

ning table

laptop

vase

bottle

wine glass

chair

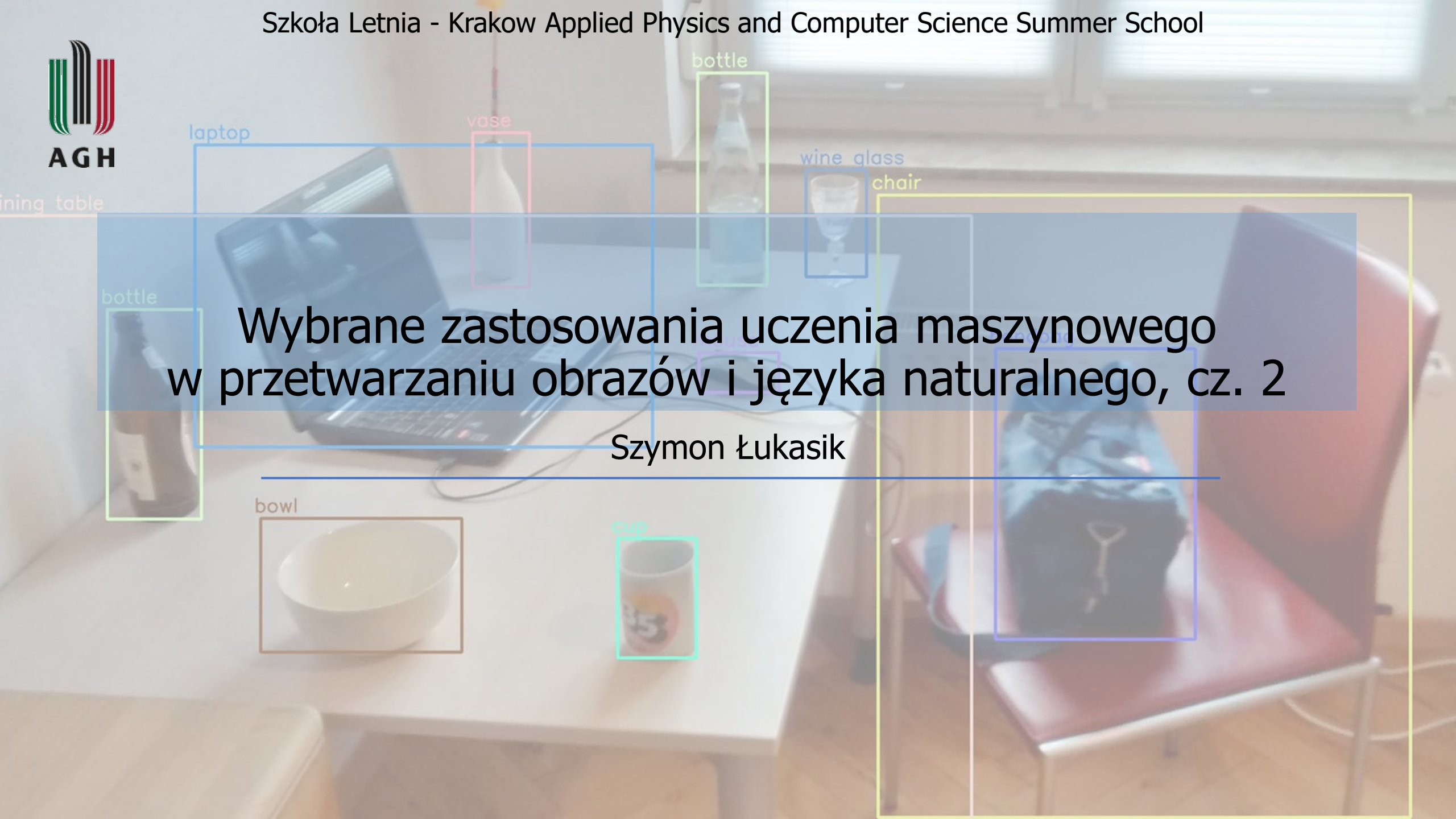
bottle

Wybrane zastosowania uczenia maszynowego w przetwarzaniu obrazów i języka naturalnego, cz. 2

Szymon Łukasik

bowl

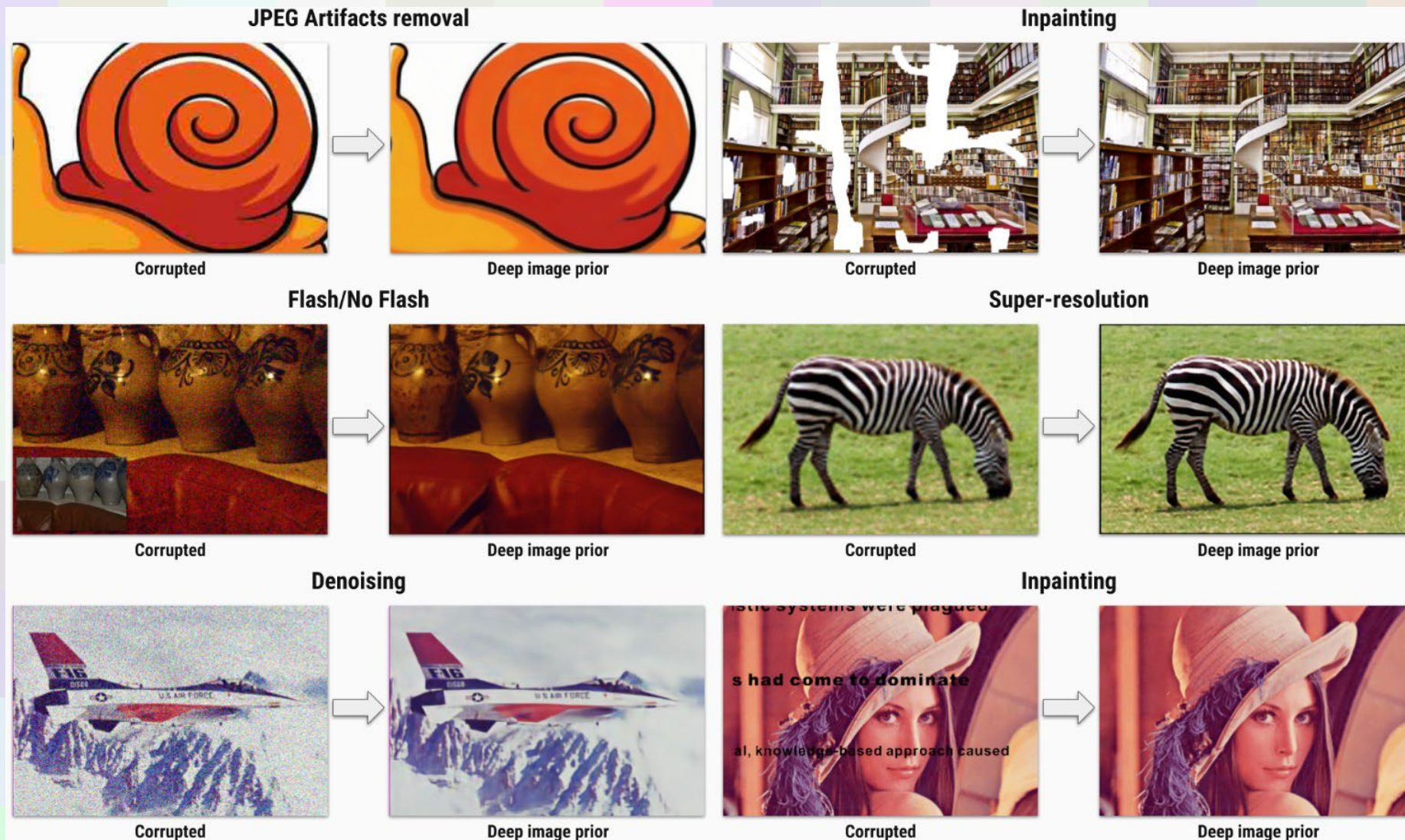
cup



Wprowadzenie

- Dynamiczny rozwój sztucznej inteligencji – w szczególności wielowarstwowych sieci neuronowych i związanego z nimi tzw. uczenia głębokiego (ang. Deep Learning) umożliwia zastosowanie podejścia bardziej elastycznego, skupionego w mniejszym stopniu na „feature engineering” (umiejętnym doborze cech)
- W ramach wykładu skupimy się na wybranych zagadnieniach zastosowań AI w przetwarzaniu obrazów
- Trójka najpopularniejszych frameworków do Deep Learning to Tensorflow, Torch(&PyTorch) i Keras (o tym na końcu)

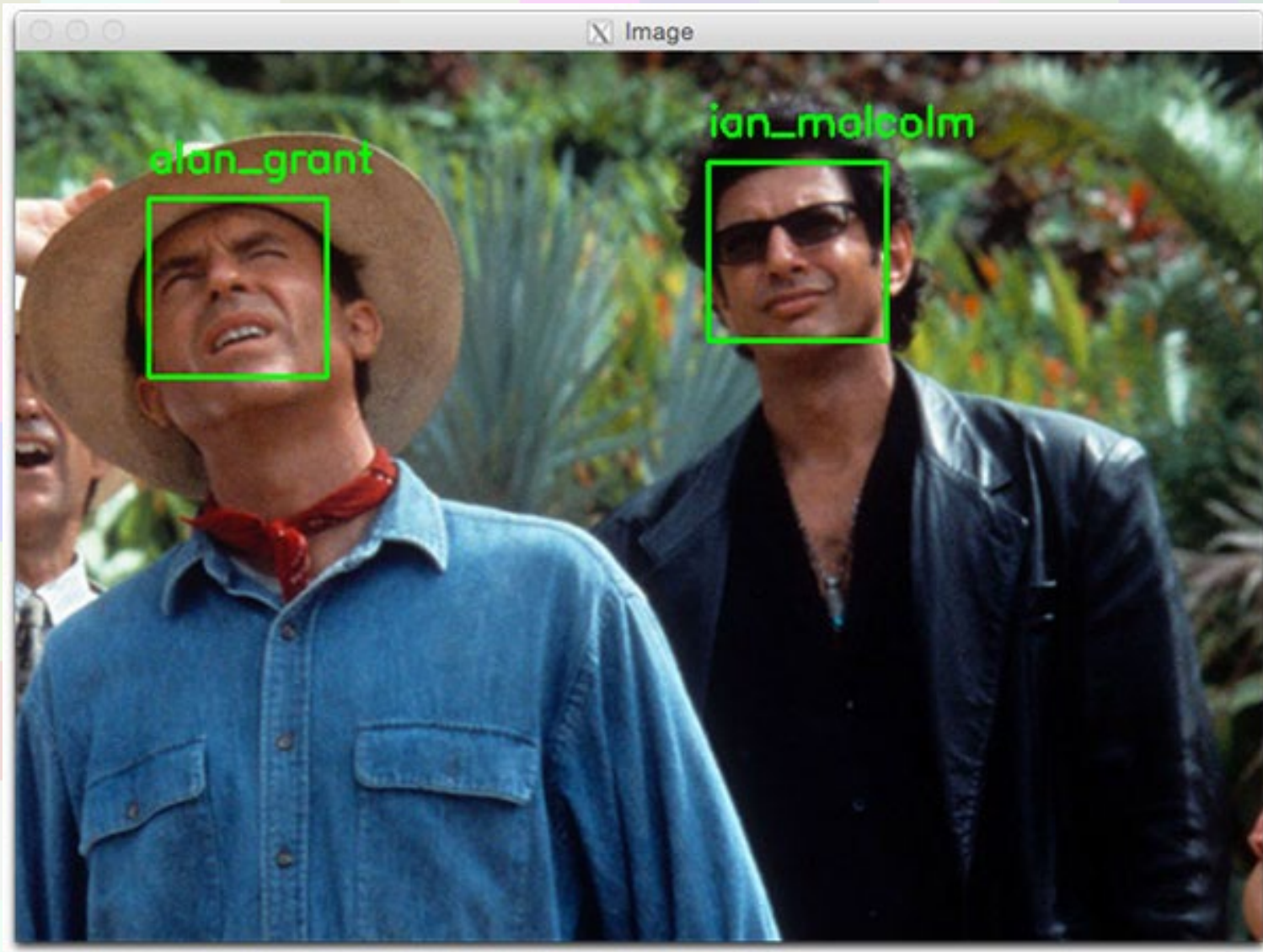
Obszary zastosowań AI w przetwarzaniu obrazów



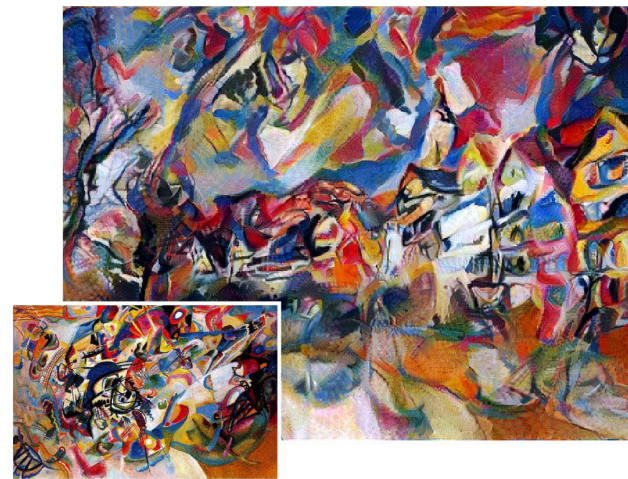
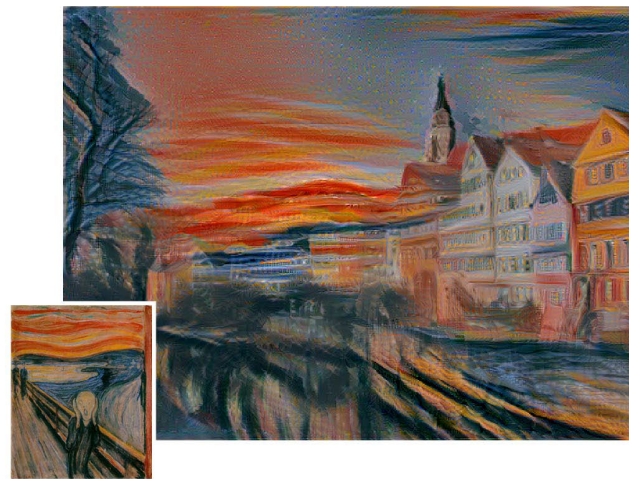
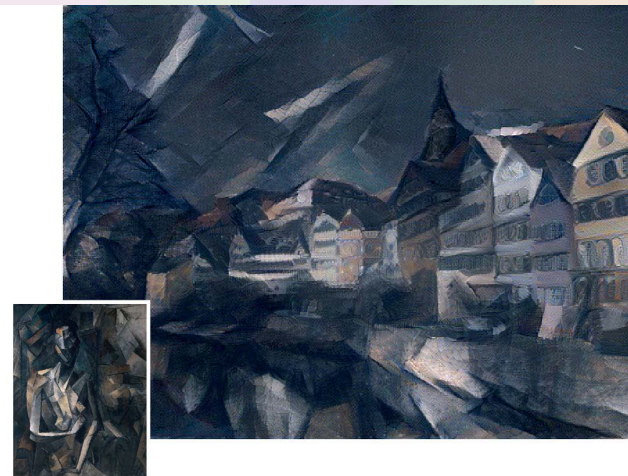
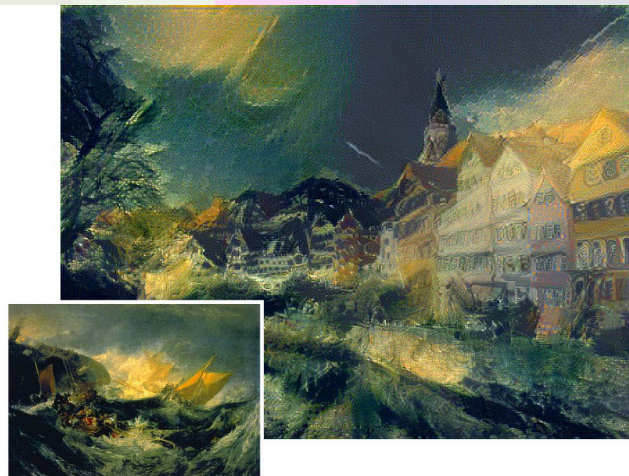
AI w przetwarzaniu obrazów – segmentacja semantyczna



AI w przetwarzaniu obrazów – klasyfikacja



AI w przetwarzaniu obrazów – style transfer



AI w przetwarzaniu obrazów – generowanie obrazów

INSTRUCTION: press +/- to adjust feature, toggle feature name to lock the feature

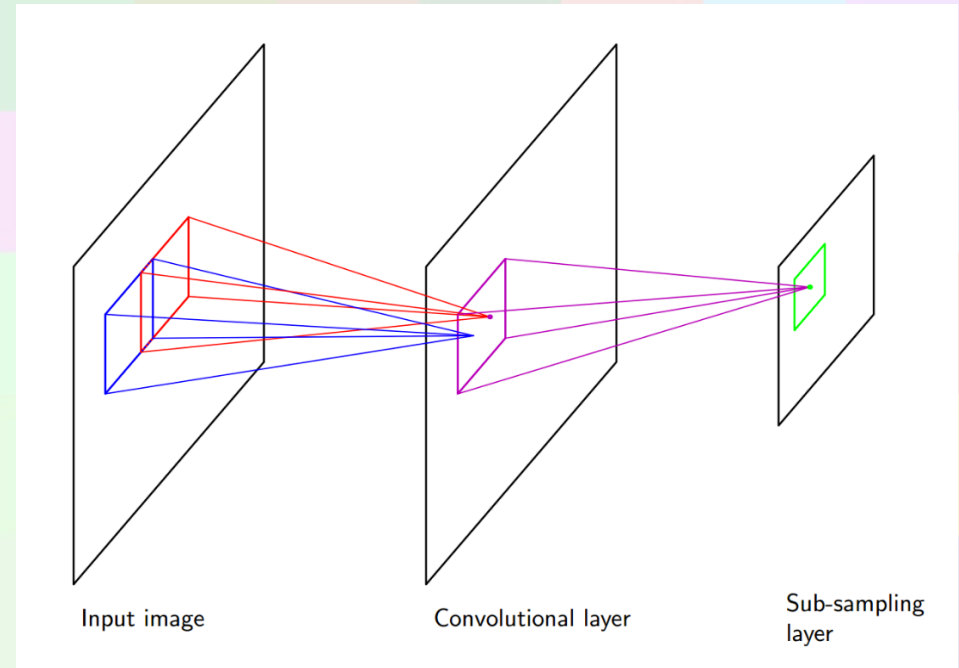


random face

Male	Age	Skin_Tone
- +	- +	- +
Bangs	Hairline	Bald
- +	- +	- +
Big_Nose	Pointy_Nose	Makeup
- +	- +	- +
Smiling	Mouth_Open	Wavy_Hair
- +	- +	- +
Beard	Goatee	Sideburns
- +	- +	- +
Blond_Hair	Black_Hair	Gray_Hair
- +	- +	- +
Eyeglasses	Earrings	Necktie
- +	- +	- +

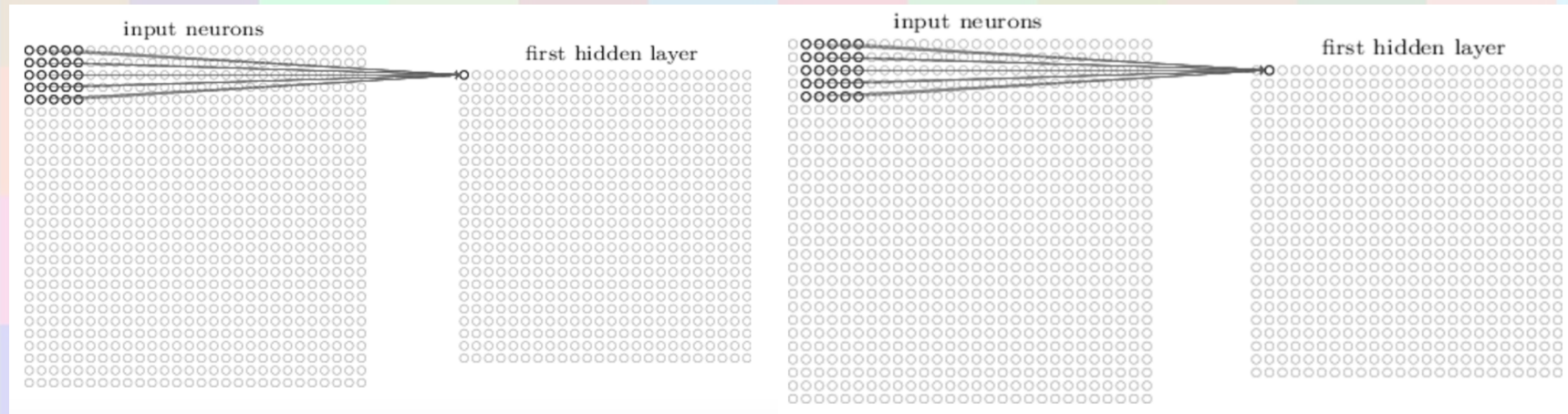
Na czym to wszystko pracuje?

- Większość z zastosowań DL w przetwarzaniu obrazów opiera się o sieci konwolucyjne (ang. Convolutional Neural Network)
- Z inżynierskiego punktu widzenia są to filtry/zestawy filtrów które uczymy w taki sposób by najlepiej wydobyć informacje z obrazu w ramach rozważanego zadania
- Warstwa konwolucyjna, inaczej mapa cech (ang. features map), składa się z jednostek analizujących niewielki obszar danych wejściowych, tj. na przykład 5x5 pikseli.
- Do każdej jednostki warstwy przypisana jest waga czyli cała warstwa analizująca 5x5 pikseli parametryzowana jest przez 25 wag oraz wartość wyrazu wolnego (ang. bias), w konsekwencji wszystkie jednostki wydobywają tą samą cechę.

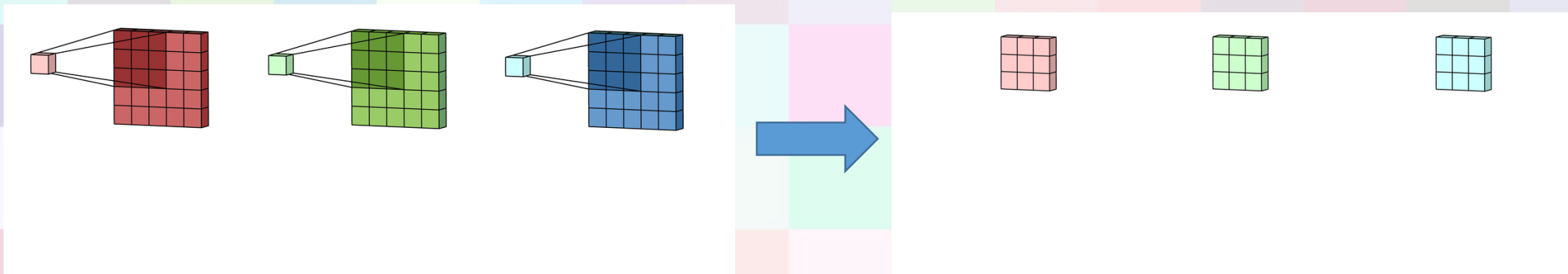


Warstwa konwolucji

Każdy element wyznaczony jest jako ważona suma wejść z dodaniem biasu

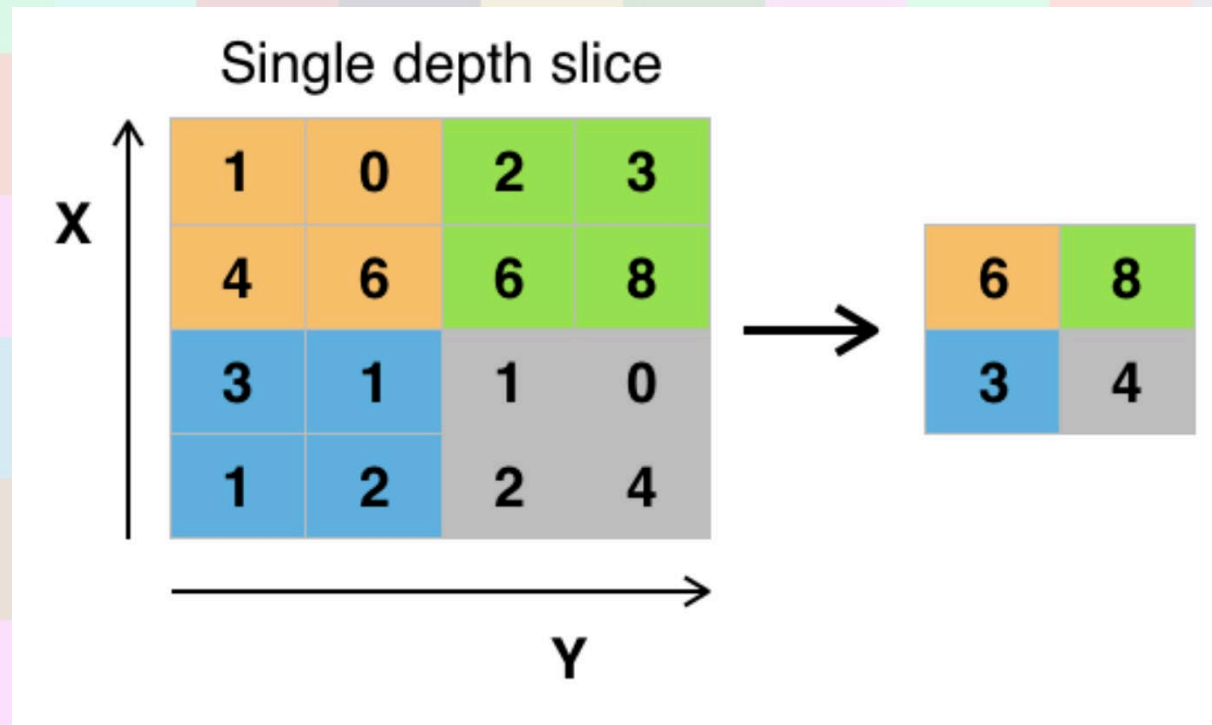


Dla obrazu kolorowego w praktyce rozważa się wstępnie każdą składową osobno – w następnym kroku agregując wynik (i dodając bias):



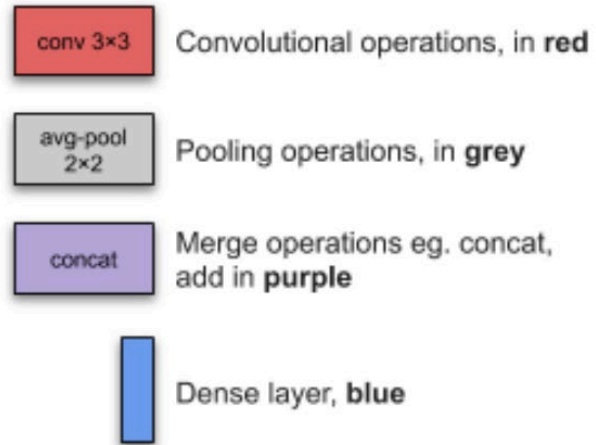
Pooling layer

Pooling layer ma za zadanie uogólnić informacje dotyczące danej cechy wydobyte przez obszar jednostek warstwy konwolucyjnej. Zwykle wykonuje operację maksimum



Pełna architektura – notacja

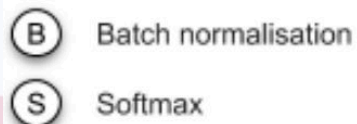
Layers



Activation Functions

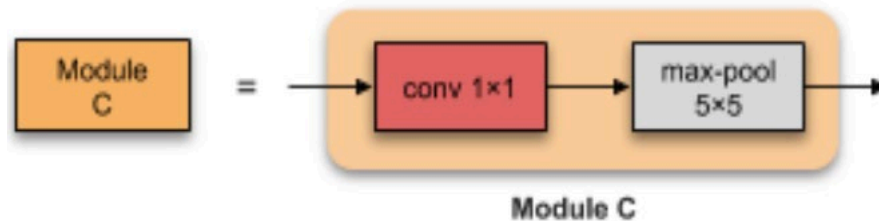
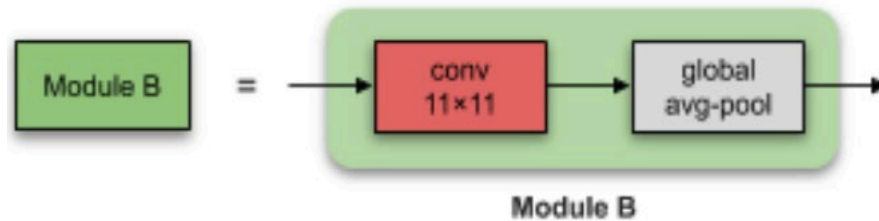
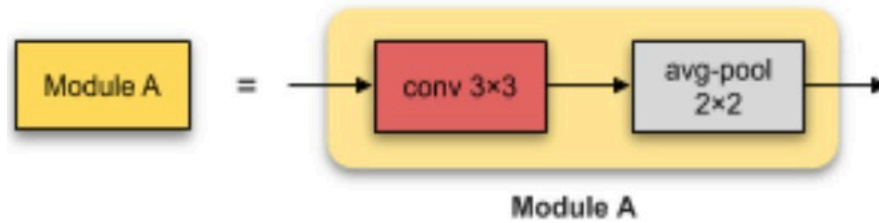


Other Functions

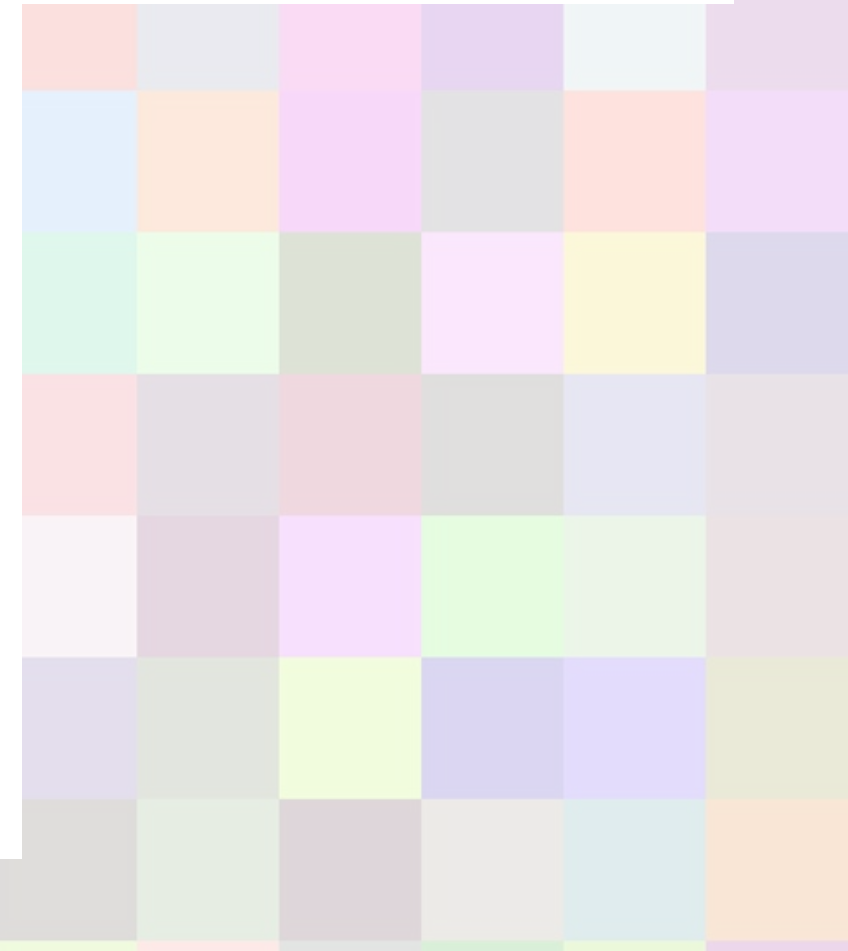
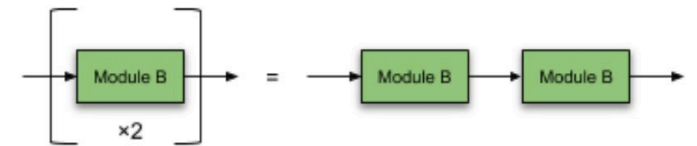


Modules/Blocks

Modules (groups of convolutional, pooling and merge operations), in **yellow, green, or orange**.
The operations that make up these modules will also be shown.

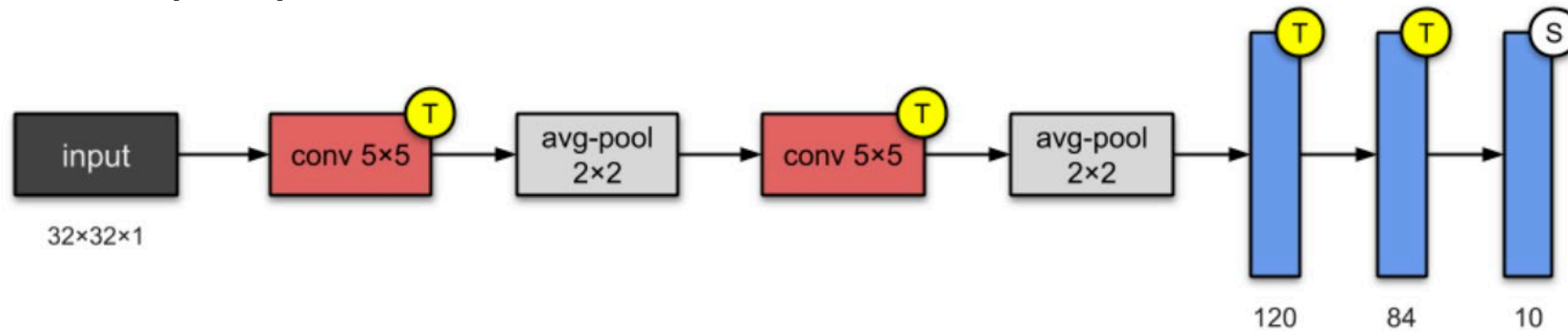


Repeated layers or modules/blocks

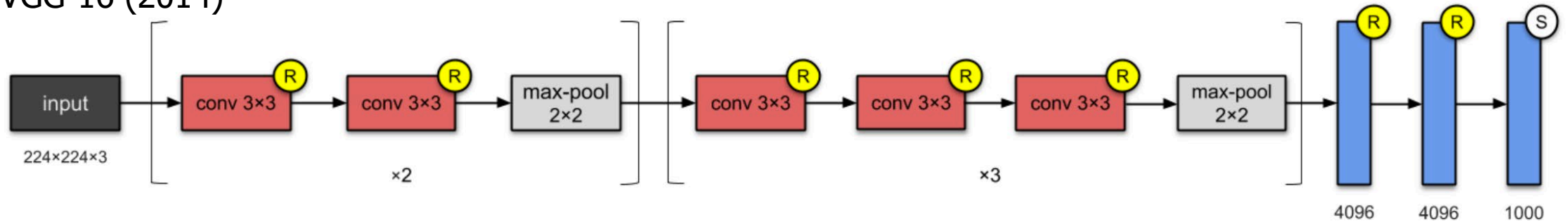


Przykłady typowych architektur

LeNet-5 (1998)



VGG-16 (2014)



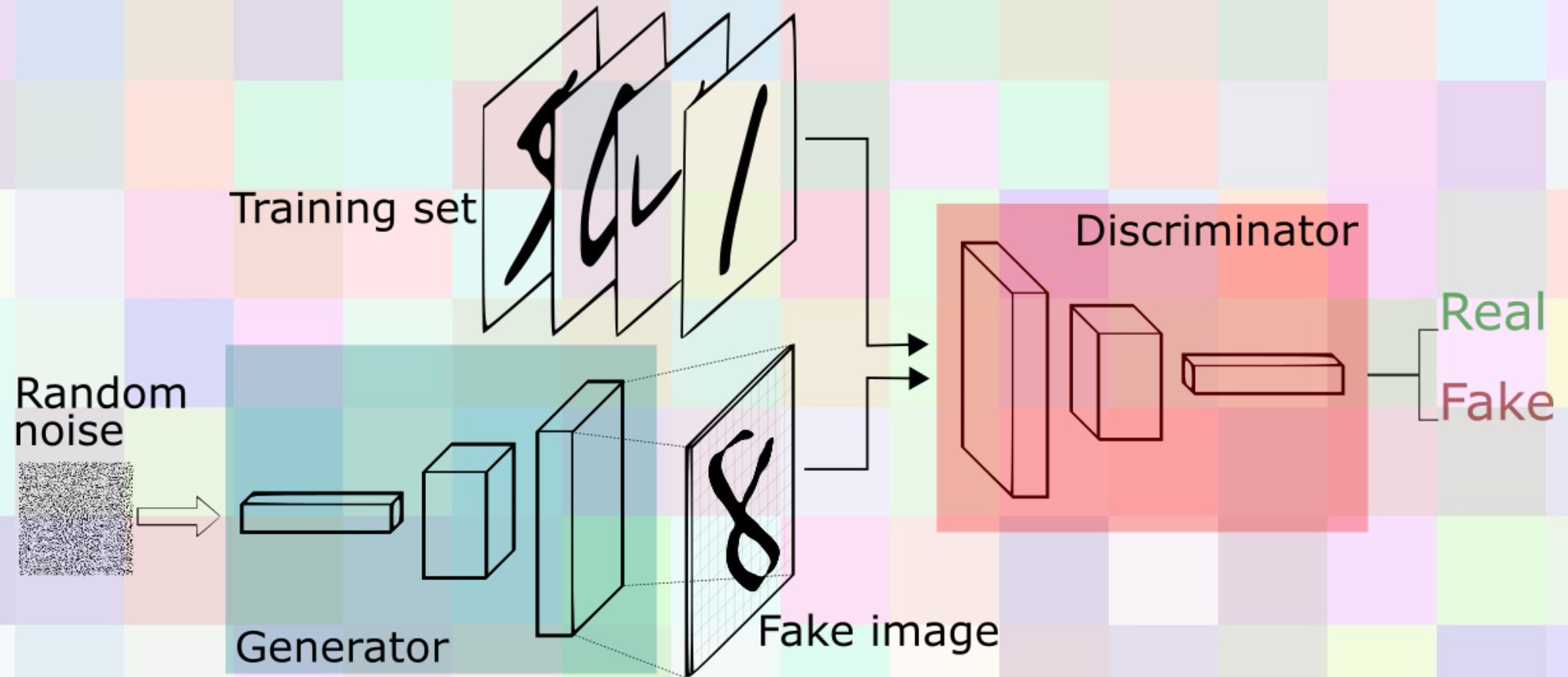
Więcej we wspomnianym zestawieniu tutaj:

<https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d>

I co dalej?

- Stworzoną sieć trzeba nauczyć – na podstawie zbioru uczącego dobrać wagi tak by zminimalizować błąd na wyjściu.
- W tym aspekcie dokonał się największy postęp w ostatnich latach – brak odpowiednich metod uczenia dla sieci z wieloma warstwami ukrytymi stanowił największy bloker rozwoju tej technologii.
- Wagi dobiera się w taki sposób by zminimalizować odpowiednią funkcję strat
- Istotny jest również dobór zbioru uczącego – powinien być on jak najbardziej reprezentatywny.
- Na następnym slajdzie zobaczycie Państwo bardzo ciekawą wizualizację dla problemu klasyfikacji siecią VGG-16. Oryginał do obejrzenia tutaj:
<https://www.youtube.com/watch?v=RNnKtNrsmg>

Ciekawostka – architektura GAN



GAN w akcji



Source: Zhu et al, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 2017

GAN w akcji

Monet ↔ Photos



Monet → photo

Zebras ↔ Horses



zebra → horse

Summer ↔ Winter



summer → winter

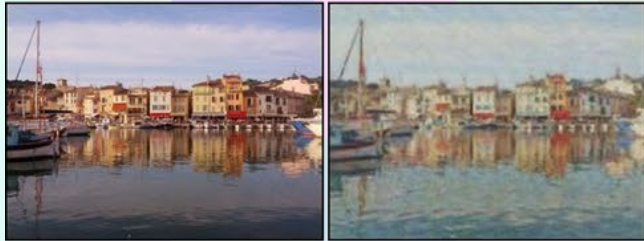


photo → Monet



horse → zebra



winter → summer



Photograph



Monet



Van Gogh



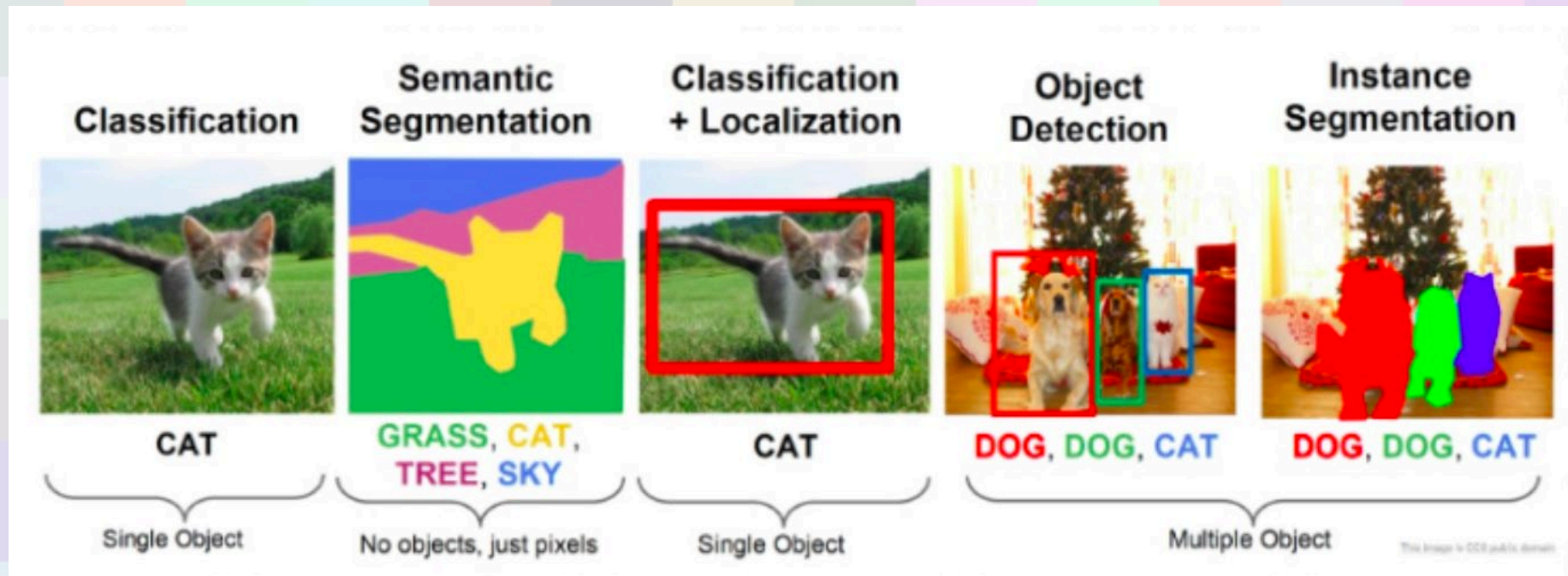
Cezanne



Ukiyo-e

Wady i zalety zastosowania DL w analizie obrazów

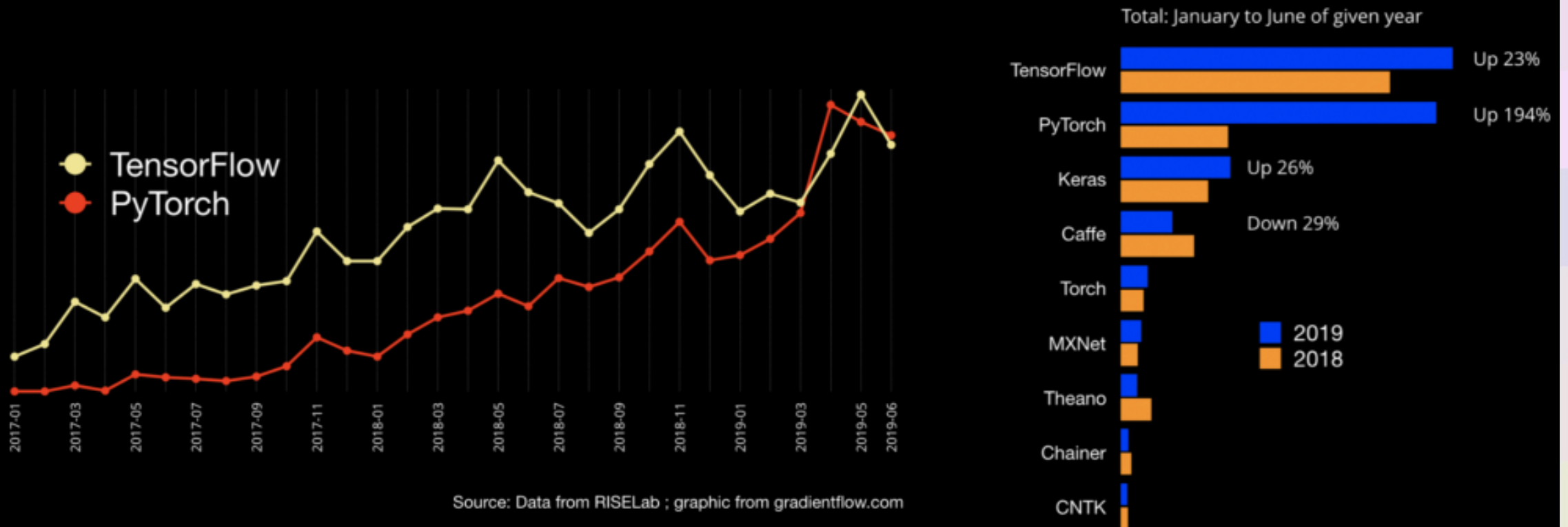
- Dzięki użyciu różnych typów warstw jesteśmy w stanie automatycznie wydobyć istotne cechy obrazów umożliwiające realizację wybranego zadania; podejście oparte o sieci neuronowe można użyć w wielu rozmaitych problemach;



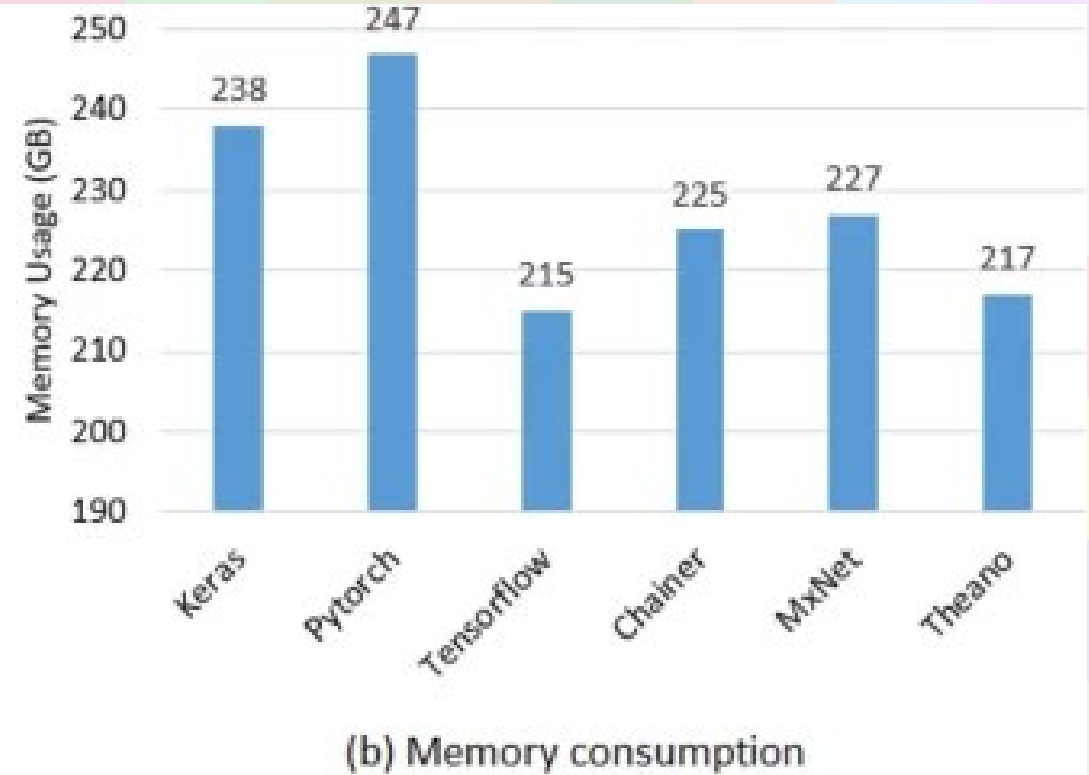
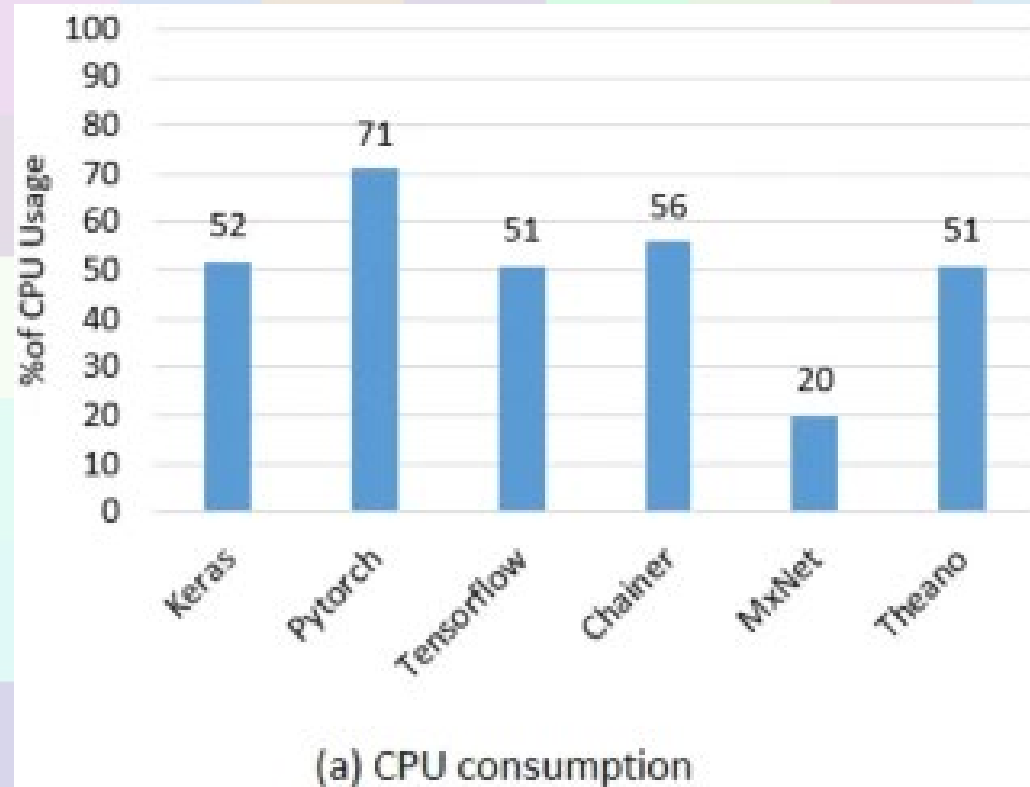
- Wady: model czarnej skrzynki, ogrom parametrów do optymalizacji, istotna rola architektury

Od czego zacząć – frameworki

Number of papers on arxiv.org that mention a given framework

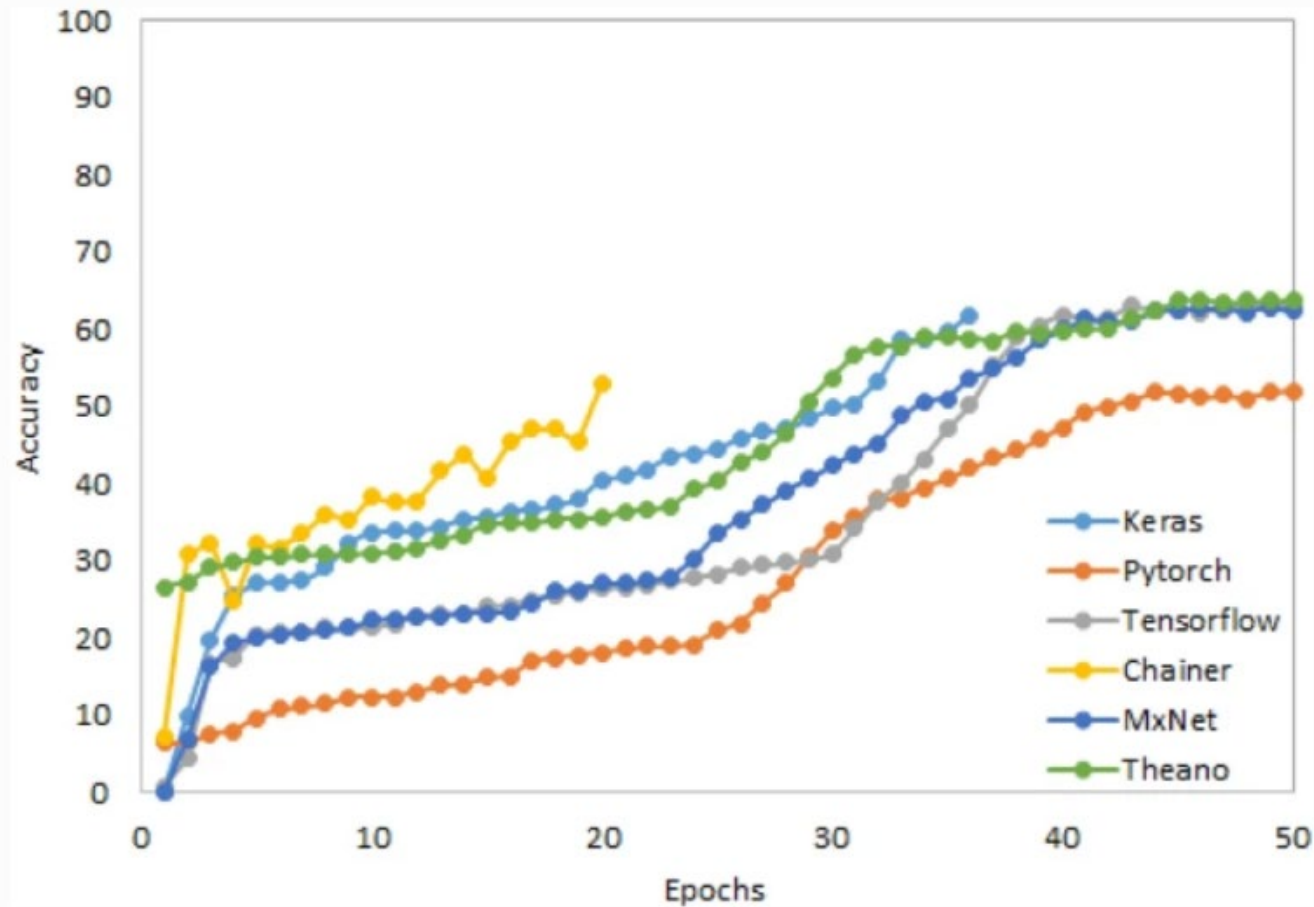


Który lepszy?



Elshawi et al., DLBench: a comprehensive experimental evaluation of deep learning frameworks, 2021

Który lepszy?



Convergence of VOC2012 for deep learning frameworks running on CPU

Elshawi et al., DLBench: a comprehensive experimental evaluation of deep learning frameworks, 2021

Żeby było łatwiej – demo Keras



<https://colab.research.google.com/github/google/objax/blob/master/examples/tutorials/mnist-tutorial.ipynb>

A gdyby użyć samych embeddingów?

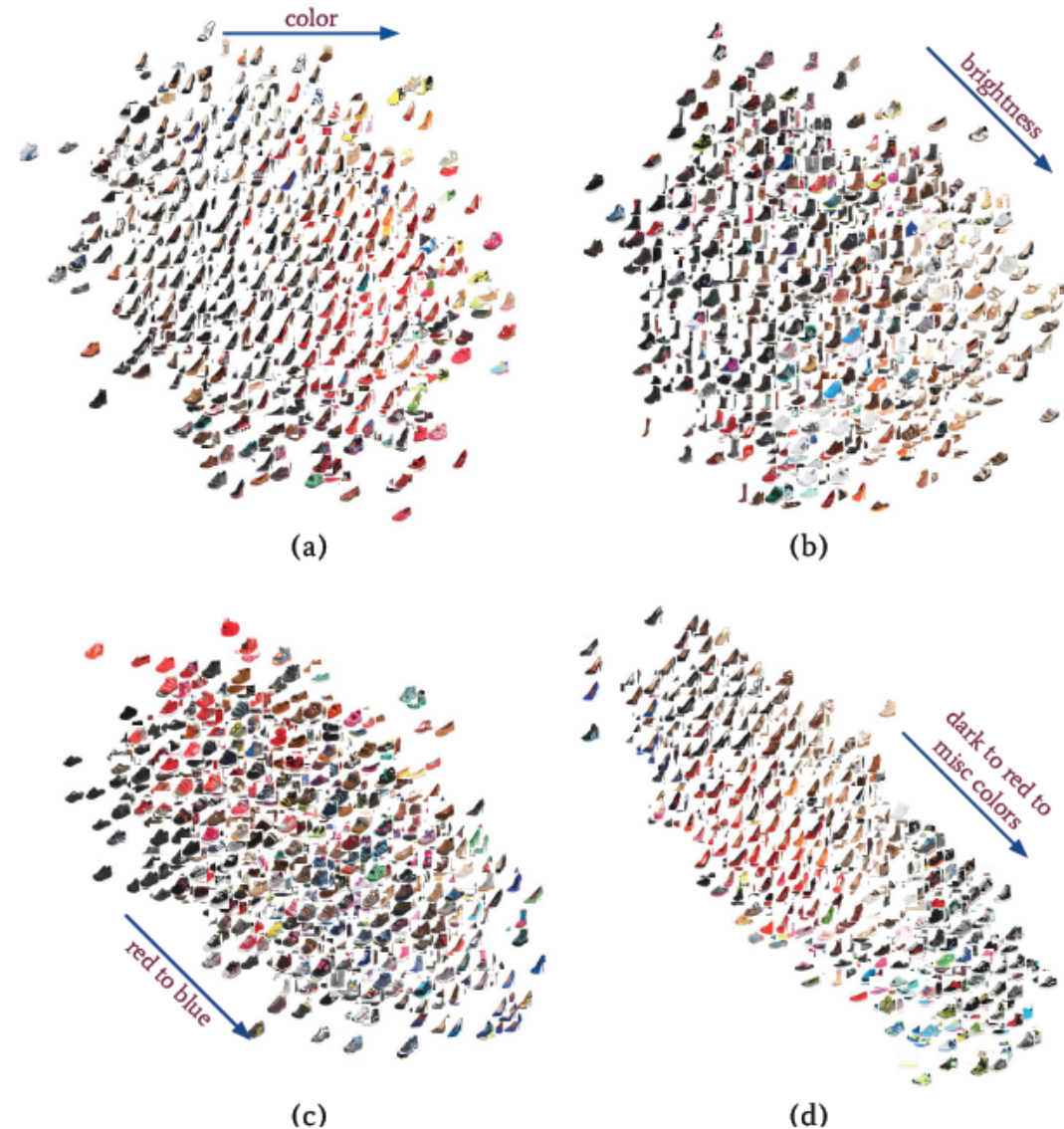
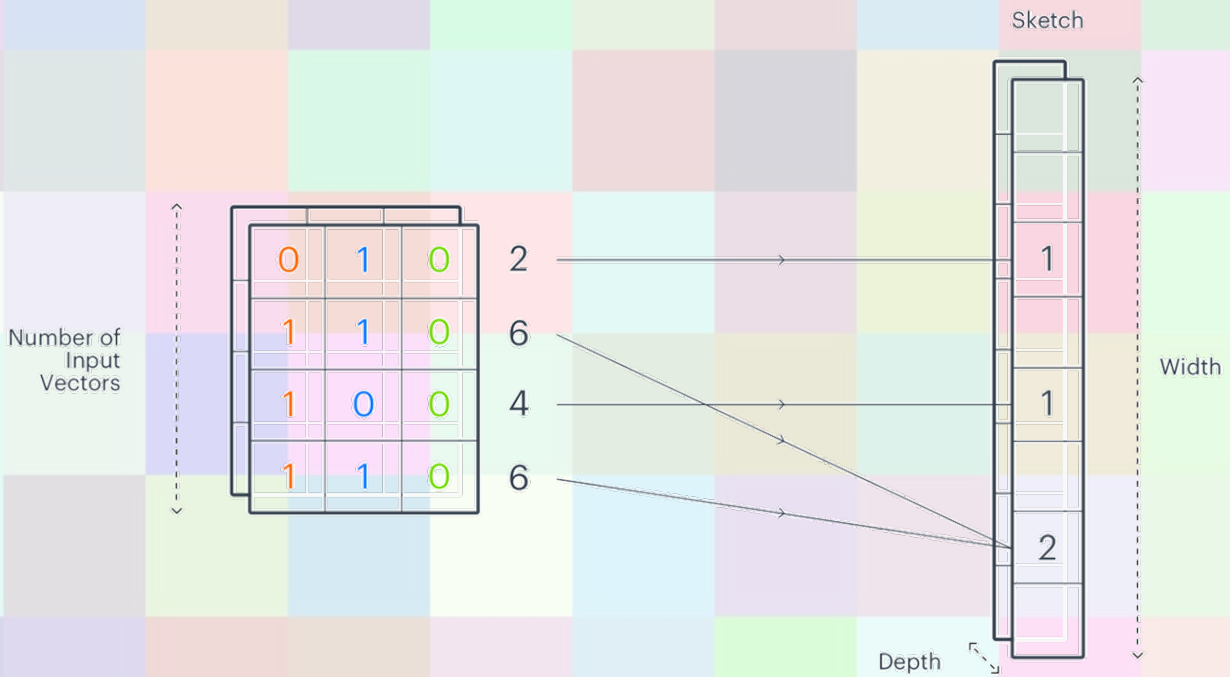


Figure 3: Qualitative Analysis of 2D projections of learnt embeddings for the UT Zappos-50k Shoes dataset: (a) A subset of the `comfort` attribute subspace learns to distinguish between colors: it is possible that humans perceive colorful shoes to be more comfortable, (b) A subset of the `open` attributes learn to reason about brightness of shoes: it is possible that humans find bright shoes to be more open, (c) A set of dimensions in the `pointy` attribute subspace embedding reasons about red versus blue colors, probably due to the fact that a number of pointy shoes in the dataset are red stilettos, and (d) A subset of the `sporty` attribute embeddings learn to reason about the shade of the color of the shoe.

A jak te modalności połączyć



<https://sair.synerise.com/emde-illustrated/>

Na zakończenie

- Istnieje mnóstwo świetnych bibliotek opartych o DL, np. Yolo, SSD, Faster-RCNN, Detectron2
- Obecnie, jakkolwiek metody tradycyjne nadal znajdują szerokie zastosowanie – zwłaszcza w zagadnieniach wymagających niewielkiego zaangażowania zasobów obliczeniowych
- Rozwój uczenia głębokiego jest również spowodowany dostępnością zasobów obliczeniowych – w tym powszechnością obliczeń na GPU



NVIDIA CUDA



Open CL



Dziękuję za uwagę!

laptop

vase

bottle

wine glass

chair

bottle

mouse

handbag

bowl

cup

Dining table