

Deep Convolutional Neural Networkによるクラウド型顔照合 の分散処理における負荷検証

A Study on Efficiency of Distributed Processing for Cloud-based Face Verification by Deep Convolutional Neural Network

加藤優[†] 福井宏[†] 山下隆義[†] 山内悠嗣[†] 藤吉弘亘[†]
中部大学[†]

Yuu KATO[†] Hiroshi FUKUI[†]
Takayoshi YAMASHITA[†] Yuji YAMAUCHI[†] Hironobu FUJIYOSHI[†]
[†]Chubu University

アブストラクト Deep Convolutional Neural Network(DCNN)は高精度な識別が可能である一方、その識別計算コストが高いという問題がある。スマートフォン、監視カメラ、レコーダなどの機器に搭載されているCPUのスペックは低く、DCNNの全処理をこれらの機器で行うことは困難である。DCNNの処理は、畳み込み層とプーリング層、全結合層を階層的に配置しているため、処理をクライアント側とクラウド側に分割させることが可能である。そこで、我々はDCNNの処理の下位層をクライアント側、上位層をクラウド側で行うクラウド型顔照合における分散処理を提案する。クラウド型顔照合は、処理の分割ができるだけでなく、クラウド側に画像を送信する必要がないため、プライバシー面で安全性の高いシステムとなる。本稿では、クラウド型顔照合において、DCNNをどのように分割すれば最適となるかの負荷検証を行う。

1 はじめに

画像認識分野において、Deep Convolutional Neural Network(DCNN)が様々なベンチマークで高精度な識別を実現している [1][2]。DCNNは、畳み込み層や全結合層の各層で乗算処理を何度も行うため、計算コストが膨大になる。そこで、クライアント側で全ての演算処理を行わずに、演算の途中結果をクラウドに送信してクラウド側でDCNNの残りの処理を行う。本稿では、顔照合を対象とし、DCNNをどのように分割すれば計算・通信リソースが最適化となるかの負荷検証を行う。

2 DCNNによるクラウド顔照合

2.1 Deep Convolutional Neural Network

DCNNは、畳み込みとプーリングを複数回繰り返すことで、エッジ勾配や色のグラデーション、またはそれらを組み合わせた複雑な重みフィルタを自動で取得する。これらのフィルタは、誤差逆伝播法を用いた勾配降下法による学習により決定される。

2.2 クラウド型顔照合システム

提案するシステムは、図1のようにDCNNの畳み込み層をクライアントとクラウドに分割する。クライアントでは、指定した畳み込み層の特徴マップをクラウドへ送信する。クラウドでは、受信した特徴マップを再度DCNNへ入力し、畳み込みとプーリングの続きを行う。DCNNの構造を分割することで、クライアントの計算コストを低減させることができる。照合処理は、DCNNの識別層で行わずに全結合層前の特徴マップを特徴ベクトル化して利用する。特徴ベクトルとクラウドに保存されている登録画像の特徴ベクトルのユークリッド距離を算出し、しきい値処理により本人かどうか判定する。

本システムでは、クライアント側のDCNNで算出した特徴マップのみを送信し、クラウド側に顔画像を保存しないため、プライバシーを配慮したシステムとなる。また、DCNNの畳み込み層における特徴マップは、フィルタの係数が既知でない場合、元画像を復元することは不可能であり、安全性の高いシステムとなる。

3 負荷検証実験

DCNNの分割における通信量、計算量の観点から検証を行う。まず、顔照合に必要なDCNNの畳み込み層の層数を調査する。次に、DCNNの第何層目から切り離れた場合が、最もクライアント・クラウドに負荷が少ないかを検証する。

クライアントのスペックはOS: OSX EI Capitan, CPU: Intel(R) Corei7-3520, メモリ: 8GBである。クラウドのスペックはOS: Ubuntu14.04.3 LTS, CPU: Intel(R) Corei7-4790, GPU: NVIDIA(R) GeForce GTX 980, メモリ: 8GBである。検証実験では、畳み込み層を1層から9層まで増やし、各構成での照合率を比較する。活性化関数にはMaxoutを使用し、プーリング層は畳み込み

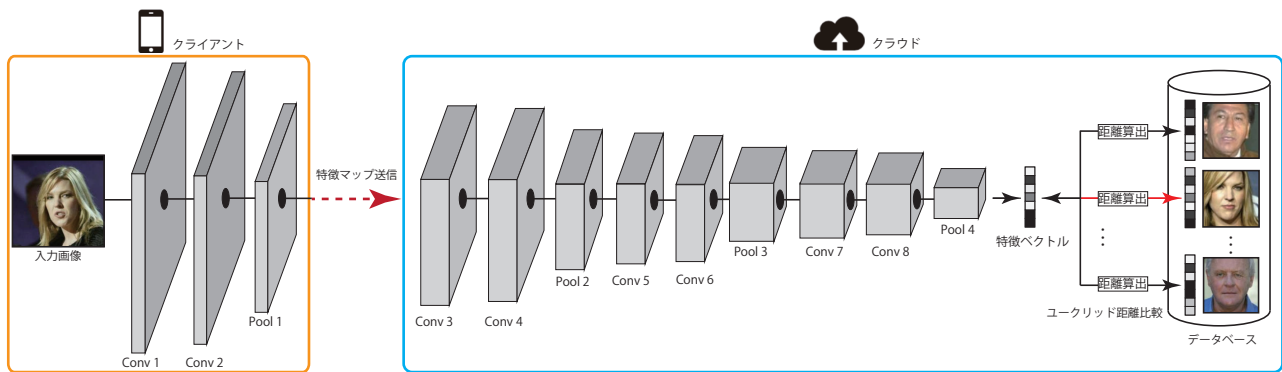


図 1: DCNN によるクラウド型顔照合の全体図

層の 2 層目, 4 層目, 6 層目及び 8 層目に配置する. また, 各層のフィルタ枚数は 1, 2 層目が 16 枚, 3, 4 層目が 32 枚, 5, 6 層目が 64 枚, 7, 8 層目が 128 枚, 9 層目が 256 枚である. DCNN の学習サンプル, 評価サンプルはそれぞれ 4,031 枚, 143 枚である.

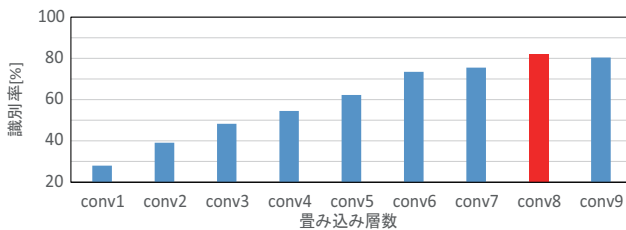


図 2: 畳み込み層数を変化させた際の精度

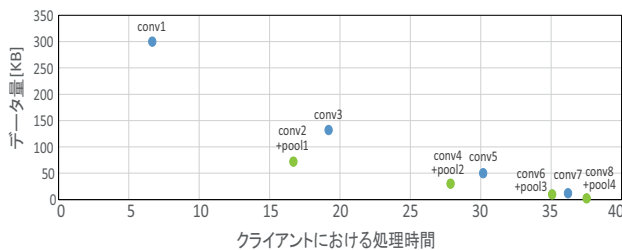


図 3: クライアントの情報送信量と処理時間

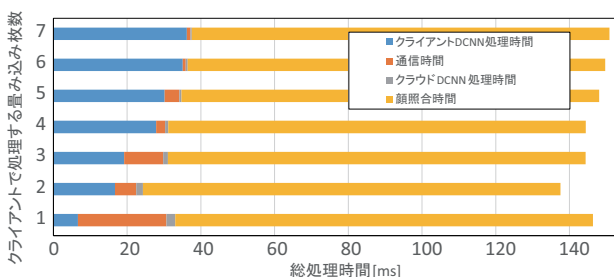


図 4: 切り離す畳み込み層の変化による総処理時間の変化

3.1 顔照合 DCNN の構成要素

図 2 に DCNN の畳み込み層数を変化させた時の識別率を示す. 畳み込み層数を増やしていくと照合精度が向上しており, 畳み込み層を 8 層とした場合の精度が最も高い. 本稿では, 畳み込み層が 8 層の DCNN を用いたクラウド型顔照合の分散処理を対象とする.

3.2 クライアントの情報送信量と処理時間

図 3 に, クライアント側で計算する畳み込み層を変化させた時の処理時間とクラウドへの情報伝送量を示す. クライアントで処理する層が多い場合, クラウドと比べ処理速度が遅いため, 総処理時間は長くなる. しかし, 切り離す層が浅い場合, 情報送信量が膨大となり, 通信時間が大きくなってしまう. 図 3 より, 2 層目の畳み込み層 (Conv2 + Pool1) で切り離した場合, クライアント側での処理時間が短く, かつデータ送信量が少なくなることがわかる.

3.3 総処理時間

図 4 は, 切り離す畳み込み層の変化による総処理時間を示す. 総処理時間は, クライアント・クラウドの DCNN の処理時間にクラウドの顔照合時間と通信時間を加算したものである. 顔照合時間は 100 人分の顔照合にかかる値であり, 通信時間は通信速度を 100Mbps(12.5KB/ms) と仮定した際の値である.

図 4 より, 前節で想定した 2 層目の畳み込み層で切り離した構成の総処理時間が最も短い. クラウド側で行う顔照合時間はどの層で分割しても大きく変わらないため, 総処理時間は, クライアントの処理時間と通信時間が大きく影響していることが確認できる.

4 おわりに

本稿では, DCNN によるクラウド型顔照合の分散処理における負荷検証を行った. クラウド型にすることで, 安全な通信を行い, クライアント側のみで処理するよりも高速に処理することが可能となった. 今後は, クラウド型顔照合のさらなる高精度化を行う.

参考文献

- [1] Y.Taigman, *et al.*, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”, CVPR, 2014.
- [2] A.Krizhevsky, *et al.*, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS, 2012.