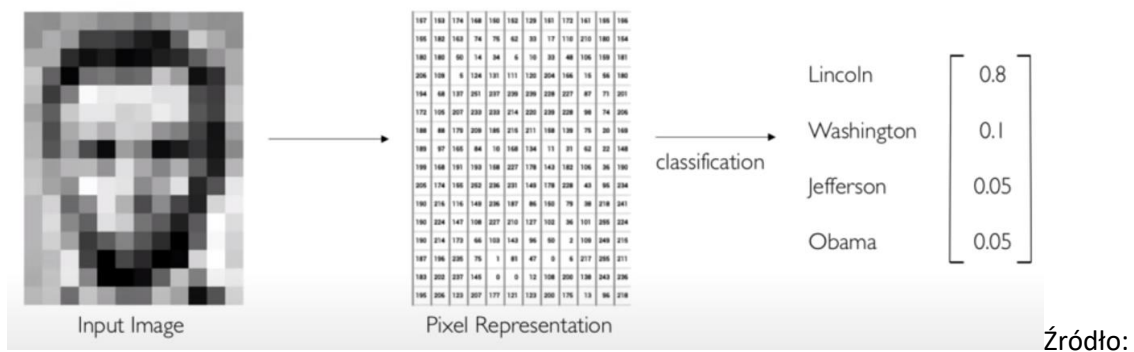


Convolutional Neural Networks

Konwolucyjne sieci neuronowe

- Dane przestrzenne (*spatial data*, obrazy, widzenie maszynowe)
- Bardzo dobrze identyfikują wzorce
- Bardzo wymagające obliczeniowo
- Różnica od MLP (*multilayer perceptron*):
 - Warstwy sieci nazywamy warstwami konwolucyjnymi (*convolutional layers*)
- Zastosowania¹:
 - Computer Vision - face recognition, image classification, analiza pisma...
 - NLP (*Natural Language Processing*) - rozpoznawanie mowy, klasyfikacja tekstu...
- Stosowane także w Uczeniu ze Wzmocnieniem (jaką akcję podjąć na podstawie aktualnego stanu gry?)

Dane wejściowe



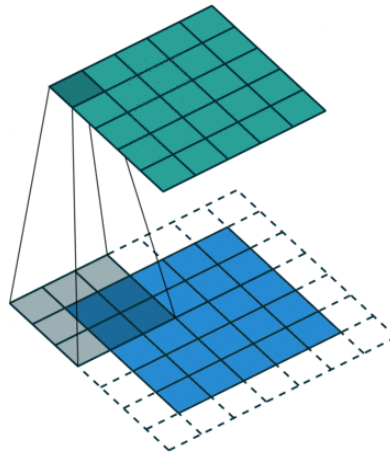
<https://www.youtube.com/watch?v=iaSUymCekI&t>

- Sieć musi wiedzieć, co takiego odróżnia poszczególne osoby od siebie
- Problemy z obrazami i MLP:
 - Na wejściu sieci otrzymujemy 1D wektor wartości pixeli. Powoduje to utratę przestrzenności obrazu.
 - MLP posiada warstwy połączone pełnie (*fully connected layers*) więc przypada pixel per neuron – ogromna liczba parametrów do nauczenia...

¹ Applications of Convolutional Neural Networks -

<http://ijcsit.com/docs/Volume%207/vol7issue5/ijcsit20160705014.pdf>

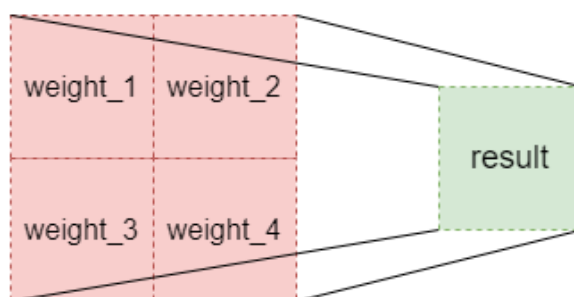
- Rozwiązanie CNN – przekazanie na wejście neuronu fragmentu obrazu (*filtr*) (np.: 3x3)



Filtr 3x3 przechodzący po obrazie

- Zaczynając od góry obrazu zdefiniowany przez nas *filtr* przechodzi po całym obrazie *pixel po pixelu*. Poprzednie warstwy przekazują te fragmenty do kolejnych itd.
- Dzięki temu zachowujemy informacje przestrzenne
- Jak wyglądają teraz obliczenia w neuronie? W dużym skrócie wyliczymy sumę ważoną pixeli w filtrze i to przekazujemy do kolejnej warstwy w celu identyfikacji konkretnej cechy (*feature*)

Konwolucja

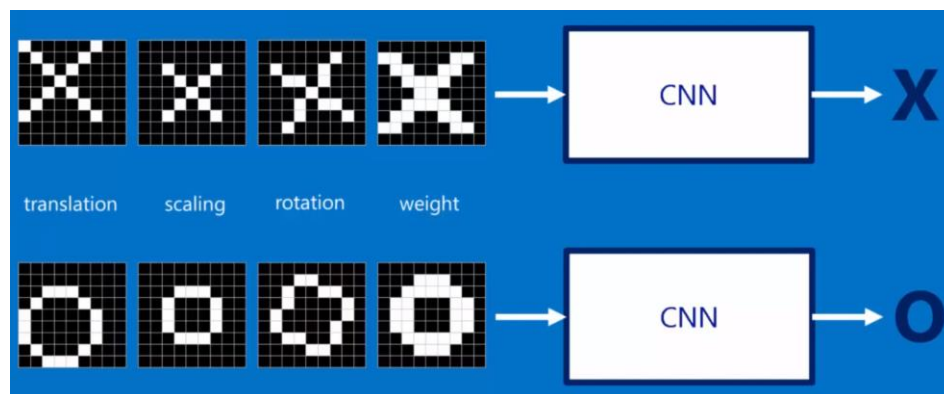


Przykład filtra 2x2

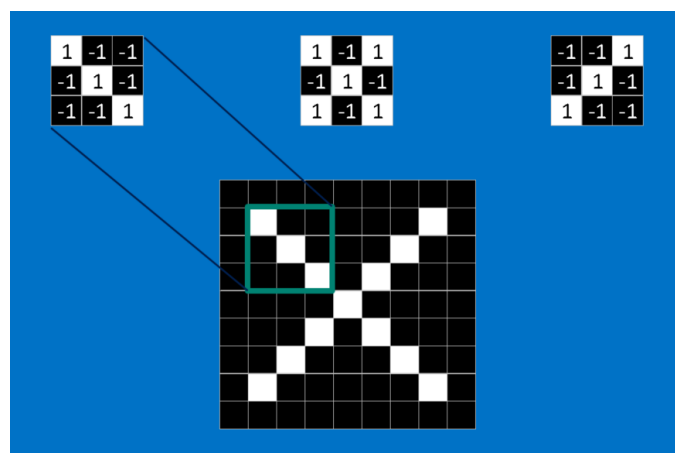
$$\begin{aligned} \text{result} = & \\ & \text{pixel_1} * \text{weight_1} + \\ & \text{pixel_2} * \text{weight_2} + \\ & \text{pixel_3} * \text{weight_3} + \\ & \text{pixel_4} * \text{weight_4} \end{aligned}$$

- Operacja aplikowania wag z filtra w celu ekstrakcji **lokalnych cech** nazywana jest **konwolucją**.
- Aby wyekstrahować wiele cech należy zastosować wiele filtrów.
- Filtr może być przesuwany co jeden pixel lub więcej - zależy to od wielkości filtra (np. Filtr 4x4 może przechodzić co dwa pixele ważne, aby objąć cały obraz filtrem)

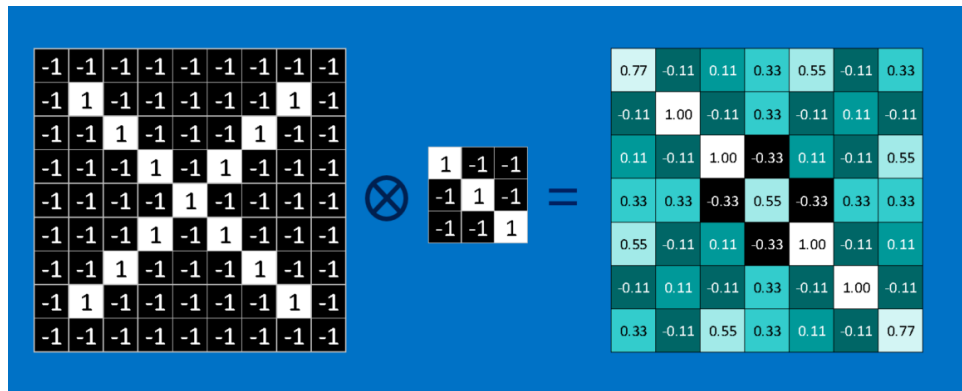
Przykład ekstrakcji cechy przez konwolucję



- Podsumowanie:
 - Celem jest rozpoznanie obrazu bez względu na to czy interesujący nas obiekt jest zdeformowany
 - CNN szuka cech fragment po fragmencie (to cechy pozwalają na odróżnienie obiektów)
 - Każda cecha jest jak mały obraz, którego szukamy w całości. Cechy powinny odpowiadać powtarzającym się wzorcom w obrazie.

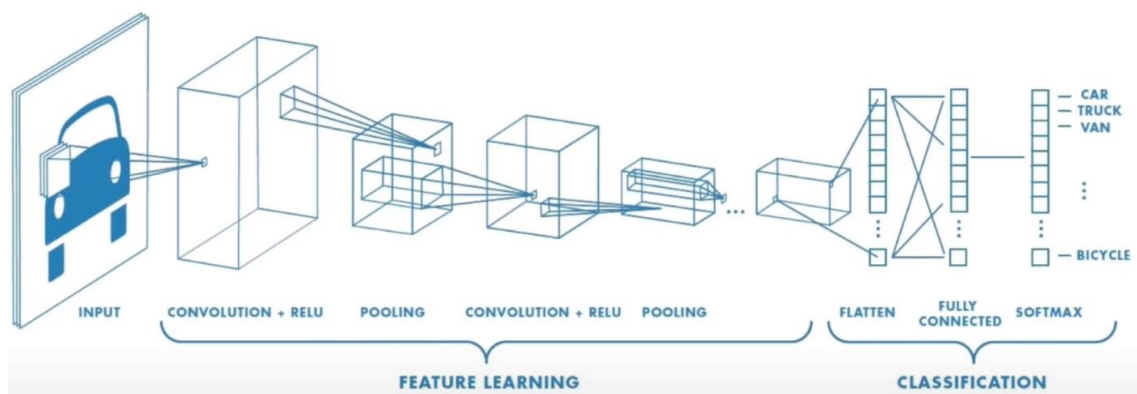


- Filtr odpowiada cecie obrazu, której CNN poszukuje.

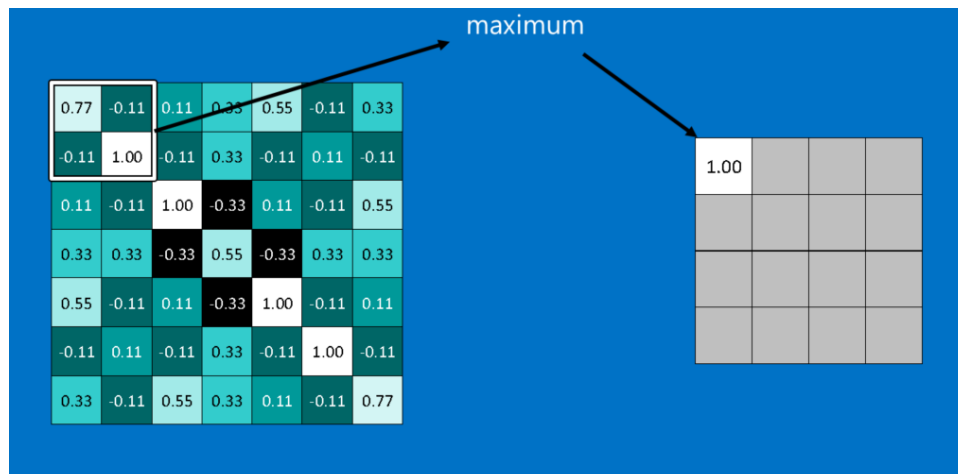


- Wyliczenie konwolucji zmniejsza wymiar macierzy. Wyliczenie konwolucji: $pixel_pierwotny \times pixel_filtra + \dots$. Dzieląc wynik przez liczbę pixeli w filtrze otrzymamy wskaźnik podobieństwa filtra do sprawdzanego fragmentu (jest to operacja opcjonalna, w p.p. interesuje nas wartość maksymalna). W wyliczonej macierzy (*feature map*) szukamy wartości największych, tam jest aktywacja odpowiadająca danemu filtrowi.
- Różne filtry dają różne mapy cech (*feature maps*)
- https://brohrer.github.io/how_convolutional_neural_networks_work.html

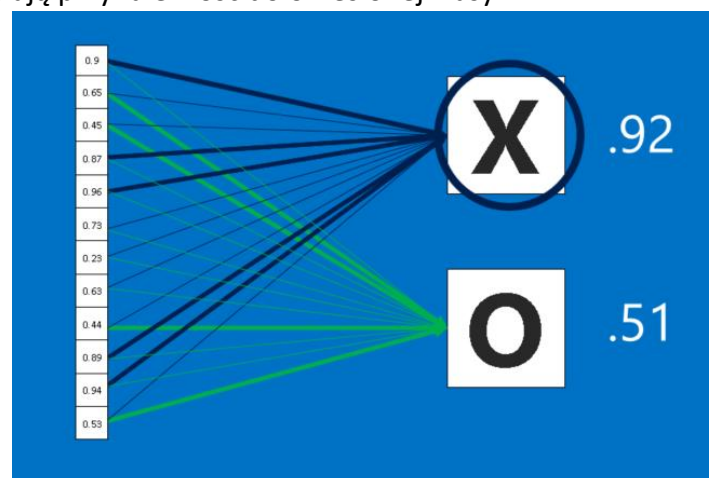
CNN



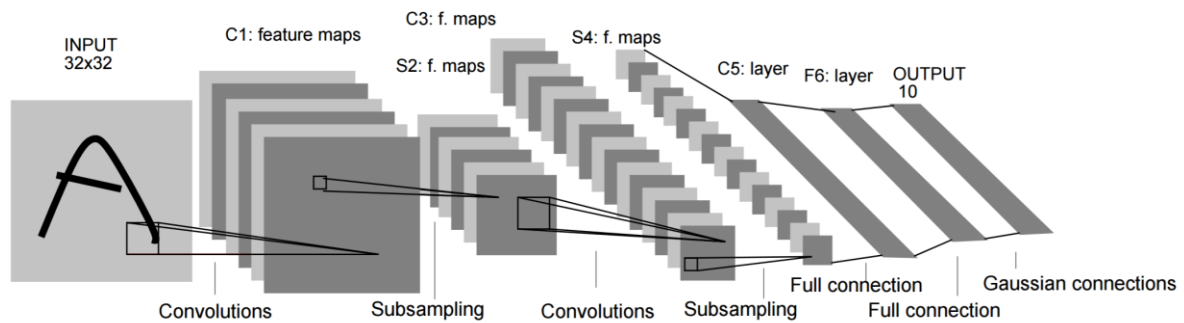
- Każdy neuron warstwy ukrytej wylicza wartość konwolucji oraz stosuje funkcję aktywacji (*ReLU*)



- **Pooling** – technika używana w CNN. Zmniejszanie obrazów zachowując ich właściwości. Polega na przechodzeniu danym fragmentem (zazwyczaj 2x2 lub 3x3) i wybieraniu wartości maksymalnej. Najczęstszy krok: 2 pixele.
 - Odciąża obliczenia
- Ostatnia warstwa jest w pełni połączona: traktuje wejście jako jednowymiarową listę. Wagi połączeń pomiędzy neuronami interpretujemy jako “głos”, do której klasy przynależy obraz.
 - Neurony specjalizują się w klasach - wyższe wartości odpowiednich neuronów określają przynależność do określonej klasy



Zadanie



Implementacja sieci LaNet-5². Zaprojektowana do klasyfikacji pisma odręcznego. Posiada dwie warstwy konwolucyjne, każda posiada warstwę *subsampling* (*pooling*).

1. Opis implementacji modelu LaNet
2. Analiza porównawcza modelu MLP z poprzednich zajęć oraz modelu LaNet:
 - a. Jakie są główne różnice, który sprawuje się lepiej dla jakiej ilości danych
 - b. Dla jakiej konfiguracji sieci MLP wyniki są zbliżone do CNN? Ile czasu trwa uczenie jednej oraz drugiej sieci aby osiągnąć podobne wyniki?

² <http://yann.lecun.com/exdb/lenet/>