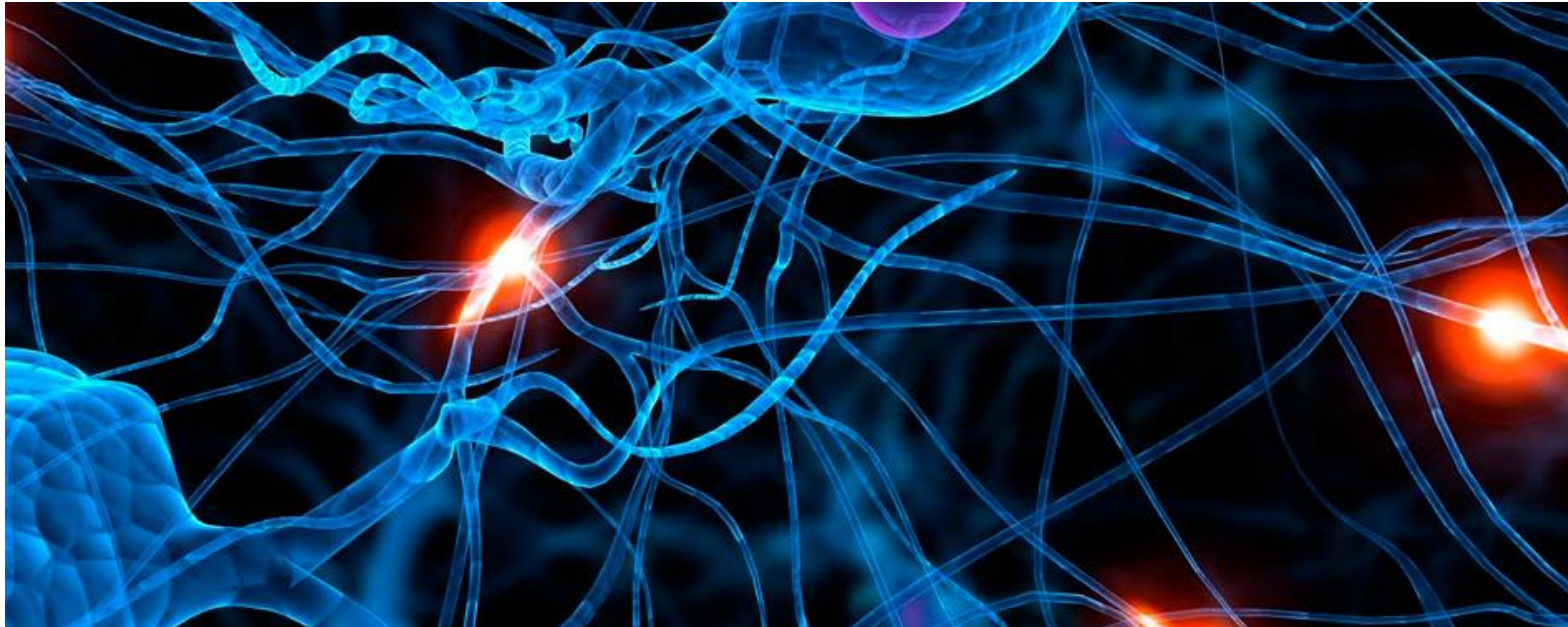


Zastosowanie sieci neuronowych do przewidywania popularności filmików

Tomasz Bocheński



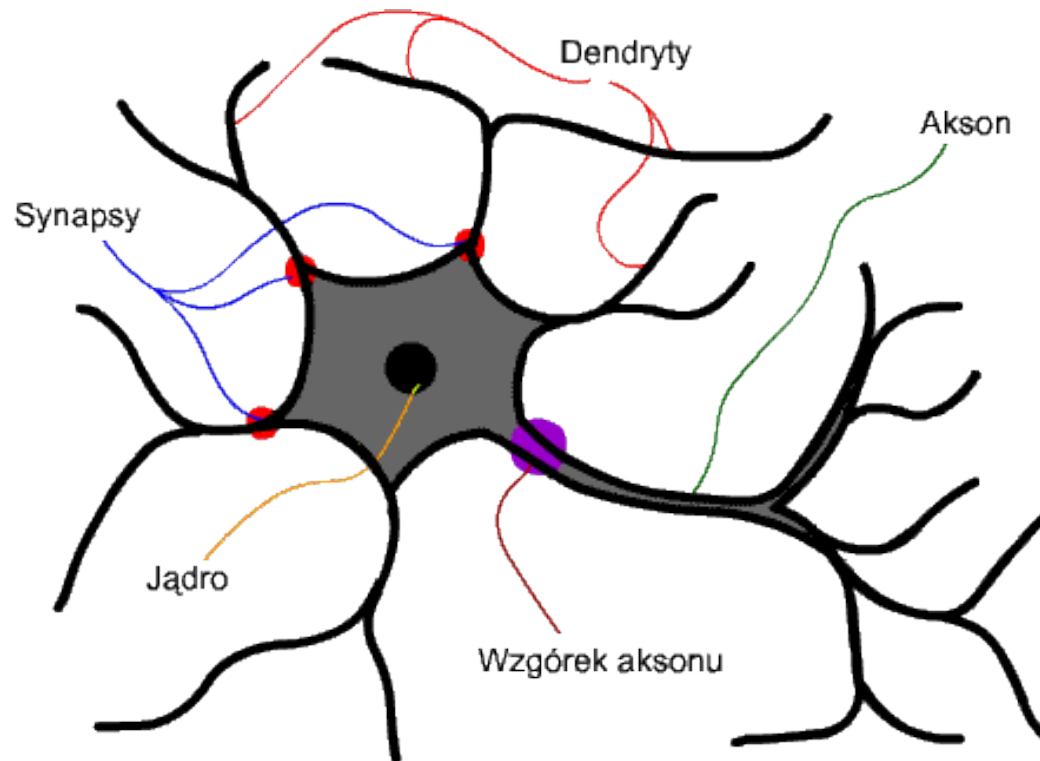
Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

Agenda

- *Sztuczne sieci neuronowe – inspiracje biologiczne*
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

Budowa neuronu

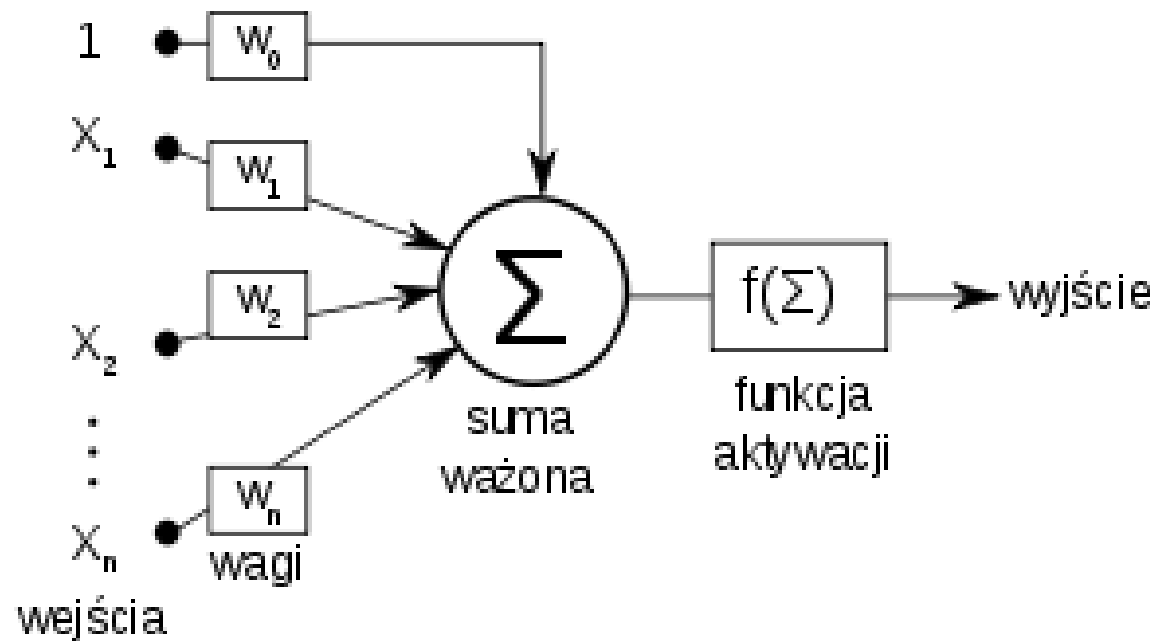


Schemat budowy neuronu [Źródło: Internet]

Z punktu widzenia sztucznych sieci neuronowych **najważniejszymi częściami składowymi neuronu są:**

- **dendryty** – wejścia neuronu
- **synapsy** – zakończenia dendrytów, mogą modyfikować trafiający do nich impuls poprzez wzmacnianie go lub osłabianie
- **jądro** – centrum obliczeniowe neuronu
- **akson** – wyjście neuronu

Budowa sztucznego neuronu



Schemat budowy sztucznego neuronu [Źródło: Internet]

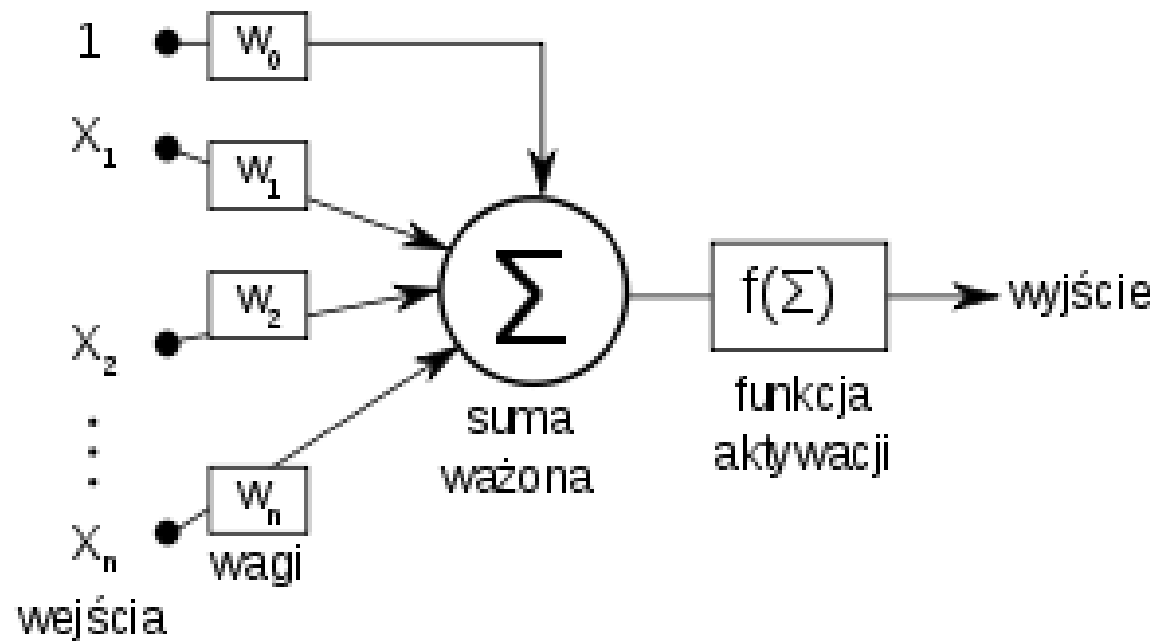
Podobieństwa między sztucznym neuronem a neuronem naturalnym:

- **wejścia** są odpowiednikami **dendrytów**
- **wagi cyfrowe** (na rysunku oznaczone przez prostokąty z literą w) są odpowiednikami **modyfikacji dokonywanych na impulsach poprzez synapsy**
- **blok sumy ważonej i blok funkcji aktywacji** jest odpowiednikiem **jądra**
- **wyjście** jest odpowiednikiem **aksonu**

Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- **Wprowadzenie do sztucznych sieci neuronowych**
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

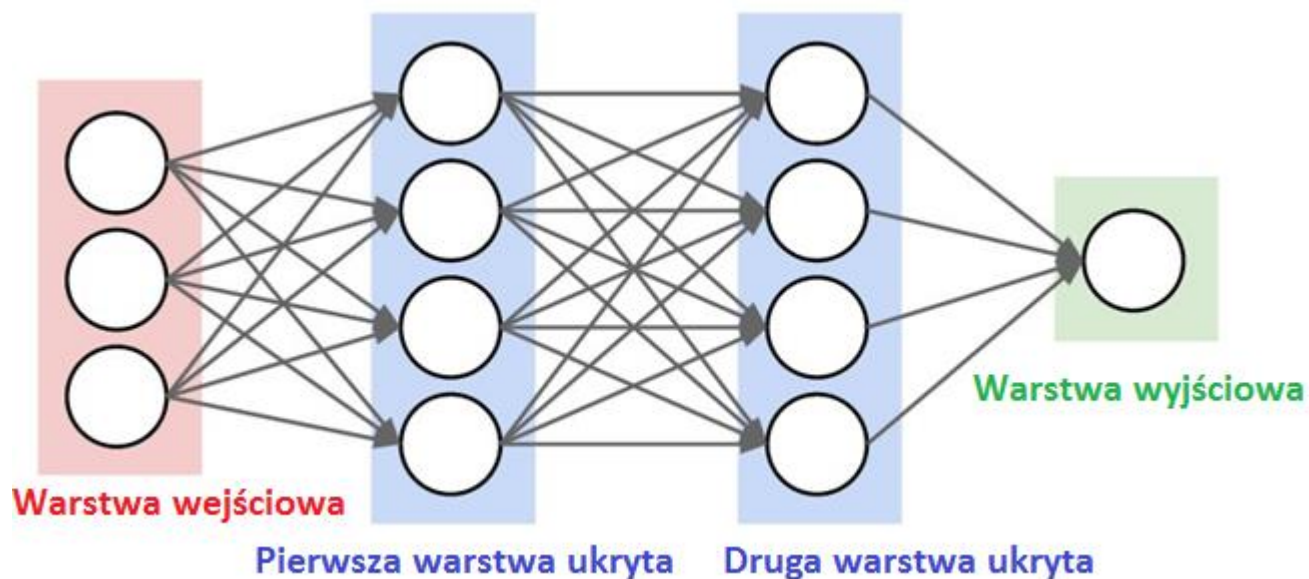
Działanie pojedynczego neuronu



Schemat budowy sztucznego neuronu [Źródło: Internet]

- Wartości na każdym z wejść ($1, x_1 \dots x_n$) są przeskalowywane przez wagę danego wejścia (w_0, w_1, \dots, w_n).
- Wszystkie otrzymane w pierwszym punkcie wyniki są sumowane w bloku sumy ważonej.
- Otrzymana w punkcie drugim wartość jest używana jako argument pewnej funkcji zwanej funkcją aktywacji. Zwykle jest to funkcja nieliniowa, np. funkcja sigmoidalna (tangens hiperboliczny). Wynik tej funkcji przekazywany jest na wyjście.

Sztuczna sieć neuronowa

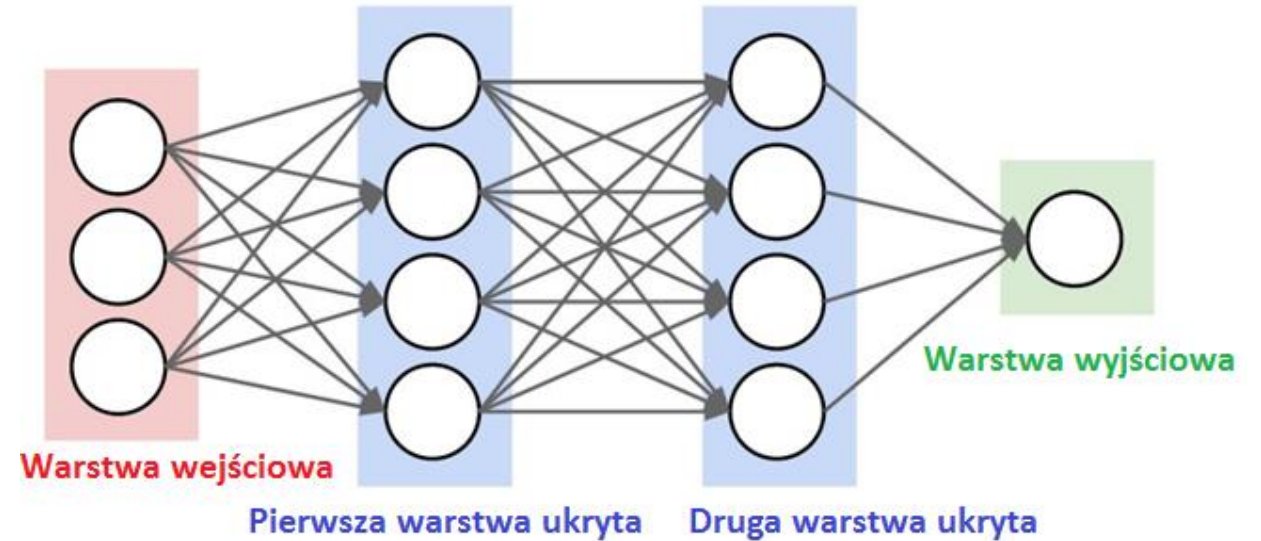
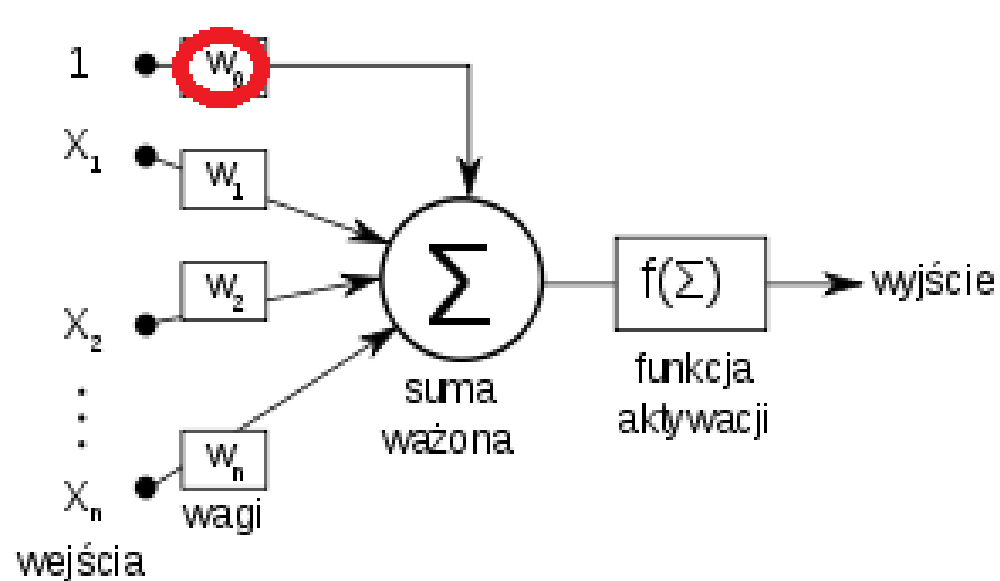


Schemat 3-warstwowej sieci neuronowej

[Źródło: Internet]

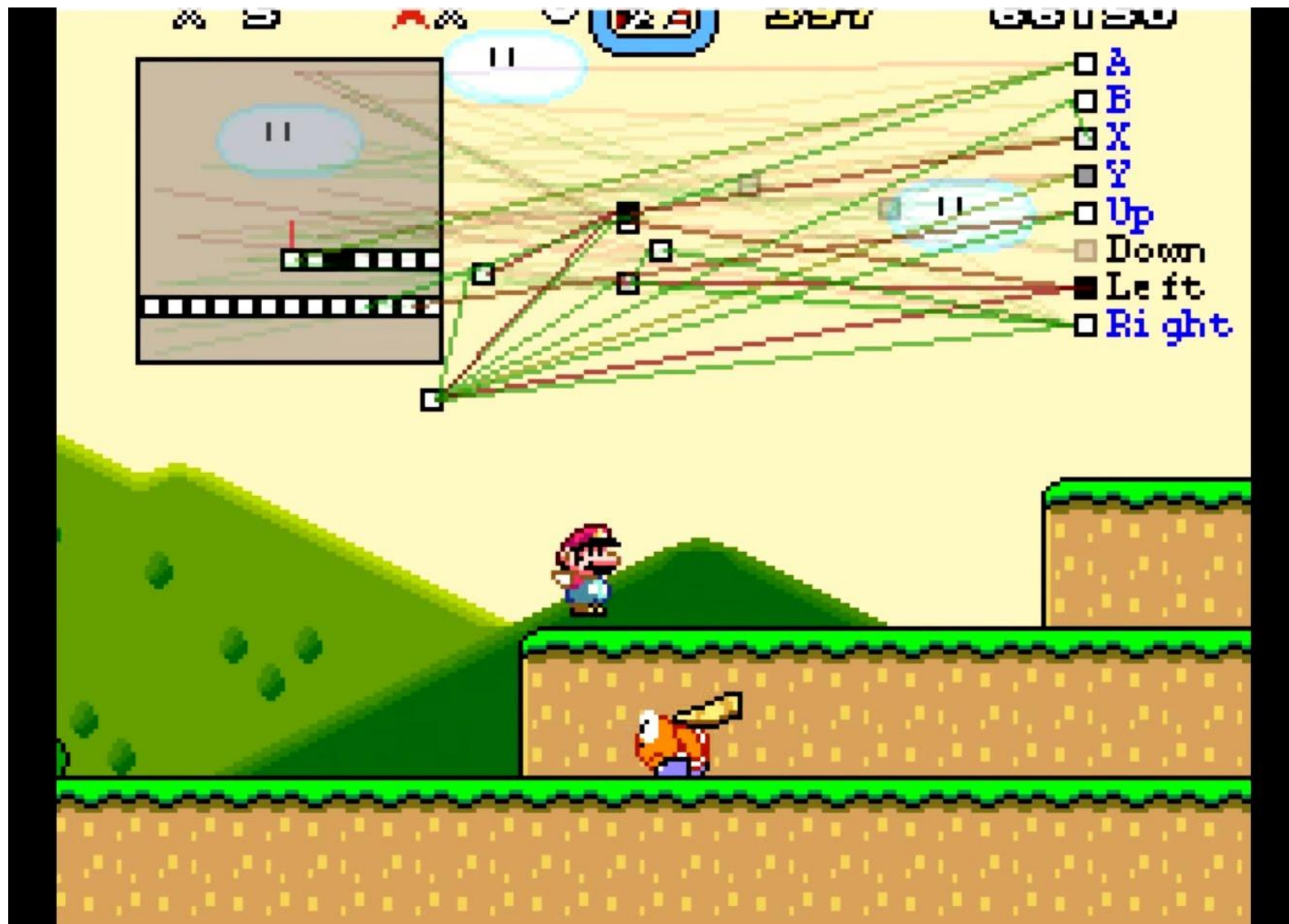
- Zaprezentowana sieć składa się z **4 warstw** (na rysunku zaznaczonych kolorowymi prostokątami): **warstwy wejściowej, pierwszej warstwy ukrytej, drugiej warstwy ukrytej oraz warstwy wyjściowej.**
- Każda z warstw (oprócz warstwy wejściowej) składa się z **neuronów.**
- Neurony w obrębie **jednej warstwy nie są ze sobą połączone**, natomiast każdy neuron połączony jest ze **wszystkimi neuronami z warstwy poprzedniej** (połączenia reprezentowane są przez strzałki) – warstwa w pełni połączona.

- Zaprezentowana sieć **składa się z 9 neuronów**: 4 w pierwszej warstwie ukrytej, 4 w drugiej warstwie ukrytej oraz 1 w warstwie wyjściowej.
- **Liczba wag wynosi 41**. Jest to liczba połączeń na schemacie: $3 \times 4 + 4 \times 4 + 4 \times 1 = 32$, zwiększona o liczbę neuronów: 9. Drugi składnik sumy wynika z faktu, że wejściem do każdego neuronu jest również pewna stała niezależna od danych wejściowych. Stała ta zaznaczona jest czerwonym kółkiem na schemacie z lewej strony. Nie została ona uwzględniona na schemacie znajdującym się po prawej stronie, ponieważ nie dotyczy ona danych wejściowych.



Uczenie sztucznej sieci neuronowej

- Aby sieć neuronowa funkcjonowała prawidłowo należy dobrać dla niej **odpowiednie wartości wag**.
- **Funkcja kosztu** – funkcja określająca jakość dobranych wag, wyznacza ona **błąd między wyjściem sieci a tym, co na tym wyjściu powinno się znaleźć**.
- Problem nauki sieci sprowadza się do problemu **znalezienia minimum funkcji kosztu**.
- Najpopularniejszym obecnie sposobem nauki sieci neuronowych jest **metoda gradientu prostego**. Polega ona na wyznaczeniu gradientu funkcji kosztu i odpowiednim uaktualnieniu wartości wag na podstawie wyznaczonego gradientu.



Tom Murphy – MarI/O [Źródło: Internet]

Agenda

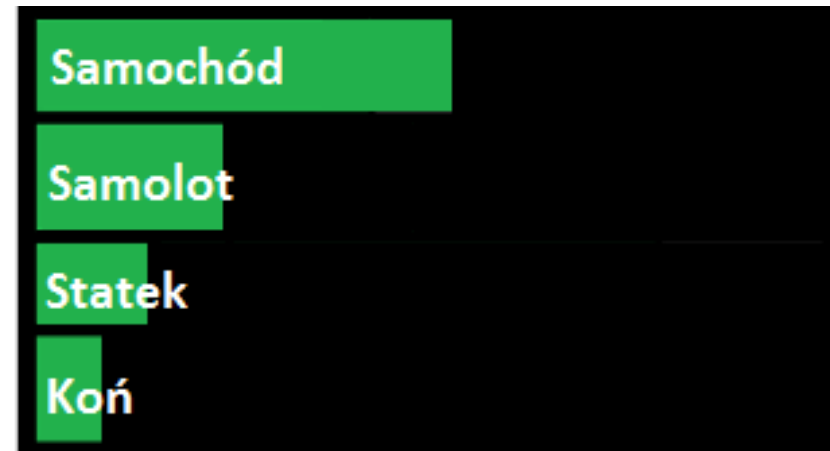
- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- ***Działanie i architektura konwolucyjnych sieci neuronowych***
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

Konwolucyjne sieci neuronowe

- W przypadku zwykłych sieci neuronowych nie istnieją ograniczenia nałożone na dane wejściowe.
- W przypadku konwolucyjnych sieci neuronowych istnieje jawne założenie: na wejście sieci podawany jest obraz, natomiast wyjściem jest wektor określający stopień przynależności danego obrazu do poszczególnych klas.



Przykładowe dane wejściowe [Źródło: Internet]



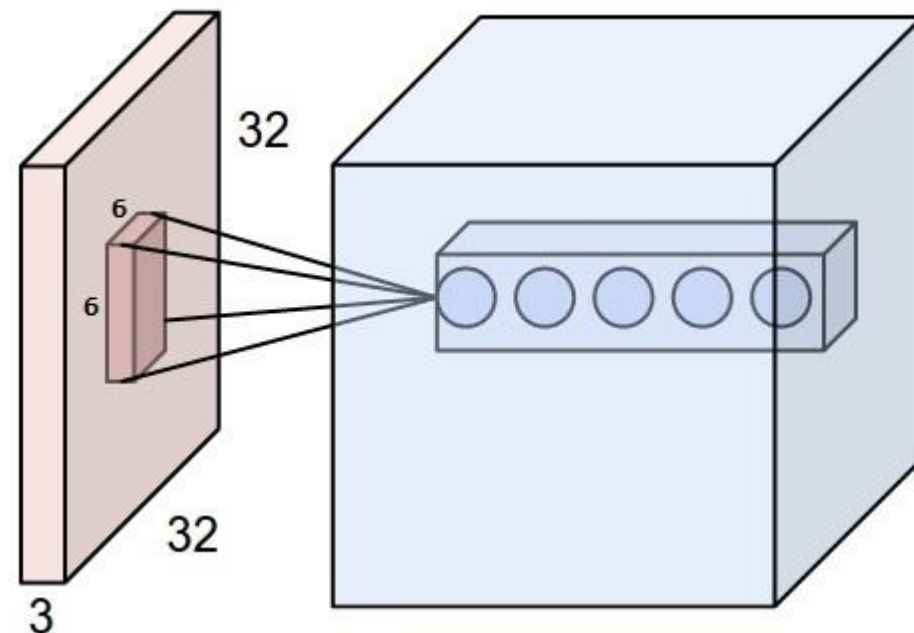
Przykładowy wektor wyjściowy

Sieci konwolucyjne a zwykłe – różnice

- Konwolucyjne sieci neuronowe składają się z **większej ilości warstw** niż zwykłe sieci (zwykłe sieci składają się najczęściej z 3 - 4 warstw, przy czym konwolucyjne sieci mają tych warstw od 10 do 20).
- W konwolucyjnych sieciach neuronowych dane wejściowe są **trójwymiarowymi wolumenami** (mają długość, wysokość oraz głębokość). Również neurony w poszczególnych warstwach tworzą **wolumeny**.
- Zwykłe sieci neuronowe składają się z warstw tego samego typu (warstw w pełni połączonych). Konwolucyjne sieci neuronowe składają się z **różnych typów warstw**:
 - **warstwy wejściowej** (INPUT) zawierającej surowy, jeszcze nie przetworzony obraz (czyli wolumen o wymiarach: długość obrazka, wysokość obrazka, 3)
 - **Warstw konwolucyjnych** (CONV)
 - **Warstw rektyfikowanej jednostki liniowej** (RELU), które dla każdego elementu wejściowego x wyznaczają wartość funkcji $\max(0, x)$
 - **Warstw próbkujących** (POOL), które przeprowadzają operację próbkowania długości oraz wysokości obrazu (głębokość pozostaje bez zmian)
 - **Warstw w pełni połączonych**

Warstwa konwolucyjna – lokalność połączeń

- W przypadku kiedy danymi wejściowymi są obrazy, łączenie wszystkich neuronów z danego wolumenu ze wszystkimi wyjściami wolumenu poprzedniego jest niepraktyczne.
- Połączenia neuronów są **lokalne w przestrzeni** (długość i wysokość), natomiast **pełne jeśli chodzi głębokość wolumenu**.
- Hiperparametry związane z lokalnością połączeń:
 - **Wielkość filtru** (F) określa wymiary przestrzeni z jaką połączony jest pojedynczy neuron

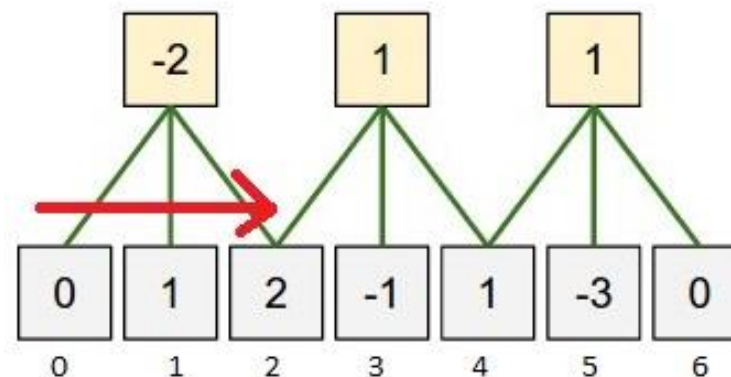


Schemat warstwy konwolucyjnej o $F = 6$
[Źródło: Internet]

Warstwa konwolucyjna – lokalność połączeń

- **Wielkość kroku** (S) określa odległość między lokalnymi obszarami obserwowanymi przez kolejne neurony
- **Wyrównanie** (P) określa o ile należy zwiększyć przestrzeń wolumenu wejściowego (długość i wysokość), wstawiając w nowo powstałe miejsca wartości 0.

Wolumen o wymiarach: F, F, g (gdzie g to głębokość wolumenu podawanego na wejście warstwy konwolucyjnej), trzymający wartości wag odpowiadające poszczególnym połączeniom między neuronem a jego lokalnym obszarem nazywany jest **filtrem**.



Połączenia między neuronami dla $S = 2$

[Źródło: Internet]

0	0	0	0
0	3	5	0
0	7	-1	0
0	0	0	0

Wolumen wejściowy o wymiarach 2×2 po zastosowaniu $P = 1$

Warstwa konwolucyjna – współdzielenie parametrów

- Współdzielenie parametrów opiera się na założeniu, że **jeśli jakaś cecha obrazu jest użyteczna do wyznaczenia w punkcie przestrzeni wolumenu, to warto jej również szukać w dowolnym innym punkcie tej przestrzeni.**
- W przypadku gdy głębokość wolumenu z neuronami wynosi K , otrzymujemy K przestrzeni. W obrębie danej przestrzeni szukane są te same cechy obrazu, jednak każda z przestrzeni szuka innych cech.

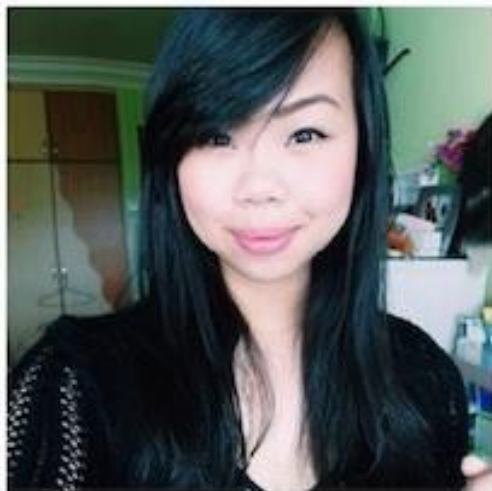
Niech: Warstwa konwolucyjna o $F = 11$ to wolumen o wymiarach: 55, 55, 96. Jaka jest liczba wszystkich wag?

- **Bez wykorzystania cechy współdzielenia parametrów:**
 $11 \times 11 \times 3 \times 55 \times 55 \times 96 = 105415200$
- **Z wykorzystaniem cechy współdzielenia parametrów:**
 $11 \times 11 \times 3 \times 96 = 34848$

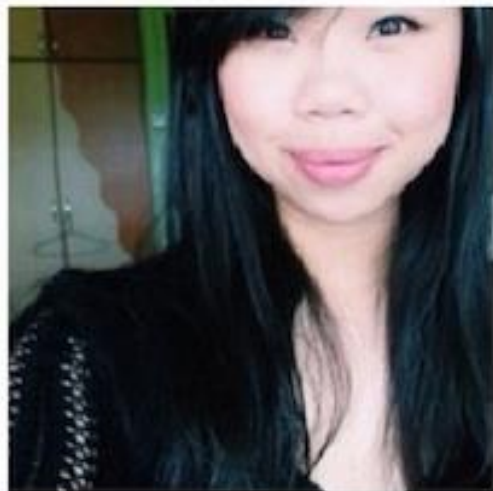


Andrej Karpathy – znajdowanie najlepszego wycinka w zdjęciu

score 53.1



score 67.3



score 44.5



score 62.8



score 66.5



score 69.6



score 52.0



score 56.3



score 45.0



score 58.3



score 45.4



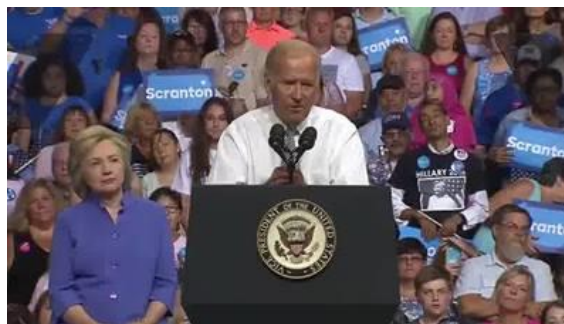
score 52.3



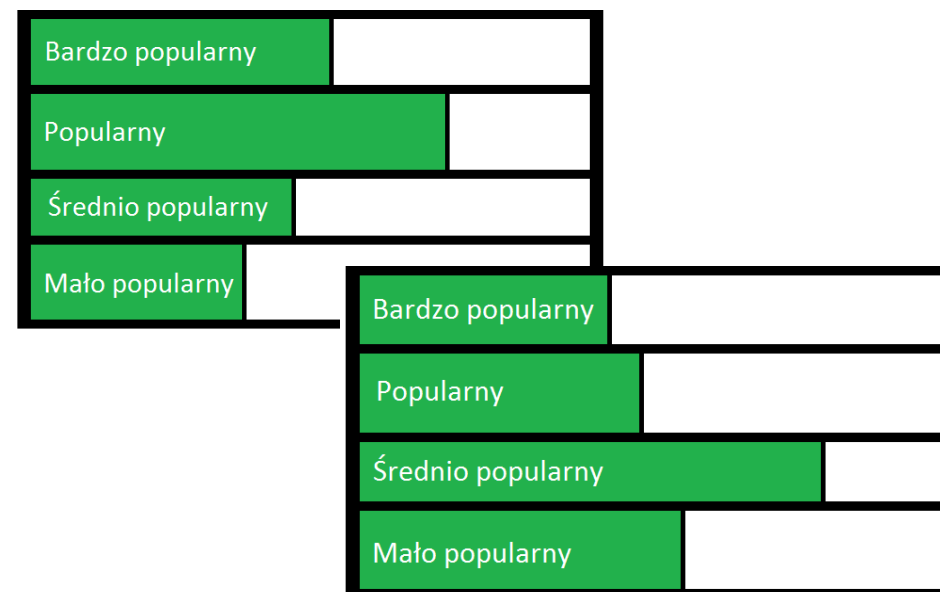
Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- ***Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików***
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

Analizowanie miniaturki filmów



Przykładowe miniaturki filmików z Facebooka



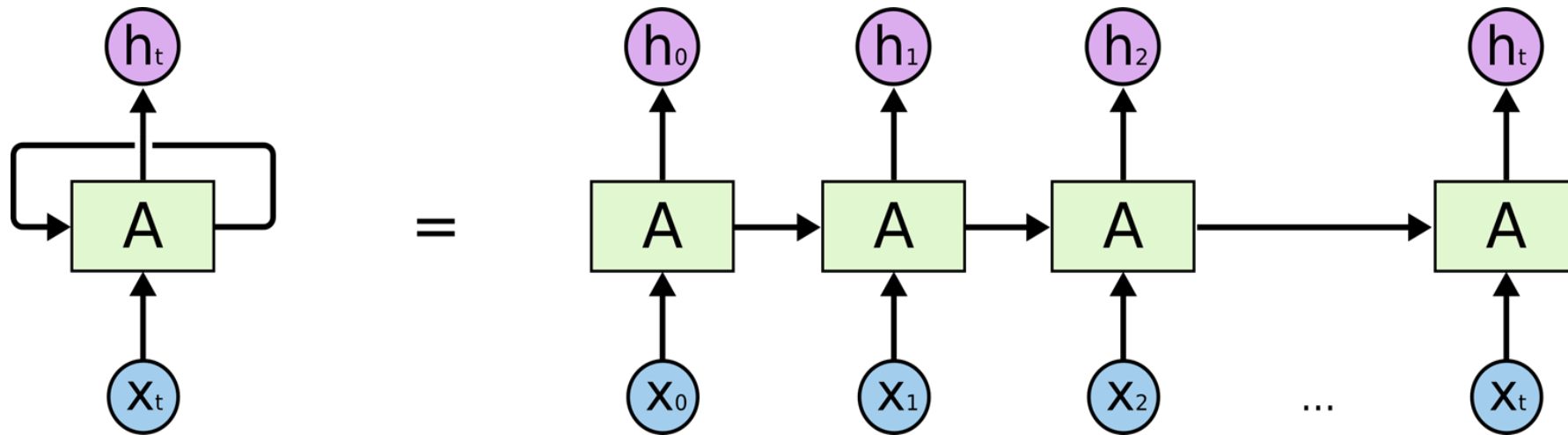
Przykładowe wektory określające stopień przynależności miniaturki do poszczególnych klas

Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- ***Działanie i architektura rekurencyjnych sieci neuronowych***
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

RNN - Rekurencyjne sieci neuronowe

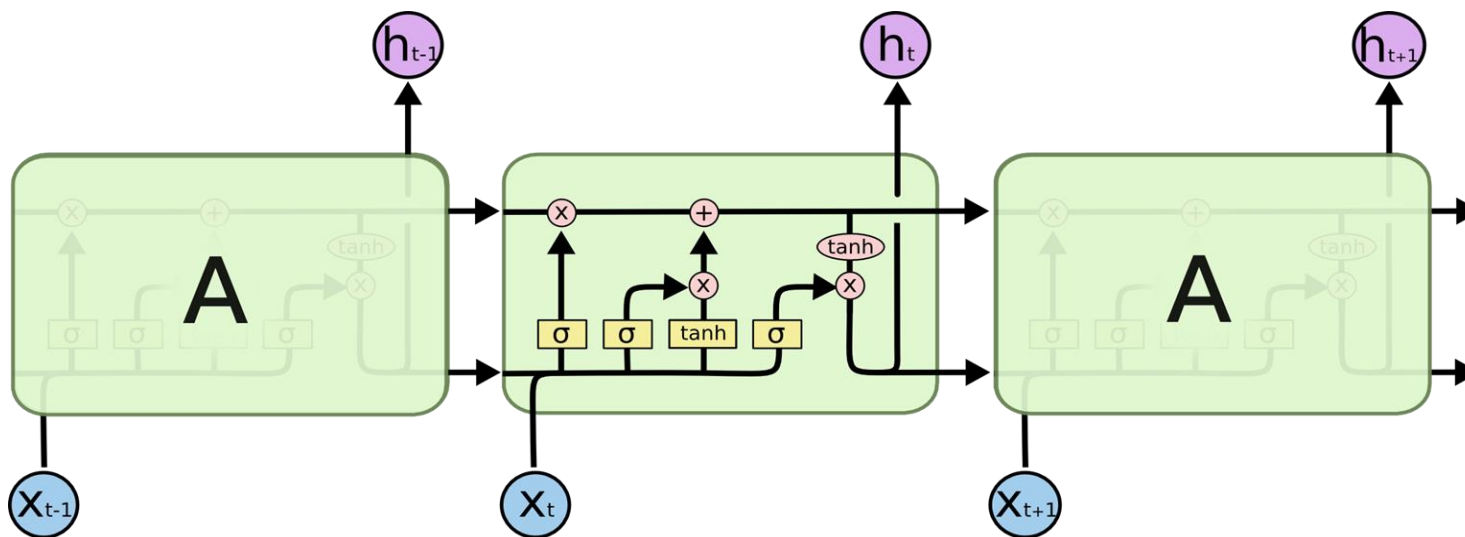
- Rekurencyjne sieci neuronowe mogą być zilustrowane jako wiele kopii tej samej sieci, z których wyjście każdej kolejnej zależy od wyjść wszystkich poprzednich.
- Architektura stworzona do przetwarzania sekwencji.
- Składają się z jednej warstwy wykonującej operacje tangens hiperboliczny.



Schemat przedstawiający rekurencyjną sieć neuronową [Źródło: Internet]

LSTM – komórki pamięci krótko-długo trwałej

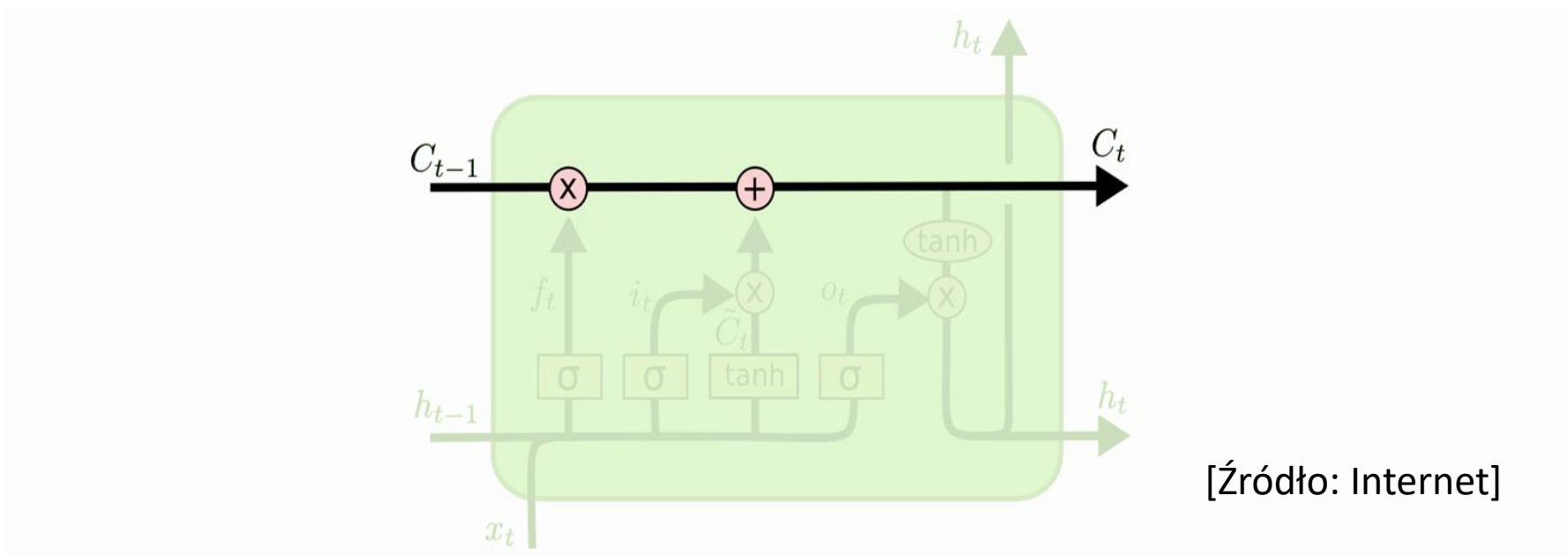
- Podobnie jak RNN, LSTM mają budowę łańcuchową.
- W przeciwieństwie do RNN, składają się one z 4 warstw: 3 bramek sigmoidalnych oraz bramki wykonującej operację tangens hiperboliczny.
- W praktyce sprawdzają się znacznie lepiej niż RNN.



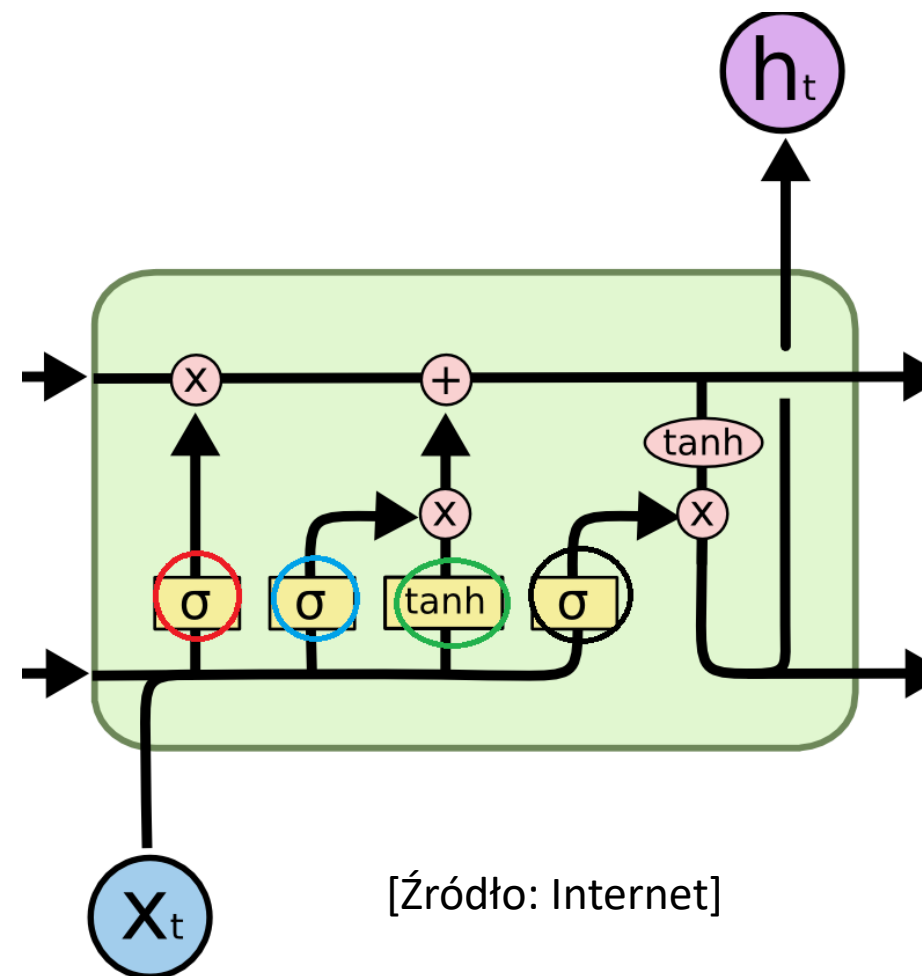
Schemat przedstawiający komórkę LSTM [Źródło: Internet]

Sposób działania LSTM

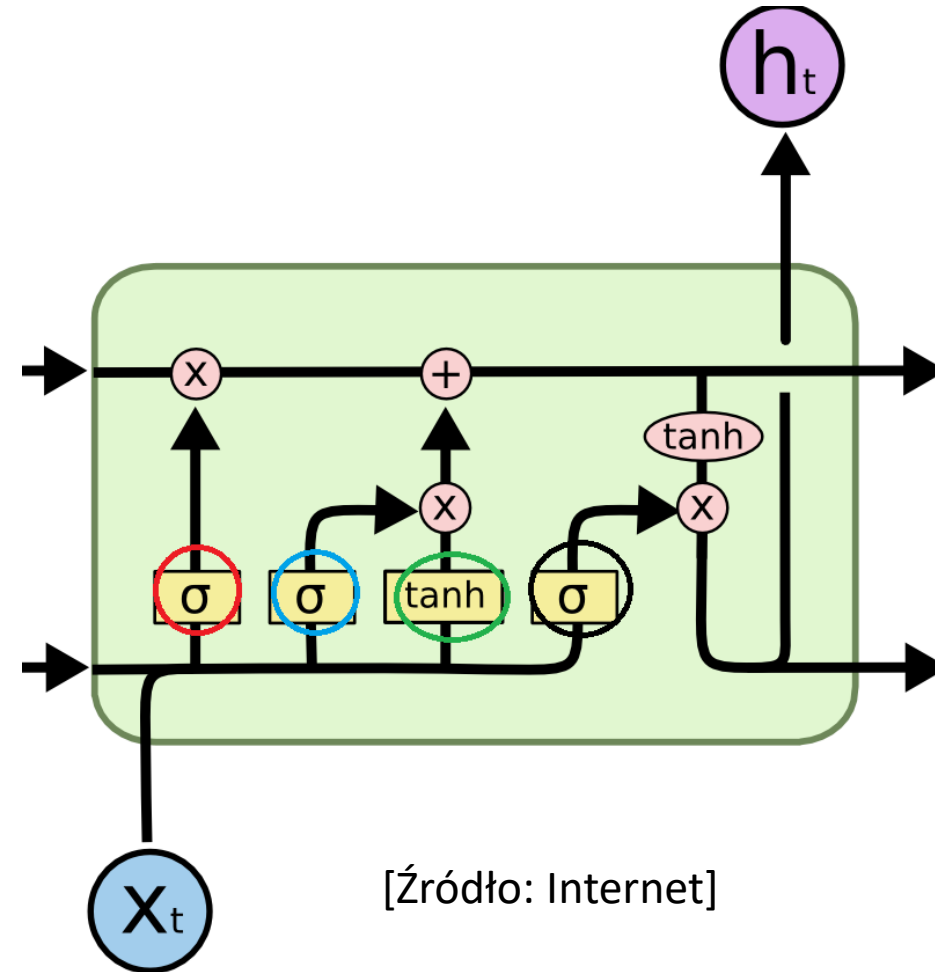
- Wyróżniona na obrazku linia to linia określająca stan komórki.
- LSTM ma możliwość modyfikowania informacji o swoim stanie poprzez usuwanie istniejących już informacji oraz dodawanie nowych.
- Jest to kluczowa umiejętność odróżniająca LSTM od RNN.



- Usuwanie istniejących informacji ze stanu komórki LSTM :
 - Bramka sigmoidalna nazywana **bramką zapominającą** (oznaczona na obrazku **czzerwonym kółkiem**) decyduje, jaką część stanu komórki należy zapomnieć.
- Dodawanie nowych informacji do stanu komórki LSTM:
 - W pierwszym kroku bramka sigmoidalna nazywana **bramką wejściową** (oznaczona na obrazku **niebieskim kółkiem**) decyduje, jaką część stanu komórki należy zmodyfikować.
 - W drugim kroku **warstwa tangensa hiperbolicznego** (oznaczona na obrazku **zielonym kółkiem**) tworzy wektor wartości, którymi uaktualniona zostanie wybrana w pierwszym kroku część stanu komórki.



- Wyznaczanie wyjścia komórki LSTM:
 - W pierwszym kroku ostatnia z **bramek sigmoidalnych** (oznaczona na obrazku **czarnym kółkiem**) decyduje, jaka część stanu komórki podana zostanie na wyjście.
 - W drugim kroku wartość stanu komórki jest poddawana transformacji tangens hiperboliczny, a jej wynik jest przeskalowywany przez wartość otrzymaną z bramki sigmoidalnej w pierwszym kroku.

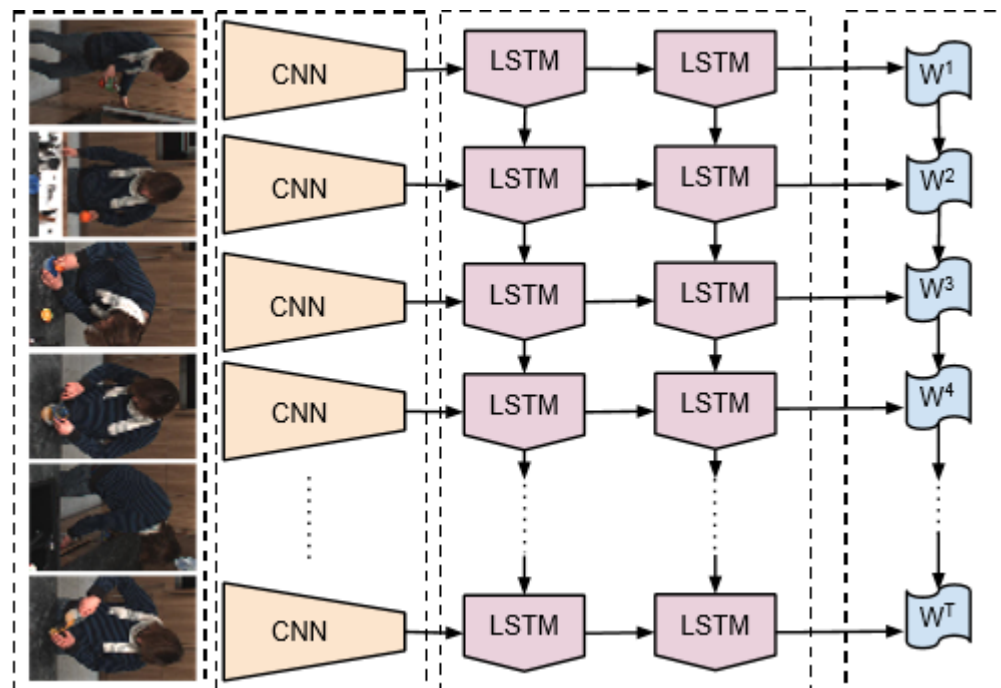


Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- **Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą**
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- Podsumowanie

LRCN – Konwolucyjne sieci rekurencyjne z pamięcią długotrwałą

- Powstały w wyniku połączenia konwolucyjnych sieci neuronowych z komórkami pamięci krótko-długo trwałej (LSTM).
- Kolejno wydobywane wizualne cechy obrazu podawane są jako sekwencja na wejście komórek LSTM.

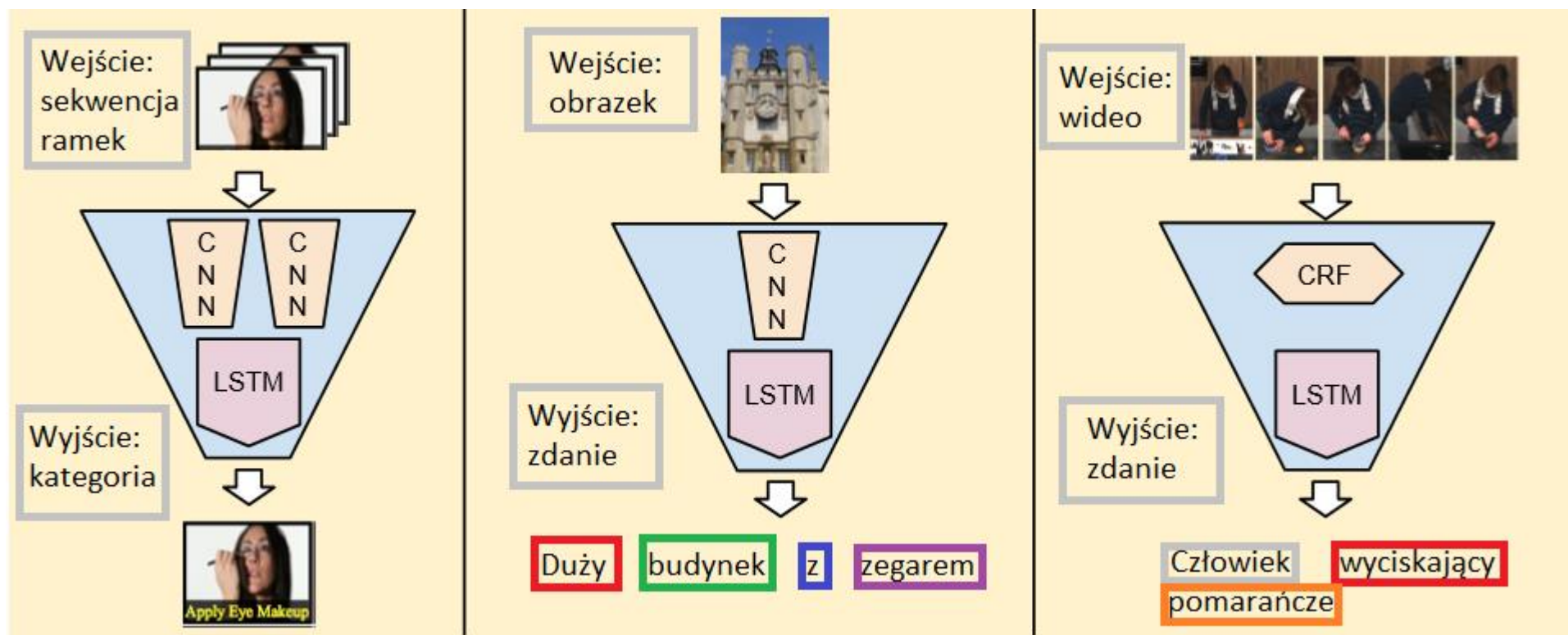


Schemat przedstawiający architekturę LRCN

[Źródło: publikacja Long-term Recurrent Convolutional Networks for Visual Recognition and Description]

Sieci LRCN powinny być wykorzystywane w przypadku gdy:

- **Wejście jest sekwencją pewnych obrazów, natomiast wyjście to wartość skalarna**
- **Wejście jest pojedynczym obrazem, natomiast wyjście to pewna sekwencja**
- **Wejście jest sekwencją pewnych obrazów, a wyjście inną sekwencją**



[Źródło: publikacja Long-term Recurrent Convolutional Networks for Visual Recognition and Description]

Agenda

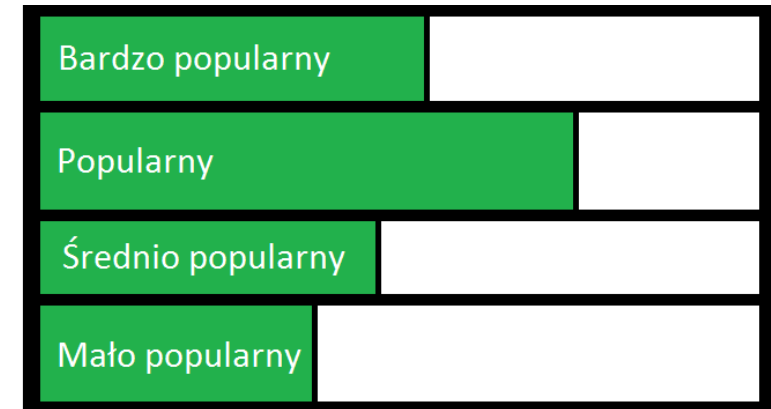
- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- *Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików*
- Podsumowanie

Wejście: sekwencja pierwszych klatek filmu

Wyjście: wartość skalarna określająca popularność danej sekwencji



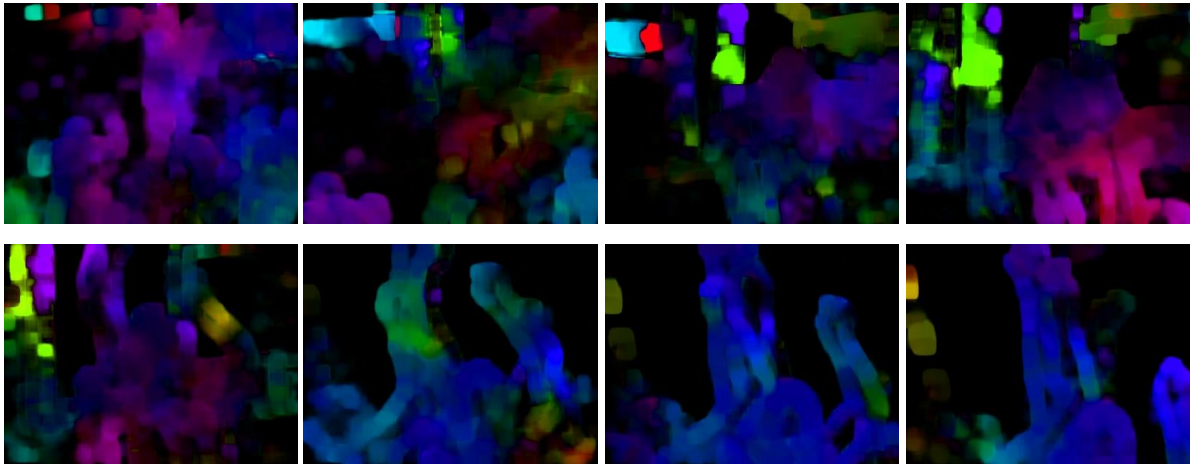
Przykładowa sekwencja pierwszych ramek filmiku z Facebooka



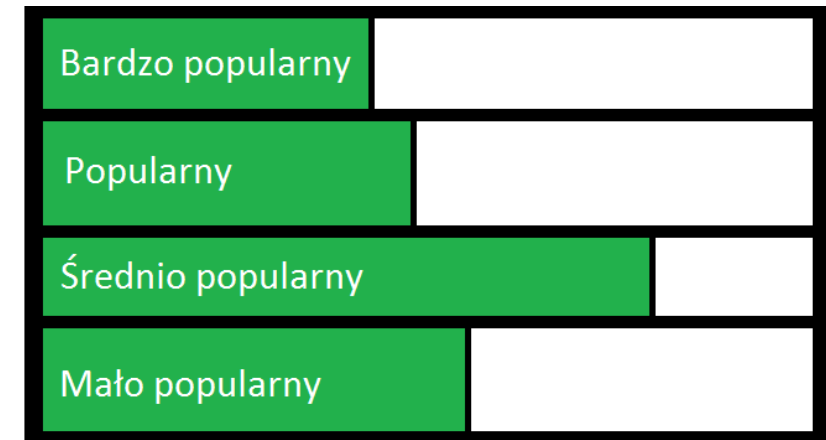
Wektor określający stopień przynależności danej sekwencji do poszczególnych klas popularności

Wejście: sekwencja obrazów przedstawiających przepływ optyczny między kolejnymi klatkami

Wyjście: wartość skalarna określająca popularność danej sekwencji



Przykładowa sekwencja obrazów przedstawiających przepływ optyczny pomiędzy kolejnymi klatkami filmiku z Facebooka



Wektor określający stopień przynależności danej sekwencji do poszczególnych klas popularności

Agenda

- Sztuczne sieci neuronowe – inspiracje biologiczne
- Wprowadzenie do sztucznych sieci neuronowych
- Działanie i architektura konwolucyjnych sieci neuronowych
- Zastosowanie konwolucyjnych sieci neuronowych w kontekście przewidywania popularności filmików
- Działanie i architektura rekurencyjnych sieci neuronowych
- Architektura LRCN – konwolucyjne sieci rekurencyjne z pamięcią długotrwałą
- Zastosowanie architektury LRCN w kontekście przewidywania popularności filmików
- *Podsumowanie*

Podsumowanie

- Sztuczne sieci neuronowe to narzędzie o potężnej mocy. Jego rezultaty potrafią niejednokrotnie dziwić oraz bawić.
- Chociaż sztuczne sieci neuronowe inspirowane są sieciami neuronowymi znajdującymi się w naszych mózgach, to jednak dużo brakuje im do tego, aby działały tak dobrze jak ich naturalne odpowiedniki. Jednak badania nad sieciami neuronowymi wciąż trwają.
- Jest wiele rodzajów sztucznych sieci neuronowych. Nie ma architektury uniwersalnej, jednak do każdego zadania da się dopasować odpowiednią dla niego architekturę.
- Chociaż sieci te różnią się budową, to ogólna idea działania jest w każdym przypadku taka sama. Pozwala to na łączenie różnego rodzaju sieci, co mogliśmy zaobserwować w przypadku architektury LRCN.

Bibliografia

- Jeffrey Donahue, Lisa Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, Trevor Darrell: *Long-term Recurrent Convolutional Networks for Visual Recognition and Description*
- Andrej Karpathy blog: *Convolutional Neural Networks for Visual Recognition*
- Andrej Karpathy blog: *What a Deep Neural Network thinks about your #selfie?*
- Lars Eidnes blog: *Auto – Generating Clickbait With Recurrent Neural Networks*
- Christopher Olah blog: *Understanding LSTM Networks*

Dziękuję za uwagę

Czy mają Państwo jakieś pytania?