DOI: 10.24427/978-83-66391-28-4_13

Sieci neuronowe w przetwarzaniu obrazów: przegląd wybranych osiągnięć

Karol PRZYBYSZEWSKI[*]

1. Wstęp

Od początku tego dziesięciolecia obserwujemy niepowstrzymywalny postęp, jaki dokonuje się w obszarze przetwarzania obrazów. Wzrost taniej mocy obliczeniowej oferowanej przez ogólnodostępne komputery oraz wzrost ilości dostępnych danych obrazowych umożliwiły szybsze eksperymentowanie i testowanie algorytmów. Spektakularnym przykładem sukcesu w tej dziedzinie są konwolucyjne sieci neuronowe, dla których pierwowzorem był neocognitron [1] przedstawiony przez Kunihiko Fukushimę w 1980 roku. Idea zaprezentowana przez Fukushimę była podstawą często cytowanej pracy Yann LeCun [2], w której została przedstawiona sieć LeNet-5, pionierska 7-poziomowa sieć konwolucyjna do klasyfikacji ręcznie pisanych cyfr. Była ona używana przez kilka banków do rozpoznawania ręcznie pisanych cyfr na zeskanowanych obrazach czeków o wielkości 32x32 piksele.

Nieprzerwany sukces konwolucyjnych sieci neuronowych jako efektywnej metody przetwarzania obrazów został ukoronowany sprzętową implementacją konwolucji na procesorach GPU [3]. Pozwoliło to na przyspieszenie obliczeń o co najmniej rząd wielkości i umożliwiło badaczom przeprowadzanie większej ilość eksperymentów w krótszym czasie. Obecnie najefektywniejsze architektury sieci neuronowych wykorzystywane do przetwarzania obrazów, takie jak: AlexNet (2012), ZFNet (2013), GoogleNet/Inception (2014), VGGNet (2014) czy ResNet (2015), oparte są o warstwy konwolucyjne. Historia powstania i rozwoju konwolucyjnych sieci neuronowych pokazuje, że od momentu odkrycia metody do momentu, kiedy w pełni zostanie wykorzystany jej potencjał, może minąć wiele lat.

W roku 2010 po raz pierwszy zostało zorganizowane wyzwanie IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) [4], którego celem było „*oszacowanie zawartości zdjęć w celu ich pobrania i automatycznej adnotacji za pomocą podzbioru dużego, ręcznie oznakowanego zestawu danych ImageNet*". Wyzwanie to kontynuowane było przez osiem lat i w dużym stopniu przyczyniło się do szybkiego rozwoju technik przetwarzania obrazów. Prace publikowane w ramach ILSVRC pokazały

[*] Politechnika Białostocka

wyraźnie ten postęp – błąd klasyfikacji obrazów (top-5) został zredukowany z 0,28 (w 2010) do 0,023 (2017) [5]. Wyzwanie Imagenet przyczyniło się również do ponownego zainteresowania sieciami neuronowymi jako efektywnymi technikami analizy danych, a w szczególności danych obrazowych [6]. Wyzwanie ILSVRC zakończyło się w roku 2017 wskazaniem dalszych kierunków rozwoju technik przetwarzania danych – nastąpiło przesunięcie ciężaru badań z rozpoznawania/klasyfikacji obiektów do badań nad maszynowym 'rozumieniem' obrazu.

W opracowaniu przedstawiłem najważniejsze obszary dziedziny przetwarzania obrazów i różnych rodzajów i architektur sieci neuronowych zbudowanych w tym celu oraz opisałem wybrane koncepcje przetwarzania obrazów oparte o sieci neuronowe.

2. Dziedzina przetwarzania obrazów

Przetwarzanie obrazów, czasami nazywane też *widzeniem maszynowym* (od angielskiego *computer vision*) to pojemna dziedzina podzielona na wiele odrębnych obszarów. W celu jak najszerszego określenia i opisania dziedziny przetwarzania obrazów użyłem danych dostępnych i utrzymywanych przez ogólnodostępny serwis PapersWithCode (www.paperswithcode.com), którego misją jest prezentacja najnowszych osiągnięć z dziedziny uczenia maszynowego. W dziedzinie przetwarzania obrazów serwis ten zidentyfikował (maj 2019):

- 364 tablice wyników (ang. *leaderboards*),
- 501 zadania (ang. *tasks*),
- 173 ogólnodostępne zbiory danych (ang. *datasets*),
- 3524 artykuły naukowe z udostępnionym kodem źródłowym (ang. *papers with code*).

Zadania przetwarzania obrazów pogrupowane są w 219 obszarów, z których 20 najpopularniejszych (największa ilość artykułów) przedstawiłem w tabeli 1.

Strona PapersWithCode uwzględnia również mniej popularne czy nawet egzotyczne obszary, takie jak:

- *Pornography Detection,*
- *Damaged Building Detection,*
- *Logo Recognition,*
- *Window Detection,*
- *Transform A Video Into A Comics.*

TAB. 1. Wybrane obszary przetwarzania obrazów
TAB. 1. Selected areas of computer vision

Nr	Obszar	Ilość zadań	Dwa zadania z największą ilością artykułów
1.	Segmentacja semantyczna (ang. *Semantic Segmentation*)	7	*Semantic Segmentation* – 417 art. *Real-Time Semantic Segmentation* – 22 art.
2.	Klasyfikacja obrazów (ang. *Image Classification*)	7	*Image Classification* – 353 art. *Few-Shot Image Classification* – 14 art.
3.	Detekcja obiektów (ang. *Object Detection*)	17	*Object Detection* – 298 art. *3D Object Detection* – 20 art.
4.	Odpowiadanie na pytania (ang. *Question Answering*)	6	*Question Answering* – 283 art. *Open-Domain Question Answering* – 15 art.
5.	Generowanie obrazów (ang. *Image Generation*)	8	*Image generation* – 158 art. *Image-to-Image translation* – 61 art.
6.	Określanie pozy sylwetki (ang. *Pose Estimation*)	9	*Pose Estimation* – 146 art. *3D Human Pose Estimation* – 25 art.
7.	Zwiększanie rozdzielczości (ang. *Super Resolution*)	4	*Super Resolution* – 124 art. *Image Super-Resolution* – 77 art.
8.	Pojazdy autonomiczne (ang. *Autonomous Vehicles*)	13	*Autonomous Driving* – 84 art. *Autonomous Vehicles* – 38 art.
9.	Rozpoznawanie i modelowanie twarzy (ang. *Facial Recognition and Modelling*)	30	*Face Recognition* – 65 art. *Face Detection* – 37 art.
10.	Obraz wideo (ang. *Video*)	35	*Object tracking* – 48 art. *Video Classification* – 30 art.
11.	Rozpoznawanie obiektów (ang. *Object Recognition*)	4	*Object Recognition* – 113 art. *3D Object Recognition* – 7 art.
12.	Wyszukiwanie obrazem (ang. *Image Retrieval*)	7	*Image Retrieval* – 105 art. *Content-Based Image Retrieval* – 9 art.
13.	Rozpoznawanie akcji (ang. *Action Recognition*)	6	*Action Recognition* – 89 art. *Action Recognition in Videos* – 13 art.
14.	Tagowanie obrazów (ang. *Image Captioning*)	2	*Image Captioning* – 96 art. *Image Paragraph Captioning* – 1 art.

Nr	Obszar	Ilość zadań	Dwa zadania z największą ilością artykułów
15.	Określanie głębokości (ang. *Depth Estimation*)	9	*Depth Estimation* – 68 art. *Monocular Depth Estimation* – 24 art.
16.	Transfer stylu (ang. *Style Transfer*)	5	*Style Transfer* – 86 art. *Image Stylization* – 7 art.
17.	Detekcja anomalii (ang. *Anomaly Detection*)	7	*Anomaly Detection* – 81 art. *Unsupervised Anomaly Detection* – 8 art.
18.	Rozumienie sceny (ang. *Scene Parsing*)	8	*Scene Understanding* – 40 art. *Scene Recognition* – 15 art.
19.	Obrazy 3D (ang. *3D*)	25	*3D Reconstruction* – 47 art. *3D Pose estimation* – 17 art.
20.	Ponowna identyfikacja osoby (ang. *Person Re-Identification*)	6	*Person Re-Identification* – 66 art. *Video-Based Person Re-Identification* – 4 art.

Źródło: opracowanie własne.
SOURCE: own elaboration.

Obszary przetwarzania obrazów wymienione w powyższej tabeli jasno pokazują że ta dziedzina ma zastosowanie w wielu praktycznych aspektach. Najczęściej używane techniki to segmentacja/detekcja oraz klasyfikacja. Istotnym zagadnieniem jest również łączenie technik przetwarzania obrazów z technikami przetwarzania języka naturalnego w takich obszarach, jak Odpowiadanie na pytania, Tagowanie obrazów czy Rozumienie Sceny. Coraz bardziej popularne są również techniki związane z tworzeniem nowych oraz modyfikacją istniejących obrazów, takie jak Generowanie Obrazów czy też Transfer stylu. Warto również zwrócić uwagę na to, że dziedzina przetwarzania obrazów nie ogranicza się jedynie do obrazów dwuwymiarowych, ale porusza również zagadnienia związane z nagraniami wideo oraz obrazami trójwymiarowymi.

3. Sieci neuronowe w przetwarzaniu obrazów

Niekwestionowanym liderem w dziedzinie przetwarzania obrazów jest koncepcja konwolucyjnych sieci neuronowych (ang. *convolutional neural networks*), czasami też tłumaczona na język polski jako *splotowe sieci neuronowe*. Ich ogromna zaleta to zdolność redukcji obrazu do formy, która jest o wiele prostsza do przetwarzania, ale zachowuje wszystkie cechy istotne do poprawnego wnioskowania. Istnieje jednak wiele innych koncepcji związanych z sieciami neuronowymi, które również dają

bardzo dobre rezultaty w obszarze przetwarzania obrazów. W poniższej tabeli wymienione zostały wszystkie istotne koncepcje związane z sieciami neuronowymi, jakie dostępne były w czasie pisania artykułu. Do opisu użyłem nazw anglojęzycznych, gdyż w wielu przypadkach nie wykształciła się jeszcze terminologia polska i dla zachowania spójności nazewniczej lepiej jest użyć jednolitego, angielskiego nazewnictwo.

TAB. 2. Istotne koncepcje sieci neuronowych związane z przetwarzaniem obrazów
TAB. 2. Important neural networks concepts related to computer vision

Nr	Nazwa	Rok publikacji
1.	Feed forward neural networks (FF or FFNN) i perceptrons (P)	1958
2.	Hopfield network (HN)	1982
3.	Kohonen networks (KN, także Self Organising (Feature) Map, SOM, SOFM)	1982
4.	Boltzmann machines (BM)	1986
5.	Restricted Boltzmann machines (RBM)	1986
6.	Radial basis function (RBF)	1988
7.	Autoencoders (AE)	1988
8.	Recurrent neural networks (RNN)	1990
9.	Long / short term memory (LSTM)	1997
10.	Bidirectional recurrent neural networks, bidirectional long / short term memory networks i bidirectional gated recurrent units (BiRNN, BiLSTM i BiGRU odpowiednio)	1997
11.	Convolutional neural networks (CNN lub deep convolutional neural networks, DCNN)	1998
12.	Liquid state machines (LSM)	2002
13.	Echo state networks (ESN)	2004
14.	Extreme learning machines (ELM)	2006
15.	Sparse autoencoders (SAE)	2007
16.	Deep belief networks (DBN)	2007

Nr	Nazwa	Rok publikacji
17.	Denoising autoencoders (DAE)	2008
18.	Deconvolutional networks (DN), również nazywane Inverse Graphics Networks (IGNs)	2010
19.	Variational autoencoders (VAE)	2013
20.	Markov chains (MC lub discrete time Markov Chain, DTMC)	2013
21.	Gated recurrent units (GRU)	2014
22.	Generative adversarial networks (GAN)	2014
23.	Neural Turing machines (NTM)	2014
24.	Deep convolutional inverse graphics networks (DCIGN)	2015
25.	Deep residual networks (DRN)	2015
26.	Attention networks (AN)	2015
27.	Differentiable Neural Computers (DNC)	2016
28.	Capsule Networks (CapsNet)	2017

ŹRÓDŁO: opracowanie własne.
SOURCE: own elaboration.

Powyższa tabela wyraźnie pokazuje aktywny rozwój architektur i koncepcji związanych z sieciami neuronowymi. Koncepcje te rozwijane są na różnych poziomach, przykładowo:

- sieci CapsNet skupiają się w dużym stopniu na zlikwidowaniu pewnych ograniczeń sieci konwolucyjnych i usprawnieniu ich zdolności bardziej efektywnej ekstrakcji istotnych cech;
- w koncepcjach, takich jak GAN czy autoenkodery, istotnym elementem jest odpowiednia kombinacja współpracujących ze sobą sieci neuronowych.

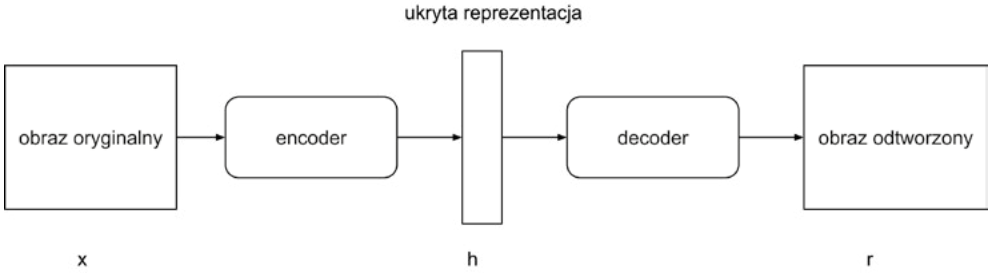
4. Koncepcje autoenkodera (AE) i GAN

W tym paragrafie przedstawione zostały koncepcja GAN [7] i koncepcja autoenkoderów, które wychodzą poza zwyczajowe zastosowanie sieci neuronowych, czyli klasyfikację, a ich zastosowanie w przetwarzaniu obrazów dało spektakularne rezultaty. GAN-y i autoenkodery są stosowane do generowania, modyfikacji czy też edycji obrazów.

Autoenkodery

Idea autoenkoderów już od wielu lat jest częścią obszaru zagadnień związanych z sieciami neuronowymi, jednak dopiero niedawne zastosowanie w przetwarzaniu obrazów dało spektakularne rezultaty, np. w postaci transferu wybranego stylu malarskiego na dowolne, amatorskie zdjęcia wykonane zwykłym aparatem.

Głównym założeniem autoenkoderów jest redukcja danych wejściowych do ukrytej przestrzeni stanów o mniejszej liczbie wymiarów, a następnie próba odtworzenia danych wejściowych z tej reprezentacji. Pierwsza część nazywa się kodowaniem, a druga – fazą dekodowania. Zmniejszając liczbę zmiennych reprezentujących dane, wymuszamy na modelu nauczenie się, jak zachować tylko istotne informacje, z których dane wejściowe można odtworzyć. Działanie to może być również postrzegane jako technika kompresji.



RYS. 1. Architektura autoenkodera

ŹRÓDŁO: opracowanie własne.
SOURCE: own elaboration.

Enkoder ma postać $h = f(x)$, natomiast dekoder: $r = g(h)$. Tradycyjnie, autoenkodery minimalizują funkcję:

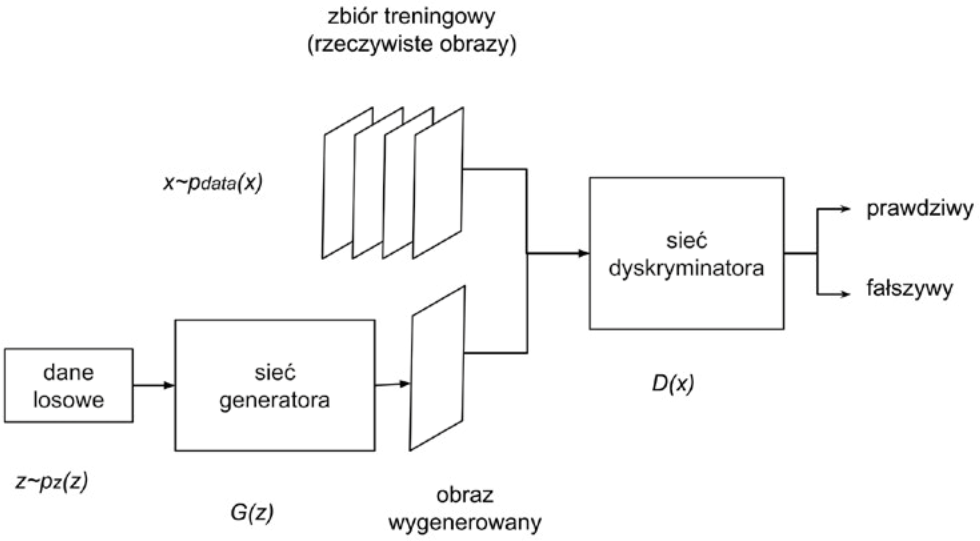
$$L(x,g(f(x)))$$

gdzie L jest funkcją kosztu karzącą $g(f(x))$ gdy nie jest ona podobna do x (przykładowo normę L^2 z ich różnicy). Koncepcja autoenkoderów okazała się bardzo pojemna, a w jej ramach można przykładowo wyróżnić:

- *Undercomplete Autoencoders* – gdzie rozmiar *h* jest mniejszy od *x* (tradycyjna, podstawowa wersja autoenkodera);
- *Sparse Autoencoders* (*SAE*) [8] – do funkcji kary dodawany jest parametr karzący za nierzadką reprezentację; *SAE* zwykle są używane do ekstrakcji cech na potrzeby innych zadań, takich jak klasyfikacja;
- *Denoising Autoencoders* (*DAE*) [9] – jako wejście otrzymują uszkodzone dane (np. zaszumiony obraz) i trenowane są w celu otrzymania oryginalnych, nieuszkodzonych danych jako wyjścia;
- *Variational Autoencoders* (*VAE*) [10] – dodaje człon regularyzujący, wymuszający odpowiedni rozkład sygnału w warstwie kodującej; w przetwarzaniu obrazów VAE używane mogą być do generowania nowych obrazów;
- oraz inne, niewymienione w tym opracowaniu.

Generative adversarial networks (GAN)

Rozpatrując koncepcję GAN z wysokiego poziomu, można stwierdzić, że architektura sieci GAN składa się z dwóch komponentów: generatora i dyskryminatora. Dyskryminator ma za zadanie określić, czy dany obraz wygląda naturalnie (to znaczy, czy jest obrazem z zestawu danych) lub czy wygląda na sztucznie utworzony. Zadaniem generatora jest natomiast tworzenie naturalnie wyglądających obrazów, które są podobne do pierwotnego rozkładu danych, czyli obrazów które wyglądają na tyle naturalnie by oszukać sieć dyskryminatora.



RYS. 2. Architektura koncepcji GAN

ŹRÓDŁO: opracowanie własne.
SOURCE: own elaboration.

Notacja przedstawiona na rysunku oznacza: $p_{data}(x)$ – rozkład prawdopodobieństwa rzeczywistych obrazów, x – próbka z rozkładu $p_{data}(x)$, $p_z(z)$ – rozkład prawdopodobieństwa generatora, z – próbka z rozkładu $p_z(z)$, $G(z)$ – sieć generatora, $D(x)$ – sieć dyskryminatora. Tak jak wspominałem wcześniej, trening sieci GAN jest realizowany jako rywalizacja między generatorem a dyskryminatorem. Można to opisać matematycznie jako:

$$min_G\, max_D\, V(D, G) = \mathbb{E}_{x\sim pdata(x)}[log(D(x))] + \mathbb{E}_{z\sim pz(z)}[log(1 - D(G(z))] \qquad (2)$$

W funkcji $V(D, G)$ pierwszym członem jest entropia pokazująca, że dane z prawdziwego rozkładu ($p_{data}(x)$) przejdą przez dyskryminator (najlepszy scenariusz). Dyskryminator stara się maksymalizować tą wartość do 1. Drugi człon to entropia pokazująca, że dane z losowego rozkładu ($p_z(z)$) przechodzą przez generator wytwarzający nieprawdziwy obraz, który jest następnie przepuszczany przez dyskryminator w celu identyfikacji autentyczności (najgorszy przypadek). Ten człon dyskryminator stara się doprowadzić do wartości 0. Całościowo zatem dyskryminator dąży do maksymalizacji funkcji V. Z drugiej strony zadanie generatora jest dokładnie odwrotne. Stara się on minimalizować funkcję V w taki sposób, aby różnica między prawdziwymi a wytworzonymi danymi była jak najmniejsza. W terminologii angielskiej do opisu tej koncepcji używa się terminu *minmax game*. Należy również wspomnieć, że koncepcja GAN ma wiele różnych zastosowań i odmian [11].

AE versus GAN

Autoenkoder kompresuje swoje dane wejściowe do wektora o znacznie mniejszych wymiarach niż dane wejściowe, a następnie przekształca je z powrotem w wektor o tym samym kształcie. GAN można opisać jako odwrócony autoenkoder – zamiast kompresować dane wielowymiarowe, dostaje wektory niskowymiarowe jako dane wejściowe, a dane o dużych wymiarach znajdują się w środku architektury sieci.

GAN nie przyjmuje rzeczywistych danych jako wejściowych, a zamiast tego otrzymuje mały wektor liczb losowych. Sieć generatora próbuje przekształcić ten mały wektor w realistyczną próbkę z danych treningowych. Następnie sieć dyskryminatora pobiera tę wygenerowaną próbkę (i kilka rzeczywistych próbek z zestawu danych) i uczy się odgadywać, czy próbki są prawdziwe czy fałszywe.

5. Sieci kapsułkowe

Konwolucyjne sieci neuronowe (CNN) odniosły wielki sukces w rozwiązywaniu problemów z rozpoznawaniem obiektów i ich klasyfikacją. Nie są jednak idealne. Jeśli na wejście sieci konwolucyjnej podamy obiekt w orientacji, której sieć nie zna lub w której obiekty pojawiają się w miejscach, do których sieć nie jest przyzwyczajona,

zadanie predykcji prawdopodobnie się nie powiedzie. CNN uczą się wzorów statystycznych na obrazach, ale nie uczą się podstawowych pojęć dotyczących tego, co sprawia, że coś rzeczywiście wygląda jak konkretny rzeczywisty obiekt (np. twarz).

W 2017 Geoffrey Hinton (i inni), zapożyczyli pomysły z neurobiologii, które sugerują, że mózg jest zorganizowany w moduły zwane kapsułkami [12] (*CapsNets*). Kapsułki te są szczególnie dobre w rozpoznawaniu cech takich, jak ułożenie (położenie, rozmiar, orientacja), deformacja, prędkość, albedo, odcień, tekstura itp. W kontekście sieci neuronowych kapsułki reprezentowane są przez grupy neuronów.

Rezultaty zaprezentowane w pracach Hintona pokazały, że *CapsNets* mają najwyższą wydajność w standardowych zestawach danych, takich jak MNIST (z dokładnością testową 99,75%) i SmallNORB (z 45% zmniejszeniem błędu w stosunku do poprzedniego najlepszego wyniku). Jednak aplikacje i wydajność tych sieci na rzeczywistych i bardziej złożonych danych nie zostały w pełni zweryfikowane. Bardzo ważną korzyścią, jaką zapewniają sieci kapsułkowe, jest przejście od sieci neuronowych typu *black-box* do tych, które reprezentują bardziej konkretne cechy, mogące pomóc nam przeanalizować i zrozumieć, w jaki sposób sieć neuronowa działa od środka. Należy również wspomnieć, że *CapsNets* używają koncepcji autoenkoderów (rekonstrukcyjna funkcja kosztu jest używana jako metoda regularyzacji), co potwierdza tezę, że idee opracowane lata temu wciąż znajdują ważne zastosowania i przyczyniają się do powstawania nowych, efektywnych rozwiązań w różnych dziedzinach eksploracji danych.

6. Powiązane prace

Gwałtowny rozwój przetwarzania obrazów przy użyciu sieci neuronowych zaowocował również dużą ilością artykułów naukowych opisujących przekrojowo wybrane zagadnienia z tej dziedziny. Warto zwrócić uwagę na artykuł [13], w którym autorzy obszernie opisują ponad 20 lat historii wykrywania obiektów. Artykuł oparty jest o przegląd ponad 400 prac obejmujących okres od 1990 do 2019 roku. Autorzy wyraźnie pokazali podział na dwie główne epoki detekcji obiektów: tradycyjne metody detekcji (do 2012) i metody oparte o głębokie uczenie maszynowe (po 2012), wśród których najpopularniejsze są koncepcje związane z sieciami konwolucyjnymi.

Bardzo dobrą i obszerną pracą opisującą współczesne architektury głębokiego uczenia maszynowego jest artykuł [14]. Autorzy kompleksowo opisują rozwój najważniejszych koncepcji w dziedzinie głębokiego uczenia maszynowego od roku 2012. Do artykułu dołączona jest również lista najpopularniejszych frameworków, SDK oraz referencyjnych zbiorów danych używanych do implementacji i ewaluacji zadań związanych z głębokim uczeniem maszynowym.

Warto również zwrócić uwagę na [15,] która skupia się na opisie historii rozwoju głębokich konwolucyjnych sieci neuronowych. Badanie skupia się na pokazaniu wewnętrznej taksonomii najnowszych, głębokich architektur CNN. Podejmuje też próbę podziału najnowszych innowacji w architekturach CNN na siedem różnych kategorii (terminologia

angielska): *spatial exploitation, depth, multi-path, width, feature map exploitation, channel boosting* oraz *attention*. Wartościowych informacji dostarcza dołączona na końcu artykułu tabela porównująca wyniki różnych architektur w odniesieniu do wyżej wymienionych kategorii.

7. Wnioski i dalsze prace

Na dzień pisania pracy nie znalazłem opracowań naukowych, które całościowo podejmują temat wysokopoziomowej taksonomii architektur i koncepcji opartych o sieci neuronowe, wraz ze wskazaniem i kategoryzacją praktycznych obszarów badawczych przetwarzania obrazów. Niniejsze opracowanie jest wstępem do dalszych prac nad opisaniem rozwoju architektur, koncepcji i modeli sieci neuronowych w oparciu o ich praktyczne zastosowania.

Lista praktycznych obszarów badawczych zaprezentowana w pracy pokazuje, że współczesne przetwarzanie obrazów to dziedzina szeroka i odważnie wkraczająca na nowe pola badań. Krótka historia kluczowych odkryć w obszarze sieci neuronowych pozwala zauważyć, że od momentu opracowania podstawowych koncepcji (tutaj: neuron) do czasu kiedy to odkrycie (tutaj: sieci konwolucyjne, autoenkodery, GAN-y, etc) osiągnie wysoki potencjał, mogą minąć dziesięciolecia.

Niniejsze opracowanie jest pierwszym z planowanej serii opisującej wysokopoziomowe koncepcje i architektury sieci neuronowych oraz łączenie tej wiedzy z praktycznymi obszarami zastosowań technik przetwarzania obrazów. W ramach dalszych prac będę starał odpowiedzieć się na takie pytania, jak:

- Jak przeprowadzić dalszą analizę struktury/architektury sieci versus zastosowania praktyczne?
- Jak szukać odpowiednich architektur dla konkretnych zastosowań?
- Jak budować nowe architektury?
- Jak skategoryzować koncepcje w ramach sieci neuronowych?

Literatura

1. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics*. 1980. 36: 193-202.
2. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998; vol. 86, no. 11: 2278-2324; DOI: 10.1109/5.726791.
3. Chellapilla K, Puri S, Simard P. *High Performance Convolutional Neural Networks for Document Processing. Tenth International Workshop on Frontiers in Handwriting Recognition*, Université de Rennes 1, Oct 2006, La Baule (France). ffinria-00112631.

4. http://image-net.org/challenges/LSVRC/2010/index.
5. http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf.
6. Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84-90. DOI: https://doi.org/10.1145/306538.
7. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. *Generative Adversarial Networks*, 2014. Conference: Advances in neural information processing systems; 2672-2680.
8. Hinton GE, Osindero S, The Y-W. 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 7 (July 2006), 1527-1554. DOI:https://doi.org/10.1162/neco.2006.18.7.1527.
9. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. 2008. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (ICML '08). Association for Computing Machinery, New York, NY, USA, 1096-1103. DOI:https://doi.org/10.1145/1390156.1390294.
10. Kingma DP, Welling M. *Auto-Encoding Variational Bayes*. Paper presented at the meeting of the ICLR, 2014.
11. Lucic M, Kurach K, Michalski M, Bousquet O, Gelly S. 2018. Are GANs created equal? a large-scale study. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA; 698-707.
12. Sabour S, Frosst N, Hinton GE. 2017. Dynamic routing between capsules. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3859-3869.
13. Zhengxia Z, Shi Z, Guo Y, Ye J. "Object detection in 20 years: A survey". arXiv preprint arXiv:1905.05055 (2019).
14. Md Zahangir A, Taha TM, Yakopcic C, Westberg S, Sidike P, Mst Nasrin S, Hasan M, Van Essen BC, Awwal AAS, Asari VK. A state-of-the-art survey on deep learning theory and architectures. *Electronics*. 2019; 8, no. 3: 292.
15. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*. 2019: 1-62.

Streszczenie

Ostatnie dziesięciolecie (2010-2019) to niepowstrzymywalny rozwój technik przetwarzania obrazów związanych z sieciami neuronowymi. Powszechne użycie internetu oraz fotografii cyfrowej dostarczyło ogromnej ilości danych do przetworzenia. Szybki rozwój sprzętu oferującego dużą moc obliczeniową (np. procesory GPU) umożliwił za to znaczne obniżenie kosztów przetwarzania danych. Oba te fakty sprawiły, że możliwe stało się szybkie i efektywne trenowanie sieci neuronowych w oparciu o wiedzę zgromadzoną w zbiorach danych obrazowych. Celem opracowania jest przedstawienie wybranych, najnowszych zagadnień z dziedziny przetwarzania obrazów

na poziomie koncepcji, bez przedstawiania dokładnego aparatu matematycznego, i rozważenie w jaki sposób koncepcje te można interpolować na inne obszary przetwarzania obrazów czy też wykorzystać do tworzenia kolejnych idei.

Słowa kluczowe: przetwarzanie obrazów, głębokie uczenie maszynowe, architektury sieci neuronowych

Summary
Neural Networks in Computer Vision: a review of selected advancements

The last decade (2010-2019) is the unstoppable development of image processing techniques associated with neural networks. The widespread use of the internet and digital photography has provided a huge amount of data to process. The rapid development of equipment offering high computing power (e.g. GPUs) has enabled a significant reduction in data processing costs. Both of these facts meant that it became possible to train neural networks quickly and efficiently based on knowledge accumulated in image data sets. The aim of the paper is to present selected, the latest issues in the field of image processing at the concept level, without presenting an exact mathematical apparatus and to consider how these concepts interpolate into other areas of image processing or use to create subsequent ideas.

Keywords: computer vision, deep learning, neural network architectures