksturo, ima lahko na drugi grobo. Zaradi tega se je pri teh raziskavah pogosto zgodilo, da je razvit pristop deloval razmeroma dobro na zbirki slik, ki so jo raziskovalci uporabili za testiranje, na slikah iz resničnega sveta pa ni mogel doseči zadovoljive natančnosti, saj raziskovalci niso predvideli vseh mogočih vizualnih variacij, ki jih lahko ista vrsta hrane zavzame na različnih slikah. Oziroma bolj točno – raziskovalci niso mogli predvideti vseh variacij, ker jih je enostavno preveč, zaradi česar so v veliki večini ročne metode dosegale natančnost razpoznavanja krepko pod 40 % (https://ieeexplore.ieee.org/document/5539907) in so bile omejene na eno živilo na sliki, kot je prikazano na spodnji sliki. Za praktično uporabnost to žal ni sprejemljiv rezultat.



Primer fotografije z enim živilom. Takih fotografij je veliko manj kot pa fotografij z več vrstami hrane. Rezultat razpoznavanja bi bila enostavna tekstovna označba »ocvrt krompir«.

Nevronske mreže na pomoč

Najboljše rezultate v razpoznavanju slik hrane so dosegli pristopi, ki uporabljajo *umetne nevronske mreže*. Kot že ime nakazuje, umetne nevronske mreže temeljijo na delovanju nevronov v človeških možganih. Sestavljene so iz slojev nevronov, skozi katere potujejo vhodni podatki (v našem primeru slike hrane) in v vsakem sloju se nevronska mreža sama nauči značilnosti slike različnih kompleksnosti – od preprostih značilnosti v začetnih slojih, kot so oblike in robovi, do najbolj zapletenih v končnih slojih, kot je vrsta jedi. Pomembno je poudariti, da se vse zgodi samodejno – nevronske mreže torej same ugotovijo, katere značilnosti so potrebne za razpoznavanje, in se jih »naučijo«, kar pomeni, da raziskovalcu ni treba ročno definirati teh značilnosti kot pri ostalih pristopih, temveč samo izhodne kategorije (vrste jedi). Učenje nevronskih mrež poteka tako, da mreža obdeluje učno zbirko slik ter v vsakem koraku učenja spreminja povezave med nevroni, tako da je funkcija izgube (angl. *loss function*) čim nižja. Ta funkcija ocenjuje, kako natančno mreža opisuje različne vrste hrane, in se meri na drugi zbirki slik, ki se razlikuje od učne zbirke. Za učenje so slike po navadi razdeljene

na tri zbirke: učno zbirko, ki se uporabi za učenje značilnosti, validacijsko zbirko, ki se uporabi za ocenjevanje natančnosti razpoznavanja v vsakem koraku učenja, ter testno zbirko, ki se uporabi za enkratno ocenjevanje natančnosti po zaključenem učenju.

Umetne nevronske mreže obstajajo že kar nekaj časa, vse od leta 1958. Razlog, da do nedavnega niso bile učinkovite pri razpoznavanju slik, je bila njihova velikost. Učenje nevronske mreže je namreč izrazito računsko zahtevno, zato so bile v preteklosti uporabljene nevronske mreže z majhnim številom slojev, ki niso bile sposobne učenja zadostnega števila značilnosti slik. V zadnjem desetletju pa so po zaslugi vedno večje računske moči računalnikov, predvsem grafično procesnih enot (GPE), ter učinkovitejših pristopov uporabljene globoke nevronske mreže, ali bolj splošno, *globoko učenje* (angl. *deep learning*). Te mreže so sestavljene iz veliko večjega števila slojev – vsebujejo lahko tudi preko 1000 slojev oziroma preko 10 milijard povezav med nevroni. Posledično so se sposobne naučiti ogromnega števila značilnosti in veliko bolje razlikujejo med različnimi vrstami hrane. Kljub temu gre za pristop, ki šteje kot ozka umetna inteligenca (angl. *narrow artificial intelligence*), kar pomeni, da je omejena na ozko definirano nalogo, v tem primeru razpoznavanje slik hrane, in ni sposobna reševanja drugih nalog. Danes sta priljubljeni predvsem dve vrsti globokih nevronskih mrež: ponavljajoče se nevronske mreže (angl. *recurrent neural networks*), ki vsebujejo časovno komponento in so primerne za probleme, kot so jezikovno modeliranje in razpoznavanje govora, ter konvolucijske nevronske mreže (angl. *convolutional neural networks*), ki so namenjene razpoznavanju slik in vsebujejo konvolucijske sloje nevronov, ki iščejo lokalne značilnosti slik. Na slednje lahko gledamo kot na sestavljene funkcije, ki so implementirane tako, da vsebujejo veliko število operacij (množenje matrik) v vsakem sloju nevronov, zaradi česar jih je mogoče naučiti veliko hitreje z uporabo GPE-jev, ki te operacije računajo vzporedno na ogromnem številu procesorskih jeder. Z globokimi konvolucijskimi nevronskimi mrežami so raziskovalci na področju razpoznavanja slik hrane dosegli stopnjo natančnosti, višjo od 90 % (https://ieeexplore.ieee.org/document/8354172) .

Razpoznavanje več vrst hrane na sliki

Zgoraj omenjene stopnje natančnosti veljajo za razpoznavanje ene vrste hrane oziroma pijače na sliki. Čim sta na sliki dve ali več vrst hrane, kot je prikazano na spodnji sliki, ta pristop odpove. Na srečo pa se lahko globoko učenje uporabi tudi za bolj napredno razpoznavanje, kjer je cilj razpoznati čisto *vsako slikovno točko (piksel)* slike. Ko vemo, katere točke na sliki pripadajo recimo goveji juhi, ni več omejitve glede števila različnih vrst hrane in pijač na sliki. Tovrstno razpoznavanje nudi nevronski mreži tudi dodatno informacijo, saj lahko le-ta ugotovi, kje na sliki je določena vrsta hrane, in se tako lažje odloči, za katero vrsto hrane gre.



Primer fotografije z več vrstami hrane. Levo: izvorna fotografija, desno: rezultat razpoznavanja na nivoju slikovnih točk (označba vsakega živila na fotografiji).

Če za globoke nevronske mreže v splošnem velja, da so računsko zelo zahtevne, pa to še toliko bolj drži za globoke nevronske mreže, ki razpoznavajo vsako točko slike – te namreč zahtevajo za velikostni razred daljši čas učenja. Zato je ta pristop postal priljubljen šele v zadnjih nekaj letih (https://pubmed.ncbi.nlm.nih.gov/29623869/) , ko je računska moč še bolj napredovala. Kljub temu pa je najboljši doslej, saj ni omejen s količino hrane na sliki oziroma z načinom zajemanja slike in omogoča samodejno računanje količine oziroma mase hrane, ker vemo, kolikšen delež slike določena vrsta hrane zajema. Nedavno je bilo za pristop razpoznavanja slik hrane na nivoju slikovnih točk organizirano mednarodno spletno tekmovanje Food Recognition Challenge (https://www.aicrowd.com/challenges/food-recognition-challenge) , kjer so najboljši algoritmi dosegli okrog 60 % natančnost (https://www.aicrowd.com/challenges/food-recognition-challenge) . Nižja natančnost glede na problem razpoznavanja ene vrste hrane na sliki je posledica zahtevnosti problema – veliko težje je pravilno razpoznati vsako točko slike kot pa celotno sliko. Kljub temu gre za vzpodbuden rezultat, ki se bo predvidoma še izboljševal v prihajajočih letih. Ni izključeno, da ne bo globoko učenje v prihodnosti doseglo stopnje natančnosti razpoznavanja, ki *presega človeško*, kar se je sicer že zgodilo pri nekaterih drugih problemih računalniškega vida, kot je recimo prepoznavanje lokacije kraja na sliki.

Odvisnost od podatkov

Globoke nevronske mreže imajo zelo pomembno pomanjkljivost – natančnost razpoznavanja je v veliki meri odvisna od števila in kakovosti slik, uporabljenih za učenje. Ni dovolj, če damo nevronski mreži na voljo samo nekaj slik, ker se bo v tem primeru lahko naučila le preprostih značilnosti hrane. Potrebno je res veliko število slik – na desettisoče, stotisoče ali celo milijone slik. Pravzaprav so globoke nevronske mreže tako požrešne, da za precej problemov še ni bila dosežena zgornja meja glede števila slik – *čim več slik, tem bolje*. Zanimivo je omeniti, da je pri razvoju rešitev s pomočjo globokega učenja še vedno ključen človeški element, saj slike, uporabljene za učenje, v veliki meri ročno posredujejo ljudje in jih naprave ne pridobivajo samodejno. Pomembno pa ni samo skupno število slik, ampak mora zbirka vsebovati tudi dovolj slik za vsako vrsto hrane oziroma pijače. Poleg tega morajo slike zajemati čim več vizualnih variacij, ki jih posamezen predmet razpoznavanja lahko zavzame na slikah. Če je recimo uporabljena zbirka sestavljena samo iz slik, zajetih ob dnevni svetlobi, in te slike vsebujejo jedi, pripravljene in servirane samo na en način, bo algoritem natančno razpoznaval samo slike, zajete pod temi pogoji – če bi posameznik posnel sliko ob drugačnih pogojih, algoritem najverjetneje ne bi pravilno razpoznal slike. Gre za problem preprileganja (angl. *overfitting*), ki je zelo razširjen in pomemben izziv pri razvoju rešitev z globokimi nevronskimi mrežami. Pri premajhnih ali preveč specifičnih zbirkah slik se lahko preprileganju izognemo s predčasno ustavitvijo učenja – nevronska mreža se tako nauči manj značilnosti, ki jih je najti samo v uporabljeni zbirki slik.

Pristopi razpoznavanja slik hrane prihodnosti

Kljub temu, da so trenutni rezultati na področju razpoznavanja slik hrane vzpodbudni, pa se jih da še izboljšati. Globoko učenje je najbolj natančen pristop, vendar ta še vedno zaostaja za človekom. Razlogi za nižjo natančnost večinoma tičijo v pomanjkanju robustnosti razvitih rešitev – hrana se lahko na slikah pojavi v veliko različnih oblikah, zaradi česar so te rešitve še bolj odvisne od velikosti in raznovrstnosti zbirk slik, ki so uporabljene za učenje. Te zbirke danes še ne zadovoljujejo dovolj teh pogojev, zaradi česar raziskovalci menijo, da bo s povečevanjem ter izboljševanjem teh zbirk v prihodnosti natančnost razpoznavanja še naraščala. Z drugimi besedami – za naslednji preskok v natančnosti ni nujno potreben razvoj novih pristopov, ampak pridobitev več in boljših vhodnih podatkov.

https://www.alternator.science/sl/daljse/kako-sodobne-naprave-samodejno-prepoznajo-hrano-s-fotografij/