

Rozpoznawanie obrazów przy pomocy głębokich sieci neuronowych- Warsztaty KFnRD ICM 2017

Mateusz Sieniawski

20 grudnia 2016

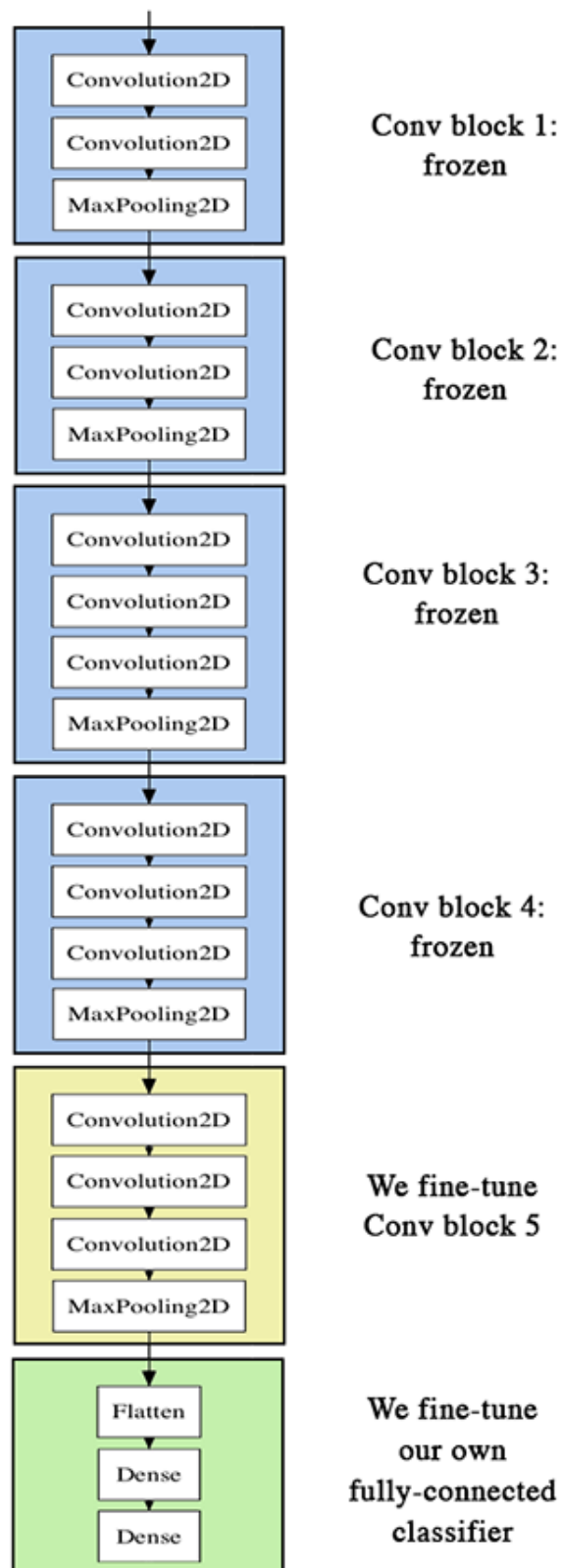
Mój background Z sieciami neuronowymi nie miałem za wiele wspólnego. Te zadanie kwalifikacyjne było de facto moim pierwszym praktycznym kontaktem z nimi. Problematyczne było dla mnie także zapoznanie się z high-endową pythonową składnią- wiem, że istnieje jakaś komenda, która robi jakąś prostą rzecz, ale nie wiem jak się nazywa i muszę googlać. Było to szczególnie uciążliwe na początku. Na szczęście, po wielu wielu godzinach uciążliwej walki, udało mi się zrobić coś, z czego jestem dumny. Może nie będzie to rozwiązanie najwyższej klasy, ale praca nad nim bardzo mnie rozwinęła.

Wszystkie kody źródłowe są tutaj: <https://github.com/Cheetar/KFnRD-ICM-2017>. W moim rozwiązaniu jest mało wizualizacji- skupiłem się w pełni na zoptymalizowaniu sieci neuronowej. Próbowałem wielu różnych architektur, a wiele z nich różniło się np. jedną warstwą. Chciałem dokładnie zrozumieć, co ma największy wpływ na wynik końcowy i w jaki sposób zbudować najbardziej optymalną sieć. Zacząłem od prostych sieci z dwoma warstwami ukrytymi, jednak osiągały one wyniki jedynie rzędu 92% (Simple network two hidden layers). Następnym krokiem było dodanie konwolucji jako pierwszej warstwy (Network with convolution). Co ciekawe, spróbowałem konwolucji z różnymi jądrami i dla większych rozmiarów osiągnąłem nieznacznie lepszy wynik. Następnie dodałem warstwę MaxPooling2D, dzięki której udało się zwiększyć precyzję sieci o 2%. (5x5 convolution with MaxPooling2D)

Moja najlepsza sieć opierała się na strukturze VGG16 (rys. 1).

Sieć uczyła się na komputerze z procesorem Intel Core i7-4700HQ 2.4GHz, co pozwalało na stosunkowo duży jej rozmiar. Na samym początku obraz jest przetwarzany przez kilka konwolucji z różnymi jądrami, a następnie przez klasyfikator z dwoma warstwami ukrytymi. Udało się osiągnąć wynik 95%. (The best network)

Kilka słów jeszcze o przeuczeniu. Było widoczne, że gdy wynik na zbiorze treningowym przekraczał już 99%, to wynik na testowym nie wzrastał/nieznacznie malał. Najlepiej jest to widoczne na najlepszej sieci, gdy wynik na zbiorze testowym jest taki sam po 10 iteracjach (96% na treningowym) jak i po 30 (99% na treningowym).



Rysunek 1: Schemat modelu VGG16