# DCNN - metody głebokiego uczenia w rozpoznawaniu obrazów

Jacek Witkowski
Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska
jacek.witkowski@gmail.com

**Streszczenie.** *W artykule zwrócono uwagę na dynamiczny rozwój Sztucznej Inteligencji w ostatnich czasach. Przedstawiono jakie może to przynieść zagrożenia i korzyści dla ludzkości. Następnie objaśniono różnice pomiędzy uczeniem nadzorowanym i nienadzorowanym oraz zaprezentowano przykładowe zastosowania Uczenia Maszynowego w obecnym świecie. Omówiono również temat sztucznych sieci neuronowych: ich budowę oraz zasadę działania. Kolejno, zaprezentowano szczególny przypadek sieci neuronowej, jakim jest sieć splotowa.*

**1. Motywacja.** Ostatnimi czasy Sztuczna Inteligencja jest wykorzystywana w coraz to większej liczbie obszarów. Zaczyna ona być obecna nawet w urządzeniach codziennego użytku. Szybki rozwój dziedziny Uczenia Maszynowego może sprawić, że świat, jaki znamy dzisiaj, w przeciągu najbliższych kilkunastu lat zupełnie zmieni swoje oblicze.

Biorąc pod uwagę to, że Sztuczna Inteligencja może być stosowana w prawie każdej dziedzinie nauki, warto zgłębić wiedzę na temat najnowszych osiągnięć związanych z Uczeniem Maszynowym. W przeciągu ostatnich 5 lat wyjątkowo szybko rozwijały się wszelkie algorytmy służące do rozpoznawania obrazów. Prawdopodobnie było to spowodowane tym, że jest to obszar, w którym można znaleźć najwięcej zastosowań dla Sztucznej Inteligencji.

Jednym z mechanizmów najczęściej wykorzystywanych do rozpoznawania obrazów oraz materiałów wideo jest sieć splotowa, o której między innymi traktuje ten artykuł.

**2. Sztuczna Inteligencja w kulturze** Temat Sztucznej Inteligencji powszechnie kojarzy się z fantastyką naukową. Zagadnienie to było poruszane w wielu książkach (zarówno zagranicznych, jak i polskich autorów, takich jak Stanisław Lem), jak również filmach (np. „I, Robot", „Matrix", „Terminator"). Przeważnie wizja przyszłości jest ukazywana w czarnych barwach: inteligentne maszyny przejmują władzę nad światem (lub przynajmniej mają takie zamiary), a ludzkość musi stawić im czoła.

Na konferencji Web Summit 2016 w Lizbonie Richard Stallman również poruszył problem zniewolenia jakie może nam przynieść szybki rozwój technologii. Jednak obawiał się on, że to nie maszyny zawładną nad światem, lecz wąska grupa osób, która wykorzysta Sztuczną Inteligencję do sterowania społeczeństwem, w taki sposób, by nikt się nie zorientował, że ulega manipulacjom. Stallman zauważył również, że możliwości śledzenia we współczesnym świecie są coraz to większe, co ma negatywny wpływ na wolność społeczeństwa.

Innym problemem związanym z Uczeniem Maszynowym, któremu ludzie będą musieli stawić czoła w przyszłości, jest moralność maszyn. By zobrazować ten problem, można posłużyć się przykładem samoprowadzących się samochodów. W sytuacji, gdy dziecko niespodziewanie wybiegnie na drogę, Sztuczna Inteligencja będzie miała

do wyboru uderzyć w dziecko lub w samochód jadący z przeciwnej strony z czterema dorosłymi osobami w środku. Wówczas pojawia się dylemat natury moralnej: czyje życie jaką ma wartość? Czy pewna śmierć dziecka jest gorszym wyborem niż możliwy zgon wielu pasażerów z drugiego pojazdu? Problem ten został przedstawiony w filmie „I, Robot", w którym robot-ratownik zamiast uratować dziecko, postanowił zająć się głównym bohaterem (granym przez Willa Smitha), gdyż ocenił, że jego szansa na przeżycie jest większa.

Problemem poruszanym przez takie sławy jak Stephen Hawking jest możliwość powstania gigantycznego bezrobocia, gdy inteligentne urządzenia zaczną zastępować ludzi. Jak powiedział: "W przypadku, gdy maszyny będą produkować wszystko, czego będziemy potrzebować, efekt końcowy będzie zależeć od tego, jak rzeczy będą redystrybuowane. Każdy będzie mógł cieszyć się życiem w luksusowym lenistwie, jeśli dochód generowany przez maszyny będzie dzielony, albo ludzie będą przygnębiająco biedni, jeśli właściciele maszyn z powodzeniem będą lobbować przeciwko redystrybucji dóbr. Póki co, raczej obserwujemy trend ku drugiej opcji, a technologia napędza wciąż wzrastającą nierówność." [1].

**3. Uczenie nadzorowane i nienadzorowane** Dzieląc uczenie maszynowe ze względu na informacje dostarczane w procesie uczenia otrzymujemy dwa rodzaje:

- uczenie nadzorowane,
- uczenie nienadzorowane.

**Uczenie nadzorowane** to proces, w którym wraz z danymi wejściowymi sieci, dostarczamy do uczonego mechanizmu również wynik spodziewany na jego wyjściu. Wówczas celem uczenia jest minimalizacja różnicy pomiędzy danymi wygenerowanymi przez mechanizm, a danymi spodziewanymi.

W **uczeniu nienadzorowanym** w procesie uczenia dostarczane są jedynie dane wejściowe (bez spodziewanego wyjścia, czyli tzw. etykiet). Wówczas celem uczenia jest modelowanie danych wejściowych (a dokładniej: rozkładu prawdopodobieństwa danych wejściowych). Przykładowo: przy uczeniu sieci neuronowej w sposób nienadzorowany, jeśli będzie ona otrzymywała obrazki ludzkich twarzy, to zacznie zauważać często występujące zależności pomiędzy pikselami (np.: krawędzie, nos, oczy, usta, jak również same twarze).

Oba rodzaje uczenia mogą być łączone. W ten sposób działa również mózg człowieka. By zobrazować omawiane podejście, warto posłużyć się przykładem. Wyobraźmy sobie, że ktoś chce nas nauczyć rozpoznawać różne rodzaje jabłek. Najprostszym podejściem byłoby pokazywanie nam wielu jabłek i mówienie jakiego rodzaju jest każde z nich. Byłoby to jednak bardzo czasochłonne, gdyż zaprezentowanie każdego jabłka wymagałoby wskazania jakiego jest ono rodzaju. Łatwiejszym sposobem byłoby zamknięcie nas w pokoju z wieloma nieopisanymi jabłkami. Wówczas zaczęlibyśmy oglądać każde z nich i po pewnym czasie zauważylibyśmy różne cechy powtarzające się na niektórych jabłkach (np. podłużny kształt, czerwony kolor, zielony kolor, zielone plamki itp.). Po wyjściu z pokoju potrafilibyśmy identyfikować cechy występujące w różnych grupach jabłek. Wiedza ta pozwoliłaby nam na nauczenie się rozpoznawania różnych gatunków na podstawie znacznie mniejszej liczby opisanych przykładów niż przy podejściu naiwnym, gdzie od początku stosowane jest uczenie nadzorowane.

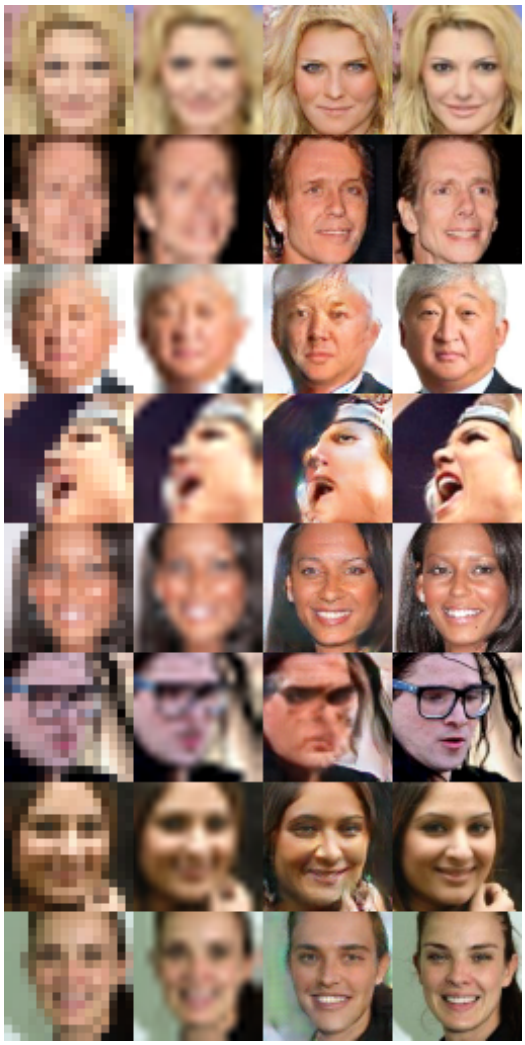
**4. Przykłady zastosowań** By lepiej zrozumieć, jakie możliwości otwiera przed nami obecny rozwój Uczenia Maszynowego, warto poznać zastosowania, jakie znajduje ta dziedzina już teraz. Jednym z nich jest zastosowanie Splotowej Sieci Neuronowej (*ang. Convolutional Neural Network, CNN*) do oceniania atrakcyjności zdjęć „selfie" [2]. Autor algorytmu, potencjalnie bardzo przydatnego dla wszystkich miłośników autofotografii, to Andrej Karpathy. W omawianym

projekcie autor:

1. Wykorzystał zdjęcia, które oznaczone były hasz-tagiem „selfie” (znalazł około 5 milionów takich zdjęć).
2. Przy użyciu innej sieci splotowej wybrał tylko te zdjęcia, na których obecna była przynajmniej jedna twarz (zostało około 2 milionów zdjęć).
3. Posortował fotografie względem liczby użytkowników obserwujących autora danego zdjęcia.
4. Podzielił posortowaną listę fotografii na grupy po 100 zdjęć (każda grupa dzięki temu zawierała zdjęcia o podobnej liczbie obserwujących danego użytkownika).
5. W obrębie każdej grupy, utworzył ranking zdjęć na podstawie ich liczby polubień. Górna połowa zdjęć była uznawana za zdjęcia dobre, a dolna - za złe.
6. Trenował sieć neuronową milionem zdjęć dobrych oraz milionem zdjęć złych. Dzięki temu mechanizm nauczył się odróżniać oba typy zdjęć.

Tak utworzona Splotowa Sieć Neuronowa była w stanie ocenić prawdopodobieństwo, że dane „selfie” jest atrakcyjne (im wynik na wyjściu sieci był wyższy, tym zdjęcie było lepiej oceniane).

Innym przykładem zastosowania Uczenia Maszynowego jest projekt Image Super Resolution [3] autorstwa Davida Garciiego. W swojej aplikacji użył on głębokiego uczenia w celu czterokrotnego powiększania obrazków: z rozmiaru 16x16 pikseli do 64x64 piksele. Porównanie wyników działania sieci neuronowej oraz interpolacji bikubicznej (standardowej metody stosowanej do powiększania obrazów w programach graficznych) zostały przedstawione na rysunku 1. Architektura zastosowana w projekcie to DCGAN (*ang. Deep Convolutional Generative Adversarial Network*, Głęboka Generatywna Antagonistyczna Sieć Splotowa [4]) wykorzystująca uczenie nienadzorowane. W uproszczeniu: autor mechanizmu w procesie uczenia podawał obrazki przedstawiające twarze. Sieć nauczyła się wówczas jakie są najczęstsze zależności pomiędzy pikselami występujące w tych obrazkach, a więc mogła również generować twarze. Następnie autor zastosował kolejny etap uczenia, w którym jako funkcję błędu wykorzystał odległość L1 pomiędzy wynikiem powiększania obrazka przez sieć, a oryginalnym obrazkiem o rozmiarze 64x64 piksele. Było to uczenie nadzorowane.



**Rysunek 1.** Powiększanie obrazków (pierwsza kolumna zawiera powiększany obraz, druga - obraz powiększony poprzez zastosowanie interpolacji bikubicznej, trzecia - obraz powiększony przez sieć splotową, czwarta - oryginalny obraz 64x64 piksele).

Do bardziej przełomowych Sztucznych Inteligencji należy projekt „AlphaGo” [5] stworzony przez firmę Google DeepMind. Owa Sztuczna Inteligencja nauczyła się grać

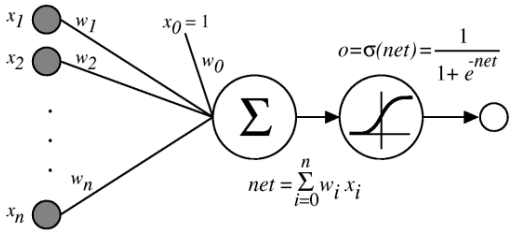
w starochińską grę planszową - Go. Mechanizm wyróżnia to, że może mieć wiele zastosowań (np. może nauczyć się grać w inne gry takie jak Space Invaders, Ms Pacman, Q*Bert) i nie był tworzony pod z góry określone zastosowanie, w przeciwieństwie do mechanizmów takich jak: IBM Watson czy IBM Deep Blue. Silnik stworzony przez firmę — DeepMind — jako dane wejściowe przyjmuje obraz w postaci mapy bitowej, a następnie do jego przetwarzania wykorzystuje głębokie splotowe sieci neuronowe (opisane w dalszej części artykułu) oraz Q-learning (szczególny rodzaj uczenia ze wzmocnieniem).

Program AlphaGo jest przełomowy również pod innym względem: jest pierwszą Sztuczną Inteligencją, która pokonała profesjonalnego gracza w Go (Fana Hui, mistrza Europy) [6] na pełnowymiarowej planszy (19x19) bez zastosowania tzw. handicapu. Pojedynek odbył się w październiku 2015 roku i zakończył się wynikiem 5:0. Niedługo potem, w marcu 2016 roku, AlphaGo zdołał pokonać Lee Sedola [7], 18-krotnego i również ówczesnego mistrza świata w grę Go. Tym samym ostatnia z popularnych gier planszowych straciła mistrza ludzkiego na rzecz programu komputerowego (podobnie jak wcześniej warcaby [8] czy szachy [9]).

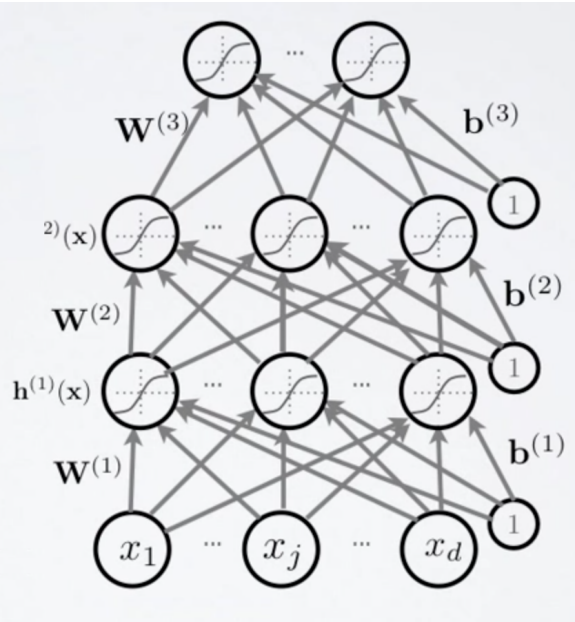
**5. Sieci neuronowe** Jednym z mechanizmów, który jest najczęściej wykorzystywany do implementacji Sztucznej Inteligencji, jest sztuczna sieć neuronowa. Jest ona inspirowana sieciami neuronowymi występującymi w biologii (np. w ludzkim organizmie, w szczególności: w mózgu). Na strukturę sieci składają się pojedyncze połączone ze sobą elementy, zwane neuronami. Budowa neuronu została przedstawiona na rysunku 2. Neurony najczęściej łączone są w tzw. sieci warstwowe. Przykładowa sieć warstwowa została zaprezentowana na rysunku 3. Jeśli liczba warstw występujących w danej sieci jest dostatecznie duża (zazwyczaj: około 10 warstw lub więcej), sieć określana jest jako głęboka.

Algorytm inferencji takiej sieci przedstawia się następująco:

1. Pobierz dane wejściowe, które są wektorem: $\overline{X} = [x_1, x_2, ..., x_n]$.
2. Pomnóż każdą współrzędną wektora wejściowego przez odpowiadającą mu wagę (określaną w procesie uczenia sieci) i zsumuj ze sobą otrzymane wartości: $net = \sum_{i=0}^{n} w_i x_i$.
3. Do otrzymanego wyniku dodaj czynnik $b$ zwany przesunięciem (*ang. bias*): $net := net + b$.
4. Otrzymaną sumę poddaj działaniu funkcji $\sigma$ zwanej funkcją aktywacji: $o = \sigma(net)$. Wynik tej operacji jest wartością wyjściową neuronu.




**Rysunek 2.** Neuron Hebba.




**Rysunek 3.** Warstwowa sieć neuronowa typu Feed-Forward. Źródło: [10]

**6. Splotowe sieci neuronowe** Szczególnym przypadkiem sieci neuronowych są sieci splotowe. Wykorzystują one operację splotu w postaci dyskretnej:

$$(f * g)[n] = \sum_{m=-\infty}^{+\infty} f[m]g[n-m] = \sum_{m=-\infty}^{+\infty} f[n-m]g[m]$$

Dzięki stosowaniu różnych masek (funkcji splatanych z przetwarzanym obrazem) w filtrach splotowych można uzyskiwać różne efekty. Istnieje wiele zdefiniowanych funkcji tego typu, które służą m.in. do:

- redukcji szumów,
- wyostrzania,
- wykrywania krawędzi.

Sieci splotowe wykorzystują to, że odpowiednio dobierając wartości maski, można uzyskiwać bardzo różne efekty. Jednak maska zamiast przyjmować z góry zadane wartości (tak jak w standardowych filtrach splotowych), otrzymuje je w procesie uczenia.

Standardowym zadaniem, do którego wykorzystywane są splotowe sieci neuronowe, jest identyfikacja obiektu obecnego na zdjęciu. Algorytm przetwarzania obrazu dzieli się na etapy:

1. Splot - zastosowanie $n$ różnych filtrów splotowych, wynikiem czego jest $n$ obrazków zwanych mapami cech.
2. Normalizację (krok opcjonalny).
3. Próbkowanie (zmniejszenie rozmiarów powstałych map cech, m.in. w celu redukcji rozmiaru przetwarzanych danych).
4. Powtórzenie kroków 1-3 wiele razy (liczba powtórzeń jest zależna od liczby warstw w sieci).
5. Przekazanie powstałych map cech do warstwy neuronów w pełni połączonej (każdy piksel trafia do wszystkich neuronów tej warstwy sieci).
6. Zastosowanie kolejnych warstw w pełni połączonych (zazwyczaj w sieci stosuje się jedną lub dwie takie warstwy).

Ostatnia warstwa ma za zadanie przedstawić na swoim wyjściu prawdopodobieństwa występowania różnych etykiet (np. kot: 50%, pies: 20%, samochód: 5%, ...).

**7. Podsumowanie** Z powodu gwałtownego wzrostu mocy obliczeniowej w ciągu ostatnich 10 lat znacząco zwiększyły się możliwości mechanizmów wykorzystujących Sztuczną Inteligencję. Wciąż rośnie liczba obszarów, w których Uczenie Maszynowe znajduje zastosowania: od automatycznych tłumaczeń tekstów, poprzez samosterujące się pojazdy, gry planszowe oraz komputerowe, medycynę, aż do wszelkiego rodzaju zagadnień związanych z obróbką obrazów czy materiałów wideo. Ostatnie z wymienionych zastosowań (przetwarzanie materiałów graficznych oraz audio-wizualnych) szczególnie często wykorzystuje głębokie sieci splotowe, czyli specyficzny rodzaj sieci neuronowych, składających się z wielu warstw, wykorzystujących filtry splotowe do osiągania założonych celów (np. identyfikacji obiektów znajdujących się na zdjęciach czy wręcz tworzenia opisów tych obrazów w języku naturalnym).

Z powodu mnogości istniejących zastosowań dla sieci splotowych oraz potencjalnie wielu dziedzin wciąż czekających na wsparcie ze strony mechanizmów tego rodzaju, warto zgłębić tę szczególną gałąź Uczenia Maszynowego. Zadaniem referencyjnym, które jest często wykorzystywane do badań nad sieciami splotowymi jest identyfikacja przedmiotów przedstawionych na obrazkach. Istnieją również bazy opisanych etykietami zdjęć takie jak CIFAR-10 i CIFAR-100 [11] czy ImageNet [12], które mogą posłużyć jako zbiór danych treningowych i testowych. W ramach swojej pracy magisterskiej stworzyłem głęboką splotową sieć neuronową realizującą wymienione zadanie referencyjne. Ze względu na niewielką moc obliczeniową urządzeń, z których mogłem korzystać, posłużyłem się bazą CIFAR-10.

**Literatura**


[1] Stephen Hawking. Stephen Hawking AMA. https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/, Lipiec 2015.

[2] Andrej Karpathy. What a Deep Neural Network thinks about your #selfie. http://karpathy.github.io/2015/10/25/selfie/, Październik 2015.

[3] David Garcia. Image Super Resolution. https://github.com/david-gpu/srez.

[4] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[5] Google Deep Mind. AlphaGO. https://deepmind.com/research/alphago/.

[6] Google Deep Mind. In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go. https://goo.gl/7Vq0kh.

[7] Google Deep Mind. Google achieves AI 'breakthrough' by beating Go champion. http://www.bbc.com/news/technology-35420579.

[8] Jonathan Schaeffer, Neil Burch, Yngvi Björnsson, Akihiro Kishimoto, Martin Müller, Robert Lake, Paul Lu, Steve Sutphen. Checkers is solved. *Science*, vol. 317 no. 5844 1518-1522, Lipiec 2007.

[9] Monroe Newborn. *Kasparov versus Deep Blue : computer chess comes of age*. Springer, Nowy Jork, 1997.

[10] Hugo Larochelle. Neural networks [7.1] : Deep learning - motivation. https://www.youtube.com/watch?v=vXMpKYRhpmI, Listopad 2013.

[11] Alex Krizhevsky, Vinod Nair, Geoffrey Hinton. Cifar-10 and Cifar-100 Datasets. https://www.cs.toronto.edu/~kriz/cifar.html.

[12] Image Net. http://www.image-net.org/.