

Uniwersytet Mikołaja Kopernika  
Wydział Matematyki i Informatyki

Paweł Marcin Chojnacki  
nr albumu: 260082  
informatyka

Praca magisterska

# **Porównanie wydajności współczesnych architektur sieci neuronowych**

Opiekun pracy dyplomowej  
dr hab. Piotr Wiśniewski

Toruń 2018



# Spis treści

<b>Słownik pojęć</b>	<b>5</b>
<b>Wstęp</b>	<b>9</b>
<b>1. Neurony</b>	<b>13</b>
1.1. Model perceptronu prostego . . . . .	14
1.2. Co to jest sieć neuronowa? . . . . .	16
1.3. Jak uczyć sieci neuronowe . . . . .	17
1.4. Elementy sieci neuronowych . . . . .	19
1.4.1. Metoda gradientu prostego . . . . .	19
1.4.2. Funkcje aktywacji . . . . .	20
1.4.3. Parametry . . . . .	21
1.4.4. Hiperparametry . . . . .	22
1.4.5. Wykres obliczeniowy . . . . .	23
1.4.6. Wsteczna propagacja błędu . . . . .	24
<b>2. Deep Learning</b>	<b>27</b>
2.1. Techniki tworzenia głębokich sieci . . . . .	27
2.1.1. Sieci neuronowe . . . . .	27
2.1.2. Convolutional Neural Networks . . . . .	27
2.1.3. Recurrent Neural Networks . . . . .	27
2.1.4. Fully Convolutional Networks . . . . .	27
<b>3. Biblioteki implementujące uczenie głębokich sieci neuronowych</b>	<b>29</b>
3.0.1. Tensorflow . . . . .	29
3.0.2. Keras . . . . .	32
3.0.3. PyTorch . . . . .	33
3.0.4. Tensorflow.js [Deeplearn.js] . . . . .	35
3.0.5. PaddlePaddle . . . . .	37
3.0.6. MXNet . . . . .	39

3.0.7. Caffe2 . . . . .	40
3.0.8. ML.NET . . . . .	42
3.0.9. CNTK . . . . .	43
<b>4. Architektura</b>	<b>45</b>
4.1. Znaczenie architektury dla wydajności sieci . . . . .	45
4.2. Przegląd najbardziej efektywnych architektur . . . . .	47
4.2.1. LeNet . . . . .	47
4.2.2. AlexNet . . . . .	47
4.2.3. VGG Net . . . . .	48
4.2.4. GoogleNet / Inception . . . . .	48
4.2.5. ResNet . . . . .	49
4.2.6. ResNeXt . . . . .	49
4.2.7. RCNN – obrazek jest zastępczy nie ma nic w intere- cie interesującego (zrobić własną wizualizację) . . . .	49
4.2.8. YOLO - You Only Look Once . . . . .	50
4.2.9. SqueezeNet . . . . .	50
4.2.10. SegNet . . . . .	50
4.2.11. Generative Adversarial Network . . . . .	50
<b>5. Testy wydajności</b>	<b>55</b>
<b>Podsumowanie</b>	<b>57</b>
<b>Spis rysunków</b>	<b>57</b>

# Słownik pojęć

## Klasyfikacja binarna

Typ zadania klasyfikacji, którego wyjście stanowi jedna z dwóch wzajemnie wykluczających się klas. Prostym przykładem jest system rozpoznający wiadomości email podzielone na dwie klasy "spam" oraz "nie spam".

## Regresja logistyczna

Metoda statystyczna używana do analizy zbioru danych, w którym istnieje więcej niż jedna zmienna determinująca wyjście. Wyjście jest prawdopodobieństwem na ile obiekt wejściowy przypomina każdą z klas modelu. Regresja logistyczna jest zazwyczaj używana do klasyfikacji binarnej.

## Funkcja kosztu

Funkcja służąca trenowaniu modelu regresji logistycznej. Pozwala na określenie odległości wartości wyjściowej sieci od celu (prawidłowej odpowiedzi). Celem sieci neuronowej jest zminimalizowanie funkcji kosztu dla sygnałów wejściowych nie spotkanych wcześniej.

## Metoda gradientu prostego (ang. gradient descent algorithm)

Algorytm pozwalający znaleźć minimum funkcji kosztu. Technika ta minimalizuje błąd w odniesieniu do parametrów modelu zadanych do trenowania przez wyliczenie optymalnych wartości wag oraz progów.

## Wykres obliczeniowy (ang. computation graph)

Dekompozycja wyrażenia (matematycznego) w pojedynczych atomowych krokach. Używany do szukania miejsc możliwego zoptymalizowania funkcji. Każdą operację wyrażenia przedstawia się jako osobny węzeł w grafie skierowanym. Wykresów obliczeniowych używa się podczas ręcznej analizy funkcji błędu oraz jako wizualizacji obliczeń w wielu bibliotekach głębokiego uczenia maszynowego.

### **Funkcja aktywacji**

Jedna z funkcji (zazwyczaj sigmoidalna lub rektyfikowalna jednostka liniowa) która przyjmuje zsumowaną wartość wszystkich wejść z poprzedniej warstwy sieci neuronowej i propaguje wartość wynikową do następnej warstwy.

### **Konwolucja**

Termin w kontekście uczenia maszynowego odwołuje się do operacji konwolucji lub warstwy konwolucyjnej. Konwolucja polega na stworzeniu macierzy będącej mieszanką macierzy wejściowej z filtrem konwolucyjnym dając zestaw wag mniejszy od oryginalnego wejścia i tworząc mapę cech.

### **Łączenie (ang. pooling)**

Technika polegająca na redukcji danych wejściowych z poprzedniej warstwy konwolucyjnej. W łączeniu zazwyczaj używa się maksymalną wartość lub średnią z obszaru puli, który z kolei jest macierzą kwadratową o zadanym rozmiarze (2x2, 3x3 lub 5x5). Aby zmniejszyć obraz należy ustalić rozmiar filtra oraz wielkość kroku. Następnie wykonuje się przejście po całym obrazie z nałożonym filtrem. Każde przesunięcie po sygnałach z warstwy poprzedniej zwraca nową wartość dla warstwy łączącej. Pozwala to ograniczyć ilość parametrów i wyodrębnić konkretne cechy obiektu.

### **Wsteczna propagacja błędów**

Najważniejszy algorytm dla znajdowania optymalnych wag w sieciach neuronowych. Przechodzi on w dwie strony sieci, pierw od warstwy wejściowej do wyjściowej wyliczane są wartości każdego neuronu. Następnie wykonywany jest powrót z ostatniej warstwy do pierwszej wyliczając pochodne cząstkowe błędów w stosunku do każdego parametru (wagi i progów).

### **Wyrzucanie połączeń (ang. dropout)**

Sposób na redukcję przeuczenia modelu względem danych treningowych przez regularyzację danych. Dokładniej polega na deaktywacji losowych neuronów w trakcie uczenia, zapobiega to sytuacjom kiedy dane treningowe są rozpoznawane zbyt dobrze, a nowe sygnały nieznane wcześniej mają dużo niższą jakość rozpoznania.

### **Epoka**

Pełne przejście przez zestaw danych treningowych, tak że każdy ele-

---

ment został przekazany do sieci jeden raz.

### **Model**

Reprezentacja obiektu utworzonego w wyniku uczenia sieci neuronowej. Z założenia model jest tym lepszy im dokładniej przewiduje klasę obiektu, którego wcześniej nie spotkał.





# Wstęp

Celem niniejszej pracy magisterskiej jest przedstawienie nowoczesnych architektur sieci neuronowych oraz przegląd najpopularniejszych bibliotek implementujących algorytmy uczenia maszynowego, a w szczególności zoptymalizowanych do pracy z uczeniem głębokim na wielu procesorach graficznych.

Głębokie sieci neuronowe, zawierające więcej niż jedną warstwę ukrytą, zostały wynalezione we wczesnych latach 60tych XX wieku. Algorytmy uczenia maszynowego były bardzo aktywnie rozwijane w latach 1950 - 1971, fundusze na badania w czasach zimnej wojny były w USA kilkakrotnie wyższe w udziale PKB. W czasie kryzysu z roku 1981, nadzieje wojska amerykańskiego na sztuczną inteligencję wygasły powodując epokę znaną jako "drugą zimę SI" (ang. second AI winter). Brak funduszy na badania znacząco zahamował rozwój systemów samouczących i do 2012 roku była to dziedzina zarezerwowana głównie dla doktorów matematyki i informatyki (głównie statystyków).

Powrót sieci neuronowych na pierwsze strony gazet popularnonaukowych spowodował Geoffrey Hinton. Wykorzystał swoje idee sprzed lat na nowym, bardzo wydajnym sprzęcie. Mowa o dwóch kartach graficznych GeForce GTX 580 3GB[9], które przez tydzień trenowały model do rozpoznania obrazów pod zawody ImageNet Competition 2012. Wcześniej zawody w rozpoznawaniu obiektów na obrazach wygrywało oprogramowanie łączące wiele algorytmów uczenia maszynowego, ale głębokie sieci neuronowe nie osiągały tak dobrych wyników, stąd nagły zwrot w kierunku narzędzia dającego tym lepsze rezultaty im więcej danych treningowych zostanie. Pierwszy raz można było zobaczyć głębokie sieci neuronowe mające trafność powyżej 80%. Po publikacji dokumentu *ImageNet Classification with Deep Convolutional Neural Networks* nastąpił gwałtowny rozwój start-upów związanych z Deep Learningiem, a firmy mające w nazwie słowa 'AI' lub 'Deep' zanotowały znaczne zwiększenie funduszy na rozwój.

Moc obliczeniowa kart graficznych i dedykowanych układów tanieje z

każdym rokiem. A zapotrzebowanie na moc obliczeniową dało przyspieszenie usługom chmurowym sprzedającym moc obliczeniową na żądanie. Zajmują się tym już nie tylko największe korporacje, ale również małe startupy jak Paperspace, Crestle, NVIDIA GPU Cloud. Kluczowe dla rozwoju dziedziny okazały karty graficzne firmy NVIDIA. Firma chcąc być liderem w rewolucji Sztucznej Inteligencji, zmieniła swój wizerunek z dostawcy rozrywki dla graczy i grafików na lidera urządzeń i oprogramowania stanowiącego podstawę systemów uczących. Przez nagłą modę, wszystkie obecnie używane narzędzia mają nie więcej niż 5 lat, oczywiście nie powstały od zera, opierają się na dużo starszych bibliotekach dla obliczeń lub narzędziach wewnętrznych wielu firm. Akademicki ruch wolnego oprogramowania pozwolił na praktykowanie dziedziny głębokiego uczenia kompletnym laikom. Przykładem może być biblioteka Caffe, napisana przez doktoranta z uniwersytetu w Berkeley. Wielu entuzjastów pomagało przez lata tworzyć wydajne implementacje algorytmów w bibliotece, którą po latach Facebook zaczął używać jako podstawowego narzędzia do przetwarzania swoich zbiorów danych. Doktor, który ją rozwijał pracuje obecnie na stanowisku dyrektora działu sztucznej inteligencji i wyznacza dalszy rozwój następcy swojego dzieła Caffe2.

Dziedzina jest tak dynamiczna, że zdarza się iż literatura w czasie publikacji papierowej książki staje się nieaktualna. Dlatego wykorzystana w tej pracy literatura to głównie publikacje elektroniczne z serwisu arXiv, będące zazwyczaj jeszcze bez recenzji naukowej oraz dokumentacja narzędzi wykorzystywanych do badania wydajności.

Ilość mocy obliczeniowej potrzebnej do stworzenia dobrego modelu powoduje konieczność przedstawienia największych problemów osób praktykujących uczenie głębokich sieci neuronowych w taki sposób by uniknąć kosztownego dochodzenia do zadowalających wyników metodą prób i błędów. W rozdziale poświęconym wydajności architektur przedstawiono sposoby na uniknięcie najczęściej popełnianych błędów oraz zbiór powszechnie przyjętych praktyk dla osiągnięcia optymalnej wydajności przy zaledwie kilku próbach. Wydajność jest tu rozumiana zarówno jako czas obliczeń potrzebny do wyuczenia modelu rozpoznawania danego zagadnienia, oraz dokładność z jaką już nauczony model potrafi rozpoznać niespotkane wcześniej dane. Oba te parametry są od siebie zależne, ponieważ koszt poprawy jakości predykcji algorytmu o ostatnie 1-2% może wynosić setki tysięcy złotych, zakłada się oczekiwaną poprawność lub korzysta z gotowych modeli dodając tylko własne klasy obiektów. Na koszt składa się energia elektryczna, serwery wyposażone w karty graficzne oraz czas oczekiwania na zakończenie obliczeń.

---

Zmiany zachodzące w rozwoju sprzętu, techniki przenoszenia cech między modelami, tzw. transfer uczenia, pozwolą na tworzenie aplikacji w czasie rzeczywistym już za kilka do kilkunastu lat.

Architektury posiadają wiele hiperparametrów, które wpływają na precyzję modelu. Po modyfikacji wartości dowolnego należy rozpocząć pracę algorytmu od początku, dlatego dobór odpowiednich wartości na początku jest niezwykle cenny. Na rok 2017, żadna otwartoźródłowa biblioteka nie posiada wbudowanych automatów do poszukiwania optymalnych hiperparametrów. Dobieranie ich dzieje się na intuicję wyrobioną dziesiątkami prób. Istnieje kilka nowatorskich technik dobierania hiperparametrów, które ze względu na to że zostały odkryte przez mało znanych badaczy, nie są jeszcze powszechnie stosowane w dużych ośrodkach badawczych. Te techniki również są na tyle nowe, że nie ma wydawnictwa branżowego opisującego ich użycie dla praktyków. Jedyny sposób na chwilę obecną to śledzenie publikacji naukowych wydawanych w serwisie arXiv. Kilka technik strojenia stałej uczenia w tym cykliczna zmiana stałej uczenia pozwala efektywnie zmniejszyć błąd klasyfikatora o 10-20%, co wiąże się z dużo wyższą wydajnością, przy znikomym nakładzie pracy.

Do prezentacji działania bibliotek, architektur i algorytmów użyty został gotowy zbiór danych. Zestaw danych treningowych pochodzi z serwisu Kaggle. Są to zdjęcia z prześwietleń rentgenowskich podzielone na dwie kategorie:

- zdrowe płuca,
- zapalenie płuc.

Dane są anonimowe, więc zgodne z prawem europejskim i zostały zweryfikowane przez Kaggle. Po zakończeniu pracy z danymi znajduje się porównanie z najlepszymi wynikami społeczności serwisu.

Badanym obszarem jest zagadnienie klasyfikacji obrazów przy pomocy głębokich sieci neuronowych. Algorytm sieci tworzy mapę cech każdej z klas. Na przykładzie zbioru zdjęć, prezentowany jest sposób wykorzystania gotowych modeli i metody optymalizacji je pod wybraną domenę. Takie podejście pozwala uniknąć wysokich kosztów tworzenia własnych modeli od zera.

Głębokie sieci neuronowe również świetnie się sprawdzają przy zadaniach nie obejmujących zagadnień prezentowanych w niniejszej pracy. Do najczęstszych zastosowań należą systemy rozpoznawania mowy, przetwarzania języka naturalnego i systemy rekomendacji produktów. Są zagadnienia

zbyt obszerne by zostały tutaj przedstawione.

Praca została podzielona na etapy tworzenia sieci neuronowej od dołu jej struktury, po wyższe poziomy abstrakcji. Rozpoczyna od przedstawienia pojedynczych elementów i struktury matematycznej sieci. Następnie opisany jest każdy element tworzący gotową architekturę. Kolejnym elementem są już gotowe architektury i biblioteki implementujące różne rodzaje algorytmów, ze wspomaganiami na programowanie rozproszone na wiele urządzeń wyposażonych w jednostki graficzne.

Wszystkie przedstawione kody źródłowe zostały napisane w języku Python 3.6 (z wyjątkiem fragmentu porównawczego w JavaScript). Python staje się dominującym językiem w uczeniu maszynowym, kosztem najpopularniejszego na chwilę obecną języka R. Python składnią jest zbliżony do pseudokodu, dodatkowo wszystkie popularne biblioteki udostępniają interfejs programistyczny dla Pythona. Książki wykorzystane w tworzeniu tego tekstu, zawierają wyłącznie kod Pythonowy (z kilkoma nawiązaniem do języków R i C++).

# Rozdział 1.

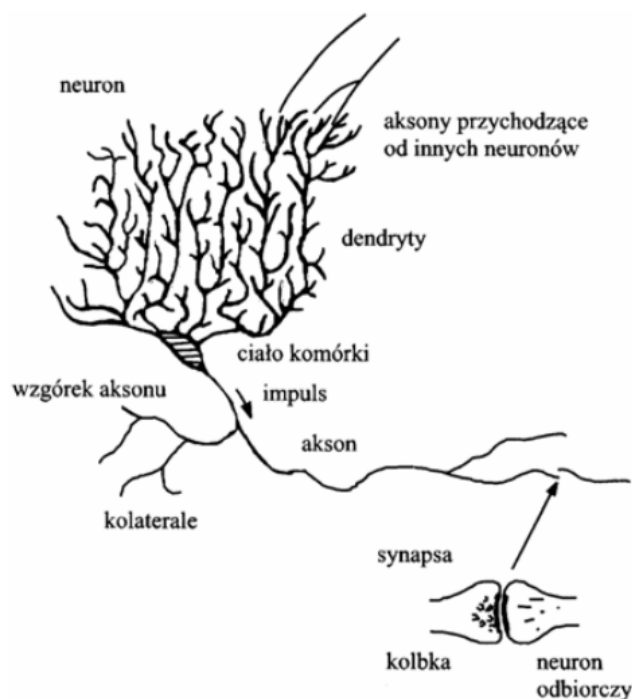
## Neurony

W niniejszym rozdziale została przedstawiona idea sztucznego neuronu McCullocha i Pittsa. Opracowany model biologicznego neuronu z 1943 roku jest tak uniwersalny, że do dziś stanowi podstawę budowy sieci neuronowych. Przedstawiono tu również w jaki sposób znaleźć mapę cech, czy mówiąc prościej jak nauczyć neurony rozpoznawać wzorce. Na końcu rozdziału opisane są możliwości jakie daje łączenie sztucznych neuronów w sztuczną sieć neuronową.

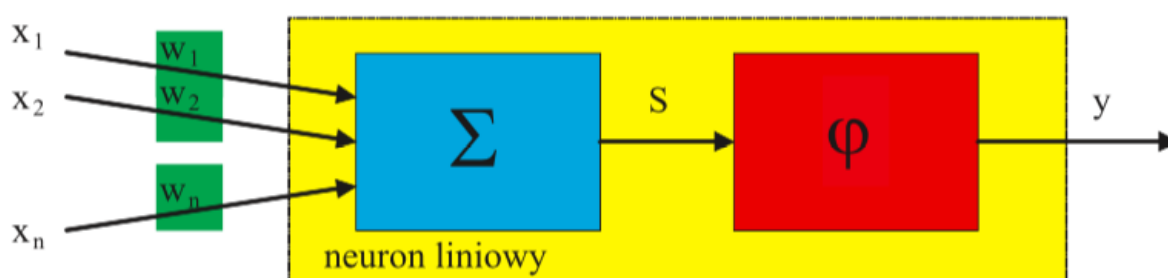
Pierwszy model sztucznego neuronu był wyobrażeniem uproszczonej biologicznej reprezentacji neuronu mózgowego z lat 40tych XX wieku. Model ten stanowił bardziej próbę implementacji idei, niż odwzorowanie prawdziwego neuronu, który jest wielokrotnie bardziej złożony, a jego zachowanie nie jest jeszcze w pełni znane. Zdefiniowany przez noblistę Ramon y Cajal'a w 1906 roku neuron jest specjalistycznym nośnikiem wszelkich informacji, a także elementem przetwarzającym wszystkie doznania i emocje, oraz sterownikiem całego ciała.

Uproszczeniem które pozwala na sprawną implementację, jest wyodrębnienie zasady przetwarzania informacji bioelektrycznej do kilku prostszych operacji. Takie uproszczenie składa się z wielu sygnałów wejściowych, wagi przypisanej każdemu z tych sygnałów, oraz pojedynczej wartości wyjściowej. Jak wielkie jest to uproszczenie widać przy porównaniu rysunku prawdziwego neuronu (również w uproszczeniu) oraz sztucznego neuronu. Taka budowa choć bardzo banalna, daje wiele możliwości. Podstawową zaletą jest prostota odwzorowania w urządzeniach elektronicznych. Przed nadejściem ery komputerów osobistych, stosowano maszyny nazywane perceptronami. Nazwa perceptron jest obecnie używana zamiennie z neuronem.

Model obliczeniowy został stworzony na podstawie algorytmu logiki prowej. Nazwa neuronu sztucznego została przyjęta w kręgach akademickich



Rysunek 1.1: Wizualizacja uproszczonego modelu neuronu mózgowego.



Rysunek 1.2: Wizualizacja prostego sztucznego neuronu.

i biznesowych, stąd często osoby nie znające budowy mózgu ani działania sieci neuronowych mylnie nazywają zagadnienie Sztuczną Inteligencją, upraszczając skomplikowane procesy biologiczne do wzmacniania lub osłabiania sygnałów przekazywanych przez synapsy. Sieci neuronowe naśladują tylko jeden rodzaj pamięci, pamięć deklaratywną i w odosobnieniu nie tworzy inteligencji.

## 1.1. Model perceptronu prostego

Podstawowym elementem z których buduje się sieci neuronowe są sztuczne neurony, nazywane również perceptronami. Mają one być w rzeczywistości

bardzo uproszczonym odwzorowaniem komórek nerwowych występujących w mózgu. Takie uproszczenia pozwalają na łatwą implementację modelu matematycznego, który ma reprezentować nasz obiekt i być tani w replikacji. Nawet po takim uproszczeniu jest on w stanie skutecznie naśladować "uczenie". Sztuczny neuron jest funkcją matematyczną

$$f(x, w) \rightarrow y$$

. Można go opisać za pomocą modelu złożonego z:

- określonej liczby wejść  $n \in \mathbb{N}$ ,
- wagi, skojarzonej z każdym z wejść  $w_i \in \mathbb{R}, i = 1..n$ ,
- wybranej funkcji aktywacji  $\phi : \mathbb{R} \rightarrow \mathbb{R}$

Charakterystycznym elementem budulca sieci jest wiele wejść i tylko jedno wyjście, dlatego tak łatwo stworzyć model będący funkcją matematyczną. Dane wejściowe oraz wyjściowe mogą przyjmować wartości z ograniczonego przedziału. Wartości przekazywane na wejściu i wartość wyjściowa zazwyczaj przyjmują znormalizowane wartości z przedziału  $x \in [-1, 1]$  dla każdego z wejść, oraz  $y \in [-1, 1]$  dla wyjścia. W uproszczeniu można przyjąć  $y = \sum_{i=1}^n (w_i * x_i)$ .  $w_i$  są nazywane wagami (dawniej wagami synaptycznymi) i podlegają zmianom w trakcie uczenia neuronu. Wagi stanowią zasadniczą cechę sieci neuronowych działających jako adaptacyjne systemy przetwarzania informacji. Zsumowana wartość jest wejściem dla funkcji aktywacji neuronu. Funkcja aktywacji zwyczajowo ma kształt sygmoidy, ale stosowane są obecnie również funkcje nazywane rektyfikownymi jednostkami liniowymi. Funkcja progowa ma za zadanie symulację zachowania przekaźnika synapsy, po przekroczeniu określonego progu aktywuje się dane zachowanie lub jak na przykładzie obrazów rozpoznanie cechy.

Ta prosta jednostka stanowi dziś podstawę budowy każdej sieci neuronowej. Aby funkcja zwracała oczekiwane wyniki, wagi powinny być poprawnie ustawione. Początkowo wagi ustawiano ręcznie za pomocą operatora (osoby przełączającej fizyczne kable), który wcześniej przeliczał je dla odpowiednich parametrów wejścia

wyjścia. W latach 50tych perceptron stał się pierwszym modelem umiającym samodzielnie wyliczyć poprawnie wagi definiujące zadaną klasę na podstawie przykładów. Wagi w zależności od wartości mogą sygnał wejściowy wzmocnić gdy waga jest większa od 1, lub stłumić gdy waga jest

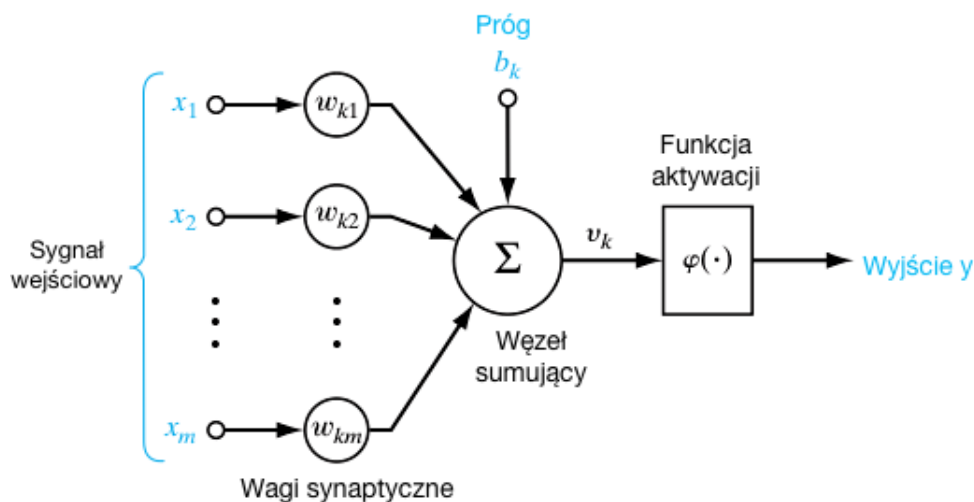
mniej niż 1. To pozwala wyuczonemu już perceptronowi na porównanie cechy obiektu wejściowego z tym co potrafi rozpoznać.

W jaki sposób sztuczny neuron jest w stanie rozpoznać sygnał wejściowy? Do wyjaśnienia zjawiska w literaturze zazwyczaj prezentuje się przedstawienie modelu w notacji wektorowej.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \quad \text{- wektor wyznaczający przestrzeń wejść, oraz}$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} \quad \text{- wektor wyznaczający przestrzeń wag.}$$

W ten notacji można wyrazić wyjście neuronu jako  $y = W^T * X$ . Wartość wyjściowa neuronu  $y$ , będzie wyższa im bliższe będzie położenie wektorów.



Rysunek 1.3: Prosta reprezentacja neuronu.

## 1.2. Co to jest sieć neuronowa?

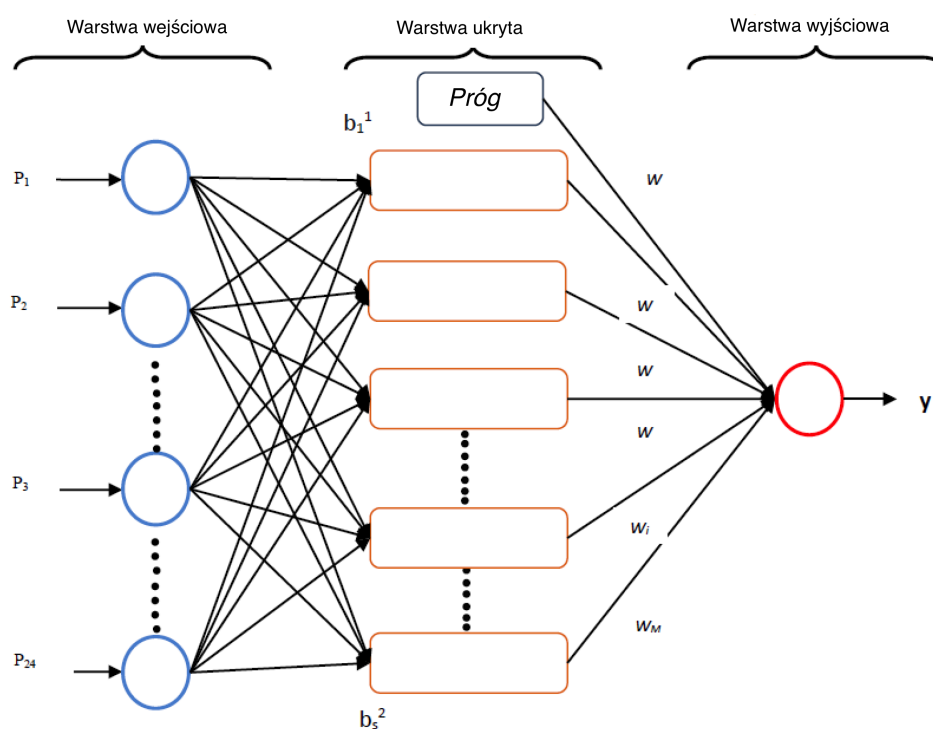
Koncepcja sztucznej sieci neuronowej to połączenie wielu neuronów w jeden obiekt. Pozwala to na rozpoznawanie bardziej skomplikowanych wzorców i



większej różnorodności typów obiektów niż klasyfikacja binarna. Zwykła sieć neuronowa składa się z trzech warstw węzłów. Warstwa pierwsza stanowi warstwę wejściową sieci i składa się z sygnału wejściowego.

Kolejna warstwa jest nazywana warstwą ukrytą, użytkownik nie ma dostępu do niej, przypomina czarną skrzynkę. Nie można zaobserwować co się wewnątrz niej dzieje, użytkownik widzi jakie są dane wejściowe, wartości które powinny zostać zwrócone (w ostatniej warstwie), ale nie ma informacji co jest i powinno się znajdować w tej warstwie podczas trenowania sieci. Ostatnia warstwa, pojedynczy węzeł nazywany jest warstwą wyjściową. Pobiera wektor wartości wyjściowych z poprzedniej warstwy (ukrytej) i na nim zostaje obliczona wartość wyjściowa (odpowiedź).

Taka architektura nazywa się siecią dwuwarstwową. Dane wejściowe, mimo że tworzą pierwszą warstwę nie są liczone w nazewnictwie. Z warstwami ukrytą i wyjściową powiązane są parametry  $w$  (ang. weight) oraz  $b$  (ang. bias), oznaczające kolejno: macierz wag oraz wektor progów.



Rysunek 1.4: Prosta sieć neuronowa złożona z dwóch warstw.

### 1.3. Jak uczyć sieci neuronowe

W obecnym momencie istnieją dwie możliwości by sieć posiadała umiejętność poprawnej klasyfikacji. Można wyznaczyć dla algorytmu zbiór opisanych sy-

gnałów wejściowych wraz z oczekiwanym sygnałem wyjściowym, ten typ tworzenia modelu sieci jest uczeniem nadzorowanym, co jest szczegółowo omawiane w tej pracy. Drugim podejściem jest podanie sygnałów wejściowych bez opisów i oczekiwanych wartości wyjściowych, to podejście nazywa się uczeniem nienadzorowanym.

## Uczenie nadzorowane

Jest to typ uczenia maszynowego, które zakłada obecność ludzkiego nauczyciela. Nauczyciel zobowiązany jest stworzyć odpowiednie dane uczące. Takie dane są parą danych, wejściowego obiektu uczącego oraz prawidłową odpowiedź wyjściową do tej danej. System na podstawie tych danych ma nauczyć się przewidywać poprawną odpowiedź dla nowych danych, ze znanej mu domeny.

Zadania uczenie nadzorowanego dzielą się na dwie kategorie, regresję i klasyfikację.

W problemie regresji próbuje się przewidzieć wyniki, które są wartościami ciągłymi, czyli mając jakieś dane próbuje się je mapować na funkcję ciągłą.

W problemie klasyfikacji algorytm ma za zadanie przewidzieć wyniki będące wartościami dyskretnymi, uściślając jest to znajdowanie klas obiektów, na podstawie danych wejściowych.

Uczenie nadzorowane ma wiele zastosowań, do głównych należą (wraz z używanym typem sieci):

- przewidywanie cen nieruchomości (zwykła sieć neuronowa),
- reklamy internetowe (zwykła sieć neuronowa),
- rozpoznawanie mowy (rekurencyjna sieć neuronowa),
- tłumaczenie maszynowe w translatorach (rekurencyjna sieć neuronowa),
- samochody autonomiczne (sieci hybrydowe lub inne niestandardowe sieci),
- rozpoznawanie obiektów na obrazach (konwolucyjna sieć neuronowa).

Ostatnie z wyżej wymienionych zastosowań jest tematem badanym w tej pracy.

Uczenie nadzorowane dzieli się również binarnie ze względu na strukturę dostarczonych danych. Pierwszy rodzaj danych, dane zawierające konkretną strukturę, zazwyczaj tabela, są bardzo dobrze obsługiwane przez większość znanych wcześniej algorytmów uczenia maszynowego i nie wymagają ogromnej mocy obliczeniowej. Choć sieci neuronowe świetnie się sprawdzają przy tego typu danych, używanie ich ma sens dopiero gdy danych jest bardzo dużo (wielkość tabel przekraczające miliony rekordów). Drugi typ danych, nieustrukturyzowane zdjęcia, pliki audio, tekst, jest znacznie trudniejszy do rozpoznania przez komputer, za to dużo bardziej naturalny dla ludzi. Dzięki głębokim sieciom neuronowym i wielkiej mocy obliczeniowej komputerów, dokładność algorytmów uczących się na tego typu danych znacząco wzrosła, z 70% do ok 95-99% dokładności zależnie od ilości danych.

## Uczenie nienadzorowane

Uczenie maszynowe bez nadzoru nauczyciela jest drugim typem uczenia maszynowego polegającym na wyciąganiu wniosków z danych bez informacji zwrotnej o poprawności wniosków oraz bez określania etykiety danych. W algorytmach tego typu nie ma funkcji wyznaczającej poprawność utworzonego modelu. Najczęstszym zadaniem stawianym algorytmom uczenia nienadzorowanego jest wyznaczanie klastrów ze zbioru danych. W tej pracy temat uczenia bez nauczyciela nie jest poruszany.

## 1.4. Elementy sieci neuronowych

Sieci neuronowe do poprawnego działania potrzebują więcej niż tylko połączenia ze sobą neuronów. Należy dobrać odpowiednie algorytmy korygujące błędy. Aby algorytm zatrzymał się w odpowiednim momencie z osobna trzeba zdefiniować warunki stopu dla epoki, ilość epok, rozmiar danych. Każda warstwa musi posiadać określoną funkcję aktywacji. Ostatnim opisanym elementem sieci są hiperparametry, czyli właściwości sterujące algorytmem by wiedział w jaki sposób wyznaczać pozostałe parametry.

### 1.4.1. Metoda gradientu prostego

Algorytm stosowany do szukania minimum lokalnego płaszczyzny wyznaczonej funkcją. Jest to bardzo prosta metoda optymalizacji stosowana do wyznaczania wag i progów. Płaszczyzna złożona z osi poziomych  $w$  i  $b$  oraz

oś pionowa będącą wartością funkcji kosztu  $J(w, b)$ . Algorytm mając początkowo losowe wagi i progi, rozpoczyna w tym losowym miejscu sprawdza wartość funkcji kosztu, następnie wykonuje małe przesunięcie w najbardziej stromym kierunku w dół płaszczyzny. Operacja zejścia powtarzana jest do zejścia do minimum lokalnego.

```
i = 0;
while i < x do
    w = w -  $\alpha * \frac{\partial J(w,b)}{\partial w}$ ;
    b = b -  $\alpha * \frac{\partial J(w,b)}{\partial b}$ ;
    i ++;
end
```

**Algorithm 1:** Schemat algorytmu

### 1.4.2. Funkcje aktywacji

Budując sieć neuronową należy funkcje aktywacji dla warstw ukrytych oraz warstwy wyjściowej. Funkcje aktywacji są znane również jako charakterystyka neuronu. Charakterystyka neuronu jest elementem, który pośredniczy między zsumowanym pobudzeniem neuronu, a jego danymi wejściowymi.

Najczęściej używaną charakterystyką była do niedawna funkcja sigmoidalna ponieważ ma wiele zalet:

- zapewnia łagodny gradient między wartościami 0 a 1,
- ma gładką i prostą do liczenia pochodną,
- jednym parametrem można dobrać kształt krzywej.

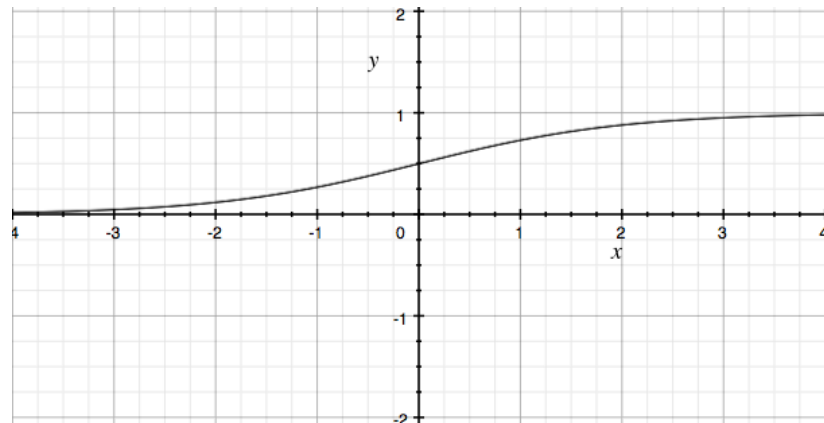
Definicja funkcji sigmoidalnej  $f(x) = \frac{1}{1+e^{-x}}$ . Obecnie nie jest zalecane używanie funkcji sigmoidalnej poza przypadkami na warstwie wyjściowej w przypadku klasyfikacji binarnej, kiedy należy stwierdzić prawdopodobieństwo klasy obiektu. Dużo wydajniejszym odpowiednikiem jest następna funkcja.

Sprawną funkcją aktywacji jest tangens hiperboliczny. Jego wartości zawierają się między -1 i 1. Definicja funkcji tangensu hiperbolicznego  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ . Jak widać na załączonych rycinach ta funkcja jest przesunięciem sigmoidy względem osi y, tak że przechodzi przez punkt (0, 0). Dla warstw ukrytych charakterystyka neuronu będąca tangensem hiperbolicznym ma lepszą skuteczność. Mediana wartości wychodzących z warstw ukrytych wynosi 0. Nie trzeba wtedy stosować skalowania względem sigmoidy, co znacząco ułatwia uczenie.

Zalecaną charakterystyką neuronu w dużych architekturach jest rektyfikowana jednostka liniowa. Definicja funkcji jest w porównaniu do pozostałych jest banalna  $relu(z) = \max(0, z)$ . Jedynym problemem przy używaniu ReLU jest wyliczenie pochodnej kiedy argument  $z$  jest mniejszy od 0. W praktyce nie sprawia to większych problemów bo wartość zostanie ustawiona na 0, ale opracowano funkcję która jest pozbawiona tej niedogodności.

Nieszczelna liniowa jednostka rektyfikowana (ang. leaky Rectified Linear Unit) definiowana  $f(x) = \max(x, 0.01x)$ . Pozwala to zmniejszyć efekt kiedy pochyła funkcji schodzi do 0, spowalniając uczenie. Niestety ta funkcja nie jest powszechnie używana w praktyce.

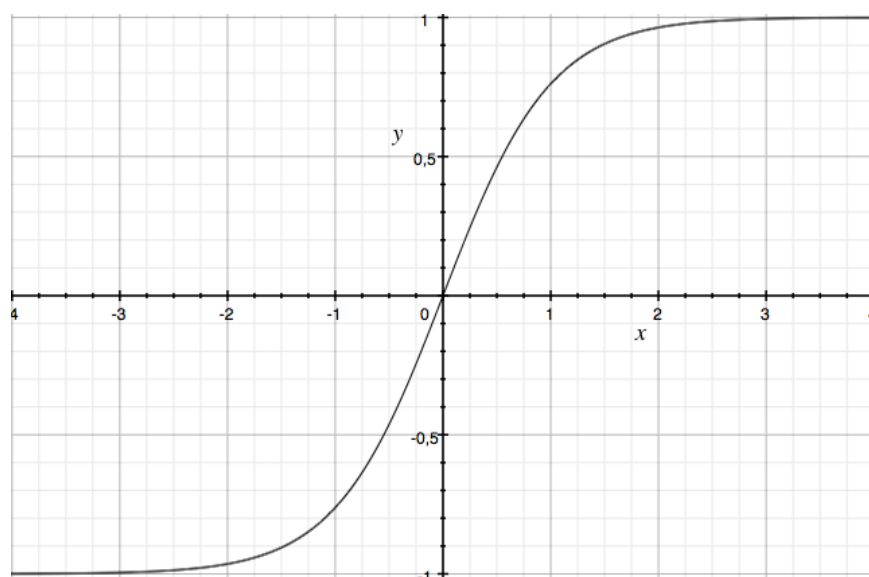
Należy podkreślić że w nowoczesnych sieciach neuronowych każda warstwa może posiadać indywidualnie ustanowioną funkcję aktywacji w zależności od architektury. Domyślnym wyborem funkcji aktywacji jest ReLU.



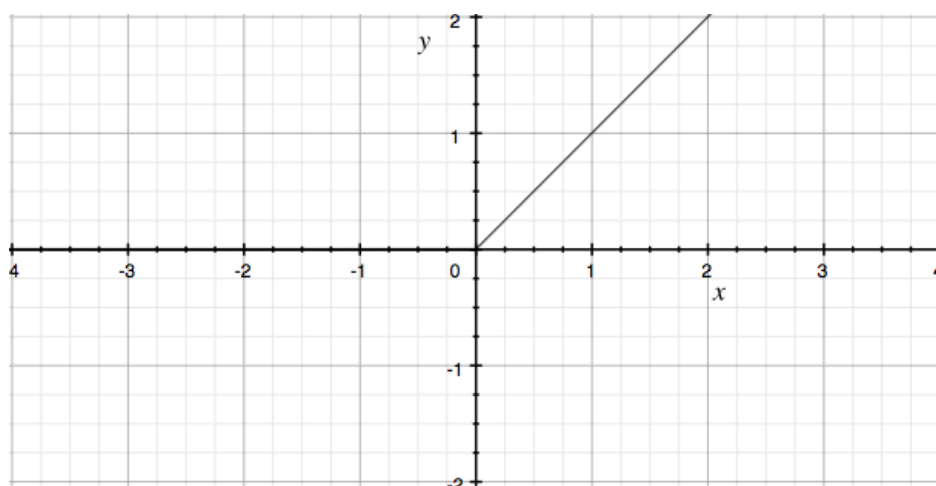
Rysunek 1.5: Wykres funkcji sigmoidalnej

### 1.4.3. Parametry

Najważniejszy element sieci neuronowej. Składa się na niego zbiór wag i progów (bias). Wagi wyznaczają sposób działania sieci, która ma nauczyć się dobrać odpowiednie parametry przy pomocy algorytmu uczącego, odpowiednio dobranych hiperparametrów oraz zbioru danych uczących. Wagi muszą być dobrane w sposób umożliwiający neuronom wykonanie czynności, których się od nich wymaga. Ze względu na ilość wag w głębokich sieciach, gdzie dla każdego z tysięcy neuronów może istnieć kilkaset wejść, proces musi być automatyczny. Sieć musi wiedzieć kiedy każdy z neuronów zwiększa swoją dokładność czy nie. Wagi i progi powinny być elastyczne i zmieniać się bardzo szybko na początku procesu uczenia i tylko w niewielkim stopniu zmieniać wartości kiedy algorytm kończy proces uczenia.



Rysunek 1.6: Wykres funkcji tangensu hiperbolicznego

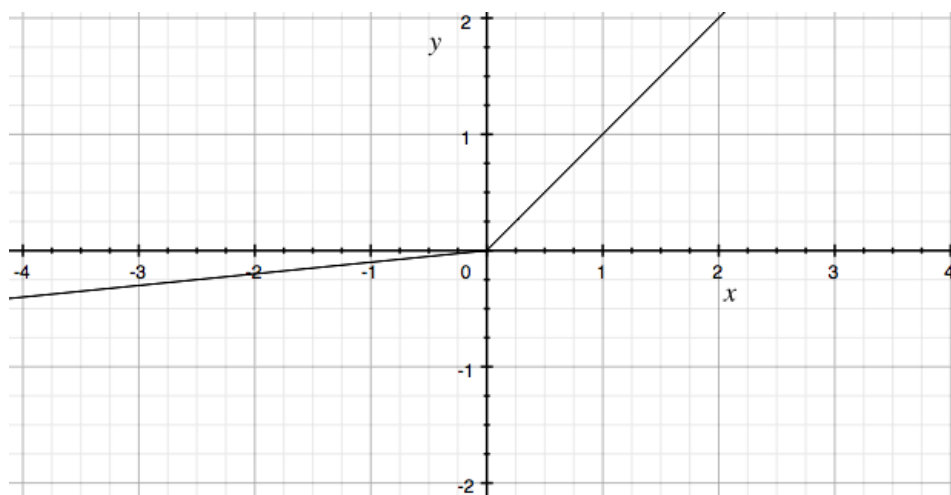


Rysunek 1.7: Wykres rektyfikowanej jednostki liniowej

#### 1.4.4. Hiperparametry

Tworzenie dobrej sieci neuronowej wymaga nie tylko parametrów ale również dobrego doboru hiperparametrów. Elementy "ręcznie" dobierane dla algorytmu uczenia. Najczęściej ustawiane parametry dla sieci neuronowej to:

- stała uczenia,
- ilość iteracji uczenia w jednej epoce,
- ilość ukrytych warstw,
- ilość ukrytych jednostek,



Rysunek 1.8: Wykres niezszerzonej rektyfikowanej jednostki liniowej

- wybór funkcji aktywacji (ReLU, tangens, sigmoida),
- parametry regularyzacji,
- rozmiary próbek uczących,

i wiele innych mniej ważnych hiperparametrów

Są to parametry kontrolujące sposób wykonania algorytmu uczącego i w ostateczności to one mają największy wpływ na zwykłe parametry wagi i progi (bias). Nie został ustalony jednolity sposób doboru hiperparametrów. Zwyczajowo stosuje się metodę prób i błędów na podstawie tego jak algorytm się zachowuje.[6] Za każdą zmianą wartości należy uruchomić algorytm i sprawdzić jak zachowuje się funkcja kosztu, jeśli maleje lepiej niż przy poprzednich wartościach można stroić dalej w tym kierunku. Dopieranie odpowiednich wartości niestety jest procesem empirycznym i należy wyrobić odpowiednią intuicję aby jak najtrafniej dobierać parametry od początku. Wartości te też nie są stałe, zmieniają się w zależności od dziedziny badanego zagadnienia. Być może najbliższe lata przyniosą dobry i spójny przewodnik dobierania najlepszych wartości, pozwoli to znacznie zautomatyzować proces.

#### 1.4.5. Wykres obliczeniowy

Aby opisać organizację obliczeń podsekcję o wstecznej propagacji błędów, należy zformalizować operacje obliczeń. Robi się to za pomocą wykresu obliczeniowego. Węzły w grafie reprezentują skalary, wektory, macierze i tensory. Drugim elementem w grafie są operacje czyli proste funkcje z jedną lub

większą ilością zmiennych. Graf obliczeniowy przydaje się kiedy istnieje zdefiniowana funkcja bądź zmienna, którą należy zoptymalizować, w przypadku sieci neuronowych oczywistą jest funkcja kosztu.

Dla przykładu chcę obliczyć funkcję kosztu  $J$  o zmiennych  $a, b, c$  zdefiniowaną  $J(a, b, c) = 3(a + bc)$ . Obliczenie tej funkcji składa się z trzech kroków, pierwszym jest policzenie  $b * c$ .

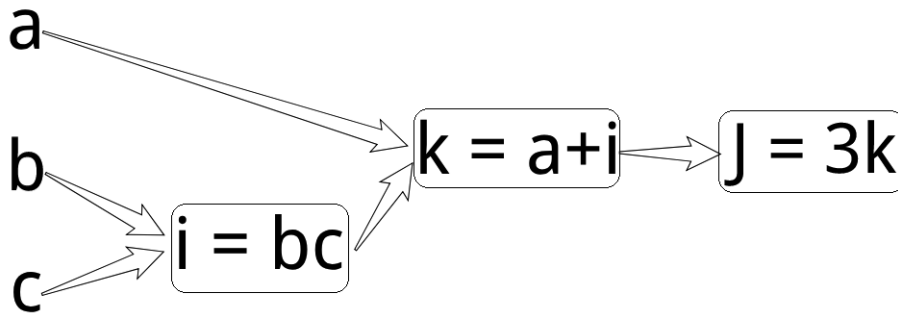
$$i = b * c$$

Następnym krokiem jest dodanie zmiennej  $a$ .

$$k = a + i$$

Trzecim i ostatnim krokiem jest wyjście funkcji kosztu  $J = 3k$ . Graf obliczeniowy powyższej funkcji znajduje się na rysunku poniżej.

Do minimalizowania funkcji kosztu algorytm oblicza pochodne w odwrotnej kolejności do przedstawionej na poprzednim przykładzie grafu.



Rysunek 1.9: Wykres obliczeniowy dla funkcji  $J(a, b, c) = 3(a + bc)$

#### 1.4.6. Wsteczna propagacja błędów

Algorytm wstecznej propagacji błędów został opracowany w latach 70tych, jednak używany dopiero ponad 10 lat później kiedy Rumelhart, Hinton i Williams wykazali skuteczność metody w strojeniu wag w ukrytych warstwach sieci.

Przekazując sygnał na wejście sieci zostaje on propagowany w przód przez wszystkie warstwy do warstwy wyjściowej. Losowy neuron w warstwie ukrytej otrzymuje wartości wyjściowe z neuronów poprzedniej warstwy, mnoży każdy otrzymany sygnał z wagą przypisaną danemu wejściu i przekazuje



zsumowany wynik do funkcji aktywacji. Każdy neuron wykonuje takie operacje do czasu trafienia na ostatnią warstwę gdzie sieć podaje już oczekiwany wynik. Cały ten proces nazywa się propagacją w przód. Mając oczekiwany wynik na warstwie wyjściowej, zazwyczaj prawdopodobieństwo przynależności do każdej z uczonych klas należy wyliczyć funkcję błędu (ang. loss function). Funkcja ta pozwala określać odchylenie wyliczonego wyniku od właściwej odpowiedzi. Celem sieci jest zminimalizowanie błędu przez wyliczenie pochodnej z funkcji kosztu w stosunku do wag modelu.

Wsteczna propagacja błędu jest narzędziem używanym przez algorytm gradientu spadkowego aby wyliczyć gradient funkcji błędu. W trakcie przejścia wstecz przez sieć neuronową następuje korekcja wag. Przechodząc od wartości wyjściowej należy zmodyfikować wagi warstwy poprzedzającej tak by zmniejszyć wartość funkcji błędu.

Istnieją również analogiczne metody korekcji wag, jak szybka propagacja (ang. quick-propagation) oraz bardziej skomplikowane i ograniczone założeniami matematycznymi metoda gradientów sprzężonych i metoda Levenberga-Marquardta. Metody te są dużo szybsze od wstecznej propagacji błędu, ale przez swoje ograniczenia prawie nigdy nie są używane w praktyce. [16]



## Rozdział 2.

# Deep Learning

### 2.1. Techniki tworzenia głębokich sieci

W tym rozdział zostają przybliżone trzy interesujące techniki głębokiego uczenia. Zostały wybrane ze względu na ich obszerną ilość zastosowań oraz potencjał, który można wykorzystać przy rozpoznawaniu wzorców na obrazach. Następnie opiszę kilka bibliotek implementujących wsparcie dla większości architektur sieci przedstawionych w pracy. Biblioteki są wybrane na podstawie przekroju platform i języków mi znanych.

#### 2.1.1. Sieci neuronowe

Sztuczne sieci neuronowe zostały opisane w poprzednim rozdziale, tutaj uwaga skupia się na podziale na różne architektury, głębokości oraz rodzaj dodatkowych warstw, takich jak filtr cech.

#### 2.1.2. Convolutional Neural Networks

Sieci konwolucyjne stosowane głównie do sygnałów wejściowych będących obrazem (zbiorem pikseli).

#### 2.1.3. Recurrent Neural Networks

#### 2.1.4. Fully Convolutional Networks



## Rozdział 3.

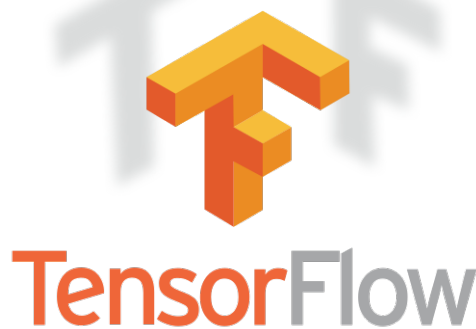
# Biblioteki implementujące uczenie głębokich sieci neuronowych

Nagły wzrost zainteresowania sieciami neuronowymi, spowodował pojawienie się wielu bibliotek implementujących Deep Learning i przy okazji wiele pobocznych zadań Machine Learningu. Obecnie wiele z największych firm technologicznych próbuje wypromować własny stos technologiczny oparty o licencję Open Source. Otwartość nie wynika z dobroduszości i chęci dzielenia się ze społecznością, a Wspólnym mianownikiem wszystkich narzędzi jest kompletne API dla użytkowników języka Python. Jest to język w którym najszybciej można znaleźć literaturę wprowadzającą w zagadnienie.

Na uwagę zasługuje fakt pojawienie się kilku bibliotek napisanych wyłącznie dla JavaScript'u. Ich zastosowanie jest nastawione na użytkowników gotowych modeli i ogranicza się głównie do importu zminiaturyzowanych wag na telefon by szybko odpowiadać na sygnały użytkownika takie jak zdjęcie z aparatu, bądź komendę głosową.

### 3.0.1. Tensorflow

Tensorflow jest pierwszą otwartoźródłową biblioteką dla obliczeń numerycznych używając grafów przepływu danych. Umożliwia praktykom uczenia maszynowego wykonywać intensywne obliczenia na danych przez wydajną implementację powszechnie używanych algorytmów głębokiego uczenia. Węzły w grafie przepływu są reprezentowane jako operacje matematyczne, zaś wierzchołki przedstawiają wielowymiarowe macierze (tensory) zapewniające komunikacje między wierzchołkami i węzłami. Architektura tworzonych



Rysunek 3.1: Logo biblioteki Tensorflow

sieci w Tensorflow jest łatwa w modyfikacji, dlatego jeden model można wykorzystać do pracy z procesorami CPU, GPU, urządzeniami mobilnymi i chmurą mieszaną. [3]

Pierwotnie Tensorflow umożliwiał pracę tylko z językiem Python. Od wersji 1.0 dodano eksperymentalne API do Javy i Go. Ta wersja zadebiutowała z innowacyjnym narzędziem Tensorflow Debugger. Jest on narzędziem tekstowym do debugowania działających aplikacji.

Należy zaznaczyć że biblioteka jest wieloplatformowa i działa na wszystkich popularnych systemach operacyjnych:

- Windows
- Linux
- Mac
- Android
- iOS
- Raspberry PI

Zestaw narzędzie zawiera Tensorflow Board graficzne oprogramowanie do analizy utworzonych modeli. Łatwość użycia i przyjęcie modelu Open Source sprawia że jest to najczęściej używana biblioteka podczas prac naukowych, o czym świadczy ilość cytowań[1]. Wiele firm wspiera projekt używając kodu i dodając do niego własne rozszerzenia, warto wymienić choćby kilku większych kontrybutorów:

- Google,
- ARM,

- Twitter,
- Ebay,
- Intel,
- Qualcomm

Główne zalety:

- Największe zasoby dokumentacji i tutoriali,
- Oglądanie w czasie rzeczywistym procesu uczenia przez Tensorboard,
- Największa społeczność skupiona wokół projektu spośród wszystkich bibliotek,
- Wspiera uczenie rozproszone na wielu maszynach,
- Odmiany Tensorflow Lite i Tensorflow JS umożliwiają zastosowanie tych samych modeli na urządzeniach mobilnych i przeglądarkach

Tensorflow ma dwie wady, działające na niekorzyść osób zaczynających pracę z sieciami neuronowymi:

- Wymaga sporej ilości kodu do utworzenia działającego modelu, przez co uważany jest za bibliotekę niskopoziomową w porównaniu do pozostałych,
- Kompletne wsparcie istnieje tylko dla języka Python,
- Jest dużo wolniejszy od CNTK i MXNet [13]

Dokumentacja: [https://www.tensorflow.org/api\\_docs/python/](https://www.tensorflow.org/api_docs/python/)

Kod źródłowy: <https://github.com/tensorflow>

Licencja: Apache 2.0

---

### 3.0.2. Keras

```
from keras.models import Sequential
from keras.layers import Dense
model = Sequential()
model.add(Dense(units=64, activation='relu',
                input_dim=100))
model.add(Dense(units=10, activation='softmax'))
model.compile(loss='categorical_crossentropy',
              optimizer='sgd',
              metrics=['accuracy'])
model.compile(loss=keras.losses.categorical_crossentropy,
              optimizer=keras.optimizers.SGD(lr=0.01,
              momentum=0.9, nesterov=True))
model.fit(x_train, y_train, epochs=5, batch_size=32)
model.train_on_batch(x_batch, y_batch)
loss_and_metrics = model.evaluate(x_test, y_test,
                                  batch_size=128)
classes = model.predict(x_test, batch_size=128)
```

Listing 3.1: Skrypt najprostszego modelu sekwencyjnego (Keras w 30 sekund)



Rysunek 3.2: Logo biblioteki Keras

Keras jest frameworkiem udostępniającym wysokopoziomowe API dla głębokich sieci neuronowych. Jego źródła zostały otwarte w 2015 roku. Społeczność skupiona wokół projektu wyróżnia się bardzo pozytywnym podejściem dla osób z poza środowiska naukowego i jest otwarta na propozycje rozwoju biblioteki. Nazwa Keras pochodzi od greckiego słowa oznaczającego róg, geneza nazwy wywodzi się z Odyseji.

Założeniem François Chollet’a, czyli autora jest zwiększenie szybkości przeprowadzania eksperymentów, dlatego Keras ma wysokopoziomowe API udostępniające uruchomienie algorytmów głębokiego uczenia w kilku liniach



kodu. Zalety wymieniane przez autora:

- Wspieranie łatwego i szybkiego prototypowania (przez modularność, przyjazny interfejs, możliwości rozszerzenia),
- Wspiera konwolucyjne sieci neuronowe, rekurencyjne sieci neuronowe i umożliwia mieszanie obu,
- Działa na CPU i GPU.

Działanie opiera się na wykorzystaniu rdzenia Tensorflow, CNTK lub Theano jako bibliotek wykonujących obliczenia. Potwierdza to idee stworzenia narzędzia dla szybkiego prototypowania modeli.

Do zalet można zaliczyć:

- Najszybsze prototypowanie ze wszystkich rozwiązań na rynku,
- Uproszczony interfejs zachęca do eksperymentów osoby nie związane z Data Science,
- Wbudowane wsparcie do pracy na GPU,
- Wspiera strumieniowanie danych przez Spark,
- Działa nie tylko na kartach NVIDIA, ale też Google TPU, AMD

Przy dłuższej pracy i bardziej skomplikowanych modelach wysokopoziomowość ogranicza elastyczność w działaniu. Nie ma tak wielu możliwości kontroli sieci jak pozostałe biblioteki, dlatego głównie poleca się ją dla nowicjuszy i eksperymentów przy standardowych typach sieci (konwolucyjnych, rekurencyjnych).

Dokumentacja: <https://keras.io/>

Kod źródłowy: <https://github.com/keras-team/keras>

Licencja: MIT

### 3.0.3. PyTorch

Oprogramowanie wywodzące się ze starej biblioteki Torch. Otwartoźródłowa paczka uczenia maszynowego dla języka Python. Główna biblioteka używana w Facebook'u do przetwarzania języka naturalnego.

PyTorch dostarcza dwie wysokopoziomowe funkcjonalności:

- Obliczenia na tensorach z akceleracją obliczeń na GPU,

- 
- Głębokie sieci neuronowe zbudowane na automatycznie różniczkowalnym systemie taśmowym

Główne zalety:

- Proces modelowania jest prosty i przejrzysty dzięki architekturze opartej na programowaniu imperatywnym,
- Dynamiczne generowanie grafu sieci,
- Deklaratywna równoległość danych, czyli dzielenie sygnałów wejściowych na wiele mniejszych,
- Najbardziej "Pythoniczna" składnia,
- Zawiera sporą ilość pretrenowanych modeli i architektur, pozwalając na kombinację ich w nowe struktury,
- Graf obliczeniowy jest generowany w trakcie działania programu, co pozwala na używanie dowolnego debuggera i nawet drukowanie obecnego stanu obliczeń na ekran, jest to spory potencjał na narzędzia interaktywne
- Służy także jako alternatywa dla numpy z możliwością obliczeń na GPU,
- Specjalnie napisane zarządzanie pamięcią GPU, pozwalające na niesamowicie wydajne zarządzanie pamięcią w porównaniu do konkurencji

PyTorch jest jednym z najmłodszych narzędzi, dlatego ma kilka poważnych wad:

- Nie jest zalecany do zastosowań produkcyjnych. Obecna wersja jest numerowana 0.4, wersja 1.0 była zapowiedziana na lato 2018. Na wrzesień 2018 brak informacji kiedy nastąpi premiera,
- Brak gotowych narzędzi do monitoringu i wizualizacji procesu uczenia i podglądu grafu sieci

Język Pyro do programowania probabilistycznego stworzony przez firmę Uber opiera się na PyTorch.

Dokumentacja: <https://pytorch.org/docs/stable/index.html>

Kod źródłowy: <https://github.com/pytorch/pytorch>

Licencja: Brak określonej. Prawa autorów zastrzeżone, oprogramowanie *is*:



Rysunek 3.3: Logo biblioteki PyTorch

#### 3.0.4. Tensorflow.js [Deeplearn.js]



Rysunek 3.4: Logo biblioteki Tensorflow.js

W 2017 roku został opublikowany projekt o nazwie Deeplearn.js z celem wprowadzenia uczenia maszynowego i głębokiego uczenia do przeglądarek internetowych bez konieczności korzystania z zewnętrznego API.

JavaScript kojarzony jest przez starych programistów jako powolny interpretowany język ograniczony do pracy w obrębie jednego procesu procesora CPU. Aby uniknąć restrykcji wydajnościowych zaimplementowano wykonanie operacji na WebGL, przeglądarkowemu interfejsowi do OpenGL.

W maju 2018 zespół odpowiedzialny za rozwój DeepLearn.js został połączony z zespołem programistów pracujących przy Tensorflow, a biblioteka zmieniła nazwę na Tensorflow.js. Dostarcza całej mocy Tensorflow do przeglądarki czy dowolnego interpretera kodu Javascript jak Node.js.

Tensorflow.js nie ma wspólnej bazy kodu z Tensorflow, również filozofia działania jest inna. Należy podkreślić że zbieżność nazw wynika z połączenia pracy dwóch zespołów w jedną jednostkę. Interfejs programistyczny udostępniany przez framework javascriptowy udostępnia zarówno niskopoziomowe elementy do uczenia maszynowego oraz wysokopoziomy interfejs wzorowany na tym z biblioteki Keras do tworzenia sieci neuronowych.

Biblioteka na podstawowym poziomie dostarcza dwa elementy:

- CoreAPI – komponent zarządzający kodem niskopoziomowym,
- LayerAPI – komponent zbudowany na CoreAPI, udostępnia abstrakcje wyższego poziomu

Podstawową jednostką danych jest tensor, czyli zbiór wartości liczbowych ułożonych w tablicę jedno lub więcej wymiarową. Tensor definiuje się

---

konstrukturem `tf.tensor`, gdzie należy określić tablice oraz kształt tensora. Istnieje wiele metod pozwalających na stworzenie tensorów różnych kształtów takich jak `tf.zeros` do wypełnienia podanego kształtu zerami, oraz takie jak `tensor1d`, `tensor2d`, `tensor3d` i `tensor4d` dla ułatwienia czytelności przy tworzeniu najczęściej używanych rozmiarów. Tensory są niezmiennie co oznacza że raz stworzony tensor nie może zmieniać wartości.

Zmienne (ang. Variables) są inicjalizowane wartościami tensorów i można modyfikować ich wartości, używa się ich do trzymania i aktualizacji wartości w trakcie trenowania modelu.

Operacje (ang. Operations [Ops]) to metody pozwalające na modyfikacje danych. Tensorflow.js udostępnia całą gamę operacji do pracy z machine learningiem. Działają one na tensorach przez zwracanie wyniku operacji jako nowy tensor.

Powyższe struktury danych należą do CoreAPI, druga warstwa operacji LayerAPI operuje na wyższym poziomie abstrakcji i zapewnia zupełnie inne bloki do budowy sieci neuronowych. Najważniejsze atuty tej warstwy to wysokie podobieństwo do kodu tworzonego w Keras, możliwość naśladowania stylu programowania znanego z języka Python. Poniżej prezentowany jest kod porównawczy między Pythonową wersją tego samego kodu i JavaScriptową reprezentacją z użyciem Tensorflow.js. Kod wykonujący intensywne operacje napisany jest asynchronicznie. Największą zaletą może się jednak okazać brak konieczności używania NumPy, którego w JavaScript brakuje. Wszystkie operacje matematyczne wymagane do uczenia maszynowego z NumPy zostały zaimplementowane w bibliotece.

```
import * as tf from '@tensorflowjs/tfjs';
const model = tf.sequential();
model.add(tf.layers.dense({units: 1, inputShape:
  [1]}));
model.compile({optimizer: 'sgd', loss:
  'meanSquaredError'});
const xs = tf.tensor2d([[1], [2], [3], [4]], [4, 1]);
const ys = tf.tensor2d([[1], [3], [5], [7]], [4, 1]);
await model.fit(xs, ys, {epochs: 1000});
model.predict(tf.tensor2d([[5]], [1, 1])).print();
```

Listing 3.2: Proste operacje w JavaScript z Tensorflow.js

```
import keras
import numpy as np
model = keras.Sequential()
model.add(keras.layers.Dense(units=1,
                              input_shape=[1]))
model.compile(optimizer='sgd',
              loss='mean_squared_error')
xs = np.array([[1], [2], [3], [4]])
ys = np.array([[1], [3], [5], [7]])
model.fit(xs, ys, epochs=1000)
print(model.predict(np.array([[5]])))
```

Listing 3.3: Prosty model w języku Python używając Keras

Dokumentacja: <https://github.com/tensorflow/tfjs>

Kod źródłowy: <https://js.tensorflow.org/>

Licencja: Licencja: Apache 2.0

### 3.0.5. PaddlePaddle



Rysunek 3.5: Logo biblioteki Paddle

Pełna nazwa biblioteki PARallel Distributed Deep LEarning. Oprogramowanie skierowane głównie na jeden rynek, mało znane w Europie / USA, jest to zaś najpopularniejszy framework w Chinach. Tworzony i utrzymywany przez firmę Baidu, znany jako Chiński klon Google. Założony w 2000 roku przeniósł pomysł na wyszukiwarkę internetową na chiński rynek i od tej pory z powodzeniem kopiuje wszystkie technologie i pomysły Google, w tym autonomiczne samochody. Jedynym problemem jaki może napotkać osoba nie znająca języka chińskiego jest częste zgłaszanie uwag i komentarzy do kodu bez opisu w języku angielskim. Tutoriale udostępniane w internecie pochodzą głównie od Baidu, stąd bardzo mała popularność poza rodzimym rynkiem. Ze strony domowej można wyczytać często powtarzające się słowa

---

o prostocie użycia; prostocie działania:

Przechodząc do treści technicznych PaddlePaddle opisywany jest jako zestaw narzędzi i bibliotek do uczenia głębokiego bazujący na języku programowania do uczenia głębokich sieci neuronowych. Biblioteka jest bardzo elastyczna i pozwala na tworzenie dowolnego kształtu modeli. Jak w Tensorflow tak i tutaj istnieje graficzne narzędzie wizualizacji nauczania Visual DL; narzędzie pozwala wyświetlić wydajność trenowania i statystyki danych takie jak dokładność, wartości funkcji kosztu, rozkład parametrów, próbki obrazu i dźwięku, wykres ONNX modelu.

Na podstawie biblioteki istnieje cała społeczność i inicjatywa EasyDL, mająca ułatwić firmom bez specjalistów od algorytmów uczenia maszynowego na wykonanie szybkich i sprawnych modeli bez konieczności posiadania terabajtów danych. Społeczność skupia się głównie w ośrodkach akademickich i startupach w Chinach, gdzie PaddlePaddle jest standardem.

Same możliwości biblioteki są potężne, pozwala na budowę:

- Botów dzięki technologii Word2Vec,
- Systemów rekomendacji,
- Klasyfikatorów obrazów,
- Translacji maszynowej języków z użyciem głębokiego uczenia,
- Detekcji obiektów na obrazie filmów z użyciem Single Shot Multibox Detector,
- Analizatora nastrojów

Przy bliższym spotkaniu, widać że jest to klon Tensorflow, który ze względu na brak konkurencji w Chinach jest dużo szybciej rozwijany i jego możliwości w zakresie uczenia na farmach serwerów przewyższają pierwowzór. Z produktów opartych na Baidu korzysta prawie miliard osób, stąd bardzo duży nacisk na optymalizację i własne rozwiązania technologiczne przekładają się na wysoką wydajność.

Dokumentacja: [http://paddlepaddle.org/documentation/docs/en/0.15.0/getstarted/index\\_en.html](http://paddlepaddle.org/documentation/docs/en/0.15.0/getstarted/index_en.html)

Kod źródłowy : <https://github.com/PaddlePaddle/Paddle>

Licencja : Apache2.0



Rysunek 3.6: Logo biblioteki MXNet

### 3.0.6. MXNet

Biblioteka do głębokiego uczenia stworzona przez fundację Apache. Wspiera największą ilość języków ze wszystkich dostępnych bibliotek:

- Python,
- C++,
- Cloujure,
- Julia,
- Perl,
- R,
- Scala

O popularności i powszechne przyjęciu świadczy implementacja na chmurach Microsoft Azure, Intel i Amazon Web Services. Główne zastosowania do których MXNet jest używany to rozpoznawanie mowy i pisma ręcznego, przetwarzanie języka naturalnego i przewidywanie zdarzeń. Użycie skupia się w kręgach przedsiębiorstw, trudno zatem znaleźć badania akademickie przeprowadzane z wykorzystaniem tej biblioteki.

Kolejnym atutem MXNet jest przenośność modeli. Są one zoptymalizowane by nie zajmować dużo pamięci, pozwala to na przeniesienie modelu wytrenowanego w chmurze do urządzenia mobilnego jak smartfon. Łatwe serwowanie modeli, skalowanie na żądanie z mieszaniem GPU i CPU jest zdecydowanie największymi atutami dlatego Amazon wybrał tę bibliotekę jako pierwszą na swoje centra obliczeniowe.

Gluon jest API wysokiego poziomu dostarczany wraz z MXNet. Jego zadaniem jest dostarczyć czystej, spójnej i prostej składni bez obniżania wydajności działania aplikacji. Istnieje również druga warstwa API wysokopoziomowego (można wybierać, z której chce się korzystać) o nazwie Module. Ta warstwa skupia się na pracy z Symbol, paczką do przetwarzania wyrażeń symbolicznych.[2] Jego najlepszymi atutami są:

- 
- uproszczenie składni budowanych sieci neuronowych, przypominające pseudokod,
  - elastyczna struktura zapewniająca kod imperatywny,
  - dynamiczny graf sieci, umożliwia zmiany struktury sieci w trakcie działania programu,
  - cache'owanie sieci neuronowej poprawia wydajność, w tym celu używa się modelu sieci HybridSequential

Dokumentacja: <https://mxnet.apache.org/api/>

Kod źródłowy: <https://github.com/apache/incubator-mxnet>

Licencja: Apache 2.0

### 3.0.7. Caffe2



Rysunek 3.7: Logo biblioteki Caffe2

Biblioteka wywodzi się z architektury Caffe, narzędzia dla praktyków głębokich sieci neuronowych udostępniającego przejrzystą składnię API w czasach przed pojawieniem się bibliotek wspieranych przez wielkie korporacje. Pierwsza wersja napisana przez Yangqing Jia w trakcie doktoratu na uniwersytecie Berkeley, obecnie jest dyrektorem działu AI w Facebook'u, gdzie rozwijany jest Caffe2. [7] O wieku świadczy dobór języków dostępnych do wykorzystania w bibliotece:

- C++
- Python



- MATLAB

Caffe2 jest rozwinięciem tego modelu, również pracuje na GPU, CPU, jest skalowalny, posiada spore wsparcie społeczności. Różnica jest w modyfikacji architektury obliczeniowej. Główną przewagą jest prosta skalowalność na wiele maszyn i urządzenia mobilne. Ta modularność i elastyczność nazywana jest odframeworkowaniem (ang. un-framework) poprzedniej wersji. [8]

Cała baza została napisana w języku C++ i była testowana latami przez Facebook'a do rozpoznawania obiektów na obrazie, analizy wiadomości do rekomendowania reklam, systemów rekomendujących na podstawie profilu użytkownika. Wydajność jest to jedną z największych zalet, po opracowaniu odpowiedniej architektury, cały kod wdrożeniowy na produkcję może być napisany w C++. Na jednym procesorze NVIDIA DGX-1 prędkość przetwarzania wynosi 231 obrazów na sekundę.[?]

Caffe2 posiada zoo modeli, gotowe do użycia pretrenowane modele dla najczęściej spotykanych zastosowań bez potrzeby wydawania zasobów na tworzenie modelu od zera. Pełna lista modeli, z którymi można eksperymentować jest pokaźna (na dzień obecny największa ze wszystkich dostępnych zwierząt w ZOO):

- AlexNet,
- GoogleNet,
- RCNN ILSVRC13,
- Densenet 121,
- Detectron,
- Inception v1,v2,v3,
- Resnet 50,
- SqueezeNet,
- VGG 19,
- ZFNET 512

Dojrzałość produktu sprawiła że istnieje wiele narzędzi programistycznych ułatwiających pracę z biblioteką. Programiści aplikacji mobilnych mają

---

do dyspozycji wtyczki integracyjne do Visual Studio, Android Studio, Xamarin. Dla pozostałych Caffe2 jest dostępne na każdą popularną platformę desktopową oraz w usługach chmurowych. Kod napisany na jedną platformę jest przenoszalny na pozostałe bez konieczności modyfikacji.

Od kwietnia 2018 roku, repozytorium Caffe2 jest przenoszone do katalogu PyTorch. Facebook zdecydował się na zwiększenie kompatybilności przez stopniowe połączenie bazy kodu. Docelowo Caffe2 ma być modulem dla rozwiązań mobilnych, a PyTorch dla całej reszty.

Dokumentacja: <https://caffe2.ai/docs/>

Kod źródłowy: <https://github.com/caffe2/caffe2>

Licencja: Apache 2.0 (Caffe 1 – BSD)

### 3.0.8. ML.NET



Rysunek 3.8: Logo biblioteki ML.NET

Nowe oprogramowanie od Microsoftu, data wydania pierwszej wersji pogłądowej 7 maj 2018. Microsoft tworząc to oprogramowanie wybrał jako grupę docelową programistów skupionych w ekosystemie .NET. API biblioteki udostępnione jest w językach C++ oraz C, przy czym kompletna dokumentacja jest tylko dla języka C. Docelowo ma stać się częścią .NET Core. Oprogramowanie od początku istnienia powstaje jako otwartoźródłowe, a Microsoft przy każdym wydaniu zachęca do zgłaszania uwag oraz propozycji dalszego rozwoju lub nowych funkcjonalności. Jest to odejście od tradycyjnego modelu utrzymywania oprogramowania w tajemnicy do czasu premiery. Nastawienie na łatwość integracji z oprogramowaniem biznesowym, chmurą Azure oraz łatwość obsługi przez programistów pracujących w technologiach Microsoftu ma zachęcić do wykorzystania oprogramowania na szerszą skalę kiedy wyjdzie ono z fazy rozwojowej i zastąpi CNTK.

W odróżnieniu od pozostałych na chwilę obecną ML.NET jest biblioteką głównie przeznaczoną do uczenia maszynowego, wyewoluowaną z wewnętrznych rozwiązań Microsoftu od lat używanych w zespołach Windows, Bing, Azure. Microsoft przyznaje że jest to przepisywanie większego systemu do świata otwartego oprogramowania. [?] Możliwości biblioteki można rozszerzać przez podłączenie modeli z innych bibliotek, wspierane od początku są Tensorflow, Caffe2, CNTK oraz Accord.NET. Sporym atutem jest natywne wsparcie w Azure każdej nowej wersji natychmiast po premierze. Wersja 0.3 przyniosła wsparcie dla formatu modeli ONNX. Narzędzia do głębokiego uczenia pojawiły się w wersji 0.5, zaprezentowanej 12 września 2018. Integracja modeli z Tensorflow umożliwia przeniesienie środowiska z testowego do modelu E2E (ang. Enterprise-2-Enterprise) Przedsiębiorstwo-do-przedsiębiorstwa: Jest to jeszcze zbyt niedojrzała biblioteka by opisywać jej architekturę, nie jest nawet zalecana do rozwiązań produkcyjnych, programiści .NET prawdopodobnie będą z niej chętniej korzystali po pojawieniu się pierwszej kompletnej wersji, co ma nastąpić do wydarzenia BUILD 2019. Obecnie traktowana jest jako sposób na przyciągnięcie do ekosystemu .NET Core osób tworzących rozwiązania dla przedsiębiorstw. ML.NET ma jeszcze wiele braków i wad (konieczność programowania skalowania, brak wsparcia platform mobilnych i wyłączność na chmurę Azure), które uniemożliwiają traktowanie go jako poważnego narzędzia komercyjnego w przeciwieństwie do kolejnego narzędzia Microsoftu CNTK.

Dokumentacja: <https://docs.microsoft.com/en-us/dotnet/machine-learning/>

Kod źródłowy: <https://github.com/dotnet/machinelearning>

Licencja: MIT

### 3.0.9. CNTK

Nazwa kompletna Microsoft Cognitive Toolkit: Jest to zestaw narzędzi stworzonych w dziale badań i rozwoju Microsoftu, zajmującego się Deep Learningiem. W 2015 zgodnie z nowym modelem biznesowym, nastawionym na częściowe otwieranie kodu źródłowego narzędzi używanych wewnątrz Microsoftu, CNTK został opublikowany na GitHub'ie w kwietniu tego roku. Microsoft wcześniej korzystał z powodzeniem z biblioteki do trenowania modeli usług krytycznych biznesowo. Architektura jest skupiona wokół głębokiego uczenia sieci neuronowych, a dopiero później dodano pozostałe popularne algorytmy uczenia maszynowego.

CNTK posiada zaimplementowane najczęściej używane narzędzia do głę-



Rysunek 3.9: Logo biblioteki CNTK

bokich sieci neuronowych, takie jak struktury sieci konwolucyjnych czy rekurencyjne sieci neuronowe. Także algorytmy spadku gradientowego i wsteczna propagacja błędów zostały zaimplementowane z automatycznym zrównoleganiem obliczeń z pojedynczego CPU na wiele procesorów GPU na wielu komputerach w sieci. Biblioteka ta jest jednym z backendów do biblioteki Keras, jednak w przeciwieństwie do Keras'a działa tylko na systemach Windows i wybranych dystrybucjach Linux'a.

Niesamowicie wartościowym atutem jest z pewnością dokumentacja dostarczana przez Microsoft. Nie tylko jest ona przejrzysta i spójna z resztą dokumentacji ekosystemu .NET, ale również dostarcza dziesiątek tutoriali krok po kroku jak budować zaawansowane modele, czyścić dane i wyciągać z nich wnioski. Przystawia nawet w jaki sposób przygotowane już rozwiązania przenieść na chmurę Azure i umożliwić korzystanie z nich klientom. Dla osób nie mających podłoża w Data Science, są gotowe przepisy, czyli przygotowane wcześniej szablony do działania na Azure z interfejsem graficznym instruującym jak przygotować dane aby otrzymać gotową usługę.

Dokumentacja: <https://www.microsoft.com/en-us/cognitive-toolkit/>

Kod źródłowy: <https://github.com/Microsoft/CNTK>

Licencja: MIT

## Rozdział 4.

# Architektura

### 4.1. Znaczenie architektury dla wydajności sieci

Elementy sieci neuronowych można ułożyć na nieskończenie wiele sposobów, wystarczy dodać kolejną warstwę neuronów lub zmodyfikować wielkość warstwy i wyniki będą zgoła inne. Celem szukania odpowiedniej dla danego zadania architektury jest otrzymanie najlepszej wydajności klasyfikacji przy skończonym i możliwie jak najkrótszym czasie uczenia. Chcąc uzyskać dobre rezultaty bez eksperymentów od zera, należy wybrać architekturę, która została wykorzystana do rozwiązania podobnego zadania bądź przetestować kilka gotowych architektur z wyuczonymi modelami, a następnie sprawdzić jak się zachowują dla posiadanych zbiorów danych na różnych hiperparametrach.

Stworzenie efektywnej architektury przesądziło o wynikach zawodów "ImageNet Competition 2012". Pierwsze miejsce w klasyfikacji obrazów wyłącznie z użyciem głębokiej sieci neuronowej zapoczątkowało prawdziwą rewolucję nie tylko wśród społeczności uczenia maszynowego, ale całego przemysłu technologicznego. Pierwszy zespół, któremu udało się zejść z błędem rozpoznania do mniej niż 25% Geoffrey Hinton, zwany ojcem chrzestnym głębokich sieci neuronowych (ang. Godfather of Deep Learning) udowodnił że jego przekonania, którymi się kierował w trakcie kilkudziesięcioletniej kariery naukowej są słuszne. Poprzeczką, której profesor Hinton nie mógł wcześniej przeskoczyć były ograniczenia obliczeniowe procesorów, kiedy technologia GPU wraz z możliwością programowania równoległego w CUDA umożliwiły przetestowanie idei, podejście do rozpoznawania obrazów zmieniło świat przemysłu, a wkrótce zmieni życie każdego człowieka.

Rok 2017 został uznany za czas kiedy ImageNet Competition został na stałe rozwiązany. Błędne rozpoznanie algorytmów wśród najlepszych drużyn wyniosło zaledwie 2%. Przez pięć lat od czasu debiutu AlexNet powstało wiele architektur sieci neuronowych, bibliotek wyspecjalizowanych w szybkim tworzeniu nowych prototypów, a te dające obiecujące wyniki zostały podzielone względem zastosowania. Obecnie wyróżniane typy sieci neuronowych są sklasyfikowane względem typu danych wejściowych. Dla sygnału wejściowego w postaci pikseli najlepiej sprawdza się konwolucyjna sieć neuronowa. Warstwy konwolucyjne potrafią bardzo dokładnie wyodrębnić cechy obrazu na różnych poziomach złożoności co jest wyjaśnione w dalszej części pracy.

Polskim ekspertem i popularyzatorem sieci neuronowych od wielu lat jest profesor doktor habilitowany inżynier Ryszard Tadeusiewicz. Osoba bardzo zasłużona w naukach technicznych. Jego książki stały się częścią bibliografii niniejszej pracy. Profesor w swoich publikacjach istotnie zwraca uwagę na architekturę jako element konieczny w budowaniu

Architektury różnią się ilością warstw ukrytych, typem neuronów i funkcji aktywacji w każdej warstwie. “Właśnie taka [warstwowa] struktura sieci wyjątkowo łatwo i wygodnie da się wytwarzać zarówno w formie modelu elektronicznego, jak i da się symulować w formie programu komputerowego. Dlatego badacze przyjęli właśnie strukturę warstwową i od tej pory stosują ją we wszystkich sztucznych sieciach neuronowych. [...] W związku z tym wszyscy tak postępują, nie martwiąc się ani przesłankami biologicznymi, ani dowodami wskazującymi, że architektura sieci bardziej wymyślnie dostosowanej do charakteru zadania może znacznie lepiej realizować stawiane zadania.”

Tak napisał w 2007 roku, kiedy nie istniały jeszcze złożone głębokie sieci neuronowe, największe w tej porzy miały 5 warstw ukrytych, a zbiory danych na których pracowano były stosunkowo niewielkie w porównaniu z dzisiejszymi zasobami skatalogowanych obrazów. Prof. Tadeusiewicz uważa że jest to najlepsze ułożenie neuronów ze względu na prostotę oraz dowód że badacze znaleźli podobne struktury w niektórych częściach ludzkiego mózgu (w tym w oku, które jest ewolucyjnie złożone z tej samej tkanki co mózgowa).

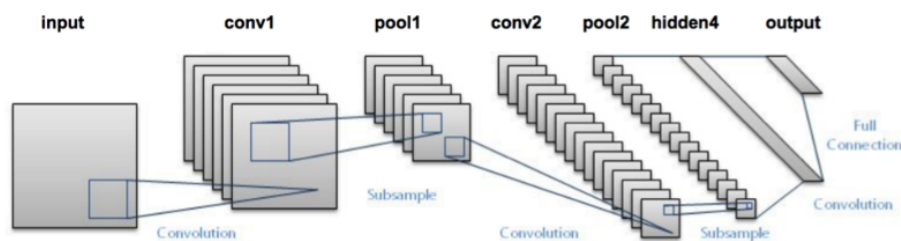
Dowodem na znaczenie ułożenia neuronów w poprawnej strukturze jest ich rosnąca skuteczność od kiedy badacze bardzo dużo eksperymentują z różnymi kombinacjami. Coroczne zawody ImageNet, w których udowadnia się że struktura sieci ma znaczenie są świetnym przykładem że jest to pro-

ces iteracyjny i jak podkreślano wcześniej, metoda empiryczna jest obecnie jedynym sposobem na znalezienie idealnego modelu.

## 4.2. Przegląd najbardziej efektywnych architektur

### 4.2.1. LeNet

[Yann LeCun, Leon Bottou, Yoshua Bengio, Patrick Haffner, 1998] - . Pierwsze praktyczne zastosowanie sieci konwolucyjnych jeszcze w latach 90'tych. Używana była do prostych zadań: czytanie kodów pocztowych, liczb.[10]



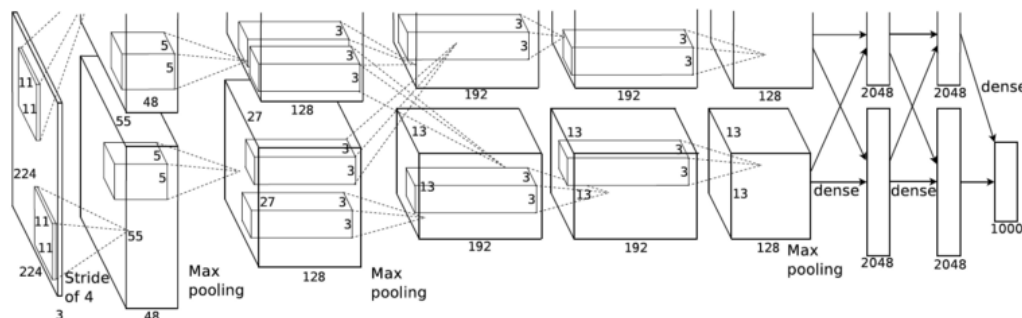
Rysunek 4.1: Architektura sieci LeNet

### 4.2.2. AlexNet

[Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, 2012] - pierwsza prawdziwa konwolucyjna sieć neuronowa (CNN), która pomogła zmienić opinię na temat uczenia głębokiego. Została zaimplementowana z użyciem biblioteki CUDA. Pierwsza sieć neuronowa, ucząca się z powodzeniem dla dużego zbioru danych. Została wytrenowana na zbiorze danych złożonym z ponad 15 milionów obrazów podzielonych na 22000 klas. W testach top-1 i top-5 uzyskała wartości kolejno 37,5% i 17%, co pozwoliło wygrać konkurs ImageNet Large Scale Visual Recognition Challenge. Sieć neuronowa zawiera 60 milionów parametrów i 650 000 neuronów. Architektura tworzy 8 warstw, gdzie pierwsze 5 to warstwy konwolucyjne, a pozostałe 3 są warstwami w pełni połączonymi, ta ostatnia ma oczywiście 1000 wyjść z funkcji softmax. AlexNet znacząco przewyższyła swoją wydajnością poprzednich uczestników i wygrała zawody redukując błąd top-5 do 15,32%. Drugie miejsce to błąd ok 26.2% (nie była to CNN). Sieć jest głęboką modyfikacją architektury Yann'a LeCun'a. AlexNet była zaplanowana na dwie karty graficzne, stąd rozdzielenie przepływu informacji na 2 części. Trenowanie sieci na 2

## 4.2. PRZEGLĄD NAJBARDZIEJ EFEKTYWNYCH ARCHITEKTUR

GPU było nowością na te czasy. Sieć została wytrenowana na zbiorze ImageNet. Do wyliczenia funkcji nieliniowych były używane ReLU (tutaj po raz pierwszy okazało się że ReLU działa dużo szybciej niż tangens hiperboliczny). Sieć o której będą uczyły się dzieci na lekcjach historii. [9]



Rysunek 4.2: Architektura sieci AlexNet

### 4.2.3. VGG Net

[Simoyan i Zisserman, 2014] - sieć złożona z 16 warstw konwolucyjnych, która charakteryzuje się małymi filtrami i dużą głębokością sieci. W trakcie trenowania sieci wyjście jest ustawione na ustalony rozmiar (224 x 224 x 3). Przetwarzanie wstępne obejmuje odjęcie mediany wartości RGB dla każdego piksela. Zdjęcie jest przetwarzane przez stos warstw konwolucyjnych, gdzie używane są filtry o bardzo małym polu widzenia (3x3) [najmniejszy możliwy rozmiar by móc rozpoznać kierunek]. Operacja Max-pooling jest wykonana na polu 4 pikseli, co pokazuje że jest to pobieranie jak najmniejszych cech z obrazu. Ukryte warstwy są wyposażone w nieliniową funkcję aktywacji ReLU. Po stosie warstw konwolucyjnych, następuje nałożenie 3 warstw w pełni połączonych (to takie duże warstwy zawierające wszystkie cechy?). Pierwsze dwie mają 4096 kanałów, trzecia już tylko 1000 (po jednym kanale na klasę obiektu). Obecnie architektura ta jest dość popularnym wyborem dla wyodrębniania cech ze zdjęć. Konfiguracje wag dla zbioru obrazów z ImageNet są dostępne online. W tej sieci problemem jest 140 milionów parametrów, którymi czasem trzeba zarządzać. [14]

### 4.2.4. GoogleNet / Inception

[Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabino-



vich, 2014] - produkt jak nazwa wskazuje wyszedł z firmy Google. Charakteryzuje się przełomową poprawą użycia zasobów wewnętrznych. ImageNet jest zaprojektowana na jak największą szerokość i głębokość jednocześnie utrzymując stałe zużycie mocy obliczeniowej. Optymalizacja jakości opiera się na zasadzie Hebbian’a (jest to teoria neuronauki, która mówi o adaptacji neuronów w trakcie nauki). Błędy klasyfikacji wynoszą top-1: 17,2% i top-5: 3,58% (dla v3). Sieć składa się łącznie z 22 warstw. Sieć jest modułowa, gdzie każdy moduł pełni rolę wielopoziomowego ekstraktora cech przeliczającego konwolucje na macierzach rozmiarów 1x1, 3x3, 5x5. Zaletą tej architektury są bardzo małe rozmiary wag (mniej niż 100MB dla v3). Kolejną ważną zaletą jest ilość parametrów, tylko 4 miliony co jest wynikiem 12 razy lepszym od AlexNet. Jest to jedna z najbardziej złożonych sieci względem architektury modułowej. W wersji naiwnej każda warstwa posiada 4 ekstraktory cech (3 konwolucyjne i jedną 3x3 max pooling), następnie wartości te są składane i wysyłane do warstwy wyżej. Do poprawienia wydajności obliczeniowej pozbyto się warstwy “w pełni połączonej”. [15]

#### **4.2.5. ResNet**

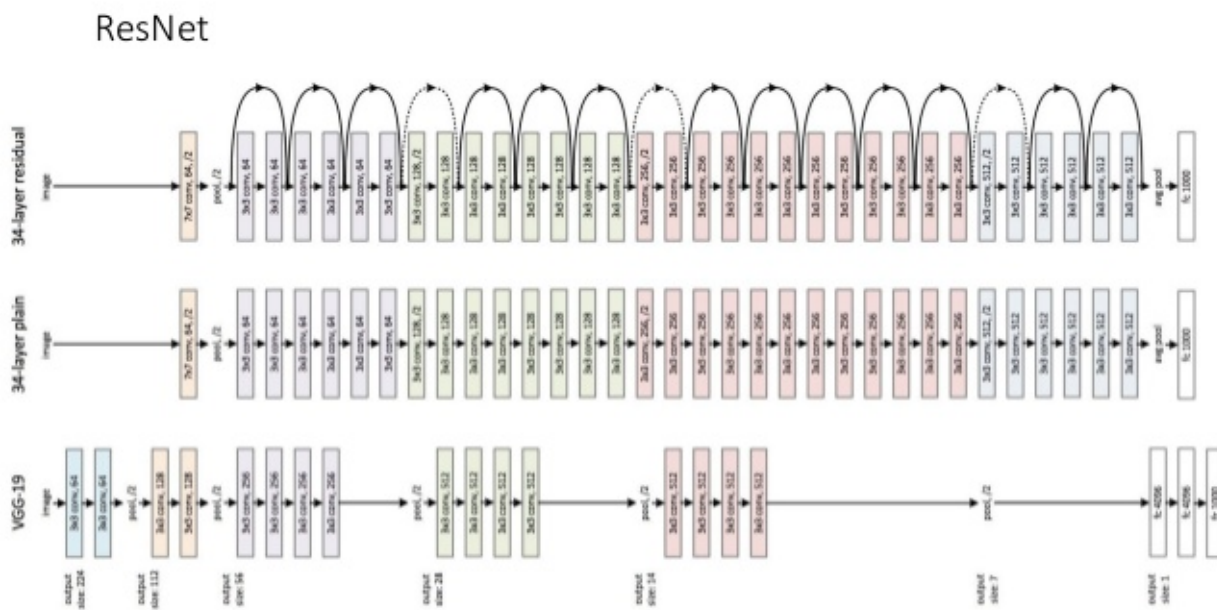
[Kaiming He et al, 2015] - architektura oparta na mikro modułach. Charakteryzuje się możliwością odrzucania połączeń i posiada ogromny moduł na normalizację zbioru iteracji (batch normalization). Normalizacja jest pomysłem zaczerpniętym z rekurencyjnych sieci neuronowych. Ta technika pozwala stworzyć sieć ze 152 warstwami ukrytymi przy zachowaniu złożoności mniejszej niż VGG. [17]

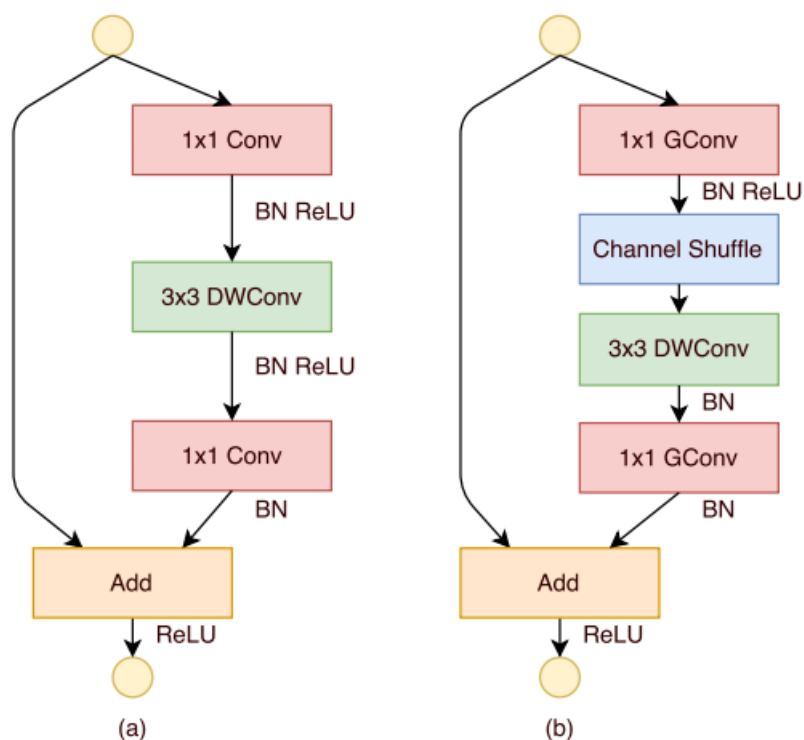
#### **4.2.6. ResNeXt**

[Saining Xie, Ross Girshick, Piotr Dollar Zhuowen Tu Kaiming He, 2017] - modularyzowalna [17]

#### **4.2.7. RCNN – obrazek jest zastępczy nie ma nic w internecie interesującego (zrobić własną wizualizację)**

[Authors, Year] - [12]

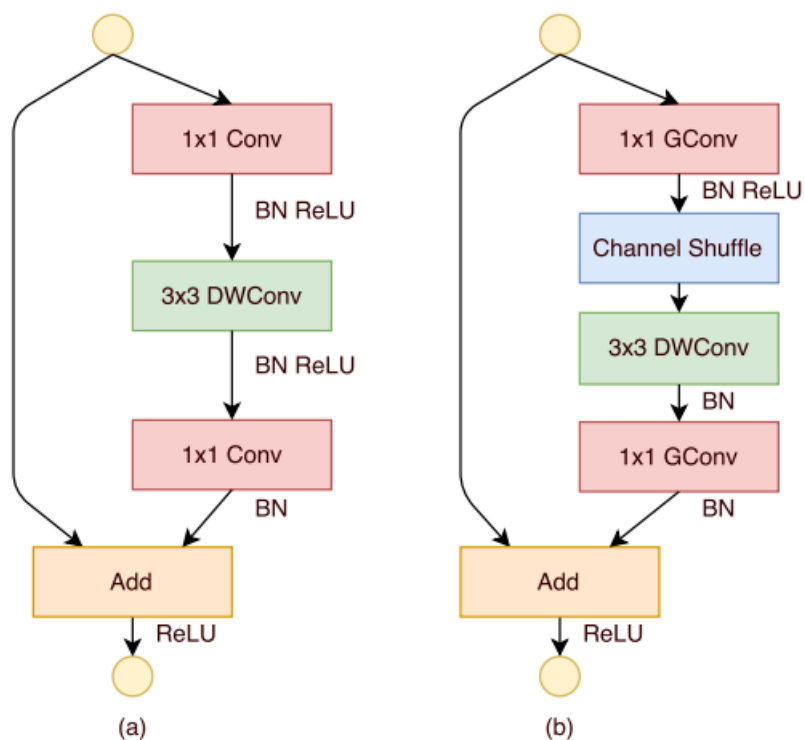




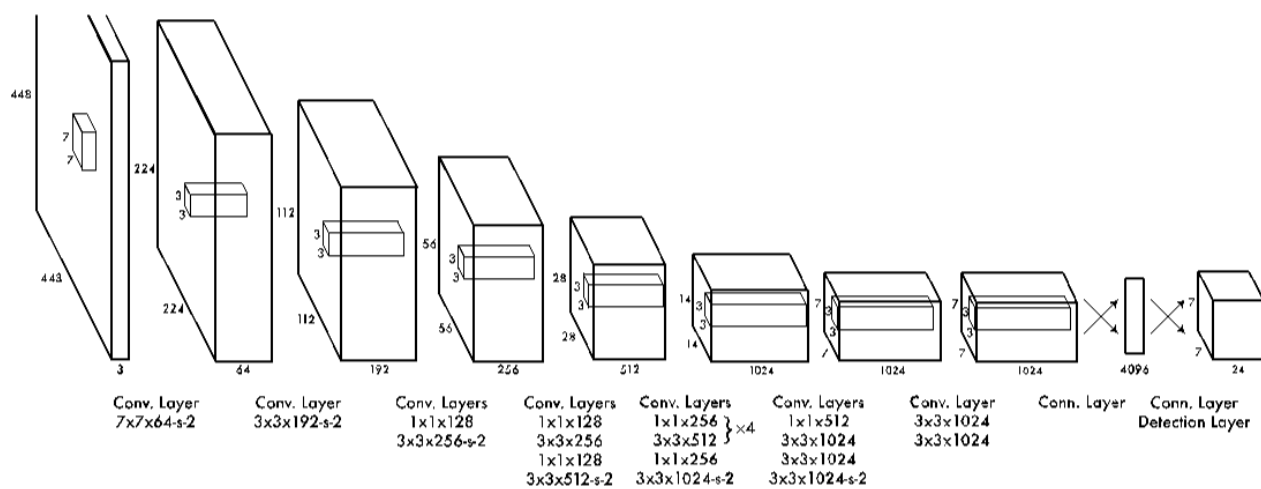
Rysunek 4.4: Architektura sieci ResNeXt

przez największe obecnie autorytety w dziedzinie. GAN obudził duże nadzieje na szybkie i "twórcze" działania algorytmów, jego najczęstszym zastosowaniem jest tworzenie komputerowych "dzieł sztuki". Jak to działa? Jedna sieć generuje kandydatów, a druga ich ocenia wytworzone obiekty. W ten sposób obie sieci uczą się od siebie wzajemnie. - [4]

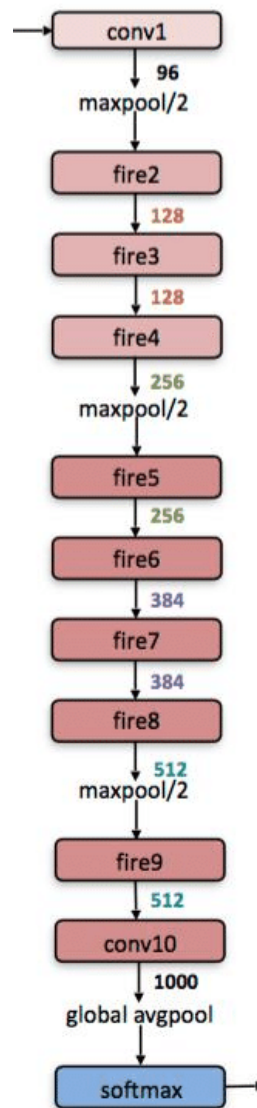
## 4.2. PRZEGLĄD NAJBARDZIEJ EFEKTYWNYCH ARCHITEKTUR



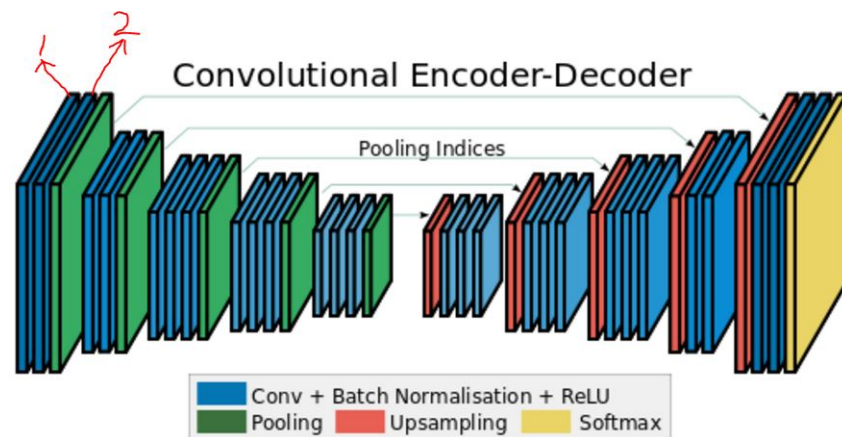
Rysunek 4.5: Architektura sieci RCNN



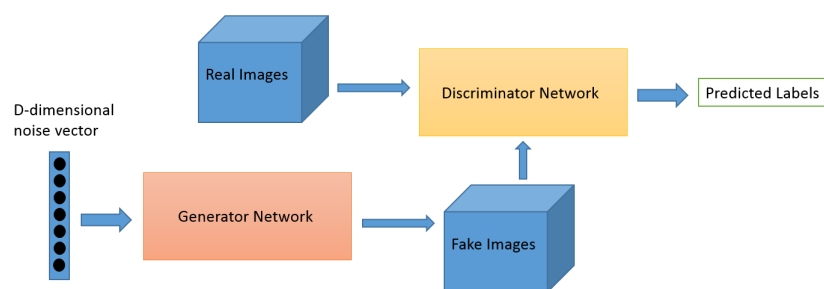
Rysunek 4.6: Architektura sieci YOLO



Rysunek 4.7: Architektura sieci SqueezeNet



Rysunek 4.8: Architektura sieci SegNet



Rysunek 4.9: Architektura sieci GAN

## Rozdział 5.

### Testy wydajności





# Podsumowanie



# Spis rysunków

1.1. Wizualizacja uproszczonego modelu neuronu mózgowego. . .	14
1.2. Wizualizacja prostego sztucznego neuronu. . . . .	14
1.3. Prosta reprezentacja neuronu. . . . .	16
1.4. Prosta sieć neuronowa złożona z dwóch warstw. . . . .	17
1.5. Wykres funkcji sigmoidalnej . . . . .	21
1.6. Wykres funkcji tangensu hiperbolicznego . . . . .	22
1.7. Wykres rektyfikowanej jednostki liniowej . . . . .	22
1.8. Wykres nieszczelnej rektyfikowanej jednostki liniowej . . . .	23
1.9. Wykres obliczeniowy dla funkcji $J(a,b,c) = 3(a+bc)$ . . . . .	24
3.1. Logo biblioteki Tensorflow . . . . .	30
3.2. Logo biblioteki Keras . . . . .	32
3.3. Logo biblioteki PyTorch . . . . .	35
3.4. Logo biblioteki Tensorflow.js . . . . .	35
3.5. Logo biblioteki Paddle . . . . .	37
3.6. Logo biblioteki MXNet . . . . .	39
3.7. Logo biblioteki Caffe2 . . . . .	40
3.8. Logo biblioteki ML.NET . . . . .	42
3.9. Logo biblioteki CNTK . . . . .	44
4.1. Architektura sieci LeNet . . . . .	47
4.2. Architektura sieci AlexNet . . . . .	48
4.3. Architektura sieci ResNet . . . . .	50
4.4. Architektura sieci ResNeXt . . . . .	51
4.5. Architektura sieci RCNN . . . . .	52
4.6. Architektura sieci YOLO . . . . .	52
4.7. Architektura sieci SqueezeNet . . . . .	53
4.8. Architektura sieci SegNet . . . . .	53
4.9. Architektura sieci GAN . . . . .	54



# Bibliografia

- [1] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *CoRR*, abs/1505.07293, 2015.
- [2] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *CoRR*, abs/1512.01274, 2015.
- [3] Ahmed Menshawy Giancarlo Zaccone, Md. Rezaul Karim. *Deep Learning with TensorFlow*. Packt Publishing, 2017.
- [4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
- [5] Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR*, abs/1602.07360, 2016.
- [6] Rachel Thomas Jeremy Howard. Fast.ai making neural nets uncool again, 2017.
- [7] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [8] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell.

- Caffe: Convolutional architecture for fast feature embedding. *CoRR*, abs/1408.5093, 2014.
- [9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [10] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.
- [11] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.
- [12] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.
- [13] Shaohuai Shi, Qiang Wang, Pengfei Xu, and Xiaowen Chu. Benchmarking state-of-the-art deep learning software tools. *CoRR*, abs/1608.07249, 2016.
- [14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.
- [16] Ryszard Tadeusiewicz. *Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#*. Polska Akademia Umiejętności, 2007.
- [17] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016.