

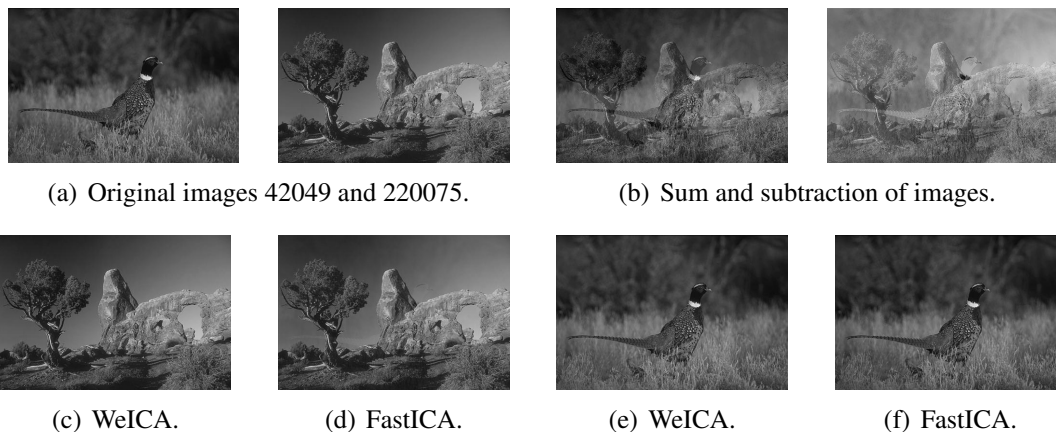
W czasie bardzo szybko rozwijających się usług telekomunikacyjnych, sieciowych, i gwałtownego rozwoju tzw. big data, przesyłane jest coraz więcej danych. Mogą to być pakiety GPRS/3G/4G/LTE, dane strumieniowe takie jak filmy czy pliki muzyczne (np. w serwisach VoD), czy dane z sieci społecznościowych takich jak np. Facebook czy Twitter.

Dostęp do tak dużej ilości zróżnicowanych danych pociąga za sobą konieczności rozwoju metod uczenia nienadzorowanego, które są często wykorzystywane w pierwszej fazie analizy danych. Istnieje potrzeba tworzenia nowych lub optymalizacji już istniejących algorytmów dedykowanych do ich przetwarzania.

W ramach niniejszego projektu, zajmiemy się stworzeniem oraz zoptymalizowaniem najbardziej popularnych algorytmów nauczania nienadzorowanego opartych na podejściu probabilistycznym. W szczególności skoncentrujemy się na metodach estymacji gęstości (np. GMM), grupowaniu danych opartym o analizę gęstości (np. DBSCAN) niezależnej analizie składowych (np. ICA).

W pierwszej części będziemy dążyć do przekształcenia istniejących problemów optymalizacyjnych, tak aby uzyskać rozwiązania dane jawnymi wzorami albo aby zmodyfikowane problemy mogły być skutecznie minimalizowane numerycznie. W drugiej części zajmiemy się zastosowaniem metod nienadzorowanych w problemach nadzorowanych, w szczególności na lepszym wstępnym uczeniu oraz inicjalizacji głębokich sieci neuronowych.

Powyższe podejście może zostać użyte do rozwiązania problemu ICA. Załóżmy, że mamy dwa zdjęcia, patrz Ilustracja 1(a). Ale ktoś udostępnia nam ich zaszumione wersje, patrz Ilustracja 1(a). Naszym celem jest odzyskanie oryginalnych obrazów. Wynik naszego algorytmu oraz klasycznej metody FastICA prezentujemy na Ilustracji 1.



Rysunek 1: Porównanie rozdzielania oryginalnych obrazów przez naszą metodę (WeICA) i FastICA.