XXVIII Seminarium

ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE 2018

Oddział Gdański PTETiS

doi: 10.32016/1.60.12

ROZPOZNAWANIE OBIEKTÓW PRZEZ GŁĘBOKIE SIECI NEURONOWE

Arkadiusz KWASIGROCH[1], **Michał GROCHOWSKI**[2]

1. Wydział Elektrotechniki i Automatyki, Politechnika Gdańska
   tel.: 58 347 17 42 e-mail: arkadiusz.kwasigroch@pg.edu.pl
2. Wydział Elektrotechniki i Automatyki, Politechnika Gdańska
   tel.: 58 347 29 04 e-mail: michal.grochowski@pg.edu.pl

Streszczenie: W referacie zaprezentowane zostaną wyniki badań nad rozpoznawaniem obiektów w różnych warunkach za pomocą głębokich sieci neuronowych. Przeanalizowano działanie dwóch struktur – ResNet50 oraz VGG19. Systemy rozpoznawania obrazu wytrenowano oraz przetestowano na reprezentatywnej, bazie zawierającej 25 tys. zdjęć psów oraz kotów, która znacznie upraszcza analizowanie działania systemów ze względu na łatwość interpretacji zdjęć przez człowieka. Zbadano wpływ pojawienia się nietypowych zdjęć na wynik klasyfikacji. Ponadto przeanalizowano zdjęcia niepoprawnie sklasyfikowane i porównano je z interpretacjami człowieka. Uzyskano bardzo wysokie wyniki klasyfikacji. Do oceny systemów użyto miar statystycznych takich jak: dokładność, czułość, swoistość, krzywe ROC

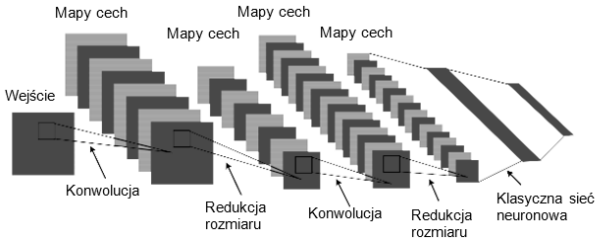
Słowa kluczowe: głębokie uczenie, sieci neuronowe, sztuczna inteligencja, przetwarzanie obrazu

1. INFORMACJE OGÓLNE

W ostatnich latach nastąpił dynamiczny rozwój głębokich sieci neuronowych – dużych modeli, które wyewoluowały z klasycznych sieci neuronowych. Algorytmy uzyskują bardzo wysokie wyniki w wielu zadaniach takich jak: rozpoznawanie obrazu, rozpoznawanie mowy, analiza teksu pisanego, synteza mowy [1]. Podczas badań przeanalizowano działanie głębokich sieci neuronowych do analizy obrazu. Zaletą głębokich sieci neuronowych jest możliwość automatycznej ekstrakcji cech obrazu, w przeciwieństwie do klasycznych algorytmów rozpoznawania obrazu, których dobór wymaga wiedzy eksperckiej bazującej na tzw. inżynierii cech. Dzięki tej umiejętności, głębokie sieci neuronowe są w stanie wyodrębnić z badanego obrazu istotne cechy, które mają największy wpływ na możliwości klasyfikacji. W przeprowadzonych badaniach przeanalizowano działanie modeli głębokich sieci neuronowych. Działania sieci przeanalizowano ilościowo oraz jakościowo. Ponadto, celem pracy było przybliżenie czytelnikom głębokich sieci neuronowych dla przykładu klasyfikacji obrazu.

2. GŁĘBOKIE, SPLOTOWE SIECI NEURONOWE

Odmianą głębokich sieci neuronowych służącą do analizy obrazu są splotowe sieci neuronowe. Typowa splotowa sieć neuronowa składa się z kombinacji 3 podstawowych typów warstw: warstwa splotowa, warstwa aktywacji oraz warstwa redukująca rozmiar (rys. 1). Dodatkowo wykorzystuję się klasyczne warstwy neuronowe tzw. warstwy pełnego połączenia.



Rys. 1. Splotowa sieć neuronowa [2]

Splotowa sieć neuronowa przyjmuje na wejście obraz, który następnie jest przetwarzany przez kolejne warstwy. Warstwa splotowa wykonuje wielokrotnie operację splotu dyskretnego na obrazie wejściowym tworząc na wyjściu mapy cech. Operacja splotu dyskretnego w obszarach analizy obrazu wykorzystywana jest do filtracji. Następnie, przefiltrowane obrazy trafiają na nieliniową funkcję aktywacji, która przetwarza każdy piksel. Dodatkowo, po niektórych warstwach aktywacji umieszczana jest warstwa redukująca rozmiar, która zmniejsza liczbę pikseli w przetwarzanych obrazach. Wyjście ostatniej warstwy splotowej trafia na klasyczną sieć neuronową.

Uczenie klasycznych sieci neuronowych z dużą liczbą warstw jest bardzo trudne, często niemożliwe. W głębokich sieciach neuronowych wprowadzono szereg modyfikacji, które umożliwiły efektywny trening tak dużych modeli.

Warstwy splotowe można interpretować jako zbiór neuronów połączonych z niewielką ilością neuronów w warstwie poprzedzającej, w przeciwieństwie do warstw klasycznych gdzie neuron połączony jest ze wszystkimi neuronami w poprzedniej warstwie. Ponadto, neurony posiadają grupowo współdzielone wagi (w ramach jednego filtru splotowego). Wymienione właściwości umożliwiają znaczną redukcję ilości parametrów, co umożliwia skuteczne uczenie takich struktur. Ponadto, szeroko stosowana funkcja aktywacji ReLU opisana wzorem $f(x)=\max(0,x)$ redukuje problem znikającego gradientu, ze względu na niemożliwość nasycenia się funkcji, jak ma to miejsce w klasycznych funkcjach aktywacji (funkcja sigmoidalna czy tangens hiperboliczny) [3].

Cechą odróżniającą głębokie sieci neuronowe od klasycznych systemów klasyfikacji obrazów jest możliwość automatycznej ekstrakcji cech. Poprawna ekstrakcja cech jest kluczowym elementem każdego systemu klasyfikacji

obrazu. Warstwy splotowe działają jako ekstraktor cech, które następnie podawane są na klasyczny klasyfikator neuronowy. W przypadku klasycznych systemów algorytmy ekstrakcji cech i ich parametry dobierane są przez badacza.

3. ARCHITEKTURY

3.1. Architektura VGG19

W badaniach przystosowaliśmy popularną architekturę VGG19 [4] do rozwiązywanego zadania. Oryginalna architektura została przez nas zmodyfikowana poprzez dodanie kolejnych warstw splotowych. W rezultacie architektura zawiera 19 warstw splotowych oraz dwie warstwy klasyczne. Łączna liczba parametrów wynosi 241 mln parametrów. Ponadto w sieci wykorzystuje się warstwy takie jak maxpooling – odpowiedzialne za zmniejszanie rozmiaru map cech oraz warstwa dropout [5], która zwiększa możliwości generalizacji sieci poprzez trening z ograniczoną liczbą neuronów.

3.2 Architektura ResNet50

Postanowiono sprawdzić działanie architektury ResNet50, ze względu na raportowane wysokie wyniki uzyskiwane przez tę sieć [6]. Sieć neuronowa składa się z 50 warstw splotowych i zawiera 24 mln parametrów. Nie używa się w niej warstw neuronów klasycznych – wyjście ostatniej warstwy splotowej trafia bezpośrednio na neuron sigmoidalny. Sieć ResNet50 korzysta z metody *batch normalization* w celu usprawnienia procesu uczenia [7]. Działanie sieci ResNet polega na przetwarzaniu residuów przy pomocy bloków opisanych wzorem $y=f_1(f_2(x))+x$, gdzie y oznacza wyjście bloku, x wejście bloku, f_1 oraz f_2 są kolejnymi warstwami sieci neuronowej. Liczba warstw w bloku może być dowolnie dobierana.

4. BAZA DANYCH

W celu analizy działania głębokich sieci neuronowych wykorzystano bazę danych, która zawierała zdjęcia psów oraz kotów. Zbiór danych dostępny jest na portalu Kaggle [8]. Baza danych wykorzystywana jest do testowania działania systemów klasyfikujących obraz. Ponadto zbiór zdjęć umożliwia łatwą analizę działania głębokich sieci neuronowych ze względu na łatwość interpretacji zdjęć zwierząt przez człowieka. Baza danych posiada 25 tys. zdjęć zwierząt różnej rozdzielczości, równo podzielonych na dwie klasy. Użyta baza danych jest odpowiednim zbiorem do uczenia głębokich sieci neuronowych ze względu na następujące cechy: duża liczba zdjęć, która umożliwia wydzielenie zbiorów treningowych oraz testowych; równa liczba zdjęć w obu klasach; bardzo zróżnicowane zdjęcia o różnych rozmiarach; zdjęcia wykonane w różnych warunkach oświetleniowych oraz otoczeniu. Baza danych zajmuje 545 MB. Rozmiar obrazów jest bardzo zróżnicowany od 60x40 do 1024x768.

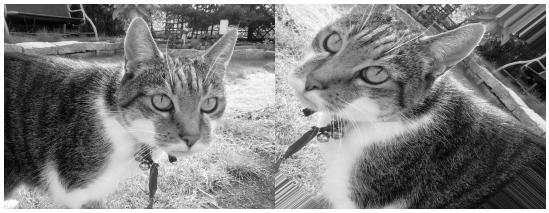
Zbiór danych podzielono na zbiór treningowych oraz testowy. Zbiór treningowych zawiera 22 tys. zdjęć, zbiór testowy 3 tys. zdjęć. Zbiór treningowy oraz testowy zostały równo podzielone na dwie klasy .

5. PROCES UCZENIA

5.1. Wstępne przetwarzanie danych

Każda z badanych sieci neuronowych przyjmuje na wejście kolorowe zdjęcie o rozmiarze 224x224px. W celu dopasowania rozmiaru obrazów do rozmiaru wejścia sieci neuronowych przeprowadzono skalowanie rozmiaru obrazu metodą najbliższego sąsiada (*nearest neighbour)*. Następnie przeskalowano wartości pikseli tak aby zawierały się w przedziale od 0 do 1.

W celu zwiększenia możliwości generalizacji sieci neuronowej wykorzystano rozszerzenie danych zbioru treningowego. Rozszerzanie danych polega na dodaniu niewielkich modyfikacji do zdjęcia, które nie powodują zmiany klasy obiektu znajdującego się na zdjęciu, natomiast znacznie zmienia się reprezentacja obrazu zapisanego w pamięci komputera w postaci macierzy. W pracy zastosowano następujące modyfikacje: odbicia w pionie oraz w poziomie, obroty, rozciąganie i zwężanie, przesunięcia. Losowa modyfikacja stosowana była przed każdym podaniem zdjęcia na sieć. Przykład rozszerzania danych widoczny jest na rys. 2.



Rys. 2. Przykład rozszerzenia danych. Lewa – oryginalny obraz, prawa – obraz wygenerowany za pomocą rozszerzania danych

W celu usprawnienia procesu optymalizacji sieci neuronowej przeprowadzono normalizację danych, która polega na zmianie średniej oraz odchylenia standardowego, tak aby wynosiły odpowiednio zero oraz jeden.

5.2. Przebieg uczenia

Uczenie głębokich sieci neuronowych przebiega w podobny sposób do uczenia płytkich sieci neuronowych. Ze względu na dwuklasowy problem klasyfikacji wybrano binarną entropię krzyżową jako funkcję celu, której wartość zależna jest od błędów popełnianych przez sieć neuronową. Minimalizacja funkcji celu prowadzi do minimalizacji błędów popełnianych przez sieć. Optymalizacja funkcji celu odbywa się przy pomocy metod gradientowych. W tym celu należy wyznaczyć gradient funkcji celu względem wszystkich wag. Obliczanie gradientu odbywa się przy pomocy metody wstecznej propagacji błędu, która wyznacza cząstkowe pochodne funkcji celu względem wszystkich parametrów sieci neuronowej. Funkcja celu optymalizowana jest przy pomocy algorytmu Mini-batch Stochastic Gradient Descent, która modyfikuje wagi na podstawie obliczonych gradientów i współczynnika uczenia. Modyfikacja wag odbywa się po podaniu na wejście ustalonej liczby przykładów uczących (uśrednianie gradientów).

5.3. Uczenie transferowane

Podczas uczenia sieci neuronowych wykorzystano technikę, która nazywa się uczeniem transferowanym. Metoda polega na wykorzystaniu sieci neuronowej, która potrafi rozwiązywać pewne zadanie, do nauki rozwiązywania innych zadań. Analogicznie do umiejętności ludzi np. osoba potrafiąca rysować znacznie łatwiej nauczy się malować niż osoba, która rysować nie potrafi. Aby wykorzystać technikę transfer learning, należy przetrenować sieć bardzo dużym zbiorem danych (np. z 1 mln przykładów uczących) Następnie taka sieć wykorzystywana jest jako punkt startowy w uczeniu innych zadań klasyfikacji obrazu. W badaniach wykorzystano przeuczone sieci dostępne w

bibliotece Keras. Metoda transfer learning jest bardzo użyteczna w zadaniach, w których dostarczona ilość danych jest niewystarczająca do przeprowadzenia efektywnego uczenia oraz dobrania milionów parametrów sieci neuronowej [9].

5.4. Sprzęt i oprogramowanie

Uczenie głębokich sieci neuronowych wymaga silnych jednostek obliczeniowych. Obliczenia prowadzone są z wykorzystaniem kart graficznych i technologii CUDA 8.0, która umożliwia obliczenia równoległe. Obliczenia równoległe znacznie przyspieszają obliczenia względem obliczeń prowadzonych na procesorze (nawet do kilkudziesięciu razy). Podczas obliczeń wykorzystano jednostkę komputerową wyposażoną w następujące komponenty: karta graficzna GeForce GTX 980Ti z pamięcią 6GB, procesor Intel Core i7-4930K, pamięć RAM 16GB.

Do obliczeń wykorzystano język Python 2.7 wraz z pakietem bibliotek. Najistotniejsze z bibliotek: Theano 0.9.0 – umożliwia tworzenie grafów obliczeniowych oraz ich automatyczne różniczkowanie, Keras 1.2.0 – biblioteka służąca do szybkiego prototypowania architektur sieci neuronowych.

6. WYNIKI I ANALIZA JAKOŚCIOWA

6.1. Wykorzystane miary statystyczne

W celu oceny algorytmów klasyfikacji wykorzystano następujące miary statystyczne: dokładność, czułość, swoistość oraz krzywe ROC. Dokładność jest stosunkiem liczby poprawnie sklasyfikowanych obrazów do liczby wszystkich obrazów w bazie. Czułość i swoistość są wykorzystywane w aplikacjach medycznych, w których można wyróżnić klasę pozytywną (obecność choroby) oraz klasę negatywną (brak choroby). Cechy tych wskaźników można jednak przełożyć na inne zadania, w których sztucznie określa się klasę pozytywną oraz negatywną. W naszych badaniach klasę „kot" przyjęliśmy jako klasę negatywną, natomiast klasę „pies" jako pozytywną. Wykorzystując te pojęcia można zdefiniować czułość oraz swoistość. Czułość jest stosunkiem poprawnie sklasyfikowanych przykładów pozytywnych (psów) do całkowitej ilości przykładów pozytywnych w zbiorze. Natomiast swoistość jest stosunkiem poprawnie sklasyfikowanych przykładów negatywnych (kotów) do całkowitej ilości przykładów negatywnych zbiorze.

Krzywe ROC są krzywymi opartymi na mierze czułości oraz swoistości. Kreślenie takiej krzywej polega na zmianie progu klasyfikacji neuronu wyjściowego i pomiar obu wartości. Następnie na wykresie nanosi się punkty, w których współrzędna X odpowiada wartość (1 - swoistość), a współrzędna Y wartości czułości. Krzywe ROC wskazują na pewność sieci neuronowej w podejmowanej decyzji. Miarą tej pewności i możliwości klasyfikacji jest obszar pod krzywą (AUC), który może przyjąć maksymalna wartość równą 1, co oznacza najlepszą jakość klasyfikacji.

6.2. Uzyskane wyniki

Podczas eksperymentów uzyskano wyniki przedstawione w Tabeli 1.

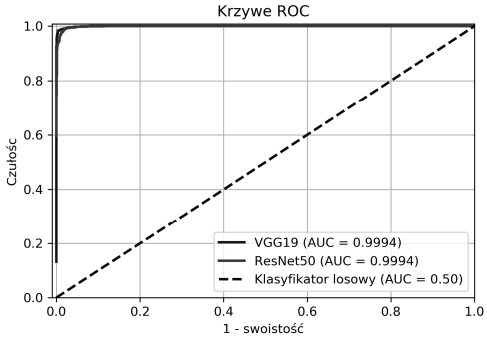
Tabela 1. Wyniki klasyfikacji

Wskaźnik	VGG19 (rozszerzanie danych)	ResNet50 (rozszerzanie danych)	VGG19 (bez rozszerzania danych)	ResNet50 (bez rozszerzania danych)
Dokładność (%)	99.12	99.24	98.88	98.44
Czułość (%)	99.12	99.28	99.36	97.92
Swoistość (%)	99.12	99.20	98.40	98.96
Czas treningu	3h	5.5h	3h	5.5 h

Każda z sieci uzyskała bardzo wysokie wyniki klasyfikacji. Sieć ResNet uzyskała najwyższy wynik, co jest zgodne z wynikami raportowanymi przez innych badaczy. Uczenie takiej sieci trwało jednak znacznie dłużej od uczenia zmodyfikowanej sieci VGG19.

Brak rozszerzania danych spowodowało pogorszenie wyników klasyfikacji.

Wykreślono krzywe ROC badanych klasyfikatorów. Krzywe widoczne są na rys. 3.



Rys. 3. Krzywe ROC badanych systemów

Z wykresu z rys. 3. można odczytać bardzo wysokie wyniki klasyfikacji psów oraz kotów. Ponadto sieci neuronowe są pewne swoich odpowiedzi tzn. neuron sigmoidalny zwraca wartości bliskie zero bądź jeden, a nie pośrednie, co mogłoby wskazywać na niepewność sieci neuronowych.

6.3. Analiza jakościowa

Przeprowadzono analizę jakościową badanych systemów w celu lepszego poznania działania oraz zachowania się algorytmów głębokiego uczenia. Analiza jakościowa jest bardzo istotna ze względu na to, że algorytmy posiadają charakter czarnej skrzynki.

Przeanalizowano zdjęcia błędnie sklasyfikowane przez sieć neuronową. Podczas obserwacji zauważono następujące zależności – zbiór obrazów błędnie sklasyfikowanych obrazów jest prawie taki sam dla obu badanych sieci. Dodatkowo zauważono, że część błędnie sklasyfikowanych zdjęć jest również trudna do poprawnego sklasyfikowania przez człowieka.

Rys. 4. Przykład kotów sklasyfikowanych jako psy. Klasyfikacja zwierzęcia sprawia problem nawet ludziom

Na rys. 4 zamieszczono przykłady takich zdjęć. Na rys. 5 umieszczono błędnie sklasyfikowane zdjęcia, które sprawiają trudność ludziom ze względu na złą jakość zdjęć czy tło zlewające się ze zwierzętami.

Rys. 5. Przykład psów sklasyfikowanych jako koty

Przenalizowano reakcję systemu na nietypowe wejścia np. zwierzę z ciałem psa i głową kota, kot i pies na jednym obrazie. Zauważono, że algorytm znacznie większą uwagę zwraca na głowę zwierzęcia, dopiero potem na jego ciało. Dodatkowo sprawdzono możliwość sieci do generalizacji wiedzy – przebadano reakcję sieci na postacie z kreskówek. Każda z sieci poprawnie klasyfikowała postacie z kreskówek takie jak pies Reksio czy kot Tom. Natomiast ludzie klasyfikowani byli jako psy. Przykłady zaprezentowano na rys. 6.

Rys. 6. Nietypowe wejścia sieci neuronowej

7. PODSUMOWANIE

Głębokie sieci neuronowe osiągnęły bardzo wysokie wyniki klasyfikacji obrazów, potrzebują jednak dużej ilości danych uczących oraz wydajnych jednostek obliczeniowych GPU. Analiza jakościowa działania sieci neuronowej umożliwia przeanalizowanie sposobu wnioskowania sieci neuronowej. Dalsze badania w kierunku zrozumienia sposobu wnioskowania głębokich sieci neuronowych pozwoli na budowanie algorytmów budzących zaufanie, co jest istotne w zastosowaniach medycznych [10][11] oraz finansowych. Kolejnym wyzwaniem jest stworzenie algorytmów wydajnych obliczeniowo, co umożliwi stosowanie algorytmów w znacznie większej liczbie urządzeń, które nie są wyposażone w wydajne jednostki obliczeniowe. W tym celu istotnym będzie opracowanie algorytmów automatycznego doboru struktur optymalnych pod względem wydajności np. ilości zużywanej pamięci czy ilości wykonywanych operacji matematycznych.

8. BIBLIOGRAFIA

1. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, p. 436, 2015.
2. Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 253–256.
3. X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” *AISTATS ’11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, vol. 15, pp. 315–323, Jun. 2011
4. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
5. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
6. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
7. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *arXiv:1502.03167,* 2015
8. https://www.kaggle.com/c/dogs-vs-cats (czas dostępu: 05.2017)
9. A. Krizhevsky, I. Sutskever, and H. Geoffrey E., “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pp. 1–9, 2012.
10. A. Kwasigroch, B. Jarzembinski, and M. Grochowski, "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy", in *International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 111–116.
11. A. Kwasigroch, A. Mikolajczyk, and M. Grochowski, ‘Deep neural networks approach to skin lesions classification — A comparative analysis’, in *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*, Poland, 2017, pp. 1069–1074.

OBJECT CLASSIFICATION WITH DEEP NEURAL NETWORKS

Deep neural networks are modern algorithms from the family of artificial intelligence, that are widely used these days for task of an image analysis. In this paper, we present results of research on deep neural network for image recognition. We tested 2 different neural architectures, namely: modified VGG19, ResNet50. In order to improve the classification results we employed two methods called dropout and transfer learning. The systems were trained on the dataset containing 22 000 training images and 3000 test images. The dataset used contains different pictures of animals (cats and dogs). The dataset of animals make analyses of network performance easier, because they are easy to interpret by human. The employed systems were tested qualitatively and quantitatively. The influence of atypical inputs were examined, also. Moreover, the analysis of improperly classified images was performed. We achieved high classification results. In order to evaluate the classification performance we utilized the following set of statistical measures: accuracy, specificity, sensitivity and ROC curves.

Keywords: deep learning, neural networks, artificial intelligence, image processing.