

Co widzą głębokie sieci neuronowe?

Patryk MIZIUŁA*

W ciągu ostatnich kilku lat świat naukowo-techniczny nauczył się uczyć tzw. głębokie sieci neuronowe rozpoznawania treści obrazów. Rezultaty są spektakularne: dobrze nauczony model potrafi znaleźć na obrazku wszystkie zwierzęta i rozróżnić ich gatunki, przerobić zwykłe zdjęcie tak, żeby wyglądało na namalowane przez Picassa, czy domalować brakujący kawałek przedmiotu, którego nigdy wcześniej „nie widział”. A wszystko opiera się na prostym przepisie: weź model matematyczny (nieskomplikowany pojęciowo, ale o wielkiej liczbie parametrów), dodaj jak najwięcej mocy obliczeniowej (w praktyce kart graficznych), poczekaj.

Obraz może być przechowywany w wersji elektronicznej jako zbiór pikseli (obraz rastrowy) lub krzywych matematycznych (obraz wektorowy). Obrazy wektorowe mają lepsze własności niż rastrowe (np. skalują się bez utraty jakości), niestety, potrafią odwzorowywać w zadowalający sposób jedynie względnie proste obiekty (np. logotypy). Dlatego w fotografii i uczeniu maszynowym używa się obrazów rastrowych. To, co my widzimy na monitorze jako kolorowy rastrowy obrazek, dla komputera jest *tensorem*, czyli trójwymiarową tablicą, o wymiarach $n \times m \times 3$, składającą się z pikseli. Trzeci wymiar odpowiada trzem kanałom kolorów – czerwonemu, zielonemu i niebieskiemu. Każdy z $n \cdot m \cdot 3$ pikseli jest liczbą ze zbioru $\{0, 1, \dots, 255\}$. Im większa liczba, tym większe „natężenie” danego koloru w danym miejscu obrazka.

W artykule *Głębokie uczenie maszyn*, Δ_{18}^1 , Paweł Gora opisuje mechanikę głębokich sieci neuronowych. W niniejszej notce opowiem o kilku ich popularnych i efektywnych zastosowaniach w zagadnieniach, w których danymi są obrazki.

Klasyfikacja

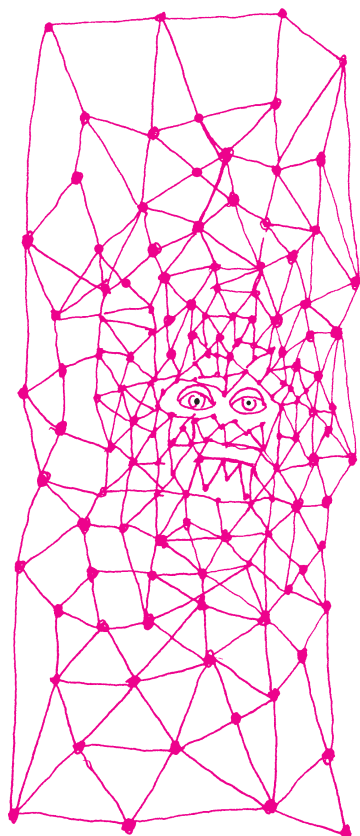
W zadaniu klasyfikacyjnym algorytm wie, że na obrazku jest obiekt należący do jednej ze znanych kategorii (np. pies, kot lub chomik) i jego zadaniem jest wskazanie, do której z nich rzeczywiście należy. Ścisłej rzecz ujmując, obrazkowi przypisywane są prawdopodobieństwa należenia do poszczególnych kategorii (np. pies 70%, kot 20%, chomik 10%).

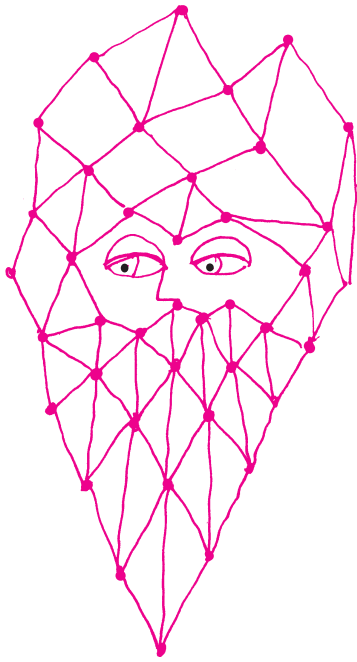
Ciekawym przykładem problemu klasyfikacji był konkurs *Right Whale Recognition* zorganizowany w 2015 roku przez NOAA Fisheries. Organizacja ta dostarczyła zdjęcia lotnicze ukazujące wszystkie 447 żyjących sztuk waleni biskajskich (*Eubalaena glacialis*). Zdjęcia wykonane zostały o różnych porach dnia, przy różnej pogodzie, z różnych kątów itd. Każdy waleń występował średnio na około 10 zdjęciach. Celem konkursu było stworzenie algorytmu rozpoznającego, który z 447 waleń jest na danym zdjęciu. Nieskromnie wspomnę, że firma *deepsense.ai* wygrała ten konkurs, budując algorytm przewyższający skutecznością specjalistów zatrudnionych do „ręcznego” rozpoznawania tych wielorybów. Nasz program w 87% przypadków trafnie wskazywał wieloryba, natomiast dla 95% zdjęć prawidłowy wieloryb znajdował się w pierwszej piątce waleń uznanych za najbardziej prawdopodobne.

Wykrywanie obiektów

W tym zagadnieniu celem jest wykrycie na zdjęciu wszystkich obiektów danego rodzaju. Można np. zażądać od algorytmu, żeby narysował ramki wokół wszystkich samochodów, czyli tak naprawdę żeby podał współrzędne lewych górnych i prawych dolnych rogów wszystkich ramek.

Na zdjęciu przedstawiającym zwykłą ulicę samochody mogą być różnych typów (osobowe, dostawcze, zabawkowe), różnych marek, wielkości, kolorów. Mogą być widziane od przodu, z boku, z tyłu, częściowo zasłonięte, tylko częściowo mieszczące się w kadrze itd. Dla człowieka to nie kłopot, on rozumie, czym jest samochód, więc go rozpozna. Okazuje się, że głębokie sieci neuronowe też





nie mają z tym problemów. Spisują się na tyle dobrze, że są szeroko używane np. w autonomicznych – czyli niesterowanych przez człowieka – pojazdach.

Segmentacja

Segmentacja polega na wskazaniu granic nieregularnych obszarów zdjęcia przedstawiających interesujące nas obiekty. Innymi słowy, każdy piksel obrazka oznaczony jest jako należący bądź nienależący do interesującej nas kategorii.

Dobrym przykładem jest wskazanie na zdjęciu satelitarnym, gdzie dokładnie sięgają pola uprawne albo które dokładnie wiodą drogi. Oprócz narzucającego się zastosowania, czyli sprawnego sporządzania elektronicznych map, monitorowania sytuacji po klęskach żywiołowych itp. segmentacja odgrywa rolę m.in. w medycynie, a dokładniej w diagnostyce obrazowej. Algorytmy wskazują na zdjęciach z prześwietleń (również na „zdjęciach” w 3D) zasięg zmian rakowych lub innych patologicznych zmian, objętość danego typu tkanki itp. Trudno porównać, czy robią to lepiej od lekarzy (to dyskusja na osobny artykuł), ale na pewno robią to sprawniej, wyręczając ludzi w żmudnych zadaniach.

Uczenie ze wzmocnieniem

Od jakiegoś czasu prowadzone są intensywne badania nad automatami grającymi w proste gry z platformy Atari jedynie na podstawie obrazków ukazujących bieżący stan gry. Innymi słowy, algorytmy nie znają „sensu” ani mechaniki gry, nie wiedzą też np., czym w grze skutkuje wybranie akcji „wciśnij na klawiaturze strzałkę w prawo”. Jedyne, co otrzymują, to „zrzut ekranu” z gry po każdym swoim ruchu. To musi wystarczyć im do nauki.

Aktualnie bazujące na obrazach algorytmy, wykorzystujące tzw. model uczenia ze wzmocnieniem, są już w stanie pojąć w grach Atari proste zależności typu „trzeba najpierw zdobyć kluczyk, żeby móc otworzyć drzwi”. Oczywiście daleko im jeszcze do korzystających z innego typu danych sztucznych kolegów ogrywających ludzi najlepszych na świecie w szachy, go czy pokera, ale postęp w czasie jest wyraźny i zapewne za parę lat sytuacja mocno się zmieni.

Przeniesienie stylu

Przeniesienie stylu jest jednym z najbardziej efektywnych zastosowań głębokich sieci neuronowych w ostatnich latach. Styl jest „pobierany” z obrazu referencyjnego i „aplikowany” do treści innego obrazu. Rezultaty są fenomenalne. Na przykład po prawej widzimy Monę Lisę stylizowaną na obraz Picassa.

Innych istniejących zastosowań głębokich sieci neuronowych (nie tylko w analizie obrazu) można by wymienić jeszcze wiele. Dziś największym problemem w tego typu zagadnieniach jest wciąż niedostateczna moc obliczeniowa. Niemniej, jeśli prawo Moore’a pozostanie prawdziwe jeszcze przez jakiś czas, czyli jeśli dostępna moc obliczeniowa wciąż będzie się podwajać co dwa lata, dzięki głębokim sieciom neuronowym za kilkanaście lat prawdopodobnie będziemy żyć w zupełnie innym, inteligentnie zautomatyzowanym świecie.

